

Decoding viral infections and cellular signalling via mathematical modelling

Frederic Grabowski

Doctoral thesis advisor:

Tomasz Lipniacki

Doctoral thesis auxiliary advisor:

Marek Kočańczyk



Institute of Fundamental Technological Research,
Polish Academy of Sciences

Warsaw 2025

Results presented in this thesis were obtained in the outstanding research environment of the Department of Biosystems and Soft Matter of the Institute of Fundamental Technological Research, Polish Academy of Sciences.

Support from the National Science Centre and the Norwegian Financial Mechanism are kindly acknowledged.

I would also like to thank my wife for her continuous support, and my cat for watching over my laptop when I wasn't.



Abstract

In my PhD thesis, I employed mathematical and computational modelling to investigate virus–host interactions, epidemic dynamics, immune escape and information processing in cell systems.

The first part of my research focuses on computational modelling of viral infections at the cellular level. I developed a spatial, stochastic, agent-based model to study the interactions between respiratory syncytial virus (RSV) and an infected monolayer of cells. This model, calibrated using diverse experimental data from our laboratory (obtained from immunofluorescence imaging, ELISA, dPCR, and Western blotting), captures the mutual antagonism between the virus and the innate immune response at both the single-cell and the population level. It allowed me to distinguish four cell states: naïve, infected, interferon-producing, infected virus-spreading, and interferon-responding. Next, I extended my analysis to influenza A virus (IAV). While the spread of RSV is mainly controlled by type I interferon, the spread of IAV is mitigated by both type I and type III interferons, which play different roles. Using a simple SIR model, I argue that type III interferons can suppress IAV by promoting faster cell death of infected cells, a mechanism that RSV evades by blocking apoptosis.

The emergence of the COVID-19 pandemic has led to restrictions in IPPT PAN experimental laboratories, but also shifted my focus to epidemic modelling, which shares many similarities with virus spread modelling, and resulted in four well-cited publications. We introduced and analysed several models to capture key aspects of the pandemic. First, using a modified SEIR model with an Erlang-distributed incubation period, I studied the interplay between epidemic spread and human population behaviour. Second, through stochastic simulations, I demonstrated that a small fraction of superspreaders can lead to epidemic spread with a high basic reproduction number (R_0), which can be underestimated during the early epidemic phase, before superspreading events occur. Later, I focused on specific SARS-CoV-2 variants: Alpha and Omicron. In particular, using a Bayesian approach, I estimated the time of emergence of the Omicron variant by modelling the accumulation of mutations during its evolution.

In the final section, I focus on cell signalling and information theory. Motivated by my work on RSV infection and earlier studies in Prof. Lipniacki's laboratory, I explored the utility of non-identifiable models in a Bayesian framework using Markov chain Monte Carlo (MCMC) simulations, demonstrating that even only partially constrained models can yield predictive insights. I also investigated the bitrate capacity of the MAPK/ERK pathway, which is critical for continuous coordination during processes like wound healing. The thesis includes two publications on the numerical estimation of mutual information (MI). The first introduces a comprehensive benchmark for evaluating existing MI estimators. This benchmark leverages the fact that MI is preserved under diffeomorphisms, enabling the construction of various pairs of random variables with known MI for a systematic comparison of estimator performance. The second article extends this work by analyzing pointwise mutual information profiles and mixture distributions, further refining the benchmark and introducing a Bayesian estimator of mutual information.

Streszczenie (Polish)

W mojej rozprawie doktorskiej skupiłem się na matematycznym i numerycznym modelowaniu interakcji między wirusami a komórkami gospodarza, dynamice epidemii, ucieczce immunologicznej oraz przetwarzaniu informacji w układach komórkowych.

Pierwsza część badań dotyczy modelowania infekcji wirusowych na poziomie komórkowym. Opracowałem przestrzenny, stochastyczny model agentowy do analizy interakcji wirusa RSV (ang. respiratory syncytial virus) z monowarstwą zarażonych komórek. Model, który został skalibrowany na podstawie danych eksperymentalnych uzyskanych w naszym laboratorium (obrazowanie immunofluorescencyjne, ELISA, dPCR i Western blot), odtwarza wzajemny antagonizm między wirusem a nieswoistą odpowiedzią odpornościową, zarówno na poziomie pojedynczych komórek, jak i całej populacji. Wyróżniłem cztery kluczowe stany komórkowe: komórki w stanie wyjściowym, komórki zarażone produkujące interferon, komórki zarażone rozprzestrzeniające wirus oraz komórki odpowiadające na interferon. Następnie rozszerzyłem analizę na wirusa grypy A (ang. influenza A virus, IAV). Podczas gdy rozprzestrzenianie RSV jest hamowane głównie przez interferon typu I, w przypadku IAV kluczową rolę odgrywają zarówno interferony typu I, jak i III, pełniące odrębne funkcje. Korzystając z prostego modelu SIR, wykazałem, że interferony typu III mogą ograniczać zarażenia IAV poprzez indukcję śmierci zarażonych komórek – mechanizm ten jednak nie działa w przypadku RSV, który skutecznie blokuje apoptozę.

Pojawienie się COVID-19 skierowało moją uwagę na modelowanie epidemii, które pod wieloma względami przypomina modelowanie rozprzestrzeniania wirusów na poziomie komórkowym. W efekcie powstały cztery często cytowane publikacje, w których opracowano i przeanalizowano kilka modeli opisujących kluczowe aspekty pandemii. Po pierwsze, wykorzystując zmodyfikowany model SEIR z rozkładem Erlanga dla okresu inkubacji, zbadałem zależności między dynamiką epidemii a zachowaniami populacji. Po drugie, na podstawie symulacji stochastycznych wykazałem, że niewielka frakcja super-roznościeli może prowadzić do epidemii o wysokiej podstawowej liczbie odtwarzania (R_0), której wartość bywa niedoszacowana w początkowej fazie epidemii, zanim wystąpią zdarzenia wielokrotnych zarażeń. W kolejnych etapach badań skupiłem się na nowych wariantach SARS-CoV-2, Alfa i Omikron. W szczególności, stosując bayesowskie podejście, oszacowałem moment pojawienia się wariantu Omikron, modelując akumulację mutacji w trakcie jego ewolucji.

Ostatnia część pracy dotyczy sygnalizacji komórkowej oraz teorii informacji. Zainspirowany badaniami nad infekcją RSV oraz wcześniejszymi pracami w laboratorium prof. Lipniackiego, przeanalizowałem użyteczność nieidentyfikowalnych modeli w podejściu bayesowskim, stosując symulacje metodą łańcuchów Markowa Monte Carlo (MCMC). Pokazałem, że nawet modele z nie w pełni zidentyfikowanymi parametrami mogą dawać precyzyjne przewidywania. Zbadałem również przepustowość informacyjną szlaku MAPK/ERK, który pozwala komórkom na koordynację przemieszczania się podczas procesów takich jak gojenie ran. Rozprawa zawiera też dwie publikacje dotyczące numerycznej estymacji informacji wzajemnej (MI). Pierwsza publikacja wprowadza benchmark estymatorów korzystając z faktu, że dyfeomorfizmy zachowują MI, co pozwala na konstrukcję różnych par zmiennych losowych o znanym MI. Druga publikacja skupia się na profilach punktowej informacji wzajemnej oraz rozkładach mieszanych, poszerza benchmark i wprowadza bayesowski estymator informacji wzajemnej.

Contents

1.	Background	1
1.1.	Compartmental models of infection	1
1.2.	Epidemic modelling	2
1.3.	Viral infections in cell populations	3
1.4.	Modelling delays in compartmental models	3
1.5.	Non-identifiability and information capacity of signalling pathways	4
2.	Research hypotheses	7
3.	Summary of main results	11
3.1.	Dissecting interactions between RSV and innate immunity	11
3.2.	Differential role of type I and type III interferons	13
3.3.	Estimation of R_0 of COVID-19	15
3.3.1.	Incubation period	15
3.3.2.	Super-spreaders	16
3.4.	Population response to the SARS-CoV-2 epidemic	17
3.5.	Emergence of new SARS-CoV-2 variants	19
3.5.1.	Variant Alpha	19
3.5.2.	Variant Omicron	20
3.6.	Applicability of partially-constrained models to signalling pathways	22
3.7.	Numerical study of MAPK/ERK channel capacity	26
3.8.	Benchmarking mutual information estimators	28
3.9.	Pointwise mutual information and mixture models	29
4.	Discussion and further directions	33
5.	List of original articles included in the thesis	35
6.	List of software written as a part of the thesis	37
7.	Conferences & seminars	37
8.	Funding	37
9.	References	39
10.	Original articles included in the thesis	43
11.	Author contributions	253

1. Background

1.1. Compartmental models of infections

Viral infections have been studied quantitatively for a long time. The classic Susceptible, Infectious, Removed (SIR) model, introduced in the 1920s [1], [2], [3], remains widely used due to its simplicity, broad applicability, and extensibility. The model describes the dynamics of an epidemic by categorizing individuals into one of three exclusive states at any given time: *Susceptible* (S) individuals can become *infectious* (I) after contact with infectious individuals, and infectious individuals eventually transition to the *removed* (R) state – which includes individuals who have either recovered and acquired immunity or died. In both cases, it is assumed that they are no longer infectious and cannot be infected again. The dynamics are commonly described with a compartmental model formulated as a set of coupled ordinary differential equations (ODEs):

$$S' = -\beta I S$$

$$I' = \beta I S - \gamma I$$

$$R' = \gamma I,$$

where the prime (') denotes differentiation with respect to time. Parameter β determines the infectivity of the disease, and γ determines the recovery (removal) rate. The initial conditions usually assume a small fraction of infectious individuals, $I(t = 0) > 0$, and no recovered individuals, $R(t = 0) = 0$. Notably, by defining the total population as $N = S + I + R$, we see that $(S + I + R)' = 0$, meaning the population size remains constant over time. The model can also be recast as a Markov chain and then simulated stochastically, for example using the Gillespie algorithm. In this case, S , I , and R take non-negative integer values, and the following transition rules apply:



Example ODE trajectories and sample stochastic simulations of a corresponding stochastic model are shown in Figure 1.

The main advantage of the SIR model is its simplicity, which makes it easy to extend. A common variation is the SEIR model, which introduces an additional exposed (E) individuals who have been infected but are not yet infectious. The model's formulation is also highly general, allowing it to be applied to both human, animal or plant populations and viral or bacterial infections in cell populations. In this thesis, adaptations of the SIR model were used to study differential roles of type I and type III interferons in influenza A (IAV) and respiratory syncytial virus (RSV) infections of cells, analyze the impact of super-spreading events on estimation of R_0 in COVID-19, investigate the effect of different incubation period distributions, and finally to compare advantages of emerging SARS-CoV-2 strains, in terms of transmission and fitness.

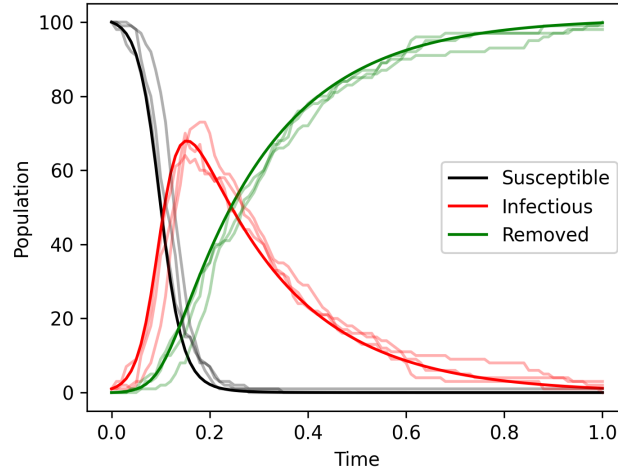


Figure 1. Trajectories of the SIR model for $N = 100$, $I(0) = 1$, $\beta = 0.5$, $\gamma = 5$. Bold lines show the ODE solution, transparent lines show three stochastic runs of the Gillespie algorithm.

1.2. Epidemic modelling

The COVID-19 pandemic has had profound global health and economic impacts. After the emergence of SARS-CoV-2 in Wuhan, China, at the end of 2019, the virus rapidly spread worldwide. By January 2020, the first cases had already been reported in North America, Australia, and Europe. In Poland, the epidemic officially began with the first confirmed case in March 2020, causing major disruptions to daily life—from holiday gatherings to the operation of experimental laboratories at IPPT PAN.

A key parameter (in addition to the mortality rate) governing the speed and severity of an epidemic is the *basic reproduction number*, R_0 . The basic reproduction number is the expected number of immediate infections caused by an infected individual, assuming that all other individuals in the population are susceptible. If $R_0 < 1$, the number of infected individuals will tend to decrease, and no epidemic outbreak is expected. Conversely, when $R_0 > 1$, the epidemic may start spreading. Initially, the numbers of both infected and recovered individuals increase, and thus the fraction of susceptible individuals $S(t)/N$ decreases. This implies that the *effective reproduction number*, $R_t = R_0 S(t)/N$, also decreases. Eventually, R_t drops below 1, and the number of infected individuals starts decreasing. We see that this occurs when $S(t) = N/R_0$, linking the peak of infections with the basic reproduction number. Once the susceptible fraction is below N/R_0 , the population is said to have attained *herd immunity*. This does not mean the disease has been eradicated; rather, it indicates that new large-scale outbreaks are unlikely to occur.

In the SIR model, the basic reproduction number is given by $R_0 = \beta/\gamma$. This can be seen from the stochastic interpretation of β and γ . An individual stays infected on average for $1/\gamma$ units of time, and infects β individuals per unit of time, thus infecting on average β/γ individuals. This points to two factors influencing R_0 : the infectivity of the virus, and the period of infectiousness. However, β , and therefore R_0 , is also influenced by social factors, such as the expected number of contacts between individuals, or measures imposed to impede viral transmission (masks, social distancing, increased diagnostic capacity, contact tracing). This last factor is the primary motivation for restrictions introduced during the pandemic. More detailed models [4] can incorporate spatial effects, such as the spatial distribution of infected and susceptible individuals, social distancing, or population density.

In this thesis, I explore a dynamical model in which population response influences β [Dynamics ★], the impact of super-spreading events (heterogeneous β across population) and incubation time (i.e., time after which an infected individual becomes infectious) on the R_0 estimation [Super-spreaders ★], and the emergence of two new SARS-CoV-2 variants [Alpha ★, Omicron ★], including a two-strain model comparing the effects of increased transmissibility vs. immune evasion of the Omicron variant.

1.3. Viral infections in cell populations

The SIR model can also be adapted to study viral infections in cell populations. In this context, infected cells replicate viral RNA and produce viral proteins, which are essential for assembling new virions. Many viruses also encode non-structural proteins (not used for assembling virions) that disrupt the host immune response. Once assembled, virions are released and can infect neighbouring cells, or spread through the extracellular medium, infecting cells farther away. Other modes of infection are also possible; the respiratory syncytial virus (RSV), studied in this thesis, can infect neighbouring cells by inducing fusion of cellular membranes, creating syncytia with multiple nuclei. In response to non-self RNA, (some) infected cells produce and secrete interferons, which serve as a warning to neighboring cells, triggering transcription of interferon-stimulated genes (ISGs) encoding proteins that make cells more resistant to infection. In terms of the SIR model, the effect of (proteins of) ISGs can be compared to vaccination, as they reduce the effective infection rate, β . In this thesis, we also explore differential roles of type I and type III interferons, linking type I to ISG production and type III mainly to acceleration of cell death, which affects the removal rate γ [Lambda ★].

To model cellular dynamics at the single-cell level, compartmental models can be extended into agent-based models. In agent-based models, each cell is treated as an individual agent, and its state evolves according to a set of rules that govern stochastic state-to-state transitions. This approach allows for heterogeneity in cellular responses and modelling of spatial effects. A simple example of an agent-based model is a spatial version of the SIR model, where cells are arranged on a grid. Susceptible cells can become infected at a predefined rate if they neighbour infectious cells. Cytokine signalling can be incorporated by coupling the cell-agents to external partial differential equations (PDEs). Other approaches include the Cellular Potts model, in which a single cell occupies multiple nodes of a grid, and the stochastic evolution is governed by a global Hamiltonian [5], [6]. In this thesis, I constructed an agent-based model of RSV infection in a cell monolayer, which was constrained by experimental data from our laboratory. The model was used to dissect the mutual antagonism between virus and the innate immune response [Antagonism ★].

1.4. Modelling delays in compartmental models

A recurring issue addressed in this thesis is the challenge of accurately modelling delays in compartmental models. Consider a toy example in which exposed individuals (E) become infectious (I) after a certain time. In the classical SEIR model, this transition is modelled by the rule $E \rightarrow I$, with the rate α determining the expected time for an individual to transition from E to I. In this case, the *distribution* of transition times is exponential, (Figure 2, blue line), meaning that the most likely time for an exposed individual to become infectious is $t = 0$. However, in reality, humans can only become infectious after a minimum amount of time has passed since exposure (for COVID-19, the incubation period is rarely shorter than 2

days [7], [8]). Such a delay is a consequence of latent physiological processes which can be modelled as intermediate steps, $E_1 \rightarrow E_2 \rightarrow \dots \rightarrow E_k \rightarrow I$. When the rate for each transition is set to α/k , the resulting distribution of transition times (from E_1 to I) follows an Erlang distribution with mean $1/\alpha$, mode $(k-1)/(k\alpha)$, and variance $1/(k\alpha^2)$. For $k \rightarrow \infty$, the Erlang distribution converges to a Dirac δ distribution, corresponding to a fixed latency time of $1/\alpha$.

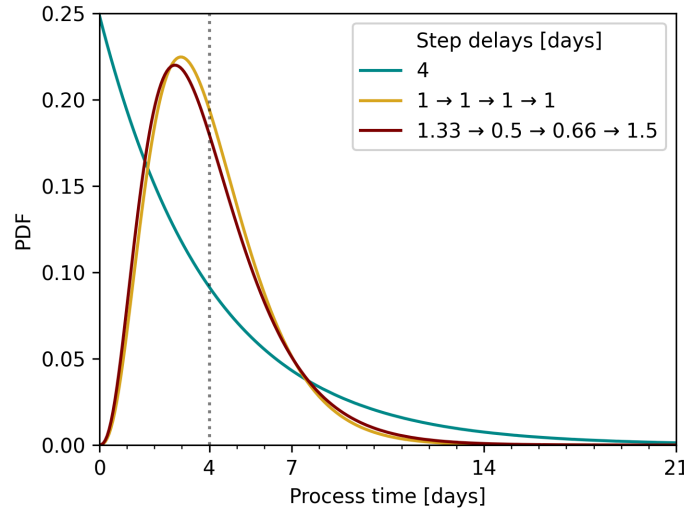


Figure 2. Distribution of completion times for a multistep process $E_1 \rightarrow \dots \rightarrow E_k \rightarrow I$, starting from E_1 . Different combinations of steps and expected transition times are shown. In each case, the expected total process time is 4 days.

We can introduce intermediate steps, even if we do not observe them directly, to achieve the correct distribution of transition times for the entire process. For example, if there are $k=4$ steps, each taking 1 day, the total process will take on average 4 days, with a mode of 3 days (Figure 2, yellow line). If we replace these intermediate steps with a single transition with an expected time of 4 days, the variance of the process time increases, and the most likely transition time will be zero. Alternatively, one could use delay differential equations with a fixed delay of 4 days, but this approach neglects the stochastic nature of the process and can easily lead to negative (unphysical) solutions. We use such intermediate steps in several articles included in this thesis [Antagonism ★, Dynamics ★, Super-spreaders ★].

1.5. Non-identifiability and information capacity of signalling pathways

Different sets of rates can lead to very similar distributions of process time (compare the yellow and red lines in Figure 2). On one hand, this justifies setting all intermediate rates to α/k when the goal is simply to model experimental delays. On the other hand, it has important consequences for model identifiability in situations where we are interested in the intermediate steps but cannot directly measure them. This is because increasing the rate of a given step can be largely compensated by decreasing the rate of another step. The mathematical consequence of this compensation is non-identifiability of model parameters. From a biological perspective, it has been suggested that such heterogeneity of parameters can be beneficial for populations of cells in terms of robustness and evolvability [9]. In this thesis, we explore the utility of such non-identifiable, dynamic models of signalling pathways [Predictive ★].

At the same time, non-identifiability also suggests that signalling pathways can be analysed without building and parameterizing explicit models. One way to approach biological data in this manner is through methods from information theory [10], [11], [12]. For example, Teng et al. [13] used hidden Markov models (HMMs) to analyze and quantify the information processed by cells responding to chemokines. In this thesis, I explore *in silico* the information rate achievable by the MAPK/ERK pathway [Limits ★]. This direction was later pursued by Paweł Nałęcz-Jawecki, who quantified the information transmitted in the MAPK/ERK pathway based on experiments [14] and analyzed, *in silico*, the information transmitted by waves of activity in a cell monolayer [Monolayer ☆].

Mutual information has several advantages over classical measures of interaction, such as Pearson or Spearman correlation. First, it is strictly positive for any pair of statistically dependent variables and exactly zero for independent variables. Second, it applies to both continuous variables (e.g., expression levels) and categorical variables (e.g. cytokine types). Third, it is invariant under continuous transformations of variables—log-transforming expression levels, for instance, does not affect the mutual information. Finally, in the context of systems biology, it quantifies how much information can be transmitted: if a signalling pathway relays 1 bit of information about the level of an extracellular cytokine, this can be interpreted as being sufficient to make a binary decision.

However, these properties come at a cost: numerically estimating mutual information from a finite sample is, in general, difficult. A useful perspective is to consider a finite sample of n points (X_i, Y_i) and an estimator E , which produces an estimate $E(\{X_i, Y_i\})$. Since mutual information is invariant to transformations, one might expect that applying transformations f, g to the variables should yield a similar result, i.e. $E(\{X_i, Y_i\}) \approx E(\{f(X_i), g(Y_i)\})$. However, it is clear that for sufficiently complicated functions f, g the estimator's performance will degrade. In an extreme case, if X and Y are valued in $\mathbb{R}^n$ with $n > 1$, continuous functions f, g can be chosen to transform any given set of n distinct points to any other set of n distinct points. While this does not mean estimation is infeasible in practical scenarios, it underscores that different estimators introduce implicit or explicit biases—much like curve fitting or selecting an appropriate deep learning architecture. This highlights the need for thorough empirical evaluation to assess the practical limitations and comparative advantages of different estimators.

To assess the performance of different mutual information estimators in practice, this thesis introduces a comprehensive benchmark covering a range of challenges, including high-dimensionality, long-tailed distributions, and cases with high mutual information [Beyond ★]. In a follow-up study, we further refine this benchmark by examining pointwise mutual information and mixture distributions, and we introduce a Bayesian estimator of mutual information [Mixtures ★].

2. Research hypotheses

1. **Hypothesis:** Cells use interferon signalling to slow down viral spread; in turn, viruses use non-structural proteins to interfere with interferon signalling. This mutual inhibition can lead to bistability at the single-cell level, in which certain cells employ interferon signalling to inform other cells about oncoming infection, while other cells are reprogrammed by the virus to produce new virions. Population averaging would mask this behavior, and thus single-cell experiments and modelling are necessary to verify this hypothesis.

Publication: *Antagonism between viral infection and innate immunity at the single-cell level.* [Antagonism ★]

2. **Hypothesis:** It is known that in human alveolar cells, prestimulation with IFN- β leads to a stronger antiviral response than prestimulation with IFN- λ 1. Surprisingly, from our own pilot experiments we learned that knockout of the IFN- λ 1 receptor IFNLR, but not knockout of the IFN- β receptor IFNAR, enhances the spread of influenza A virus (IAV). Since in IFNLR-deficient cells the increased number of IAV-infected cells (with respect to wild-type cells) did not lead to an increase in the number of cell deaths, we hypothesise that IFN- λ 1 promotes death in infected cells.

Publication: *Type III interferons suppress influenza A virus infection independently of STAT activation by triggering cell death.* [Lambda ★]

3. **Hypothesis:** For a basic reproduction number R_0 of 3, the classic SEIR model predicts a surge of COVID-19 cases, with over 30% of the population infected in the worst week. This estimate, a consequence of assuming a constant infection rate, β , appears implausible from a sociological perspective. We decided to investigate models in which people respond to high risk of infection and death by modifying their behavior (even in the absence of administratively imposed restrictions). We also consider a scenario in which, over time, individuals gradually accept the risk.

Publication: *Dynamics of COVID-19 pandemic at constant and time-dependent contact rates.* [Dynamics ★]

4. **Hypothesis:** The classical SEIR model assumes that the infection rate, β , is uniform across the population. However, it has been shown that social networks have a highly uneven distribution of connections, with a small number of people having a large number of contacts. In an epidemic, these people can become super-spreaders, significantly increasing R_0 . This effect would be hidden in the initial phase of the epidemic, i.e. before any super-spreading event occurred. We hypothesise that this could lead to an underestimation of R_0 , and explain the discrepancy between initial estimates of R_0 for COVID-19 (between 2 and 4) and the much faster spread of the epidemic in European countries (which suggested R_0 in the range of 4.7–11.4).

Publication: *Super-spreading events initiated the exponential growth phase of COVID-19 with R_0 higher than initially estimated.* [Super-spreaders ★]

5. **Hypothesis:** At the end of 2020 we observed that the SARS-CoV-2 variant Alpha has had a significant replicative advantage over previous strains and will likely spread towards global fixation. Accumulating mutations in variant Alpha sequences pointed to the emergence of potential future variants.
Publication: *SARS-CoV-2 Variant of Concern 202012/01 Has about Twofold Replicative Advantage and Acquires Concerning Mutations.* [Alpha ★]
6. **Hypothesis:** We estimated the initial date when the Omicron variant started to spread based on the number of accumulated mutations. This estimation suggested that Omicron spread with a very high reproduction number, R_0 , in the South Africa population in which most individuals have already been infected by previous variants. We thus hypothesised that the significant replicative advantage of the Omicron variant is associated with immune evasion.
Publication: *The spread of SARS-CoV-2 variant Omicron with a doubling time of 2.0–3.3 days can be explained by immune evasion.* [Omicron ★]
7. **Hypothesis:** As previously explored by Kočańczyk et al. [15], the MAPK pathway employs a combination of positive and negative feedback loops, which leads to pulsatile signals of ERK activity. This allows cells to coordinate using spatially organised waves of ERK activity during wound healing or cell movement. It was previously shown that only around 1 bit of information about the stimulus concentration is encoded in the level of ERK activity [16], [17]. We hypothesised that since the MAPK pathway processes information *continuously*, the quantity of interest is the information transmitted per unit of time, i.e. bitrate, and estimated it *in silico*.
Publication: *Limits to the rate of information transmission through the MAPK pathway.* [Limits ★]
8. **Hypothesis:** Even a single biological system can have dozens of different proteins interacting. We might be interested only in the "input", "output", or some key proteins, but in works included in this thesis (Dynamics ★, Antagonism ★), we saw that intermediate steps are crucial for replicating the system's dynamics. Inspired by previous work on sloppiness, we hypothesise that understanding the effective dynamics of a signalling pathway does not necessarily require measuring all variables. However, it might still be necessary to include processes associated with these variables in the model.
Publication: *Predictive power of non-identifiable models.* [Predictive ★]
9. **Hypothesis:** Information theory is a valuable tool for studying biological systems. However, comprehensive benchmarks of MI estimators were missing, especially comparing classical (for example, the KSG estimator [18]) to neural approaches (such as InfoNCE [19]). It is known that MI is invariant to homeomorphisms. We thus hypothesise that homeomorphisms can be used to transform pairs of random variables for which MI can be calculated analytically to obtain a variety of probability distributions, with known MI, which can be used for testing different estimators. To verify this hypothesis, we built a benchmark using information-preserving transformations of multivariate Gaussian and Student distributions.
Publication: *Beyond normal: On the evaluation of mutual information estimators.* [Beyond ★]

10. **Hypothesis:** We hypothesise that Pointwise Mutual Information (PMI), especially the distribution of PMI (PMI profile), can be a valuable tool for understanding mutual information estimation. In particular, we discovered that the PMI profile is preserved by transformations used in our previous benchmark, uncovering a limitation in our approach. We circumvent this issue by using mixture distributions, allowing us to study probability distributions with different PMI profiles.

Publication: *On the Properties and Estimation of Pointwise Mutual Information Profiles*. [Mixtures ★]

3. Summary of main results

3.1. Dissecting interactions between RSV and innate immunity

We hypothesised that during the infection of a cell monolayer, individual cells could follow qualitatively different transcriptional and translational programs due to the antagonism between viral proteins and innate immunity. To investigate this interaction, I constructed an agent-based, spatial, stochastic model that simulates RSV infection of a cell monolayer. The model was parameterised using a wide range of experimental data both at the single-cell and at the population level generated in our laboratory (immunofluorescence imaging, Western blotting, ELISA, dPCR).

The model scheme is depicted in Figure 3. A cell's state is defined by the activation step of each variable: v (infection status), $vRNA$ (viral RNA), $vProteins$ (viral proteins), $pIRF3$ (active, phosphorylated transcription factor IRF3 that regulates expression of interferon, IFN), $IFNi$ (intracellular interferon), $pSTATs$ (active phosphorylated transcription factors STAT1 and STAT2, activated by interferon and capable of triggering activation of interferon-stimulated genes, ISGs), and $ISGs$ (proteins encoded by ISGs). The activation steps change according to predefined rules, with possible steps of activation indicated as boxes. For example, an infected cell ($v = 1$) can transition towards producing viral RNA (i.e., the step of $vRNA$ can increase). Once $vRNA = 3$, RNA synthesis is complete, and the levels of $vProteins$ and $pIRF3$ can increase. The requirement for $vRNA$ to reach step 3 to initiate synthesis of viral proteins is indicated by the dark box. The intermediate steps are introduced to capture the experimentally observed $vRNA$ ‘incubation period’, as explained in the Background. Once $vProteins = 3$, neighbouring cells can become infected (v can change from 0 to 1 therein). Concurrently, when $pIRF3 = 3$, the cell starts producing (intracellular) interferon, $IFNi$. Once produced ($IFNi = 3$), interferon is secreted, increasing the amount of extracellular interferon ($IFNe$). Extracellular interferon diffuses (Figure 2B), reaching first neighboring cells, then more distant cells (diffusion is modelled by approximating a 3D process). Cells (including the secreting cell), respond to the presence of $IFNe$ by phosphorylation of STATs ($pSTATs$ increases), and, subsequently, production of (proteins of) interferon-stimulated genes ($ISGs$). Notably, partial activation of STATs can still activate $ISGs$ production, albeit at a slower rate (indicated by light grey boxes). Finally, the model incorporates several cross-inhibitory interactions. Viral proteins inhibit the activation of $pIRF3$, $IFNi$, and $pSTATs$; and $ISGs$ impede the production of $vRNA$ and $vProteins$.

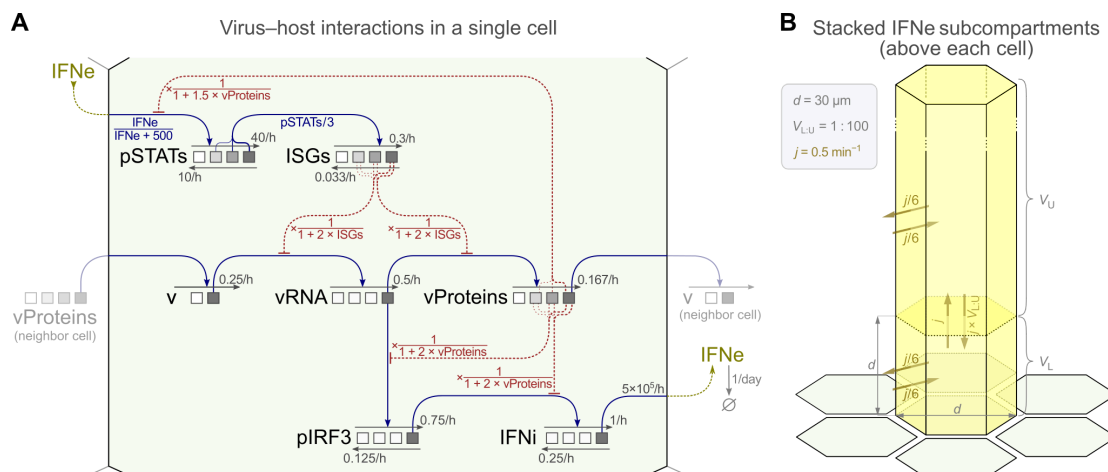


Figure 3. RSV infection of a cell monolayer, model scheme. **A.** The state of a cell evolves according to predefined rules. When a state reaches the level of activation indicated by grey boxes, it can enable (blue arrows) or inhibit (red hammer-headed arrows) other transitions. **B.** Diffusion of extracellular interferon is modelled assuming two compartments (lower and upper) above each cell. Cells secrete interferon into their respective lower compartments, but at later stages of the infection, interferon also accumulates in upper compartments. Reproduced from Antagonism ★.

After being constrained by experimental cell population-level data, the model successfully reproduces spatial patterns observed during viral infection (Figure 4), i.e. single-cell-level data. Importantly, cells producing viral proteins rarely produce interferon, and vice versa (panel A). This phenomenon arises because in cells producing viral proteins, non-structural viral proteins quickly block interferon production. Consequently, interferon is primarily produced by infected cells that detected viral RNA before the onset of viral protein production. In response to interferon, bystander cells produce ISGs, which in turn slow down accumulation of viral RNA and viral proteins, thereby extending the window of opportunity for interferon production. This interplay between virus and innate immunity would be hidden at the population level, where only the average level of ISGs and viral proteins is considered.

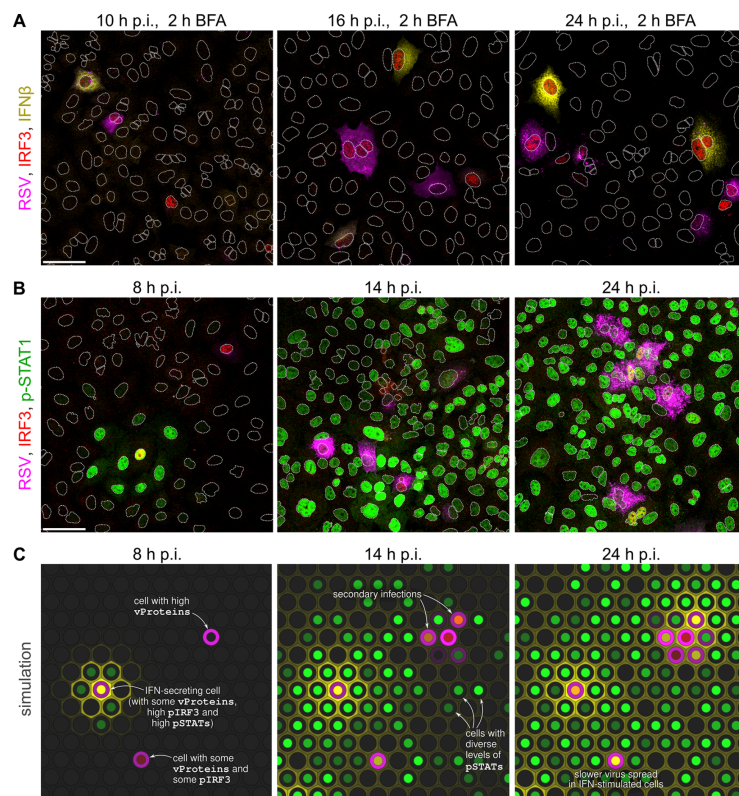


Figure 4. Comparison between immunostaining images (**A–B**) and model simulations (**C**). **A.** Cells in which non-self RNA was detected (IRF3 active, red) may start producing and, eventually, secreting interferon (yellow). **B.** Interferon leads to activation of STAT1 by phosphorylation (green) in neighbouring cells. At later stages of the infection, interferon accumulates and almost all non-infected cells are phosphorylated (active) STAT. **C.** Model simulations match the experimentally observed effects. Reproduced from Antagonism ★.

3.2. Differential role of type I and type III interferons

Both IFN- β and IFN- λ s are known to activate STAT1/2 signalling pathways, leading to the production of interferon-stimulated genes (ISGs). While a type I interferon, IFN- β , generally elicits a stronger antiviral response in respiratory epithelial cells, type III interferons, IFN- λ 1– λ 4, play a distinct and non-redundant role, although the underlying mechanisms remain unclear. In animal studies, Galani et al. [20] have shown that mice deficient in IFNLR1 (subunit of the receptor for type III IFNs) suffer higher mortality in response to influenza A virus infection.

Our experiments on A549 cells (human alveolar cell line) revealed further differences in these two interferons' roles. Prestimulation with IFN- β leads to greater suppression of influenza A virus (IAV) compared to prestimulation with IFN- λ 1. However, knock-out of IFNLR1 causes a more pronounced increase in viral spread than knock-out of IFNAR1 (subunit of the receptor for type I IFNs). Moreover, this phenomenon is specific to IAV: knock-out of IFNLR1 does not increase the spread of respiratory syncytial virus (RSV), which is known to effectively block apoptosis. Based on these findings, we hypothesized that IFN- λ s suppress IAV spread by inducing early cell death in infected cells, thereby limiting the number of secondary infections.

To test this hypothesis, the most straightforward approach would be to measure the number of cell deaths during IAV infection. However, no differences were observed in cell death rates between IFNLR1 knock-out and IFNAR1 knock-out conditions. To reconcile these observations, I employed a simple SIR model. In this model, IFN- β induces ISG production, reducing the infection rate β , while IFN- λ increases the removal rate, which we suggestively renamed λ (typically denoted as γ). An increase in λ reduces the basic reproduction number $R_0 = \beta / \lambda$, thereby limiting the overall spread of the infection. This constitutes an antiviral effect, even though it cannot be directly demonstrated by measuring cell death rates.

The model kinetics shows that the effect of increasing the removal rate λ depends on the time of measurement (Figure 5A). At different time points, an increase of λ can either lead to an increase or a decrease of the fraction of removed cells, indicating that in the analysis of experimental data, the amount of dying cells is not a reliable metric for detecting changes in λ . The fraction of infected cells changes monotonically with λ (Figure 5B), but is also affected by β (Figure 5C). This indicates that changes in the fraction of infected cells alone cannot conclusively indicate changes in λ . In contrast, the ratio of removed to infected cells (R/I) changes distinctly and monotonically with λ (Figure 5D), while being only weakly influenced by β (Figure 5E). This demonstrates that R/I is a reliable metric for experimentally testing our hypothesis.

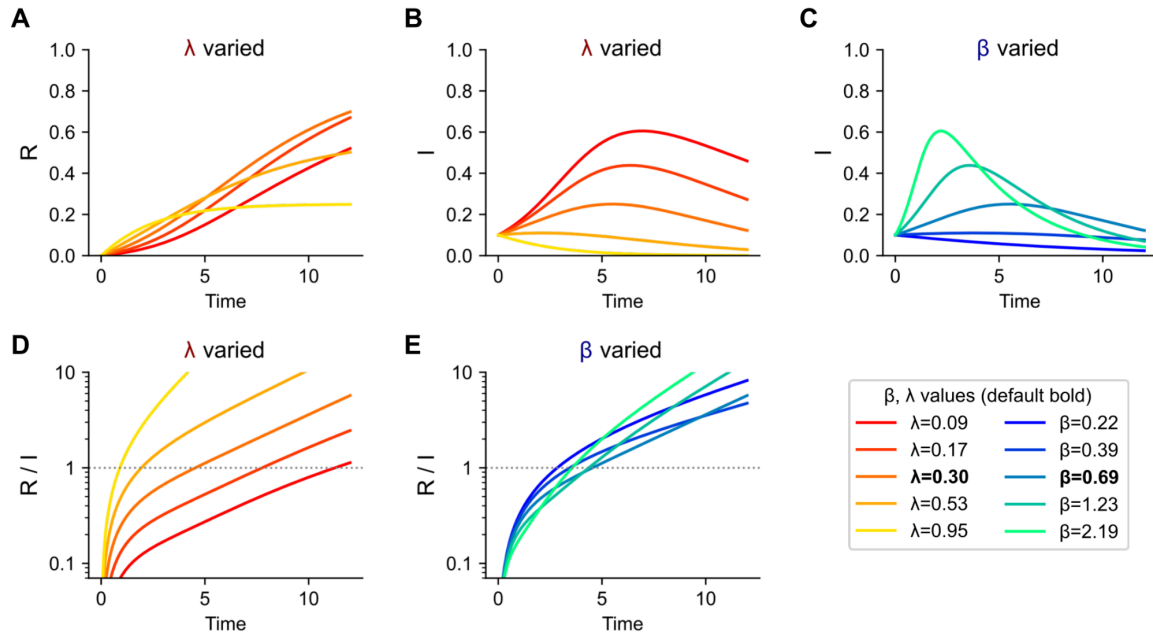


Figure 5. Fraction of removed, infected, and the ratio of removed to infected cells as a function of time, the removal rate λ and the contact rate β . Default parameter values are shown in bold. Reproduced from Lambda ★.

We performed paired measurements of viral protein levels (via Western blots) and cell death counts (via FACS analysis). Figure 6 presents the relative changes in IAV nucleoprotein (NP) and cell deaths in IFNLR1 KO compared to WT cells. While cell death levels remained relatively unchanged, there was an almost threefold increase in IAV NP levels in IFNLR1 KO cells. This observation supports the hypothesis that the absence of IFNLR1 reduces the coefficient λ , indicating that the lack of IFNLR1 leads to slower death of infected cells. However, the biological mechanism by which IFN- λ promotes selective cell death remains to be identified, representing a key limitation of this study.

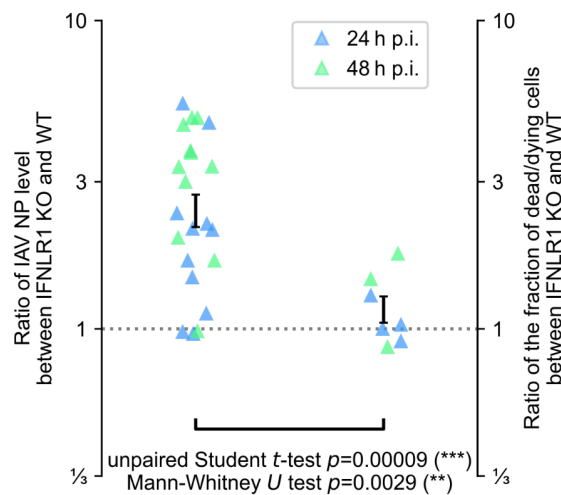


Figure 6. Relative accumulation of IAV nucleoprotein (NP, left column) and cell deaths (right column), shown as a ratio between IFNLR1 KO and WT A549 cells. Higher levels of IAV NP in IFNLR1 KO cells together with a small increase in cell deaths indicate an overall lower level of cell deaths per infected cell, compared to WT cells. Reproduced from Lambda ★.

3.3. Estimation of R_0 of COVID-19

The basic reproduction number, R_0 , is a critical parameter in understanding epidemic spread, dictating the extent to which the contact rate or fraction of susceptible individuals (through recovery or vaccinations) must be reduced to prevent further spread. Early estimates of R_0 for COVID-19 ranged between 2 and 4, suggesting that a fourfold reduction in contact rates would be sufficient to prevent epidemic growth. However, our estimates (based on data from China, Italy, Spain, France, UK, Germany, Switzerland and New York State) were significantly higher, in the range from 4.7 to 11.4. This discrepancy arises mainly from the assumed distribution of the incubation period and our estimates of the doubling time based on published data.

3.3.1. Incubation period

The incubation period of COVID-19 has been experimentally estimated to average around 5 days, with a range from 2 to 14 days [7], [8]. The SEIR model, on the other hand, assumes that the incubation period follows an exponential distribution (see Background). This assumption implies that the most likely incubation time is 0 days, which is inconsistent with observed data. To better capture the observed distribution of incubation periods, we modified the SEIR model by introducing k exposed states (Figure 7). This modification causes the incubation period to follow an Erlang distribution. By setting $k=5$ and the average incubation period $\tau=5$ days, the model produces a more realistic distribution: a low probability of infectiousness within the first 2 days and a negligible probability of incubation periods exceeding 2 weeks. Setting $k=1$ reverts to the standard SEIR model.

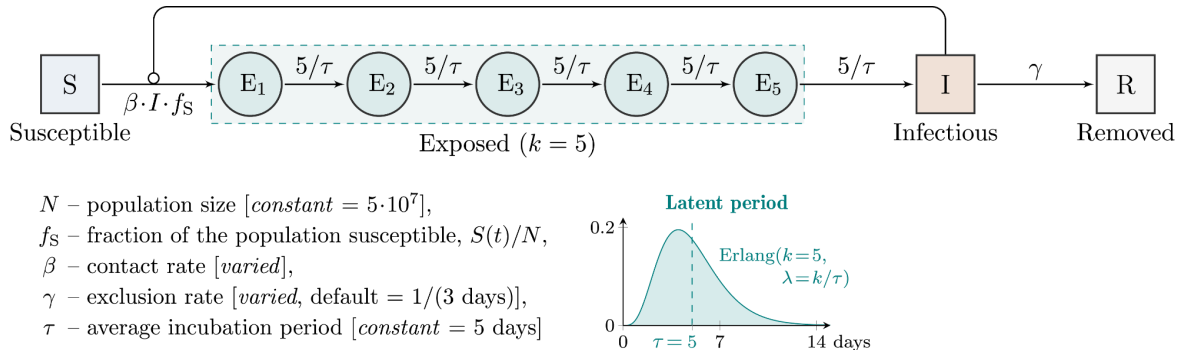


Figure 7. Modified SEIR model with multiple exposed states. The introduction of additional states leads to a realistic, Erlang distributed latent period. Reproduced from Dynamics ★.

This modification has significant implications for estimating the basic reproduction number R_0 . While the number of exposed states, k , does not directly influence the basic reproduction number (since $R_0 = \beta/\gamma$), it does affect the daily multiplication coefficient, θ , which can be directly estimated from the epidemic dynamics and used to calculate R_0 . The coefficient θ is defined as the day-to-day fractional increase in the number of new confirmed cases. For instance, if the number of new cases on a given day is 25% higher than the previous day, then $\theta = 1.25$. The daily multiplication coefficient can be computed using the following implicit formula, where $\alpha = \log \theta / \text{day}$ is the daily growth rate on a logarithmic scale:

$$\frac{\beta}{\gamma + \alpha} = \left(1 + \frac{\tau \alpha}{k}\right)^k.$$

The relationship between R_0 and θ for various values of k is shown in Figure 8A, with the corresponding distributions of incubation periods shown in Figure 8B. For the same daily multiplication coefficient, θ , a higher number of exposed states k leads to a higher R_0 . This has practical consequences for interpreting pandemic growth rates. For example, a daily multiplication coefficient of $\theta = 1.5$ could be explained by $R_0 = 6\text{--}7$ under the classic SEIR model ($k = 1$), or a much higher $R_0 = 10\text{--}11$ under our modified model with $k = 5$. Although the pandemic grows at a similar pace in both scenarios, the higher R_0 value in the modified model implies that contact rates must be reduced nearly twice as much to contain the spread. This in turn translates to more stringent public health restrictions.

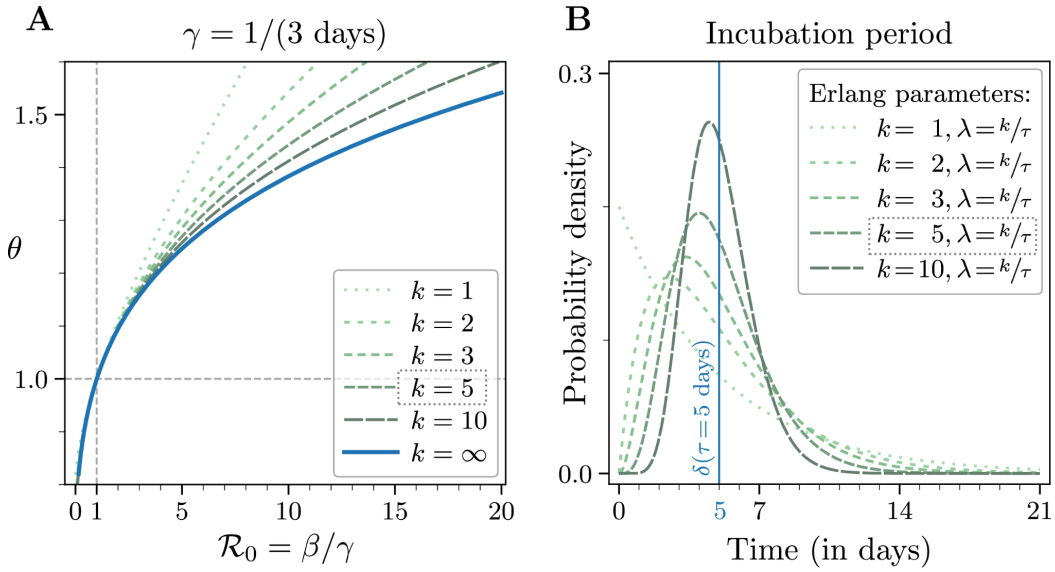


Figure 8. Influence of the number of exposed states k . **A.** Relationship between R_0 and the daily multiplication coefficient θ . **B.** Distribution of the incubation period. Higher values of k lead to a lower daily multiplication coefficient and a more narrowly distributed incubation period. The default value is $k = 5$. Reproduced from Dynamics ★.

3.3.2. Super-spreaders

Divergent estimates of R_0 can also be attributed to differences in the reported doubling times ($T_d = \log 2 / \log \theta$). Early studies of Wu *et al.* [21], [22] and Li *et al.* [23] reported doubling times of 5.2 days, 6.4 days and 7.4 days, respectively. In contrast, our estimates indicated much shorter doubling times, ranging from 1.94 to 2.68 days. These estimates were based on data collected during the two-week period following the point when the total number of cases exceeded 100 in six European countries, China and New York State.

We chose this threshold because super-spreading events, which are critical drivers of epidemic spread, become more prominent as the number of infections increases. In the very early phase of the pandemic, when fewer than 100 cases are reported, the probability of observing such events is low, potentially leading to underestimation of the doubling time. By including data after the 100th-case mark, our method captures the effect of super-spreading events, providing a more accurate estimate of doubling times.

An extreme example of super-spreading effects is illustrated in Figure 9, where 1% of the population accounts for two-thirds of all infections (hyper-spreaders). We ran multiple simulations of the onset of the epidemic, beginning with a single infected (but not

hyper-spreading) individual, and estimated the doubling times using two methods: (1) a 14-day window starting from 100 total cases and (2) a 30-day window starting from the first case. As shown in the accompanying histograms, the second method overestimates the doubling time on average by 50%, while the first method yields more accurate results on average and exhibits smaller standard deviations. In reality, the difference would likely not be as drastic, although Endo et al. [24] estimated that 80% of secondary transmissions could have been caused by as little as 10% of infectious individuals.

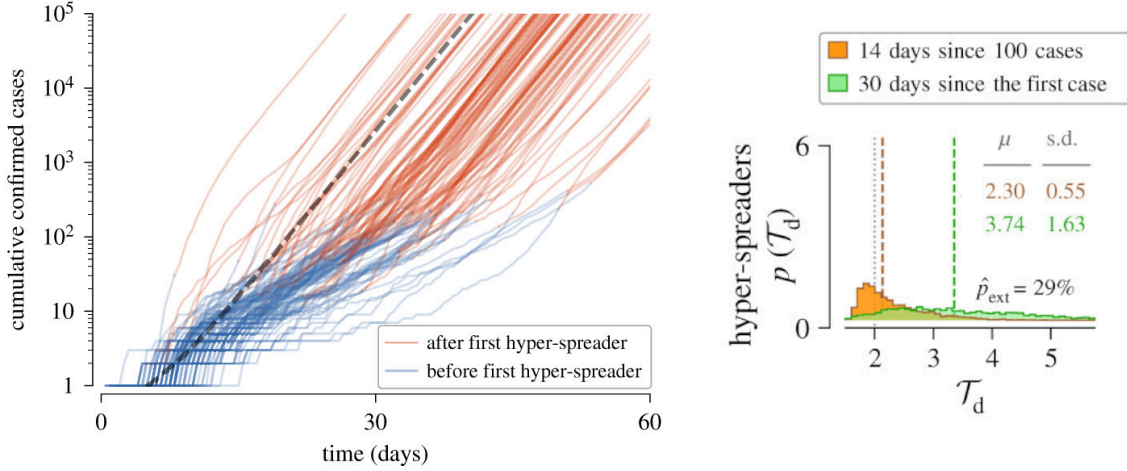


Figure 9. Left: Stochastic simulations of the system with 1% hyper-spreaders responsible for two-thirds of all infections (100 trajectories are shown), starting from one non-hyper-spreading infected individual. Dashed line shows the trajectory of the corresponding deterministic model. Right: Distribution of estimates of the doubling time T_d based on two schemes is shown. Reproduced from Super-spreaders ★.

Our estimates of R_0 (4.7–11.4) are significantly higher than the initial estimates of 2–4. These values can be compared to the measures introduced that eventually halted the exponential growth of the epidemic. By May 2020 (at the time of our initial submission), Italy and Spain successfully ended the initial exponential growth phase through nationwide quarantine measures. During this period, Google's COVID-19 Community Mobility Reports indicated an approximately 80% reduction in mobility across retail and recreation, transit stations, and workplaces in these countries. Together with increased social distancing, this reduction possibly lowered the infection rate β at least fivefold. Additionally, widespread testing and isolation of individuals who tested positive reduced the effective infectious period, $1/\gamma$. Consequently, we suspect that the effective reproduction number was reduced more than fivefold, resulting in values smaller than 1 and containment of the epidemic's spread. This suggests that the basic reproduction number R_0 in these countries was larger than 5, consistent with our estimates.

3.4. Population response to the SARS-CoV-2 epidemic

First, we explored the long-term evolution of the epidemic under the assumption that the basic reproduction number does not change over time (i.e. the epidemic is essentially ignored by the population). In this scenario, the outcome depends solely on the daily multiplication coefficient, θ . For a population of 50 million people starting with 500 confirmed cases, the maximum number of daily new cases is 0.3 million for $\theta = 1.05$ (corresponding to

a doubling time $T_d \approx 2$ weeks) and reaches 1 million for $\theta = 1.1$ ($T_d \approx 1$ week), as shown in Figure 10A. Figure 10B depicts the number of new cases in the worst month or week of the epidemic, alongside the time at which the peak occurs. For the still conservative estimate, $\theta = 1.1$, we find that, in the worst month, nearly half of the population would be infected.

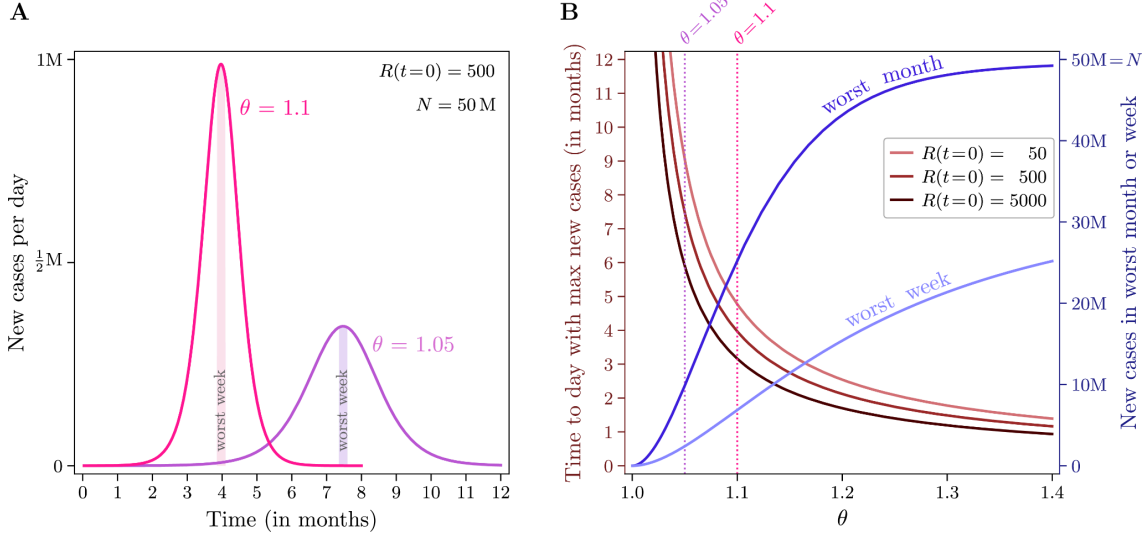


Figure 10. **A.** Simulations of the epidemic for two values of the daily multiplication coefficient: $\theta = 1.05$ (corresponding to a doubling time $T_d \approx 2$ weeks) and $\theta = 1.1$ ($T_d \approx 1$ week). **B.** Number of new cases in the worst week/month of the epidemic, and the time to the day with the highest number of new cases, both as a function of θ . Reproduced from Dynamics ★.

However, this scenario is unrealistic. With such high daily case numbers, it is reasonable to assume that people will reduce their contact rates voluntarily, for example, by increasing physical distancing. To account for such time-dependent dynamics, we introduce a relationship for the contact rate, β , that depends on the number of daily new cases:

$$\beta = \beta_0 H / (H + (\text{daily new cases})) .$$

Here, the parameter H controls the extent of voluntary isolation: as daily new cases rise, the contact rate decreases. When the number of daily cases reaches H , the contact rate is halved. For $H = 10,000$, the number of daily cases grows rapidly until it peaks at around 15,000 cases, at which point it starts to gradually decrease, over multiple years, as the fraction of susceptible individuals is slowly depleted (Figure 11A). However, this scenario would require the population to reduce contact rates for extended periods of time. In Figure 11B, we introduce a further assumption: people gradually become more negligent, accepting the risk of infection as time passes. To reflect this, we allow H to vary over time:

$$H(t) = H_0 (t + T_{\text{negl}}) / T_{\text{negl}} .$$

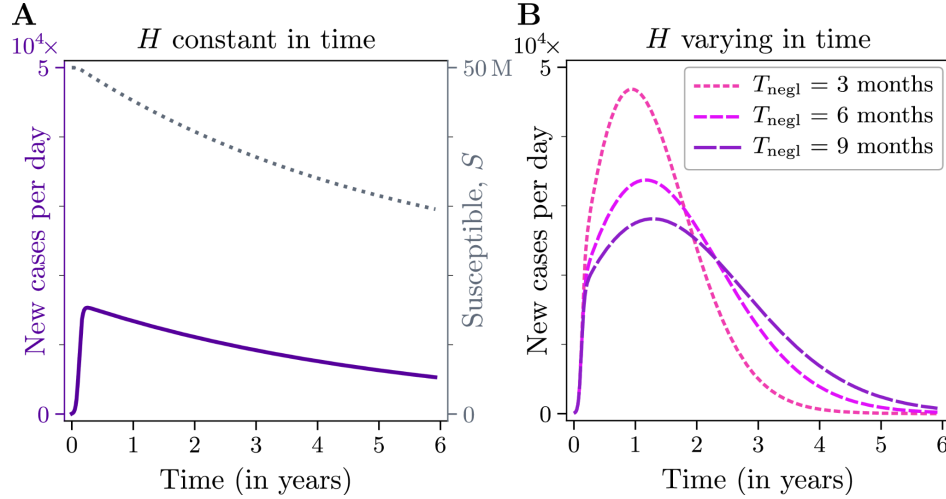


Figure 11. Dynamics of the epidemic with the population responding to high numbers of new cases by decreasing the contact rate **(A)** and, additionally, by slowly accepting the risk **(B)**. Reproduced from Dynamics ★.

This final modification results in a broader peak of 30,000 to 50,000 cases, but a shorter epidemic than in Figure 11A. Although this model is simplistic and based on arguably arbitrary assumptions about the population's response, it goes beyond the very unrealistic scenario where no population-level response occurs. Furthermore, despite its simplifications, the model's predictions were surprisingly accurate in retrospect. The maximum number of new cases predicted in Figure 11B matches historical data from Poland: the first three waves of COVID-19 saw approximately 30,000 daily new cases, while the fourth wave peaked at nearly 60,000 cases on the worst day.

However, while the model successfully captured the scale of these peaks, it did not predict multiple waves, largely driven by the emergence of new SARS-CoV-2 variants.

3.5. Emergence of new SARS-CoV-2 variants

Mutations in the virus can increase its fitness, typically by enhancing infectivity or immune evasion—thereby increasing the likelihood of infecting individuals who have recovered or been vaccinated. Mutations can also lead to outcomes such as decreased virulence [25], lowering the risk of host death, isolation, or treatment. The large number of individuals infected with SARS-CoV-2 provided ample opportunities for mutations, leading to the emergence of new strains. In our work, we analysed the emergence of the Alpha and Omicron variants. These publications have been widely cited and reported, including by *BBC News* [26], *News Medical* [27] and others [28], [29].

3.5.1. Variant Alpha

The first known sample of the SARS-CoV-2 variant Alpha, also referred to as PANGO lineage B.1.1.7 or initially as Variant of Concern (VOC) 202012/01, was collected on September 20, 2020, in Kent, United Kingdom. It began spreading rapidly and, by January 2021, accounted for 86% of all genomes sequenced in England. The rise of this new variant coincided with a sharp increase in total cases in Kent and Greater London.

In Figure 12A, we categorized sequenced genomes into four groups: variant Alpha (labelled VOC), 20A.EU1 (the previous dominant strain), other genomes carrying the D614G mutation, and all remaining genomes with D614 (both variant Alpha and 20A.EU1 carry the D614G mutation). As new, more fit variants emerge, their share of sequenced genomes grows until they reach fixation or are overtaken by the next dominant variant. To estimate the replicative advantage of variant Alpha over 20A.EU1, we analyzed the ratios of their respective sequence counts (Figure 12B). We calculated the weekly growth rate of these ratios, ν , by fitting two trend lines. Using a 9-week window, we obtained a lower estimate of $\nu = 1.88$ [95% CI: 1.75–2.01], while a 5-week window yielded a higher estimate of $\nu = 2.24$ [95% CI: 2.03–2.48]. Assuming both variants shared the same serial interval of 6.73 days, we computed the ratio of their reproduction numbers as $R_{\text{Alpha}}/R_{20\text{A.EU1}} = \nu^{6.73/7}$. This resulted in a lower estimate of 1.83 [95% CI: 1.71–1.96] for the 9-week window and a higher estimate of 2.18 [95% CI: 1.97–2.40] for the 5-week window.

This approximately two-fold increase in the reproduction number explains the rapid fixation of variant Alpha in Kent, as well as the increase in the total number of new cases.

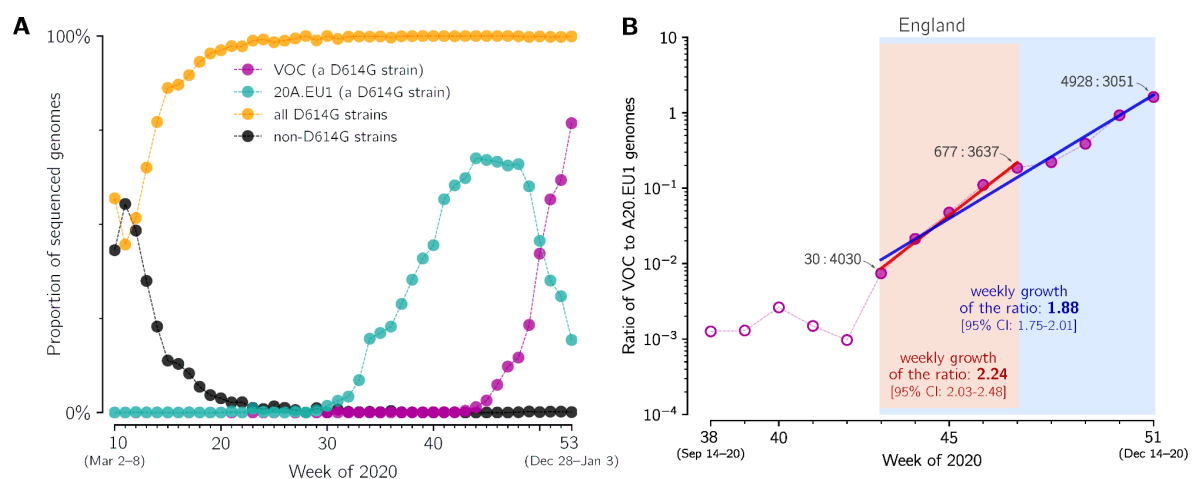


Figure 12. A. History of dominant variants of SARS-CoV-2 in England, as of Jan 3, 2021. **B.** Replicative advantage of variant Alpha over A20.EU1, the previous dominant strain in England. Reproduced from Alpha ★.

3.5.2 Variant Omicron

The SARS-CoV-2 variant Omicron was first identified in Gauteng, South Africa. Figure 13 shows the estimated case numbers for Omicron, Delta, and other pre-existing variants in Gauteng and the neighboring KwaZulu-Natal province. After a wave of Delta infections in both regions, which effectively displaced all pre-existing variants, case numbers dropped to nearly zero by mid-October (panel A). Notably, this decline occurred without significant reductions in population mobility at workplaces or retail and recreation centers (panel B), suggesting that these populations had effectively reached herd immunity.

However, the emergence of Omicron triggered a new wave of infections with a doubling time of approximately three days (panel C). This rapid spread suggests that Omicron effectively evades immunity acquired from previous infections, including variant Delta. Additionally, the rapid growth of Omicron cases in countries with high vaccination rates further suggests evasion of immunity provided by vaccination (see Figures 2–3 in Omicron ★). Immune

evasion can pose a greater threat than increased infectivity, as it expands the pool of susceptible individuals, potentially leading to larger outbreaks (see Figure 4 in Omicron ★).

The timing of the initial Omicron spreading event can be estimated by extrapolating the exponential growth trend of estimated Omicron cases (panel C). However, this method can be unreliable due to high stochasticity when case numbers are low and few Omicron genomes are sequenced. To independently verify our estimate, we analyzed the cumulative number of mutations present in sequenced genomes (panel D). Crucially, this approach does not rely on consistent genome sequencing or timely infection reporting. Instead, assuming that the virus accumulates random mutations at an approximately constant rate, we performed a Bayesian analysis to determine the most likely date when Omicron sub-variants began to diverge, pinpointing the event that initiated its spread in Gauteng. Results from this method align with the simpler estimate based on growth trends (panel C), further supporting our estimate of the doubling time.

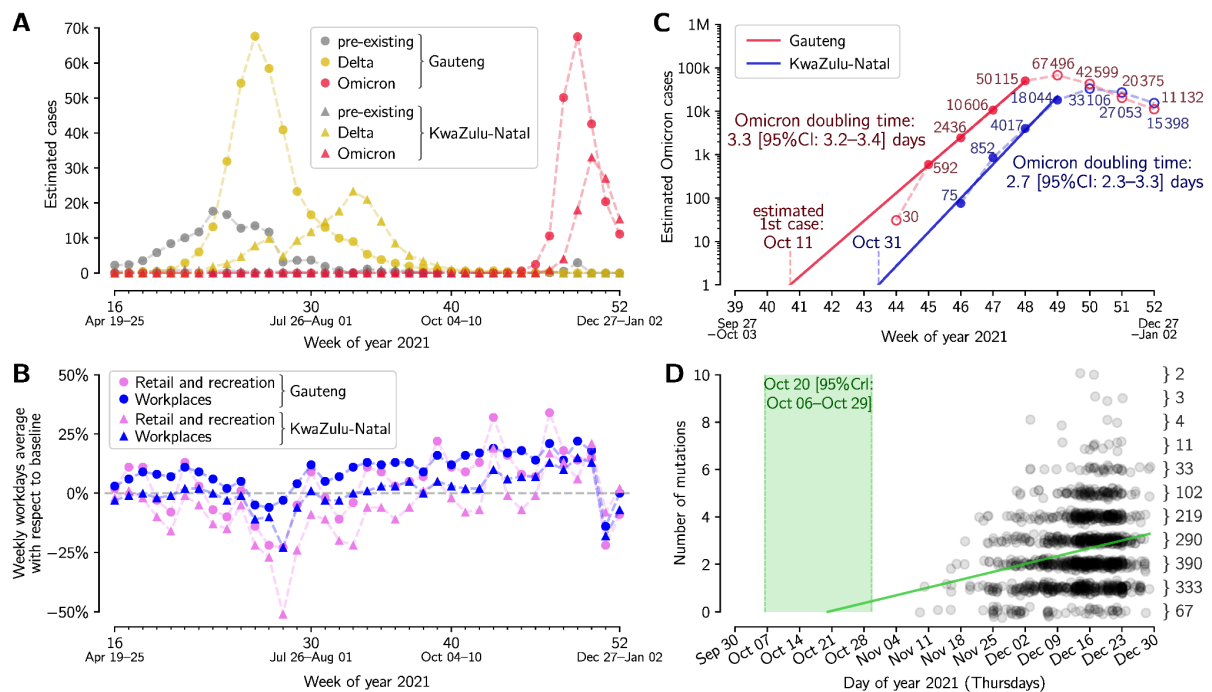


Figure 13. **A.** Timeline of dominant SARS-CoV-2 variants in Gauteng and KwaZulu-Natal provinces, South Africa, as of Jan 2, 2022. **B.** Mobility in retail and recreation, as well as workspaces, in the affected provinces. **C.** Doubling time of Omicron, based on the estimated number of new cases. **D.** Bayesian estimation of the initial spreading event date, based on accumulation of mutations in sequenced Omicron genomes. Each circle corresponds to a sequenced genome; the number of sequenced genomes for each total number of mutations is given on the right. Reproduced from Omicron ★.

3.6. Applicability of partially-constrained models to signalling pathways

A common motif between the multi-state SEIR models used to model the COVID-19 epidemic and the agent-based model of RSV presented earlier is the need to incorporate intermediate transition states. In the case of COVID-19, these additional states ensure that the incubation time and serial interval align with reported data. For RSV, the intermediate states are necessary to account for experimentally measured incubation periods—from virion entry to the production of viral RNA or proteins—as well as delays in protein signalling and transcription, such as the time from phosphorylation of IRF3 to secretion of IFN- β . These intermediate steps introduce additional model parameters, which we have previously avoided by assuming that the rates of all intermediate steps are equal. However, this assumption might not always hold, and we might not always know how many steps should be introduced. These challenges point to a deeper issue: how can we model a large, interconnected system when we are only able to measure a few variables?

Previously, we tackled this problem by starting with a simple model and gradually increasing its complexity until we were satisfied with the agreement between the model and the observed data. An alternative approach involves so-called "sloppy models" [30], [31]. The core idea behind sloppy models is that biological systems often contain many redundant counterbalances. As a result, different sets of parameters can lead to very similar system behaviour. For example, if there is insufficient protein A, the effect can often be compensated by increasing levels of a different protein, B. In fact, individual cells having different "sets of parameters" can enhance the cell population's robustness [9].

We investigated the applicability of this framework to dynamical models in systems biology by studying a simple signalling cascade (of four kinases, K_i) with a negative feedback (structurally similar to, for example, the MAPK or NF- κ B pathways). The model scheme and governing equations are shown in Figure 14a,b. Since we are primarily interested in multiplicative changes, we always transform the resulting trajectories using $\text{softlog}(x) = \log_{10}(x + 0.001)$. To generate synthetic training data, we simulated the true system with default parameters and a rectangular input signal, S_{train} (corresponding to adding an activatory cytokine at $t=1$ and washing it out at $t=5$). Then, we simulated measurement errors by adding gaussian noise to the (softlog transformed) true trajectories of kinases K_i . We generate three measurements for each K_i at time points $t=1, 2, \dots, 12$ (Figure 14c). We then use the Markov Chain Monte Carlo (MCMC) algorithm to generate sets of parameters that fit these synthetic measurements.

First, we allowed the MCMC algorithm to use only measurements of K_4 . The resulting parameter sets predict the response of K_4 to a different input signal, S_{test} , with a very small margin of error (Figure 14d). Meanwhile, the levels of activation of K_1 , K_2 and K_3 differ significantly between parameter sets, implying that we cannot reliably infer their dynamics based solely on the available measurements and model constraints. Similarly, additionally including measurements of K_2 (Figure 14e) allows for prediction of K_2 's response to the test signal, but does not improve predictions for the remaining steps, K_1 and K_3 .

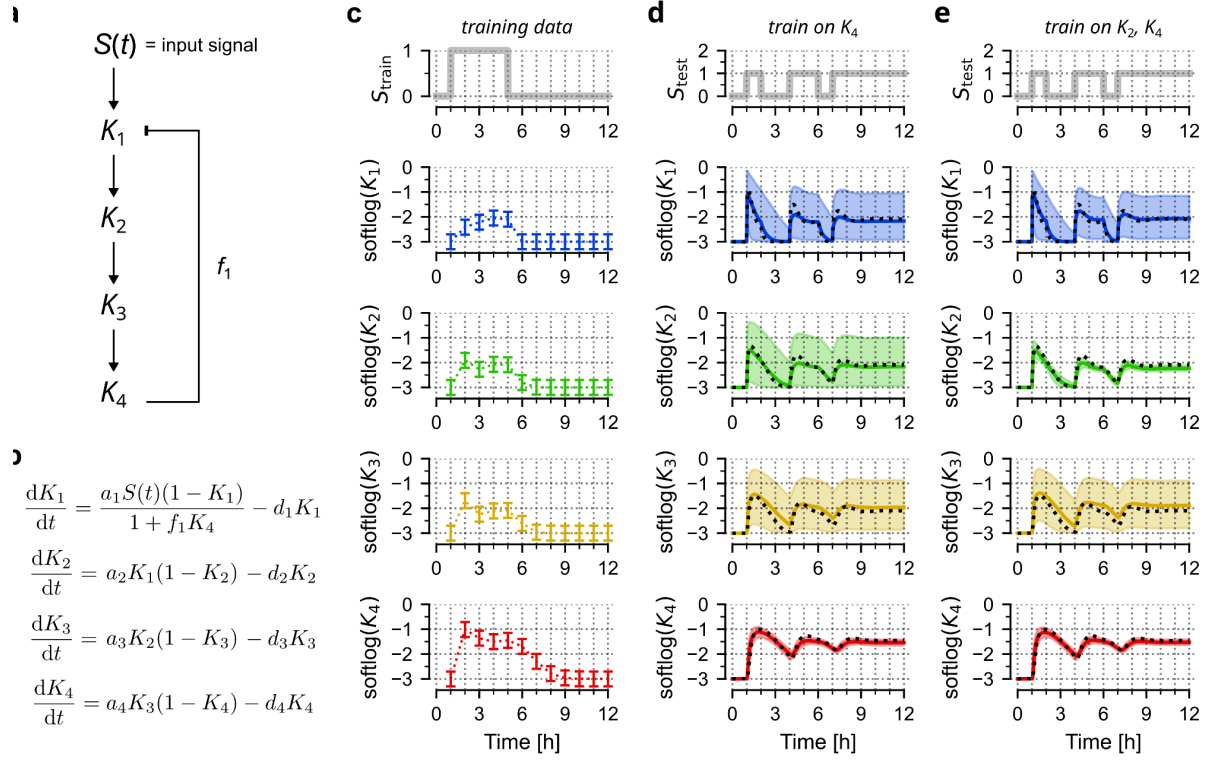


Figure 14. Toy model of a signalling cascade with negative feedback. **(a, b).** Model scheme and equations. **(c).** Simulated measurements in response to the signal S_{train} . **(d, e).** Model predictions in response to a different signal, S_{test} , after training on K_4 **(d)** and K_2 and K_4 **(e)**. Reproduced from Predictive ★.

This behavior can be understood in terms of *sloppy* and *stiff* directions in the parameter space. A change in parameters along a "sloppy" direction does not significantly alter the behavior of measured system variables, while a change in a "stiff" direction does. For instance, increasing the rate a_2 two-fold can be countered by a two-fold increase in d_2 , leaving the overall dynamics of the system largely unchanged. Thus, the direction of "increasing the activation rate of i and deactivation rate of i " is "sloppy". In contrast, the direction of "increasing all activation rates" is stiff, as it strongly affects the system's dynamics. Measuring only K_4 allowed us to approximate the overall delay of the system with significant accuracy but provides little insight into the source of this delay. This results in only one stiff direction after measuring K_4 .

The distribution of parameters after successively including measurements from additional steps is shown in Figure 15a. Notably, even after measuring all four steps—allowing for accurate predictions of the entire system—the distribution of some parameters remains broad. We quantify this uncertainty in parameter estimation by plotting the principal multiplicative deviations, $\delta_i = \exp(\sqrt{\lambda_i})$, where λ_i are the principal values of the distribution of log-transformed parameters (Figure 15b). We define stiff dimensions as those with $\delta_i \leq 1.5$, meaning that parameters obtained from MCMC vary in these directions by at most a multiplicative factor of 1.5. When only K_4 is measured, a single direction is stiff, and this is sufficient for accurate predictions of K_4 in response to previously unseen input signals. The number of stiff dimensions increases to two when K_2 is also included, and to four when all steps K_i are measured. To better visualize these results, we show optimised sparse bases spanning these "stiff subspaces" in Figure 16 (details in Predictive ★).

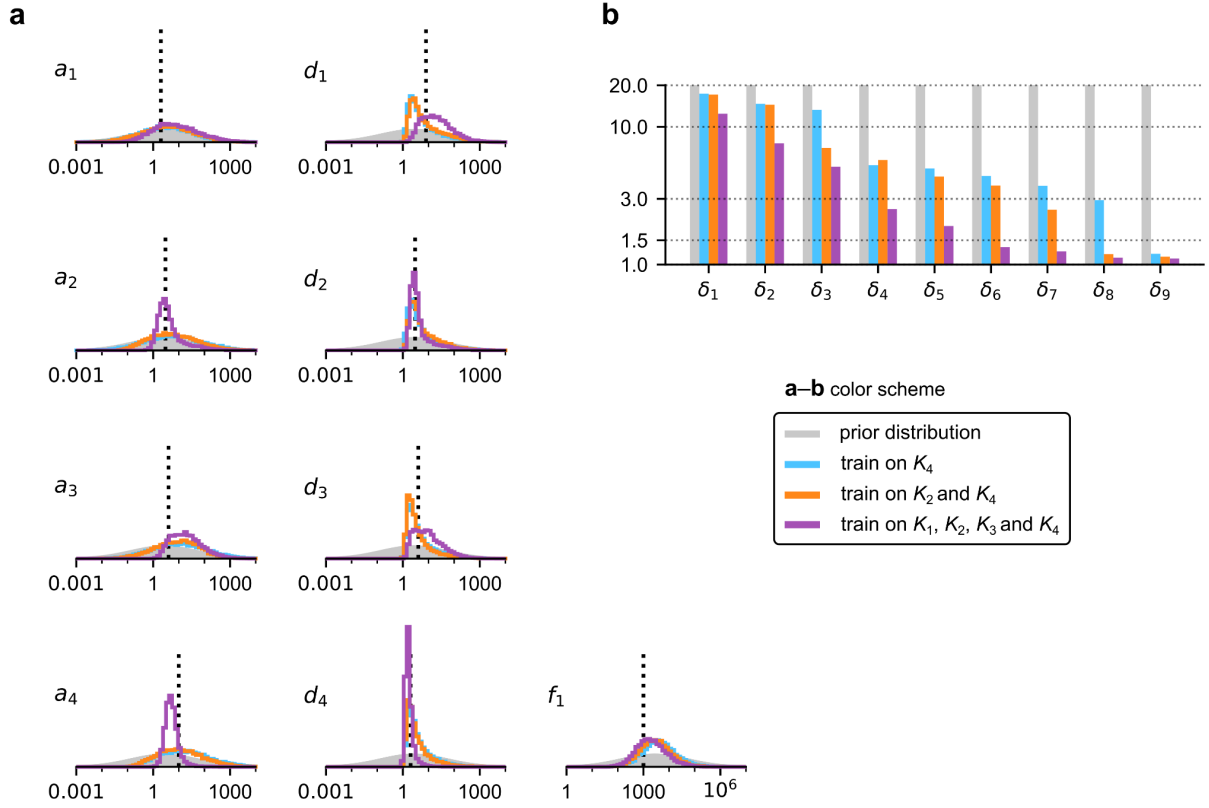


Figure 15. Successive inclusion of steps K_i in training. **a.** Histograms of parameters derived from MCMC simulations. **b.** Principal multiplicative deviations, $\delta_i = \exp(\sqrt{\lambda_i})$, where λ_i s are the principal values of the distribution of log-parameters. Reproduced from Predictive ★.

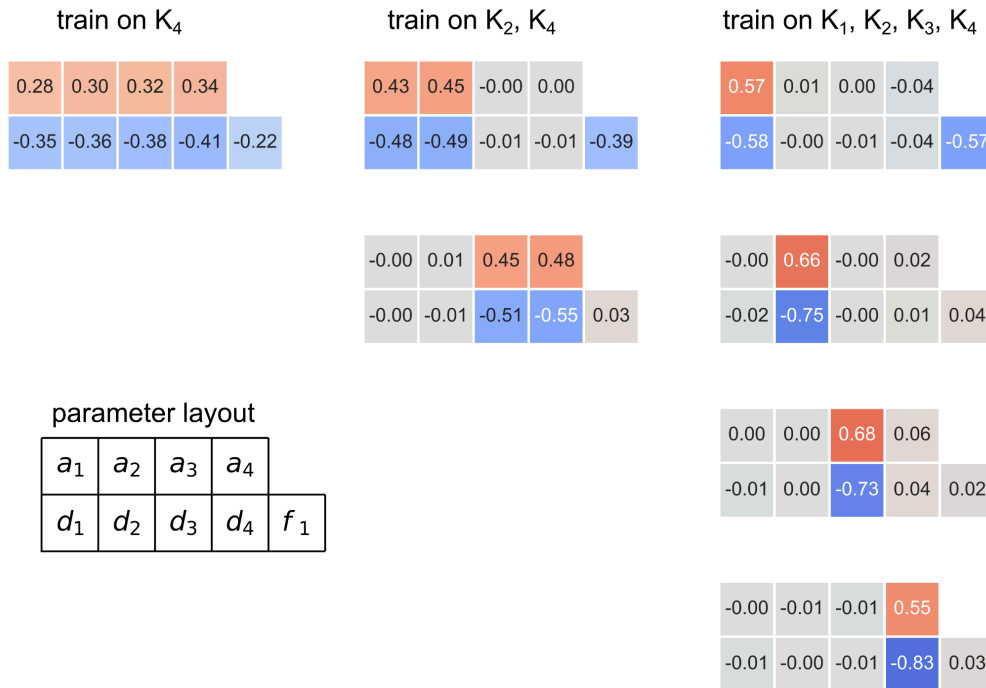


Figure 16. Normal vectors in log-parameter space spanning the same "stiff subspace" as the stiff eigenvectors (with $\delta_i \leq 1.5$ in the previous Figure), optimised to minimise the grey entries. Reproduced from Predictive ★.

Even though these models are only partially constrained, they still exhibit significant predictive power. The model trained solely on K_4 can predict the effect of a multiplicative change in parameters, even though the parameters are not constrained by the available measurements (Figure 17). Such multiplicative changes are commonly explored in systems biology, as they can be easily induced through techniques like gene knockdown or protein overexpression.

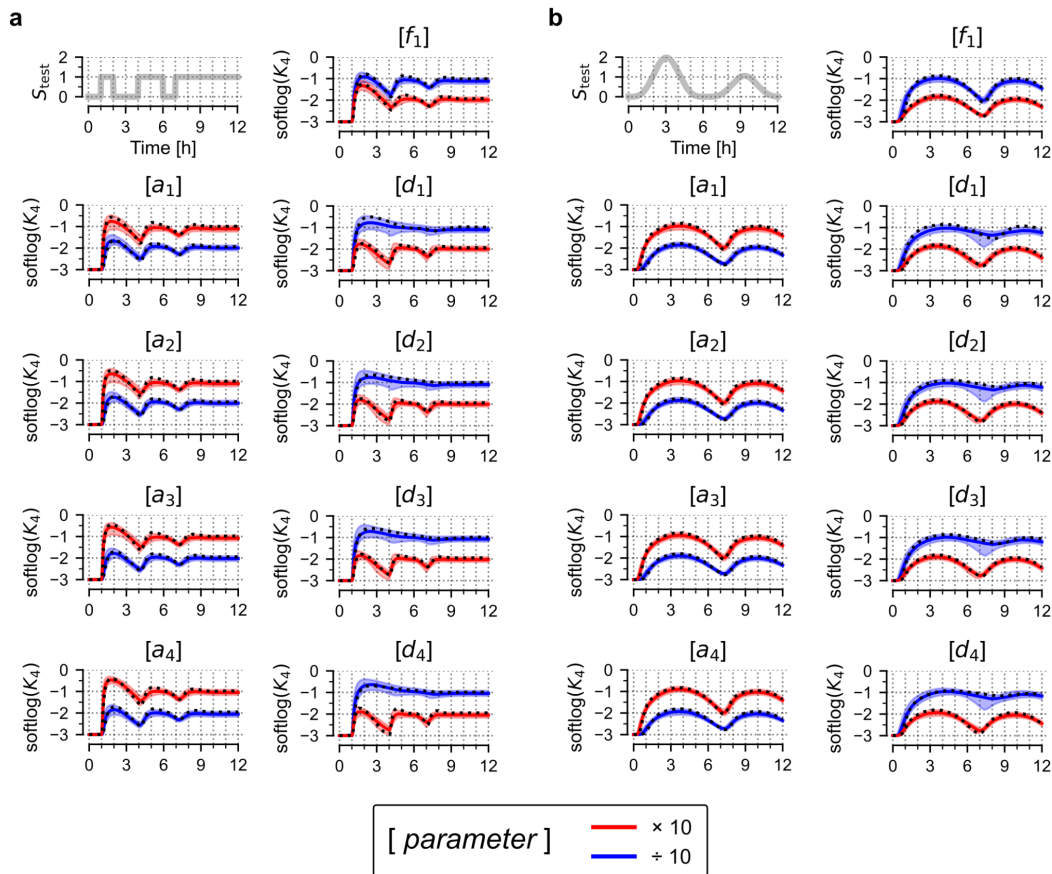


Figure 17. Model predictions of the trajectory of K_4 for 10-fold changes of parameters. Model trained only on K_4 . The changed parameter is shown in square brackets above each subpanel. Reproduced from Predictive ★.

In summary, the proposed approach allows for more interactive work with biological models [32]. Researchers can start with a general representation of the studied system and focus their measurements on the variables of interest. At any stage, the parameter sets generated by MCMC can be examined to identify which parameters or predictions are well-constrained by the experimental data. If necessary, additional measurements can be taken to further refine the set of permissible parameters.

A notable practical limitation of this approach is that MCMC can be computationally expensive for large models. However, techniques like Hamiltonian Monte Carlo [33], [34], No-U-Turn Sampler [35], stochastic variational inference [36], or simulation-based inference [37], are making sampling more efficient. For larger models, sloppiness can also be analyzed locally, using tools like the Fisher information matrix [38], [39]. Importantly, this computational bottleneck does not undermine the broader insight: biological systems often exhibit inherent redundancies, allowing accurate predictions even without fully constrained parameters.

3.7. Numerical study of MAPK/ERK channel capacity

Signalling pathways can also be investigated in terms of their information processing capacity. Kočańczyk et al. [15] demonstrated that by employing a combination of positive and negative feedback loops, the MAPK pathway is well-suited to process pulsatile signals of ERK activity. This enables cells to coordinate their movements using spatially localised waves of ERK activity during e.g. wound healing. While previous studies showed that the level of ERK activation can transmit only about 1 bit of information about stimulus concentration [16], [17], we quantify (*in silico*) the amount of information that can be transmitted through the MAPK pathway per unit of time, which we refer to as the *bitrate*.

We measured the bitrate by analyzing the system's (stochastic) responses to binary sequences of pulses. A key limitation is the pathway's refractory time—if two pulses are sent too closely together, the second pulse will be ignored. Consequently, an important parameter is the base repeating time, or the delay between subsequent pulse slots. Figure 18 shows all possible binary sequences of length 3, sent with different base repeating times. When the base repeating time is set to an hour, the pathway processes all pulses effectively. However, when the base repeating time is decreased to just half an hour, consecutive pulses are ignored. We group sequences that lead to similar responses in red frames. Importantly, *decreasing* the base repeating time *increases* the information throughput.

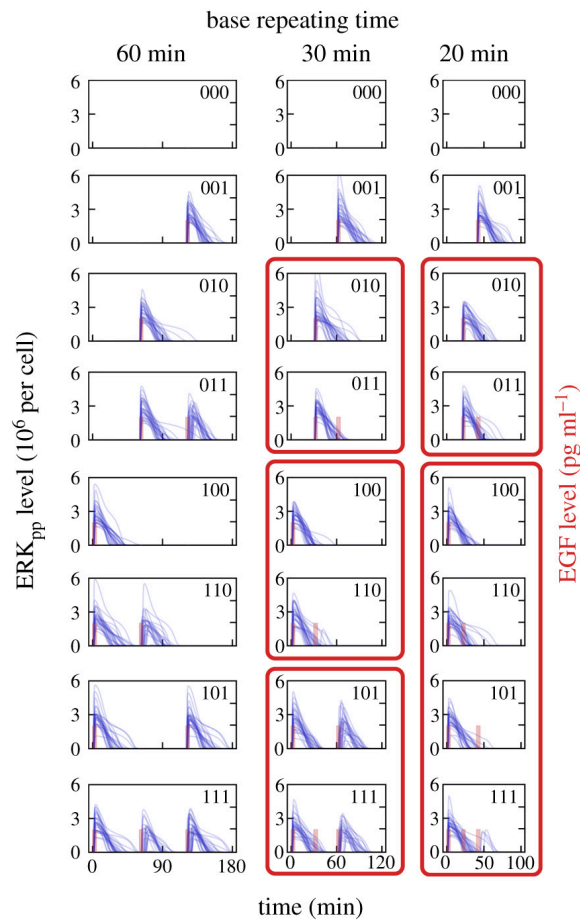


Figure 18. Simulated trajectories of the MAPK model in response to all possible binary sequences of length 3, with three different base repeating times. Sequences that elicit similar responses are grouped within red boundaries. Reproduced from Limits ★.

We can understand this by imagining that the system is able to perfectly distinguish between the red groups. In this case, with a base repeating time of 60 minutes, we can correctly recognise all pulses and thus transmit 1 bit per hour. When the base repeating time is reduced to 30 minutes, some pulses overlap, and only 6 distinct groups can be recognised. Consequently, we can transmit up to $\log_2(6)$ bits every 3×30 minutes, or approximately 1.7 bits per hour. With a base repeating time of 20 minutes, the bitrate increases further to $\log_2(4)$ bits per 60 minutes, or 2 bits per hour.

In reality, the distinction between merged inputs is not perfect. Figure 19 shows the system's response to four different inputs (00, 01, 10, and 11) for various base repeating times, measured as the log-integral of the ERK activity in response to the first and second halves of the stimulation. For a base repeating time of 60 minutes, the responses can be clearly distinguished. However, as the base repeating time decreases, the response to input 11 (red) begins to overlap with the response to input 10 (blue). The mutual information was computed using the Kraskov–Stögbauer–Grassberger (KSG) estimator [18]. By optimizing the probabilities of input sequences, we were able to maximise the channel's capacity. Panel B compares the results for equiprobable and optimised input probabilities, alongside numerical estimates and our theoretical approximation (see Limits ★ for details).

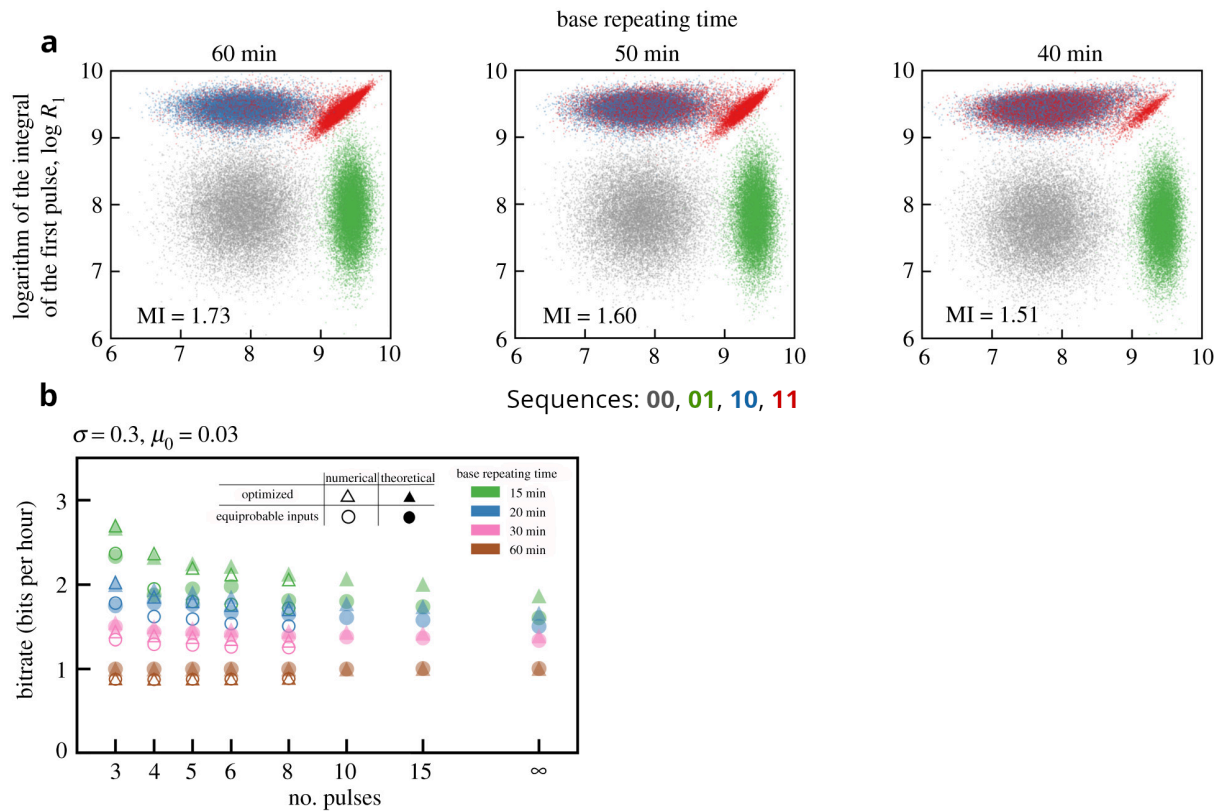


Figure 19. a. Scatter plots showing the system's response to two binary pulses (grey = 00, green = 01, blue = 10, red = 11) for three different base repeating times. **b.** Bitrate estimates as a function of sequence length, comparing results with and without optimised input probabilities. Numerical and theoretical estimates are presented. Reproduced from Limits ★.

3.8. Benchmarking mutual information estimators

To compute the bitrate of the MAPK pathway, we estimated the mutual information between binary sequences and the pathway’s responses using the KSG estimator. Although the KSG estimator is theoretically guaranteed to converge given enough samples, in real life acquiring a sufficient number of samples can be prohibitively expensive, especially for high-dimensional variables. Since then, several new estimators based on neural networks and the Donsker-Varadhan variational formula have been proposed, including MINE [40], InfoNCE [19], and NWJ [41], [42], [43]. However, mutual information estimators have typically been evaluated using simple families of random variables, such as Gaussian distributions or selected one-dimensional distributions, which can introduce selection bias. To address this gap, together with Paweł Czyż we constructed a comprehensive benchmark for evaluating mutual information estimators.

The tasks in the benchmark leverage the fact that continuous injections do not affect mutual information. This result is formalized in the following theorem:

Theorem 1. *Let $\mathcal{X}$, $\mathcal{X}'$, $\mathcal{Y}$ and $\mathcal{Y}'$ be standard Borel spaces (e.g., smooth manifolds with their Borel σ -algebras), and let $f: \mathcal{X} \rightarrow \mathcal{X}'$ and $g: \mathcal{Y} \rightarrow \mathcal{Y}'$ be continuous injective mappings. Then, for every $\mathcal{X}$ -valued random variable X and $\mathcal{Y}$ -valued random variable Y :*

$$I(X; Y) = I(f(X); g(Y))$$

Starting with a pair (X, Y) with known mutual information (e.g., multivariate normal or Student distributions), we can create more complex tasks by applying appropriately chosen transformations. This procedure is illustrated in Figure 20, where a bivariate Gaussian distribution is transformed into four distinct distributions used in the benchmark.

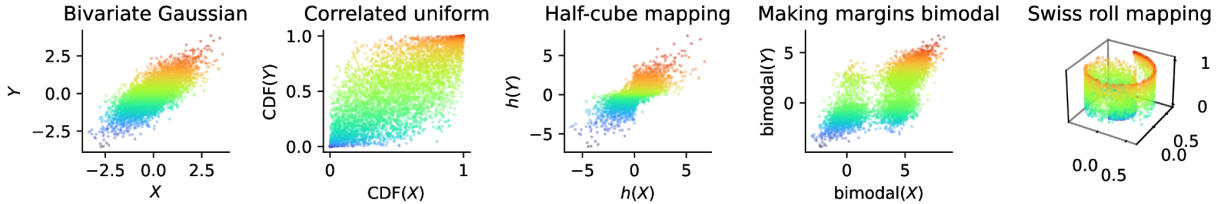


Figure 20. A bivariate Gaussian is transformed by applying continuous injective mappings to its marginals. The resulting benchmark tasks have the same mutual information as the original distribution.. Color shows the original value of Y . Reproduced from Beyond ★.

The main benchmark (Figure 21) consists of forty tasks, designed to test different factors such as dimensionality, long-tailed distributions, or robustness to diffeomorphisms. We found that, for low- and medium-dimensional problems, the classical KSG estimator remains a reliable baseline. Neural estimators perform better in high-dimensional problems, especially when interactions are sparse (i.e., only a few dimensions contribute to mutual information). The benchmark also highlights that evaluating mutual information estimators solely on Gaussian distributions can lead to overly optimistic results, as these tasks tend to be simpler. Furthermore, when the distribution is known to be approximately Gaussian, the estimator based on canonical correlation analysis [44] (CCA) significantly outperforms all other, more general methods.

3.9. Pointwise mutual information and mixture models

The benchmark proposed in Beyond ★ has an important limitation, which we explore in a follow-up article, Mixtures ★, by studying the pointwise mutual information (PMI) *profile*. To begin, we recall that the mutual information between two random variables, X and Y , can be expressed as the expected value of the PMI *function*:

$$I(X; Y) = \mathbb{E}_{(x, y) \sim P_{XY}} \text{PMI}_{XY}(x, y)$$

$$\text{PMI}_{XY}(x, y) = \log p_{XY}(x, y) - \log p_X(x) - \log p_Y(y),$$

where p_{XY} , p_X and p_Y are the probability density functions (PDFs) of the joint and marginal distributions, respectively.

The PMI profile, Prof_{XY} , is defined as the distribution (over real numbers) of the PMI function when sampling pairs (x, y) from the joint distribution P_{XY} . In other words, the PMI profile is the distribution of the random variable $T = \text{PMI}_{XY}(X, Y)$. Consequently, the expected value of the PMI profile coincides with the mutual information:

$$I(X; Y) = \mathbb{E}_{T \sim \text{Prof}_{XY}} [T]$$

For most pairs of variables the PMI profile is not known analytically. However, if the joint distribution of (X, Y) is multivariate normal, the PMI profile can be expressed in terms of canonical correlations between X and Y :

Theorem 2. *Let X and Y be random variables such that their joint distribution $P_{XY} \in P(\mathbb{R}^m, \mathbb{R}^n)$ is multivariate normal. If $k = \min(m, n)$ and $\rho_1, \rho_2, \dots, \rho_k$ are the canonical correlations between X and Y , then the profile Prof_{XY} is a generalized χ^2 distribution, specifically the distribution of the variable:*

$$T = I(X; Y) + \frac{1}{2} \sum_i \rho_i (Q_i - Q_i'),$$

where Q_i and Q_i' are independent and identically distributed (i.i.d.) variables drawn from the χ_1^2 distribution. In particular, Prof_{XY} is symmetric around its median, which is equal to the mean. Additionally,

$$I(X; Y) = -\frac{1}{2} \sum \log(1 - \rho_i^2)$$

All moments of Prof_{XY} exist, and its variance is given by $\text{Var}[T] = \rho_1^2 + \dots + \rho_k^2$.

Crucially, Theorem 1 from Beyond ★ (on the invariance of mutual information under continuous injections) can be extended to the entire PMI profile. We state the theorem here in the case of diffeomorphisms and smooth manifolds; the more general setting of Borel spaces and continuous injections is discussed in Appendix A.5 of Mixtures ★. For smooth manifolds $\mathcal{X}$ and $\mathcal{Y}$, we define the set $P(\mathcal{X}, \mathcal{Y})$ to consist of all probability measures P_{XY} on $\mathcal{X} \times \mathcal{Y}$ such that the joint measure P_{XY} , as well as the marginal measures P_X and P_Y , can be represented by smooth and positive PDFs. This ensures that the PMI function always exists and is smooth.

Theorem 3. Let $P_{XY} \in P(\mathcal{X}, \mathcal{Y})$, and let $f: \mathcal{X} \rightarrow \mathcal{X}$ and $g: \mathcal{Y} \rightarrow \mathcal{Y}$ be diffeomorphisms. Then, for $X' = f(X)$ and $Y' = g(Y)$, it holds that $P_{X'Y'} \in P(\mathcal{X}, \mathcal{Y})$ and $\text{Prof}_{XY} = \text{Prof}_{X'Y'}$.

This result has significant implications for the benchmark constructed in Beyond ★. All tasks generated by transforming multivariate normal variables will have a PMI profile characterized by Theorem 2. Conversely, pairs of variables (X, Y) with a different PMI profile can never be constructed by transforming multivariate normal variables. This limitation is illustrated in Figure 22. The blue bivariate normal distribution is transformed into a distinct multimodal distribution, but the PMI profile remains unchanged. The simple mixture distribution, shown in green, has a distinct asymmetric PMI profile and thus cannot be generated by transforming any normal distribution. However, mixture distributions commonly arise in real world problems (e.g., in our analysis of the MAPK pathway), and should, therefore, be included in any general-purpose benchmark.

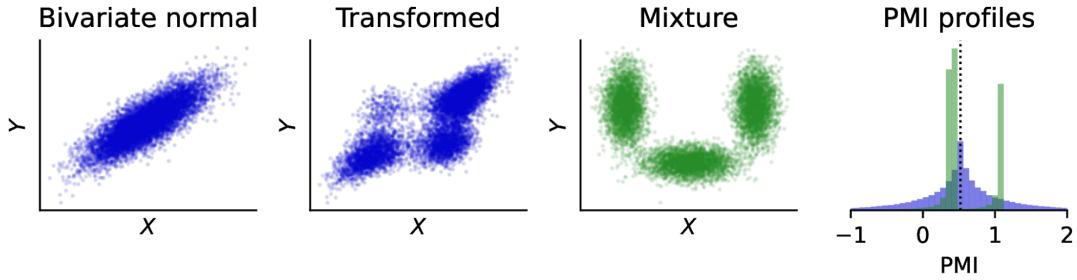


Figure 22. A bivariate normal distribution is transformed to produce a multimodal benchmarking task; both have the same generalized- χ^2 PMI profile (blue). The mixture distribution has a distinct asymmetric PMI profile (green) and thus cannot be constructed by transforming any multivariate normal distribution. Reproduced from Mixtures ★.

To address this limitation, we propose a broader family of probability distributions, called *Bend and Mix Models* (BMMs), combining marginal transformations and statistical mixtures.

Definition 1. A distribution $P_{XY} \in P(X, Y)$ is considered a *Bend and Mix Model* if we can efficiently sample from the joint distribution $(X, Y) \sim P_{XY}$ and numerically evaluate the densities $p_{XY}(x, y)$, $p_X(x)$ and $p_Y(y)$ at every point $(x, y) \in X \times Y$.

Any distribution satisfying this definition can be used as a basic building block for a BMM. In particular, multivariate normal and Student distributions both qualify. More complex distributions can be constructed using the following bending and mixing operations:

Proposition 1. If P_{XY} is a BMM and f and g are diffeomorphisms with numerically tractable Jacobians (e.g., normalizing flows), then $P_{f(X)g(Y)}$ is a BMM.

Proposition 2. Consider BMMs $P^1_{XY}, \dots, P^K_{XY}$. If $w_1, \dots, w_K$ are positive weights such that $w_1 + \dots + w_K = 1$, then the mixture distribution $P_{X'Y'} = w_1 P^1_{XY} + \dots + w_K P^K_{XY}$ is a BMM.

Unfortunately, the mutual information of a mixture distribution cannot be estimated as a weighted sum of mutual information of its components. In Mixtures ★, we prove the following tight bound on the mutual information of mixtures:

$$0 \leq I(X'; Y') \leq \sum_i w_i I(X_i; Y_i) + \log K,$$

as illustrated in Figure 23. However, the properties of BMMs have been chosen to enable efficient sampling of the PMI profile. This can be achieved by sampling (x, y) from the joint distribution P_{XY} and evaluating $t = \text{PMI}_{XY}(x, y)$. The mutual information can then be approximated using a Monte Carlo estimate $I(X; Y) = \mathbb{E}[T]$.

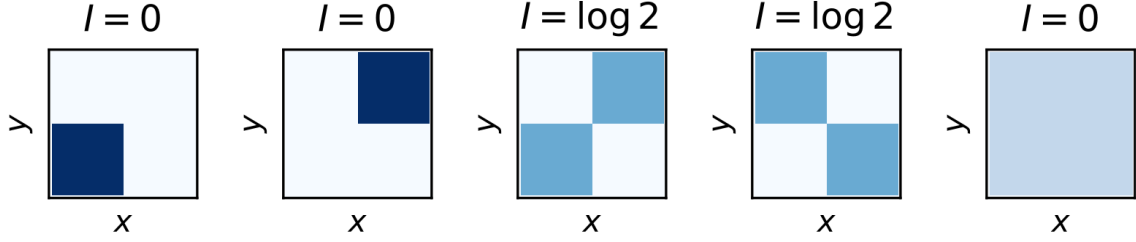


Figure 23. Mixture distributions can create and destroy information. The first two pairs of random variables have zero mutual information, but their mixture (third column) has 1 bit of mutual information. By constructing another mixture with a suitably chosen pair of random variables (fourth column), the mutual information is again reduced to zero (fifth column). Reproduced from Mixtures ★.

Using BMMs, we extend the previous benchmark. Selected tasks are shown in Figure 24 demonstrating both bending and mixing. The article also includes several case studies illustrating the utility of the PMI profile and BMMs. These include an analysis of estimator robustness to inlier and outlier noise (Sec. 3.2), and an in-depth examination of the relationship between variational estimators and the PMI profile (Sec. 3.3). We also show how one can construct distributions involving discrete variables (Appendix B).

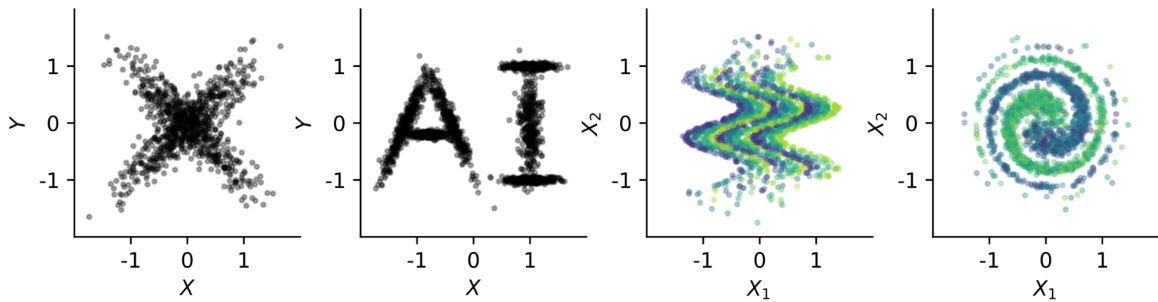


Figure 24. Example tasks constructed using BMMs. In the Waves and Galaxy tasks, the variable Y is indicated by color. Reproduced from Mixtures ★.

Finally, we show that BMMs can be used to develop model-based Bayesian estimators of mutual information. This approach is particularly useful when sample sizes are small, and uncertainty quantification is required—for instance, when one wants to compute the probability that mutual information is positive. As in Predictive ★, this involves running an MCMC algorithm to fit plausible parameter sets of the BMM to the observed samples. For each parameter set, mutual information is then estimated using the Monte Carlo sampling procedure outlined above. An example of this process is shown in Figure 25.

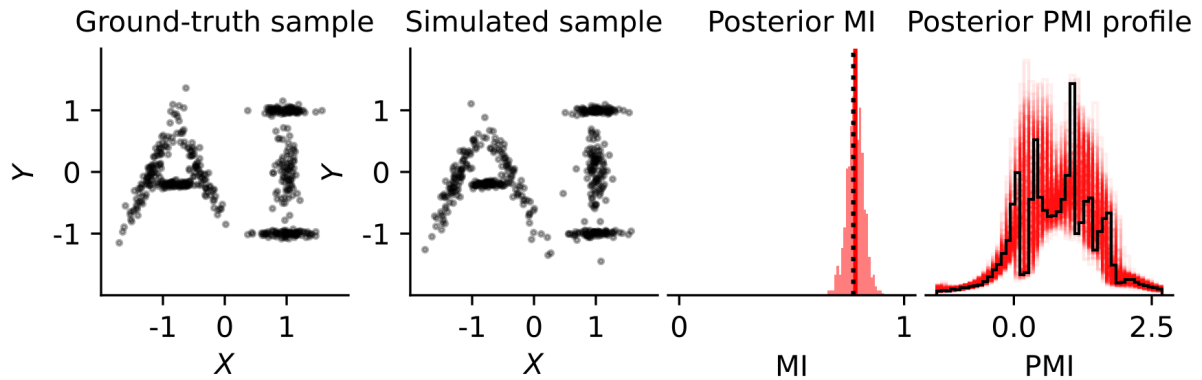


Figure 25. Bayesian estimation of mutual information and of the PMI profile. The first panel shows 500 data points sampled from the ground truth distribution. Posterior parameter sets of a Gaussian Mixture Model (GMM) are sampled using MCMC simulations. The second panel shows 500 data points sampled from an example parameter set obtained in this process. For each parameter set, the mutual information is estimated using Monte Carlo sampling, and the histogram of resulting values is shown (panel three). Since GMMs are a subclass of BMMs, the PMI profile can also be computed for each sampled parameter set, resulting in a posterior distribution of PMI profiles (panel four). Reproduced from Mixtures ★.

4. Discussion and further directions

The research presented in this thesis contributes to our understanding of viral infections, epidemic spread, and cellular signalling by introducing new insights and modelling approaches. However, these remain vast and evolving fields, and the included papers point to several directions for further investigation.

Our early model of COVID-19 explored the hypothesis that populations would initially decrease contact rates (by self-isolation or restrictions) but would gradually increase them due to increasing risk acceptance. This accurately predicted a prolonged pandemic. However, we did not anticipate the oscillations caused by the repeated emergence of new strains and delays in population response. While automated systems flagged new variants, analysis and risk assessment were frequently carried out by individual groups in the context of specific variants, often at the level of individual countries. This allowed for development of some more advanced, but *post hoc*, models based on accumulated data. Potentially, developing more comprehensive models that integrate real-time data [45], [46]—such as reported cases, emerging variants, vaccination rates, and mobility patterns—could provide a unified and continuously updated source of risk assessment, facilitating more responsive public health policies in future pandemics.

In the realm of cellular signalling, our work on the MAPK/ERK pathway highlights how information-theoretic approaches can help elucidate how cells transmit and process information responding to diverse stimuli. This line of inquiry was further pursued by Nałęcz-Jawecki et al., who quantified the bitrate of the MAPK/ERK pathway based on experimental data [14] and analyzed the capacity of MAPK/ERK waves to transmit information *in silico* [Monolayer ☆]. Currently, in collaboration with Paweł Nałęcz-Jawecki, we are applying a similar approach to measure how oncogenic mutations and pharmacological interventions can impact information flow. A particularly interesting direction is the development of tools that enable real-time intervention during experiments, such as optogenetic stimulation of individual cells. New methods, such as CORTICAL [47], could allow for direct measurement of channel capacity in this setting by continuously adjusting inputs and decoding cellular responses.

The agent-based model of RSV infection provided insights into the antagonistic interaction between viruses and innate immunity. However, it also underscored the limitations of traditional compartmental models in capturing the complexity of biological systems. The exploration of non-identifiable models and the application of information theory present promising alternatives for studying cellular signalling networks. Advances in deep learning are also likely to shape the methodologies explored in this thesis. State-of-the-art vision models already enable real-time tracking and stimulation of individual cells during experiments, allowing for a more precise understanding of cell population responses and their information-processing capabilities. Similarly, a future iteration of systems similar to AlphaFold3 [48] could predict potential interactions within a biological system, serving as a prior for the Bayesian techniques applied to non-identifiable models in this thesis. Ultimately, integrating deep learning with mechanistic modelling will provide a more comprehensive framework for understanding complex biological systems.

5. List of original articles included in the thesis

The following articles are included in this thesis and are referenced in the text using their short references and are marked with a ★ symbol. First-author equal contributions are marked with the dagger (†) symbol. Citations as reported by Scopus as of March 30, 2025.

Viral infections of cells

1. Antagonism ★

Grabowski, F.[†], Kocharczyk, M.[†], Korwek, Z., Czerkies, M., Prus, W., & Lipniacki, T. (2023). Antagonism between viral infection and innate immunity at the single-cell level. *PLOS Pathogens*, 19(9), e1011597. [citations: 2]

2. Lambda ★

Prus, W., Grabowski, F., Koza, P., Korwek, Z., Czerkies, M., Kocharczyk, M., & Lipniacki, T. (2024). Type III interferons suppress influenza A virus infection independently of STAT activation by triggering cell death. *bioRxiv*, 2024–09. **Preprint**.

COVID-19 epidemic

3. Dynamics ★

Kocharczyk, M., Grabowski, F., & Lipniacki, T. (2020a). Dynamics of COVID-19 pandemic at constant and time-dependent contact rates. *Mathematical Modelling of Natural Phenomena*, 15, 28. [citations: 34]

4. Super-spreaders ★

Kocharczyk, M., Grabowski, F., & Lipniacki, T. (2020b). Super-spreading events initiated the exponential growth phase of COVID-19 with $\mathcal{R}_0$ higher than initially estimated. *Royal Society Open Science*, 7(9), 200786. [citations: 41]

5. Alpha ★

Grabowski, F.[†], Preibisch, G., Giziński, S., Kocharczyk, M.[†], & Lipniacki, T. (2021). SARS-CoV-2 variant of concern 202012/01 has about twofold replicative advantage and acquires concerning mutations. *Viruses*, 13(3), 392. [citations: 78]

6. Omicron ★

Grabowski, F., Kocharczyk, M., & Lipniacki, T. (2022). The spread of SARS-CoV-2 variant Omicron with a doubling time of 2.0–3.3 days can be explained by immune evasion. *Viruses*, 14(2), 294. [citations: 75]

Information theory & cell signalling

7. Limits ★

Grabowski, F.[†], Czyż, P.[†], Kochańczyk, M., & Lipniacki, T. (2019). Limits to the rate of information transmission through the MAPK pathway. *Journal of the Royal Society Interface*, 16(152), 20180792. [citations: 8]

8. Predictive ★

Grabowski, F., Nałęcz-Jawecki, P., & Lipniacki, T. (2023). Predictive power of non-identifiable models. *Scientific Reports*, 13(1), 11143. [citations: 2]

9. Beyond ★

Czyż, P.[†], Grabowski, F.[†], Vogt, J., Beerenwinkel, N., & Marx, A. (2024). Beyond normal: On the evaluation of mutual information estimators. *Advances in Neural Information Processing Systems*, 36. [citations: 10]

10. Mixtures ★

Czyż, P.[†], Grabowski, F.[†], Vogt, J. E., Beerenwinkel, N., & Marx, A. (2025). On the Properties and Estimation of Pointwise Mutual Information Profiles. *Transactions on Machine Learning Research*.

Other co-authored articles, not included in the thesis

11. Asymptotic ☆

Grabowski, F., Kwapisz, J. H.[†], & Meissner, K. A. (2019). Asymptotic safety and conformal standard model. *Physical Review D*, 99(11), 115029. [citations: 19]

12. Monolayer ☆

Nałęcz-Jawecki, P., Szyc, P., Grabowski, F., Kochańczyk, M., & Lipniacki, T. (2025). Information transmission in a cell monolayer: A numerical study. *PLOS Computational Biology*, 21(2), e1012846.

6. List of software written as a part of the thesis

BMI: Framework for benchmarking mutual information estimators.

Used in **Beyond** ★, **Mixtures** ★. Available at: <https://github.com/cbg-ethz/bmi>

VIS-A-VIS: Agent-based simulator of viral infection spread.

Used in **Antagonism** ★. Available at: <https://github.com/grfrederic/visavis>

Identifiability: Framework for running and analyzing ODE systems using MCMC.

Used in **Predictive** ★. Available at: <https://github.com/grfrederic/identifiability>

CCE: Tool for optimizing channel capacity.

Used in **Limits** ★. Available at: <https://github.com/pawel-czyz/channel-capacity-estimator>

7. Conferences & Seminars

Presenting author names are underlined.

Prus, W., Grabowski, F., Koza, P., Korwek, Z., Czerkies, M., Kochańczyk, M., & Lipniacki, T. (2024) Dissecting role of type I and type III interferons at the single-cell level. 23rd European Conference on Computational Biology. *Poster*.

Czyż, P.[†], Grabowski, F.[†], Vogt, J. E., Beerenwinkel, N., & Marx, A. (2024) On the Properties and Estimation of Pointwise Mutual Information Profiles. Sixth Symposium on Advances in Approximate Bayesian Inference-Non Archival Track. *Poster*.

Grabowski, F.[†], Preibisch, G., Giziński, S., Kochańczyk, M.[†], & Lipniacki, T. (2021) SARS-CoV-2 variant of concern 202012/01 has about twofold replicative advantage and acquires concerning mutations. Seminarium badawcze "Biologia obliczeniowa i bioinformatyka". *Presentation*.

8. Funding

Interaction of natural killer cells with an infected cancer cell population: microfluidics-based experiments and single-cell-level mathematical modeling. National Science Centre (NCN): Grieg (Norwegian Financial Mechanism), 2020–2024, 2019/34/H/NZ6/00699.
PI: Prof. Tomasz Lipniacki.

Influence of oncogenic mutations on information transmission in the MAPK signaling pathway. National Science Centre (NCN): Opus 18, 2020–2024, 2019/35/B/NZ2/03898.
PI: Prof. Tomasz Lipniacki.

9. References

- [1] W. O. Kermack, A. G. McKendrick, and G. T. Walker, "A contribution to the mathematical theory of epidemics," *Proc. R. Soc. Lond. Ser. Contain. Pap. Math. Phys. Character*, vol. 115, no. 772, pp. 700–721, Jan. 1997, doi: 10.1098/rspa.1927.0118.
- [2] W. O. Kermack, A. G. McKendrick, and G. T. Walker, "Contributions to the mathematical theory of epidemics. II. —The problem of endemicity," *Proc. R. Soc. Lond. Ser. Contain. Pap. Math. Phys. Character*, vol. 138, no. 834, pp. 55–83, Jan. 1997, doi: 10.1098/rspa.1932.0171.
- [3] W. O. Kermack, A. G. McKendrick, and G. T. Walker, "Contributions to the mathematical theory of epidemics. III.—Further studies of the problem of endemicity," *Proc. R. Soc. Lond. Ser. Contain. Pap. Math. Phys. Character*, vol. 141, no. 843, pp. 94–122, Jan. 1997, doi: 10.1098/rspa.1933.0106.
- [4] B. Sartorius, A. B. Lawson, and R. L. Pullan, "Modelling and predicting the spatio-temporal spread of COVID-19, associated deaths and impact of key risk factors in England," *Sci. Rep.*, vol. 11, no. 1, p. 5378, Mar. 2021, doi: 10.1038/s41598-021-83780-2.
- [5] F. Graner and J. A. Glazier, "Simulation of biological cell sorting using a two-dimensional extended Potts model," *Phys. Rev. Lett.*, vol. 69, no. 13, pp. 2013–2016, Sep. 1992, doi: 10.1103/PhysRevLett.69.2013.
- [6] M. H. Swat, G. L. Thomas, J. M. Belmonte, A. Shirinifard, D. Hmeljak, and J. A. Glazier, "Chapter 13 - Multi-Scale Modeling of Tissues Using CompuCell3D," in *Methods in Cell Biology*, vol. 110, A. R. Asthagiri and A. P. Arkin, Eds., in *Computational Methods in Cell Biology*, vol. 110, Academic Press, 2012, pp. 325–366. doi: 10.1016/B978-0-12-388403-9.00013-8.
- [7] S. A. Lauer *et al.*, "The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application," *Ann. Intern. Med.*, vol. 172, no. 9, pp. 577–582, May 2020, doi: 10.7326/M20-0504.
- [8] N. M. Linton *et al.*, "Incubation Period and Other Epidemiological Characteristics of 2019 Novel Coronavirus Infections with Right Truncation: A Statistical Analysis of Publicly Available Case Data," *J. Clin. Med.*, vol. 9, no. 2, Art. no. 2, Feb. 2020, doi: 10.3390/jcm9020538.
- [9] B. C. Daniels, Y.-J. Chen, J. P. Sethna, R. N. Gutenkunst, and C. R. Myers, "Sloppiness, robustness, and evolvability in systems biology," *Curr. Opin. Biotechnol.*, vol. 19, no. 4, pp. 389–395, Aug. 2008, doi: 10.1016/j.copbio.2008.06.008.
- [10] R. Cheong, A. Rhee, C. J. Wang, I. Nemenman, and A. Levchenko, "Information Transduction Capacity of Noisy Biochemical Signaling Networks," *Science*, vol. 334, no. 6054, pp. 354–358, Oct. 2011, doi: 10.1126/science.1204553.
- [11] J. Selimkhanov *et al.*, "Accurate information transmission through dynamic biochemical signaling networks," *Science*, vol. 346, no. 6215, pp. 1370–1373, Dec. 2014, doi: 10.1126/science.1254933.
- [12] T. Jetka, K. Nieniałowski, T. Winarski, S. Błoński, and M. Komorowski, "Information-theoretic analysis of multivariate single-cell signaling responses," *PLOS Comput. Biol.*, vol. 15, no. 7, p. e1007132, Jul. 2019, doi: 10.1371/journal.pcbi.1007132.

- [13] Y. Tang, A. Adelaja, F. X.-F. Ye, E. Deeds, R. Wollman, and A. Hoffmann, "Quantifying information accumulation encoded in the dynamics of biochemical signaling," *Nat. Commun.*, vol. 12, no. 1, p. 1272, Feb. 2021, doi: 10.1038/s41467-021-21562-0.
- [14] P. Nałęcz-Jawecki, P. A. Gagliardi, M. Kochańczyk, C. Dessauges, O. Pertz, and T. Lipniacki, "The MAPK/ERK channel capacity exceeds 6 bit/hour," *PLOS Comput. Biol.*, vol. 19, no. 5, p. e1011155, May 2023, doi: 10.1371/journal.pcbi.1011155.
- [15] M. Kochańczyk *et al.*, "Relaxation oscillations and hierarchy of feedbacks in MAPK signaling," *Sci. Rep.*, vol. 7, no. 1, p. 38244, Jan. 2017, doi: 10.1038/srep38244.
- [16] S. Uda *et al.*, "Robustness and Compensation of Information Transmission of Signaling Pathways," *Science*, vol. 341, no. 6145, pp. 558–561, Aug. 2013, doi: 10.1126/science.1234511.
- [17] K. Tudelska *et al.*, "Information processing in the NF- κ B pathway," *Sci. Rep.*, vol. 7, no. 1, p. 15926, Nov. 2017, doi: 10.1038/s41598-017-16166-y.
- [18] A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," *Phys. Rev. E*, vol. 69, no. 6, p. 066138, Jun. 2004, doi: 10.1103/PhysRevE.69.066138.
- [19] A. van den Oord, Y. Li, and O. Vinyals, "Representation Learning with Contrastive Predictive Coding," Jan. 22, 2019, *arXiv*: arXiv:1807.03748. doi: 10.48550/arXiv.1807.03748.
- [20] I. E. Galani *et al.*, "Interferon- λ Mediates Non-redundant Front-Line Antiviral Protection against Influenza Virus Infection without Compromising Host Fitness," *Immunity*, vol. 46, no. 5, pp. 875–890.e6, May 2017, doi: 10.1016/j.immuni.2017.04.025.
- [21] J. T. Wu *et al.*, "Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China," *Nat. Med.*, vol. 26, no. 4, pp. 506–510, Apr. 2020, doi: 10.1038/s41591-020-0822-7.
- [22] J. T. Wu, K. Leung, and G. M. Leung, "Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study," *The Lancet*, vol. 395, no. 10225, pp. 689–697, Feb. 2020, doi: 10.1016/S0140-6736(20)30260-9.
- [23] Q. Li *et al.*, "Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia," *N. Engl. J. Med.*, vol. 382, no. 13, pp. 1199–1207, Mar. 2020, doi: 10.1056/NEJMoa2001316.
- [24] "Estimating the overdispersion in COVID-19 ... | Wellcome Open Research." Accessed: Jan. 29, 2025. [Online]. Available: <https://wellcomeopenresearch.org/articles/5-67/v3>
- [25] S. M. Best and P. J. Kerr, "Coevolution of host and virus: the pathogenesis of virulent and attenuated strains of myxoma virus in resistant and susceptible European rabbits," *Virology*, vol. 267, no. 1, pp. 36–48, Feb. 2000, doi: 10.1006/viro.1999.0104.
- [26] "This is how new Covid-19 variants are changing the pandemic." Accessed: Feb. 12, 2025. [Online]. Available: <https://www.bbc.com/future/article/20210127-covid-19-variants-how-mutations-are-changing-the-pandemic>

- [27] D. L. T. MD, “Researchers claim SARS-CoV-2 B.1.1.7 variant likely to become globally dominant,” News-Medical. Accessed: Feb. 14, 2025. [Online]. Available: <https://www.news-medical.net/news/20210106/Researchers-claim-SARS-CoV-2-B117-variant-likely-to-become-globally-dominant.aspx>
- [28] “Alpha Altmetrics.” [Online]. Available: <https://medrxiv.altmetric.com/details/97079244/news>
- [29] “Omicron Altmetrics.” [Online]. Available: <https://medrxiv.altmetric.com/details/118568424/news>
- [30] K. S. Brown and J. P. Sethna, “Statistical mechanical approaches to models with many poorly known parameters,” *Phys. Rev. E*, vol. 68, no. 2, p. 021904, Aug. 2003, doi: 10.1103/PhysRevE.68.021904.
- [31] R. N. Gutenkunst, J. J. Waterfall, F. P. Casey, K. S. Brown, C. R. Myers, and J. P. Sethna, “Universally Sloppy Parameter Sensitivities in Systems Biology Models,” *PLOS Comput. Biol.*, vol. 3, no. 10, p. e189, Oct. 2007, doi: 10.1371/journal.pcbi.0030189.
- [32] J. Liepe, S. Filippi, M. Komorowski, and M. P. H. Stumpf, “Maximizing the Information Content of Experiments in Systems Biology,” *PLOS Comput. Biol.*, vol. 9, no. 1, p. e1002888, Jan. 2013, doi: 10.1371/journal.pcbi.1002888.
- [33] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, “Hybrid Monte Carlo,” *Phys. Lett. B*, vol. 195, no. 2, pp. 216–222, Sep. 1987, doi: 10.1016/0370-2693(87)91197-X.
- [34] R. M. Neal, *MCMC using Hamiltonian dynamics*. 2011. doi: 10.1201/b10905.
- [35] M. D. Hoffman and A. Gelman, “The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo,” *J. Mach. Learn. Res.*, vol. 15, no. 47, pp. 1593–1623, 2014.
- [36] M. D. Hoffman, D. M. Blei, C. Wang, and J. Paisley, “Stochastic Variational Inference,” *J. Mach. Learn. Res.*, vol. 14, no. 40, pp. 1303–1347, 2013.
- [37] G. Papamakarios, D. Sterratt, and I. Murray, “Sequential Neural Likelihood: Fast Likelihood-free Inference with Autoregressive Flows,” in *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, PMLR, Apr. 2019, pp. 837–848. Accessed: Jan. 29, 2025. [Online]. Available: <https://proceedings.mlr.press/v89/papamakarios19a.html>
- [38] M. K. Transtrum, B. B. Machta, and J. P. Sethna, “Geometry of nonlinear least squares with applications to sloppy models and optimization,” *Phys. Rev. E*, vol. 83, no. 3, p. 036701, Mar. 2011, doi: 10.1103/PhysRevE.83.036701.
- [39] M. Komorowski, M. J. Costa, D. A. Rand, and M. P. H. Stumpf, “Sensitivity, robustness, and identifiability in stochastic chemical kinetics models,” *Proc. Natl. Acad. Sci.*, vol. 108, no. 21, pp. 8645–8650, May 2011, doi: 10.1073/pnas.1015814108.
- [40] M. I. Belghazi *et al.*, “Mutual Information Neural Estimation,” in *Proceedings of the 35th International Conference on Machine Learning*, PMLR, Jul. 2018, pp. 531–540. Accessed: Jan. 29, 2025. [Online]. Available: <https://proceedings.mlr.press/v80/belghazi18a.html>

- [41] X. Nguyen, M. J. Wainwright, and M. I. Jordan, "Estimating Divergence Functionals and the Likelihood Ratio by Convex Risk Minimization," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5847–5861, Nov. 2010, doi: 10.1109/TIT.2010.2068870.
- [42] S. Nowozin, B. Cseke, and R. Tomioka, "f-GAN: Training Generative Neural Samplers using Variational Divergence Minimization," in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2016. Accessed: Jan. 29, 2025. [Online]. Available: https://papers.nips.cc/paper_files/paper/2016/hash/cedebb6e872f539bef8c3f919874e9d7-Abstract.html
- [43] B. Poole, S. Ozair, A. V. D. Oord, A. Alemi, and G. Tucker, "On Variational Bounds of Mutual Information," in *Proceedings of the 36th International Conference on Machine Learning*, PMLR, May 2019, pp. 5171–5180. Accessed: Jan. 29, 2025. [Online]. Available: <https://proceedings.mlr.press/v97/poole19a.html>
- [44] H. HOTELLING, "RELATIONS BETWEEN TWO SETS OF VARIATES*," *Biometrika*, vol. 28, no. 3–4, pp. 321–377, Dec. 1936, doi: 10.1093/biomet/28.3-4.321.
- [45] S. Engebretsen *et al.*, "A real-time regional model for COVID-19: Probabilistic situational awareness and forecasting," *PLOS Comput. Biol.*, vol. 19, no. 1, p. e1010860, Jan. 2023, doi: 10.1371/journal.pcbi.1010860.
- [46] D. Baccega, P. Castagno, A. Fernández Anta, and M. Sereno, "Enhancing COVID-19 forecasting precision through the integration of compartmental models, machine learning and variants," *Sci. Rep.*, vol. 14, no. 1, p. 19220, Aug. 2024, doi: 10.1038/s41598-024-69660-5.
- [47] N. A. Letizia, A. M. Tonello, and H. V. Poor, "Cooperative Channel Capacity Learning," *IEEE Commun. Lett.*, vol. 27, no. 8, pp. 1984–1988, Aug. 2023, doi: 10.1109/LCOMM.2023.3282307.
- [48] J. Abramson *et al.*, "Accurate structure prediction of biomolecular interactions with AlphaFold 3," *Nature*, vol. 630, no. 8016, pp. 493–500, Jun. 2024, doi: 10.1038/s41586-024-07487-w.

Original articles

RESEARCH ARTICLE

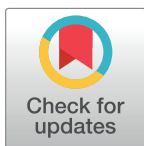
Antagonism between viral infection and innate immunity at the single-cell level

Frederic Grabowski¹✉, Marek Kochańczyk¹✉, Zbigniew Korwek¹, Maciej Czerkies¹, Wiktor Prus¹, Tomasz Lipniacki^{1,2*}

1 Institute of Fundamental Technological Research, Polish Academy of Sciences, Warsaw, Poland, **2** Department of Statistics, Rice University, Houston, Texas, United States of America

✉ These authors contributed equally to this work.

* tlipnia@ippt.pan.pl



OPEN ACCESS

Citation: Grabowski F, Kochańczyk M, Korwek Z, Czerkies M, Prus W, Lipniacki T (2023) Antagonism between viral infection and innate immunity at the single-cell level. PLoS Pathog 19(9): e1011597. <https://doi.org/10.1371/journal.ppat.1011597>

Editor: Christopher F. Basler, Icahn School of Medicine at Mount Sinai, UNITED STATES

Received: March 15, 2023

Accepted: August 2, 2023

Published: September 5, 2023

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.ppat.1011597>

Copyright: © 2023 Grabowski et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Source code of the agent-based simulator of the innate immune response to an infection with an RNA virus is available at <https://github.com/grfrederic/visavis>

Abstract

When infected with a virus, cells may secrete interferons (IFNs) that prompt nearby cells to prepare for upcoming infection. Reciprocally, viral proteins often interfere with IFN synthesis and IFN-induced signaling. We modeled the crosstalk between the propagating virus and the innate immune response using an agent-based stochastic approach. By analyzing immunofluorescence microscopy images we observed that the mutual antagonism between the respiratory syncytial virus (RSV) and infected A549 cells leads to dichotomous responses at the single-cell level and complex spatial patterns of cell signaling states. Our analysis indicates that RSV blocks innate responses at three levels: by inhibition of IRF3 activation, inhibition of IFN synthesis, and inhibition of STAT1/2 activation. In turn, proteins coded by IFN-stimulated (STAT1/2-activated) genes inhibit the synthesis of viral RNA and viral proteins. The striking consequence of these inhibitions is a lack of coincidence of viral proteins and IFN expression within single cells. The model enables investigation of the impact of immunostimulatory defective viral particles and signaling network perturbations that could potentially facilitate containment or clearance of the viral infection.

Author summary

When our cells fight a proliferating virus, the virus fights back. We created a spatial stochastic computational model to understand the impact of such mutually antagonistic relations on the progression of viral infection. The model was tuned to an array of in vitro experiments with the respiratory syncytial virus infecting a monolayer of cells originating from the human respiratory tract. It reproduces complex spatial patterns of cell signaling states. We observed that the antagonistic virus–host cell interactions elicit dichotomous cell fates: infected cells either express viral proteins or produce interferons (biochemical messengers that alert not yet infected cells).

Introduction

Innate immunity provides the first line of defense against viral infections. It acts primarily by feed-forward signaling from infected cells to not (yet) infected bystander cells: upon

(under the BSD-3-Clause license). The simulator implements the computational model described in and used throughout the article. The immunostaining imaging dataset is available at Zenodo (doi: [10.5281/zenodo.7428925](https://doi.org/10.5281/zenodo.7428925)). The ELISA, dPCR and Western Blot data are available as [supplementary files](#).

Funding: This study was funded by Narodowe Centrum Nauki (National Science Centre, Poland) grant 2018/29/B/NZ2/00668 (TL) and the Norwegian Financial Mechanism GRIEG-1 grant operated by Narodowe Centrum Nauki (National Science Centre, Poland) 2019/34/H/NZ6/00699 (TL). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

recognition of viral genetic material by intracellular receptors, some infected cells synthesize and secrete intercellular messengers, most importantly interferons (IFNs), that prompt nearby cells to enter an antiviral state, rendering them more resistant to secondary infections [1–5]. Imminently, with respect to the presence or absence of the virus and the antiviral state, a (partially) infected cell population is spatiotemporally stratified into functionally distinct subpopulations [3,6–8].

To invade a host cell and successfully replicate, viruses evolved multiple strategies to evade or counteract cellular mechanisms of the innate immune response [9]. Commonly, viral, often non-structural, proteins attenuate viral RNA sensing and interfere with the synthesis of IFNs and IFN-induced STAT signaling. For example, in respiratory viruses, NS1 of influenza A/B virus (IAV/IBV) directly and indirectly inhibits activation of IRF3, a key transcription factor of type I and type III IFNs [10,11]; NS1 and NS2 of the respiratory syncytial virus (RSV) similarly impinge on activation of IRF3 and, principally by disrupting STAT1 and STAT2 signaling, impede the induction of IFN-stimulated genes (ISGs) [12–14]; a multitude of effectively analogous interactions have been recently discovered in the case of SARS-CoV-2 [15,16]. The mutual virus–host cell interactions amplify heterogeneity of the stratified, infected cell population and modulate the progression of viral infection. The picture can be even more convoluted in the presence of immunostimulatory, replication-incompetent defective viral particles [17].

Understanding the complex interplay between viruses and the innate immune response at the system level requires experimentation at the single-cell resolution combined with data-driven single-cell level computational modeling [18,19]. Thus far, integration of quantitative data on *Ifnb1* (IFN β gene) and ISGs expression in single cells over time and stochastic modeling allowed to demonstrate that paracrine signaling has a major impact on heterogeneous cell responses to infection [2,3,20]. Recently, benefits of interferon expression heterogeneity have been studied within a spatial agent-based model of an infection with a “generic” virus [21]. A spatial agent-based model with ODE-based intracellular kinetics and stochastic transitions between cell types in a recent work of Aponte-Serrano *et al.* [22] was directly tuned to recapitulate the kinetics of plaque formation by IAV in human bronchial cells. It was demonstrated that a sufficiently fast synthesis and diffusion of (or a prestimulation with) interferon may entirely arrest plaque growth. The action of IAV NS1 was factored into a reduction of the activity of a viral RNA sensor and an IRF inducer, RIG-I. The employed generative computational approach enabled extending the model by inclusion of multiple types of immune cell types (also those involved in adaptive immunity), several diffusible cytokines, as well as cell attachment and migration [23,24]. It was proposed that locally concentrated exposure to a virus can elicit a productive infection, whereas uniform exposure to the virus is likely ineffective.

Here, we present an agent-based, spatial, stochastic model of infection propagation in a monolayer of cells. The model was developed and manually calibrated based on an array of our experiments on epithelial cells of respiratory origin (A549 cell line) infected with human RSV, a common pathogen in severe respiratory disease in both young children and the elderly [25–27]. Upon viral infection, these cells communicate through interferons of type I (IFN β) and type III (IFN λ s) [28]. By analyzing results from cell population-level techniques (Western blot, ELISA, dPCR), we delineated the structure and constrained kinetic rates of the regulatory network. By analyzing heterogeneous single-cell responses in immunofluorescence microscopy images, we quantified the degree of antagonism between the virus and host cells. The data demonstrate that, by triggering universal interferon-mediated mechanisms of the innate immune response [28,29], epithelial cells fight the virus using STAT-inducible ISGs, which inhibit the synthesis of viral RNA and proteins. RSV proteins, in turn, specifically attenuate the immune response by interfering with the activation of IRF3 [30,31], synthesis of IFNs [32],

phosphorylation of STAT1 [33], and by degrading STAT2 [34–36]. These reciprocal, antagonistic interactions give rise to a transient switch-like behavior: an infected cell is unlikely to simultaneously produce IFNs and express RSV proteins. The modes of interference of various viruses display common, recurrent motifs, rendering the constructed model general enough to study, after readjustment of transition rates, kinetics of infection with another virus.

Results

Spatial stochastic model of viral infection

The constructed agent-based, spatial, stochastic model of viral infection is outlined in Fig 1. Below, we describe modeled cell components and biological processes, including virus–host cell interactions. Model derivation and parameter calibration are described in Methods. Experimental results used to parametrize the rates of processes included in the model are shown in Figs A–E in S1 Appendix (we provide raw data as S1, S2, and S3 Dataset files). Model processes and rate parameters are given in Table A in S1 Appendix.

Agents (cells) are arranged in a regular two-dimensional lattice. The state of each agent-cell is fully characterized by seven discrete variables: the binary variable v , which indicates the cell's infection status, and six other variables: $vRNA$, $vProteins$, $pIRF3$, $IFNi$, $pSTATs$, and $ISGs$, each assuming a discrete value from the set $\{0, 1, 2, 3\}$ and in this way describing the status of the respective intracellular biochemical species (Fig 1A). The cell state evolves as a

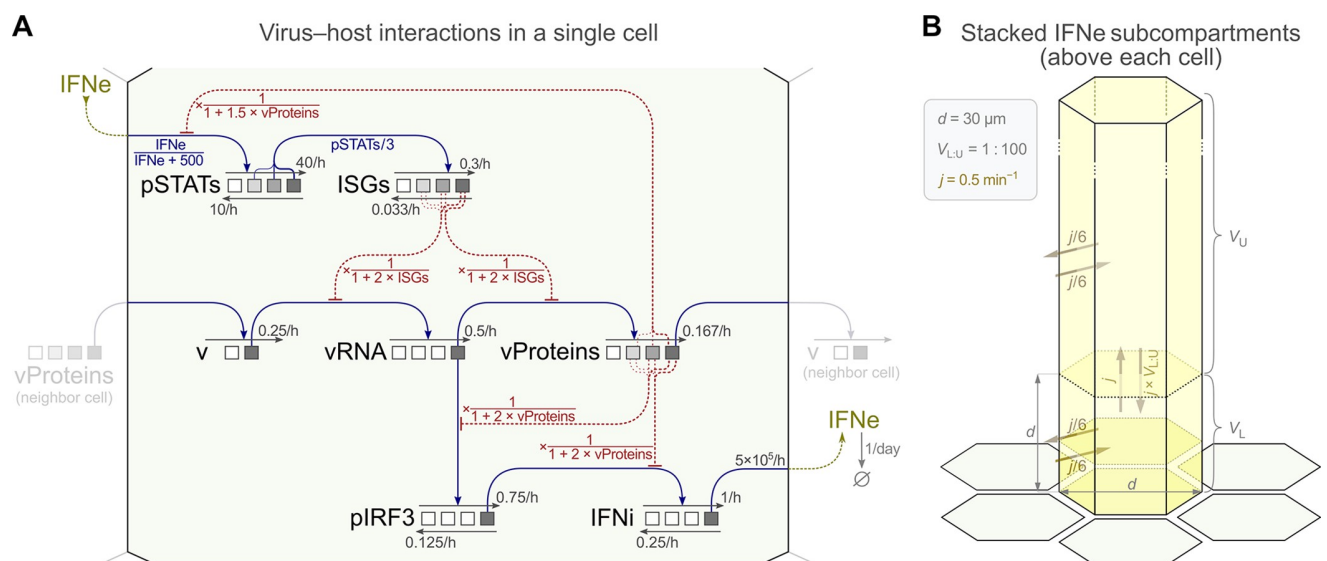


Fig 1. Computational model. **A.** Host cell–virus interactions and adjusted kinetic rate parameter values. Black arrows indicate state transitions of seven variables characterizing the cell state: v , $vRNA$, $vProteins$, $pIRF3$, $IFNi$, $pSTATs$, $ISGs$. All forward transitions (thin straight arrows pointing right) are induced by other variables (as indicated by blue arrows), while the reverse transitions (thin straight arrows pointing left) have constant rates. The corresponding rate coefficients are given in gray. For forward reactions to occur, the inducing agent variable has to be maximally active, with the exception of activation of $pSTATs$ (Michaelis–Menten kinetics) and activation of $ISGs$ (by partially active $pSTATs$); see model reactions in Table A in S1 Appendix. Red dashed hammer-headed arrows are inhibitory interactions: three of them emanate from $vProteins$, two emanate from $ISGs$. The expressions in red indicate how $vProteins$ or $ISGs$ modify the (forward) transition propensities. When an arrow has three roots (all inhibitions and $ISGs$ activation), the strength of the interaction depends on the stage of activation of the inducing variable (additionally indicated by the hue of the inducing variable). The extracellular interferon $IFNe$ is produced at the given rate (molecules / hour) when the state of the intracellular interferon is $IFNi = 3$. $IFNe$ diffuses (see panel B) and, when present in the lower subcompartment above a given cell, regulates the $pSTATs$ transition rate with Michaelis–Menten kinetics. An infected cell with $vProteins = 3$ infects adjacent (non-infected) cells ($v = 0 \rightarrow v = 1$) at the given rate (the infecting virions are implicit). **B.** Processes related to the transport of extracellular IFN ($IFNe$) and related parameter values. $IFNe$ diffuses between the lower and the upper subcompartment above each cell and between respective subcompartments above neighboring cells.

<https://doi.org/10.1371/journal.ppat.1011597.g001>

time-continuous Markov process influenced by neighboring cells, which may infect the cell ($v = 0 \rightarrow v = 1$), and by extracellular interferon, IFN_e .

Intracellular state transitions

As depicted in [Fig 1A](#), viral infection ($v = 1$) leads to the synthesis of viral RNA (increase of $vRNA$), followed by the synthesis of viral proteins (increase of $vProteins$) and the production of infectious virions (implicit in the model). Advancements of viral entities (v , $vRNA$, $vProteins$) are irreversible. The emergence of viral RNA leads to phosphorylation of IRF3 (increase of $pIRF3$ at $vRNA = 3$), which in turn leads to the synthesis of (intracellular) interferon (increase of IFN_i at $pIRF3 = 3$) and its secretion. Secreted interferon is represented as a continuous variable (IFN_e) and diffuses between subcompartments associated with individual cells (see [Fig 1B](#) and [Methods](#) for details). Extracellular interferon induces phosphorylation of STAT1 and STAT2 (represented by an increase of $pSTATs$, with saturable kinetics) that jointly trigger the synthesis of proteins coded by interferon-stimulated genes (increase of $ISGs$). In the model, we consider only one generic type of interferon that accounts for both type I and type III interferons ($IFN\beta$ and $IFN\lambda s$, respectively; our experiments indicate that in the case of infection with RSV, $IFN\beta$ plays a decisive role, see [Fig D](#) panels a and b in [S1 Appendix](#) in conjunction with [Fig A](#) panels a, b and [Fig E](#) in [S1 Appendix](#)). It is worth noting that our model includes both autocrine and paracrine interferon signaling, however, as explained later, we have found that the autocrine interaction does not play a significant role during RSV infection of A549 cells.

Virus–host cell interactions. The model includes two mutually antagonistic groups of interactions. The first group comprises a three-level inhibition of the immune response—phosphorylation of IRF3, synthesis of interferons, and phosphorylation of STAT1 and STAT2—by viral proteins (the strengths of each inhibition increase with $vProteins$). The second group consists of two interactions, in which the accumulation of viral RNA and viral proteins is slowed down due to activity of proteins coded by interferon-stimulated genes (the strength of both inhibitions increases with $ISGs$). The presence of these two reciprocally antagonistic groups of interactions ensures the mutual inhibition of the virus and the innate immune response.

Model simulations vs. immunostaining images

To characterize the incidence and co-incidence of biological entities included in the model at the single-cell level, we performed an array of experiments on A549 cell cultures infected with RSV (with or without additional $IFN\beta$ prestimulation). The cells were immunostained for RSV proteins, IRF3, phospho-Tyr701 STAT1 (p-STAT1), and $IFN\beta$, auxilarily stained with a DNA marker, and imaged using multi-channel confocal microscopy. In [Fig 2](#), representative microscopy images from experiments at a multiplicity of infection (MOI) of 0.01 ([Fig 2A and 2B](#)) are juxtaposed with model simulations ([Fig 2C](#)). The case of an MOI of 1 is shown in [Fig F](#) in [S1 Appendix](#).

[Fig 2A and 2B](#) show that at 8–10 h post infection (p.i.) with RSV, the infection may manifest through IRF3 phosphorylation (p-IRF3 is translocated to the nucleus, giving a discernible nuclear IRF3 signal, overlapping with that of NF- κ B at later time points in [Fig G](#) in [S1 Appendix](#)), expression of RSV proteins, and/or production of $IFN\beta$. In [Fig 2A](#), a 2-hour-long treatment with brefeldin A prior to fixation enabled capturing $IFN\beta$ in the cytoplasm, allowing for its reliable visualization and quantification (see [Fig H](#) in [S1 Appendix](#) for immunostaining of $IFN\beta$ without a prior treatment brefeldin A). Initially, characteristic rings of p-STAT1 positive cells are formed ([Fig 2A](#), 8 h p.i.), often revealing an infected cell being a local source of $IFN\beta$

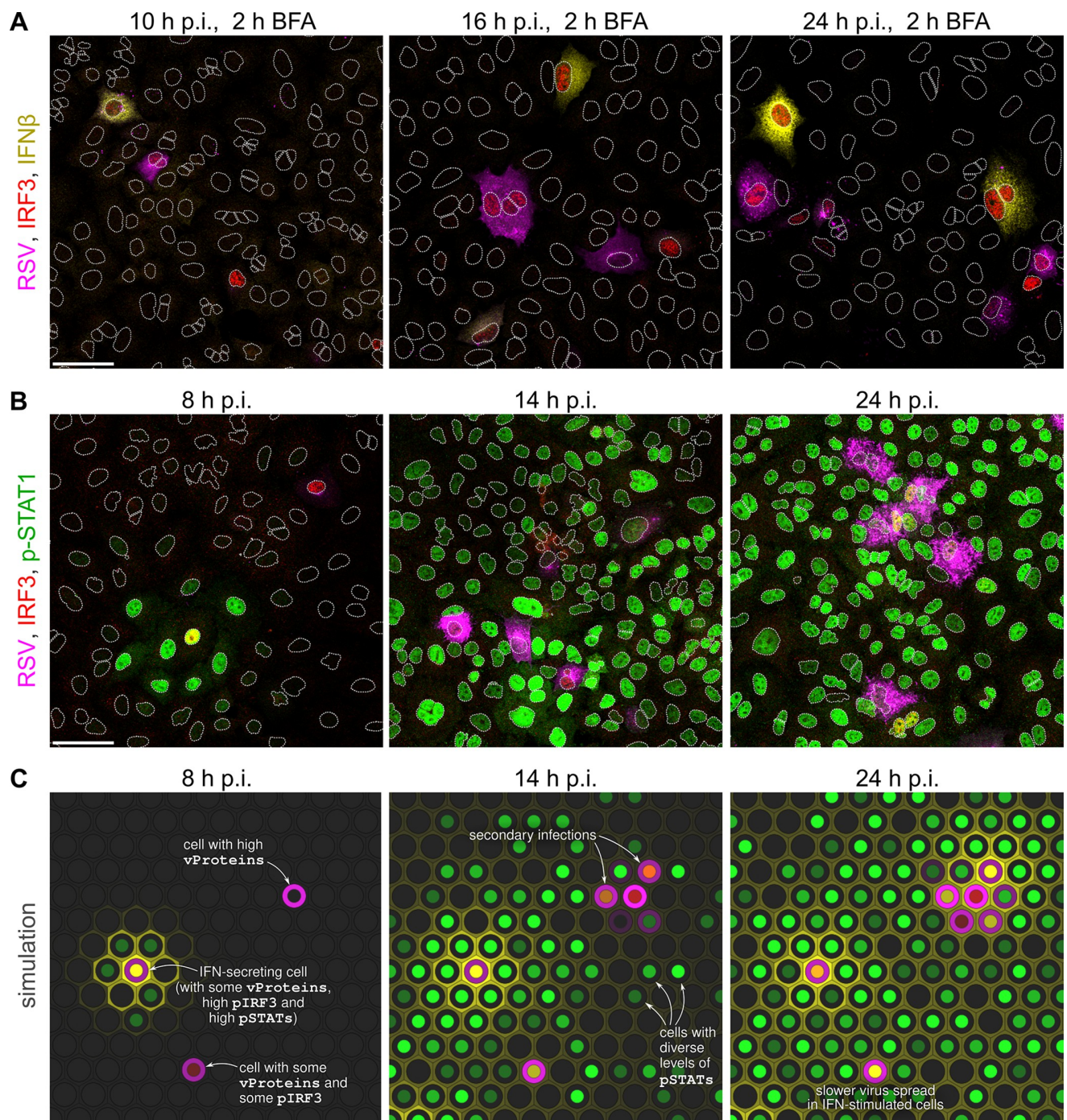


Fig 2. Images from experiments and snapshots from a simulation. A, B. Overlays of images of A549 cells at three time points post infection (p.i.) with RSV at an MOI of 0.01, immunostained for: RSV proteins (polyclonal antibody)—magenta, IRF3—red, (intracellular) IFN β —yellow (only in panel A), phospho-Tyr701 STAT1 (p-STAT1)—green (only in panel B). To improve the detection of intracellular IFN β , we used a 2-hour-long treatment with brefeldin A (BFA, in panel A) before cell fixing, which prevented secretion of IFN β produced during this treatment. When phosphorylated, IRF3 translocates to the nucleus giving a discernible signal. White dotted lines are nuclear outlines determined based on DAPI counterstaining (channel not shown). Scale bars, 50 μ m. C. Snapshots from a simulation of infection at an MOI of 0.01 in a compact monolayer of cells (subpanels show fragments of a simulated 100 $\times$ 100 lattice). The color key corresponds to pseudocolors of immunostained proteins used in panels A, B (concentration of IFN β in the lower subcompartment is indicated at hexagon borders; yellow/orange/olive colors in nuclei result from mixing of red and green at various intensities).

<https://doi.org/10.1371/journal.ppat.1011597.g002>

(Fig 2B). In subsequent time points, 14 h p.i. and 24 h p.i., secondary infections yield small clusters of RSV proteins-producing cells. In the considered time span, IFN β reaches all the cells and the spatial profile of p-STAT1, still concentrated around local sources of IFN β at 14 h p.i., is largely homogeneous across bystander cells at 24 h p.i. Fig 2C shows an example model simulation, at the same stages of infection. Contrary to immunostaining images, the simulation allowed us to study the same cell population at different time points, as well as to simultaneously record all components of the model.

After a high-MOI infection (Fig F in S1 Appendix), nearly all bystander cells display p-STAT1 already at 10 h p.i., but also very likely have a virus-infected neighbor. These cells are expected to be granted less time to upregulate their ISGs prior to infection compared to the bystander cells after low-MOI infection. As demonstrated using model trajectories (Fig I in S1 Appendix), during a low-MOI infection, the primary infected cells exhibit fast growth of the average level of vProteins and have low levels of ISGs (Fig I panel a in S1 Appendix), however the cells infected between 16 and 24 h p.i. (secondary infections) have on average a higher level of ISGs, which noticeably hamper the production of vProteins (Fig I panel b in S1 Appendix).

Antagonism between RSV and the single-cell immune response

RSV proteins terminate STAT activity. Immunostaining images obtained at 24 h p.i. show markedly lower levels of p-STAT1 in cells expressing RSV proteins (Fig 3A). To quantify this effect, we compared the distributions of the nuclear p-STAT1 signal in cells stratified into either expressing or not expressing RSV proteins ('same cells' statistics in Fig 3B). Next, we selected cells that do not express RSV proteins, divided them into either having or not having an RSV proteins-expressing cell in direct neighborhood, and compared nuclear p-STAT1 between these two groups ('neighboring cells' statistics in Fig 3B). To characterize quantitatively the extent to which the considered distributions are discordant, we calculated the signed Kolmogorov–Smirnov (sKS) statistic (see Methods for details). The sKS statistic is a real number in the range $[-1, +1]$, where values near zero imply that the distributions are similar, low values suggest the first distribution frequently has lower values than the second distribution, and high values suggest that the second distribution frequently has higher values than the first one. We use the sKS statistic rather than a pure KS statistic because in certain cases the compared distributions exchange their locations. It also allows us to differentiate between an sKS value, which is consistently low and positive or low and negative for all experimental replicates, and a noisy sKS value changing its sign between replicates. We acknowledge that the sKS can be problematic when a narrower distribution is nested within a broader one, however, this is not the case in our study. For all studied MOIs (0.01, 0.1, and 1), the value of sKS is about 0.5 for 'same cells' and close to 0 for the 'neighboring cells', which agrees well with mean model predictions (Fig 3C). However, experimental results show high variability between replicates. To investigate the potential source of this variability, we first ran the model 10,000 times on a 50×50 cell lattice (containing $n = 2500$ cells, which is typical for shown experimental replicates) and computed the 95% CrI. The resulting plot shows that most of the variability in experiments with a low fraction of RSV proteins-expressing cells can be explained by stochastic noise. Next, we ran the model on a 50×50 lattice with parameters randomly selected from log-normal distributions ($\sigma = 0.2$) centered at the default values, and again showed the 95% CrI. The high variability of obtained numerical results demonstrates the systems' sensitivity to variation of underlying biological parameters, which could explain the spread observed in experimental replicates. Snijder *et al.* showed that cell population context is a major component of the cell-to-cell variation observed during viral infection [37]. Using the model, in Fig

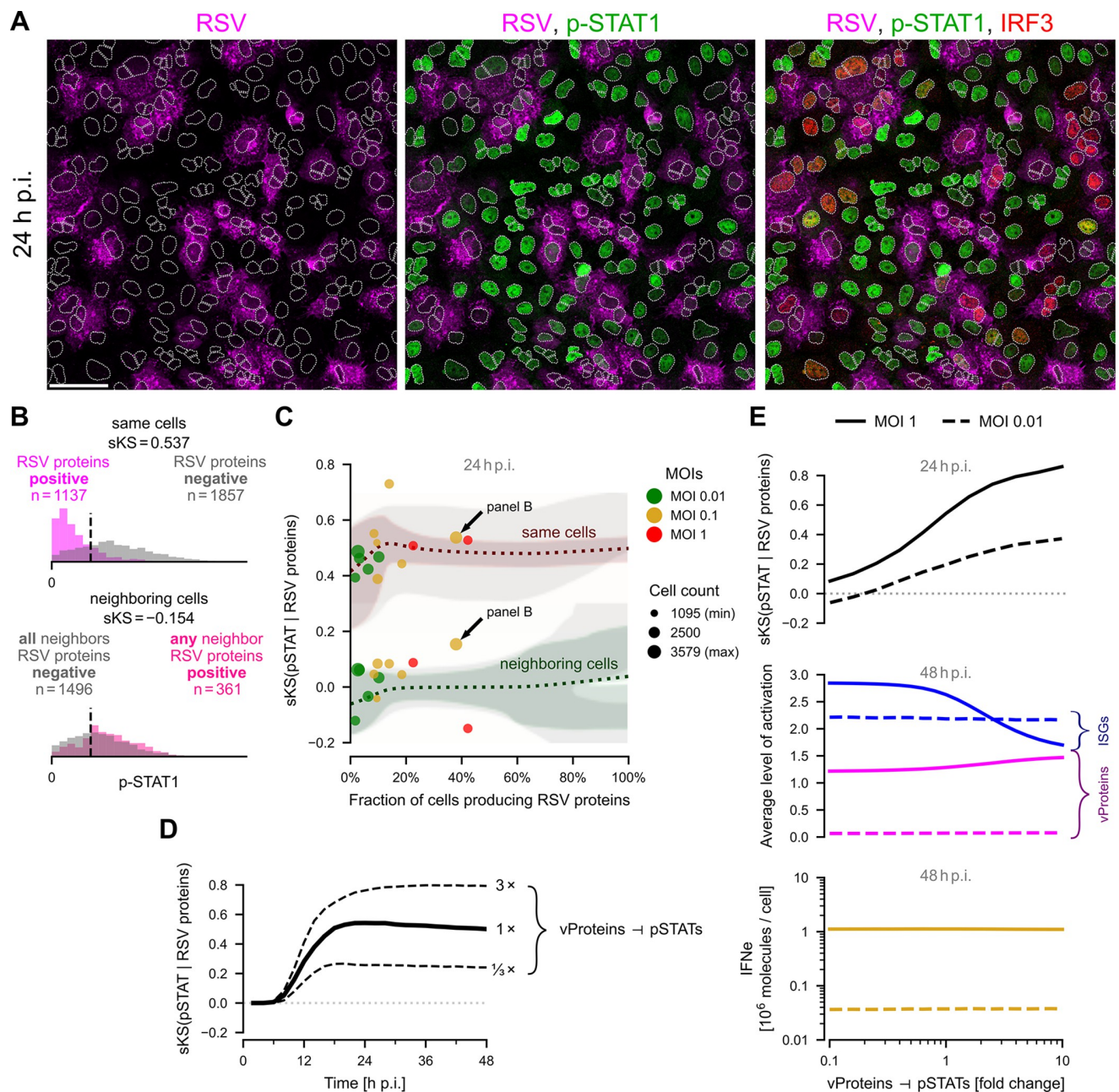


Fig 3. Relation between RSV proteins and STAT signaling. **A.** A549 cells 24 h post infection (p.i.) with RSV at an MOI of 0.1, immunostained for RSV proteins—magenta, p-STAT1—green, and IRF3—red; left to right, incremental overlays. White dotted lines are nuclear outlines determined based on DAPI counterstaining (channel not shown). Scale bar, 50 μ m. **B.** Distributions of the mean nuclear p-STAT1 intensities in cells either expressing or not expressing RSV proteins (top subpanel, ‘same cells’) and in cells not expressing RSV proteins but either having or not a neighboring cell expressing RSV proteins (bottom subpanel, ‘neighboring cells’). Pairs of histograms, cell counts, and signed Kolmogorov–Smirnov (sKS) statistics are given for a representative experiment of an MOI of 0.1 (marked in panel C with an arrow). **C.** Dependence of the ‘same cells’ sKS(pSTATs | RSV proteins) statistics and ‘neighboring cells’ sKS(pSTATs | RSV proteins) statistics on the fraction of RSV proteins-expressing cells at 24 h p.i. Marker sizes correspond to the number of cells quantified in each experiment. Experimental results are juxtaposed with model predictions. Colored dotted lines show mean predictions for default parameters, colored contours show 95% CrI computed for default parameters, and gray contours show 95% CrI computed for randomly perturbed default parameters with log-normal ($\sigma = 0.2$) noise. In both cases, CrI bands are computed based on multiple simulations each with $n = 2500$ cells. **D.** ‘Same cells’ sKS(pSTATs | RSV proteins) over the period of two days post infection at an MOI of 0.1, according to the model. The dashed lines were obtained at different strengths of the vProteins $\rightarrow$ pSTATs inhibition. **E.** Model-based analysis of the influence of the vProteins $\rightarrow$ pSTATs inhibition: ‘Same cells’ sKS(pSTATs | RSV proteins) (upper subpanel, at 24 h p.i.), average ISGs and vProteins states (middle subpanel, at 48 h p.i.), average level of extracellular interferon IFN̳ (lower subpanel, at 48 h p.i.).

<https://doi.org/10.1371/journal.ppat.1011597.g003>

3D we show that the sKS statistic for ‘same cells’ increases in time from 0 to 0.5 (stabilizing at 24 h p.i.), which indicates more pronounced virus-induced pSTATs deactivation at later times. The model also predicts that the value of the ‘same cells’ sKS increases with the strength of the vProteins $\downarrow$ pSTATs inhibition (Fig 3D and 3E), but is insensitive to variations in the strengths of the inhibition of pIRF3 and the inhibition of IFN α by vProteins (Fig J panel a in S1 Appendix). The ‘neighboring cells’ sKS statistics remain close to zero regardless of the strengths of all inhibitions originating from vProteins (Fig J panels a and b in S1 Appendix).

In agreement with the model, we observed that activation of STAT1/2 is key for attenuation of viral spread; prestimulation with IFN β substantially lowered viral load (Fig E in S1 Appendix), while IFNAR or STAT1 or STAT2 KO (Fig D panels c and a in S1 Appendix) increased viral load. Surprisingly, the model indicates that the strength of pSTATs inhibition by vProteins – although influencing the status of pSTATs in vProteins-expressing cells – very weakly influences the overall spread of infection (Fig 3E). For an MOI of 1 (but not for an MOI of 0.01) and stronger vProteins $\downarrow$ pSTATs inhibition, the average status of ISGs is lowered, but this leads to a very modest increase of the average status of vProteins at 48 h p.i. and does not influence extracellular IFN accumulation. This is because, in individual cells, inhibition of pSTATs and consecutive depletion of the build-up of ISGs typically occurs after the (irreversible) accumulation of vProteins, and is thus not advantageous to the virus according to the model. This suggests that autocrine pSTATs activation does not play a significant role during RSV infection of A549 cells.

RSV proteins inhibit IRF3 activation. Viral RNA is required for synthesizing RSV proteins as well as for activating IRF3; these two processes, however, may but do not need to occur simultaneously in the same cell. Accordingly, in our immunostaining images, there are cells expressing RSV proteins with and without active IRF3, as well as cells with active IRF3 but devoid of RSV proteins (Fig 4A). Imaging data collected at different time points (16, 20, 24, 36, 40, and 48 h p.i.) and various initial MOIs (0.01, 0.1, and 1) show that approximately 40% of RSV proteins-expressing cells have active IRF3, with a good agreement with the model (Fig 4B). Conversely, approximately 50–80% of cells with active IRF3 express RSV proteins (Fig 4C). In experiments, the percentage of IRF3-active cells is somewhat larger for a larger fraction of cells expressing RSV proteins, but this finding is not reproduced by our model. An auxiliary demonstration of the vProteins $\downarrow$ pIRF3 inhibition is provided in Fig G in S1 Appendix, where a sizable fraction of cells displays NF- κ B but no IRF3 activity at 24 h and 40 h p.i. (but not at 16 h p.i.). In Fig 4B and 4C, we observe high variability between experimental replicates, indicating that the counterbalance between the strength of the innate immune response (indicated by the fraction of cells with active IRF3) and virus spread (indicated by the fraction of cells with RSV proteins) is heterogeneous and sensitive to experimental conditions. Analogously to Fig 3C, we plot 95% CrI in Fig 4B and 4C. These CrI show that the observed variability in experimental replicates could be attributed to relatively small variations in biochemical rates, which are expected between cell populations.

To show the importance of inhibitions, we include model predictions with the vProteins $\downarrow$ pIRF3 inhibition and with all inhibitions either turned off or made 10-fold stronger (Fig 4B and 4C). This analysis highlights the key role of the vProteins $\downarrow$ pIRF3 inhibition in the regulation of the proportions of cells co-expressing active IRF3 and RSV proteins with respect to cells expressing RSV proteins (Fig 4B) or active IRF3 (Fig 4C). According to the model, the fraction of cells that display both RSV proteins and IRF3 activity increases in time (as the infection progresses) and substantially decreases with the strength of inhibition of pIRF3 by vProteins (Fig 4D, left subpanel). Consequently, the overall fraction of IRF3-active cells is lower for stronger inhibition (Fig 4D, right subpanel).

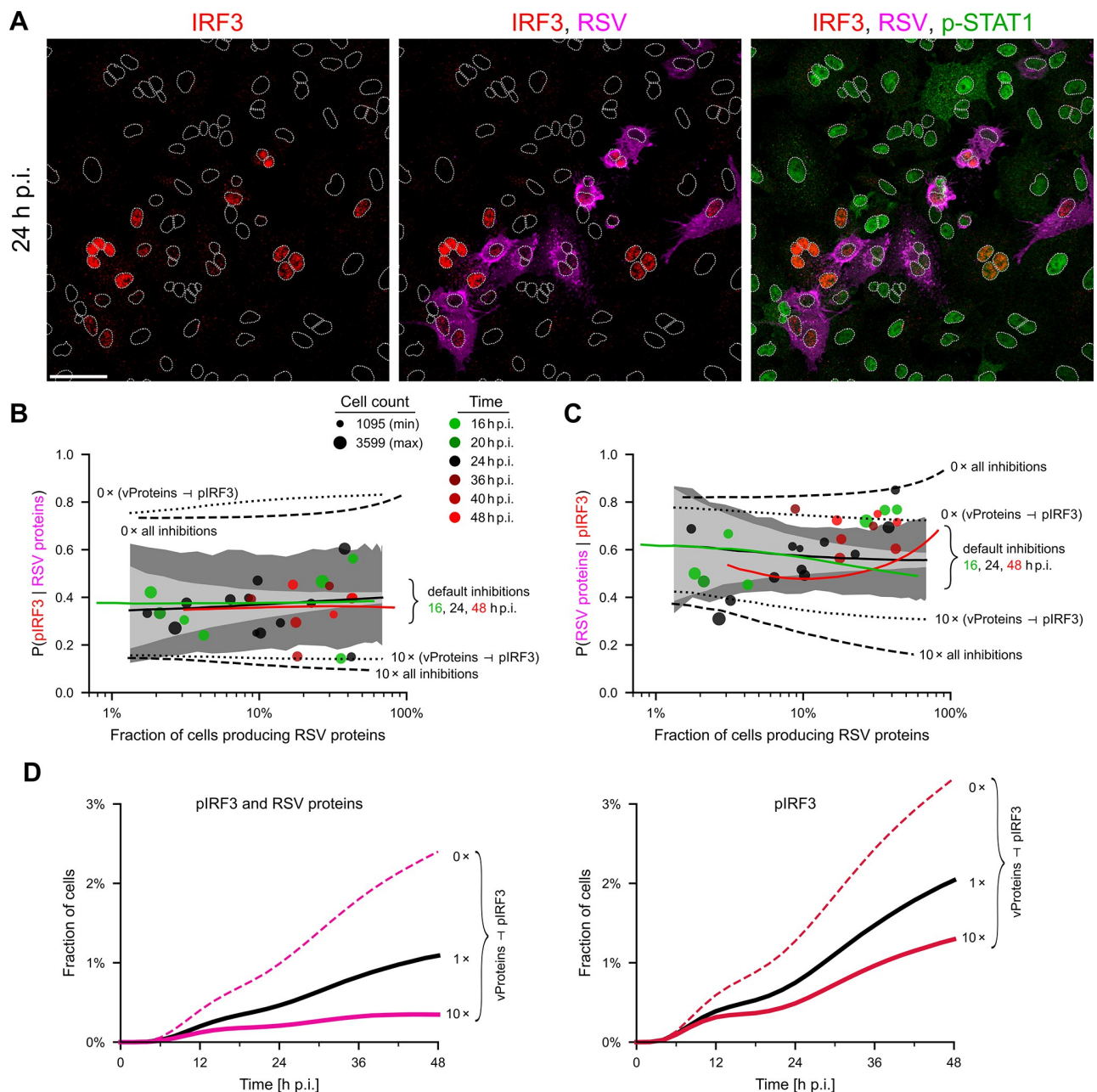


Fig 4. Relation between expression of viral proteins and IRF3 activity. **A.** Incremental overlays of images of A549 cells 24 h post infection (p.i.) with RSV at an MOI of 0.1, immunostained for IRF3 –red, RSV–magenta, and p-STAT1 –green. White dotted lines are nuclear outlines determined based on DAPI counterstaining (channel not shown). Scale bar, 50 μm . **B.** Conditional probability $P(\text{pIRF3} | \text{RSV proteins})$ as a function of the percentage of RSV proteins-expressing cells. Experiments (at 16, 20, 24, 36, 40, 48 h p.i.)–disks, model (at 16, 24, 48 h p.i.)–lines (line points were computed for a specific MOI, resulting in a given proportion of RSV-positive cells; results are averages over 1000 stochastic simulations on the 100×100 lattice). Dashed lines show model predictions (at 24 h p.i.) without the vProteins $\rightarrow$ IRF3 inhibition or with 10 times stronger inhibitions (as shown in the figure). Dotted lines show model predictions (at 24 h p.i.) without the vProteins $\rightarrow$ IRF3 inhibition or with 10 times stronger inhibitions (as shown in the figure). Light gray contours show 95% CrI computed for default parameters; dark gray contours show 95% CrI computed for randomly perturbed default parameters with log-normal ($\sigma = 0.2$) noise, both for $n = 2500$ cells. **C.** Conditional probability $P(\text{RSV proteins} | \text{pIRF3})$, notation as in panel B. **D.** Proportion of the (IRF3 & vProtein)-positive cells (left) and IRF3-positive cells (right) as a function of time for a default strength of the vProteins $\rightarrow$ pIRF3 inhibition, no inhibition, and the inhibition 10 times stronger than the default.

<https://doi.org/10.1371/journal.ppat.1011597.g004>

Mutual exclusion between the expression of RSV proteins and the synthesis of IFN β .

Viral RNA triggers production of IFN (via activation of IRF3) and is required for the synthesis of viral proteins. However, as we can see in immunostaining images in Fig 5A (as well as in Fig 2A and Fig H in S1 Appendix), the accumulation of IFN and RSV proteins rarely coincides in the same cell. The analysis of immunostaining images indicates that about 10% of cells expressing RSV proteins exhibit IFN accumulated during 2 hours of brefeldin A treatment that blocks IFN secretion (Fig 5B). In turn, 30–40% of the cells exhibiting accumulated IFN have RSV proteins (Fig 5C). Both percentages do not change significantly with initial MOIs (0.01, 0.1, and 1) and over time (experimental time points: 16, 20, 24, 36, 40, and 48 h p.i.), and are in good agreement with the model. As previously demonstrated in Figs 4B, 4C, 5B, and 5C we observe high variability between experimental replicates, which again can be attributed to stochastic noise and relatively small variability of biochemical parameters.

When comparing predictions of the model in which inhibitions $vProteins \rightarrow pIRF3$ and $vProteins \rightarrow IFNi$ are either turned off or made 10-fold stronger (dotted lines in

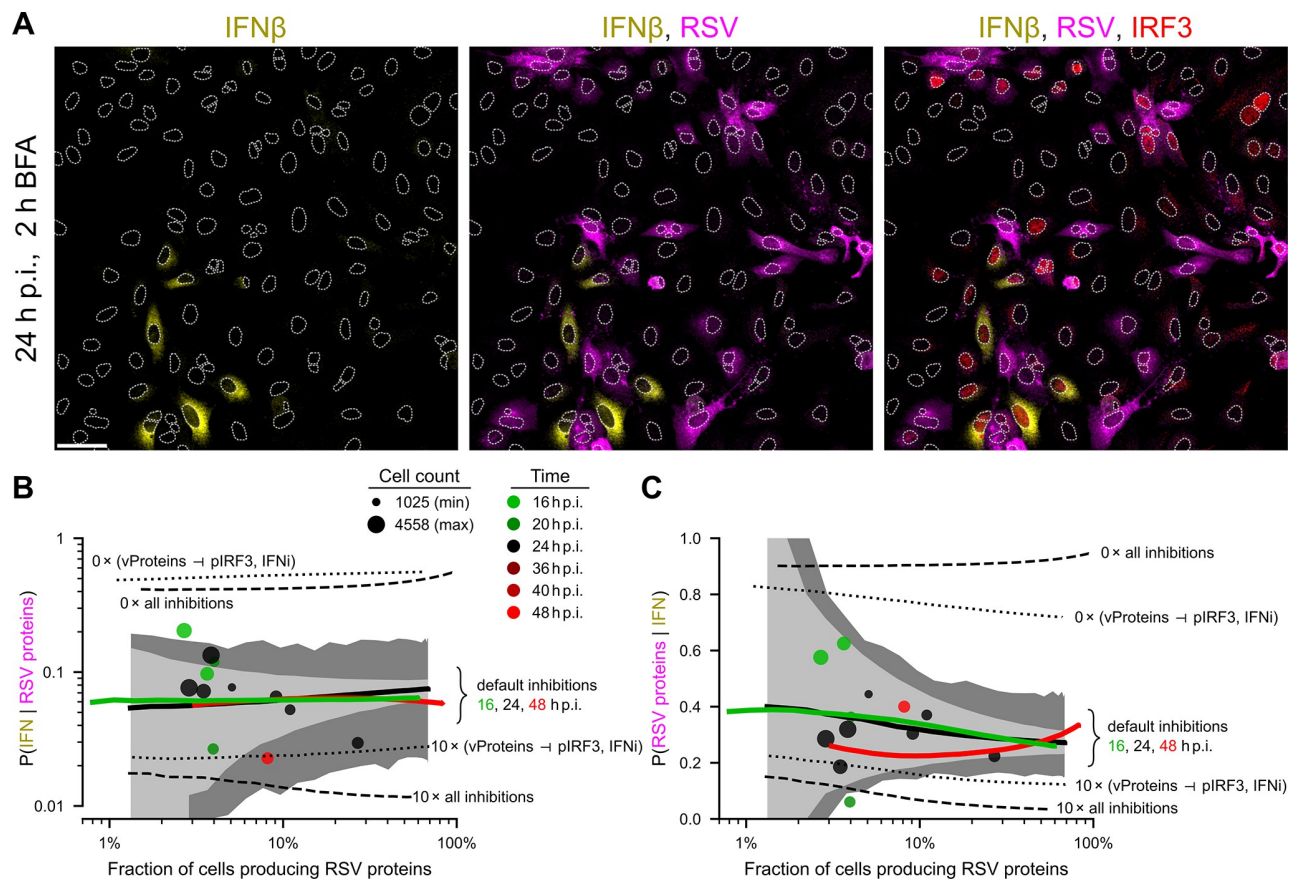


Fig 5. Mutually exclusive expression of RSV proteins and synthesis of IFN β . **A.** Incremental overlays of images of A549 cells 26 hours post infection (h p.i.) with RSV at MOI 0.1, immunostained for IFN β –yellow, RSV –magenta, and IRF3 –red. To improve detection of intracellular IFN β , we used a 2-hour-long treatment with brefeldin A (BFA) before cell fixing, which prevented secretion of IFN β produced during this treatment. White dotted lines are nuclear outlines determined based on DAPI counterstaining (channel not shown). Scale bar, 50 μ m. **B.** Conditional probability $P(\text{IFN} | \text{RSV proteins})$ as a function of the percentage of RSV proteins-expressing cells. Experiments (at 16, 20, 24, 36, 40, 48 h p.i.) –disks, model (at 16, 24, 48 h p.i.) –lines (line points were computed for a specific MOI, resulting in a given proportion of RSV-positive cells; results are averages over 1000 stochastic simulations on the 100×100 lattice). Dashed lines show model predictions (at 24 h p.i.) without all 5 inhibitions or with 10 times stronger inhibitions (as shown in the figure). Dotted lines show model predictions (at 24 h p.i.) without $vProteins \rightarrow pIRF3$ and $vProteins \rightarrow IFNi$ inhibitions, or with these two inhibitions 10 times stronger. Light gray contours show 95% CrI computed for default parameters; dark gray contours show 95% CrI computed for randomly perturbed default parameters with log-normal ($\sigma = 0.2$) noise, both for $n = 2500$ cells. **C.** Conditional probability $p(\text{RSV proteins} | \text{IFN})$, notation as in panel B.

<https://doi.org/10.1371/journal.ppat.1011597.g005>

Fig 5B and 5C) to predictions of the model in which all inhibitions are either turned off or made 10-fold stronger (dashed lines in Fig 5B and 5C), we can see the crucial role of inhibitions $vProteins \rightarrow pIRF3$ and $vProteins \rightarrow IFNi$ in achieving correct proportions of cells co-expressing IFN and RSV proteins with respect to cells expressing RSV proteins (Fig 5B) or IFN (Fig 5C).

Model-based analysis of the influence of inhibitory interactions on the progression of viral infection. The effect of antagonistic inhibitions from $vProteins$ to $pIRF3$ and $IFNi$, and from $ISGs$ to $vRNA$ and $vProteins$ (Fig 6A) on the progression of the infection is further studied using the validated model. We focus on the case of $MOI = 0.01$, in which the immune response is granted time to develop.

As shown in Fig 6B, the fraction of cells producing both IFN and RSV proteins grows in time as the infection spreads, but remains below 0.2% even at 48 h p.i. This fraction is controlled by the strengths of inhibitions $vProteins \rightarrow pIRF3$ and $vProteins \rightarrow IFNi$. When these inhibitory interactions are 10 times stronger, the fraction is close to 0, whereas their removal increases the fraction fivefold, to 1%. Correspondingly, the removal of these inhibitory interactions increases the overall fraction of cells producing IFN (at 48 h p.i.) more than twofold, to above 1% (Fig 6C). However, when 10-fold stronger inhibitory interactions are assumed, the relative effect on the fraction of IFN-producing cells is negligible, since only about 10% of cells producing IFN have $vProteins$ (consistently with data in Fig 5B). The effect of removing inhibitory interactions $vProteins \rightarrow pIRF3$ and $vProteins \rightarrow IFNi$ is forwarded through subsequent tiers of the immune response leading to an increase in the levels of $pSTATs$ (Fig 6D) and $ISGs$ (Fig 6E). The most significant increase is observed, respectively, at about 20 and 30 h p.i. Consequently, induction of a more pronounced antiviral state at about 30 hours (epitomized by an increased level of $ISGs$) leads to a decrease of $vProteins$ -producing cells after that time point (Fig 6F). In summary, the analysis presented in Fig 6B–6F dissects regulatory steps in which the virus enhances its propagation by suppressing the innate immune response.

Next, we analyze how the strengths of inhibitory interactions mediated by the immune response, $ISGs \rightarrow vProteins$ and $ISGs \rightarrow vRNA$, influence the propagation of infection and activation of the immune response. As shown in Fig 6G, an increase in the strength of the $ISGs \rightarrow vProteins$ inhibition leads to a decrease in the fraction of $vProteins$ -producing cells and a simultaneous increase in the number of interferon-producing cells. Clearly, this inhibitory interaction attenuates viral infection and promotes interferon signaling. The effect of inhibitory interaction $ISGs \rightarrow vRNA$ is different; an increase in its strength leads to a moderate decrease of the fraction of $vProteins$ -producing cells, but also to a substantial decrease of interferon-producing cells (Fig 6H). This is because viral RNA is necessary for both the synthesis of viral proteins and activation of the $IRF3 \rightarrow IFN$ pathway.

Influence of defective viral genomes (DVGs) on virus spread

In the model we have not accounted for the potential formation of defective viral particles with incomplete genomes that are capable of activating IRF3 (triggering IFN synthesis), but are incapable of producing infectious virions and may not be able to express proteins that would inhibit IFN synthesis. In experimental conditions, DVGs may emerge as an artifact during viral amplification [38] and although we cannot rule out their presence in our experiments, the obtained results can be explained by the model that neglects their presence. Since DVGs are known to be important drivers of the immune response [17], we analyze theoretically the consequence of their presence on IFN synthesis and virus spread.

For the sake of simplicity we assume that there is a fraction of cells infected solely with DVGs (f_{DVG}) that are capable of activating IRF3 but, in the absence of replication-competent

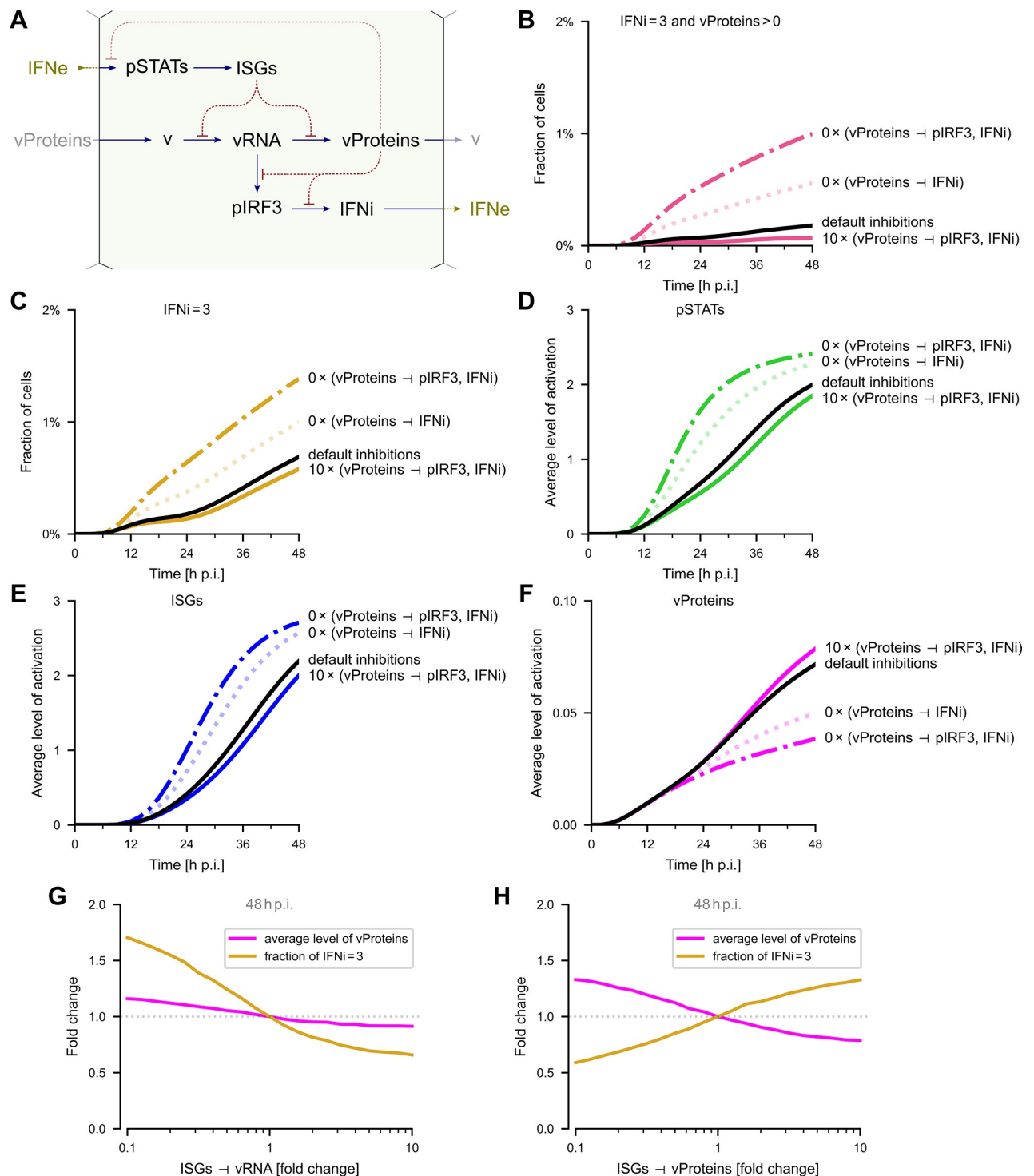


Fig 6. Influence of inhibitions on the progression of viral infection (model). A. Simplified scheme of antagonistic interactions present in the model. B–E. (B) Fraction of cells producing IFNe (IFNi = 3) and vProteins (vProteins > 0) simultaneously, (C) fraction of cells producing IFNe, (D) average activation of pSTATs, (E) average activation of ISGs, (F) fraction of cells producing vProteins; all as a function of time for an MOI of 0.01 for default inhibition strengths (black line), inhibitions vProteins $\rightarrow$ pIRF3 and vProteins $\rightarrow$ IFNi 10 times stronger (thick line), and turned off (dashed, dotted and dash-dotted lines). G. Change of the average level of vProteins and of the fraction of cells producing IFNe as a function of the ISGs $\rightarrow$ vProteins inhibition, at 48 h p.i. at an MOI of 0.01. H. Change of the average level of vProteins and of the fraction of cells producing IFNe as a function of the ISGs $\rightarrow$ vRNA inhibition, at 48 h p.i. at an MOI of 0.01.

<https://doi.org/10.1371/journal.ppat.1011597.g006>

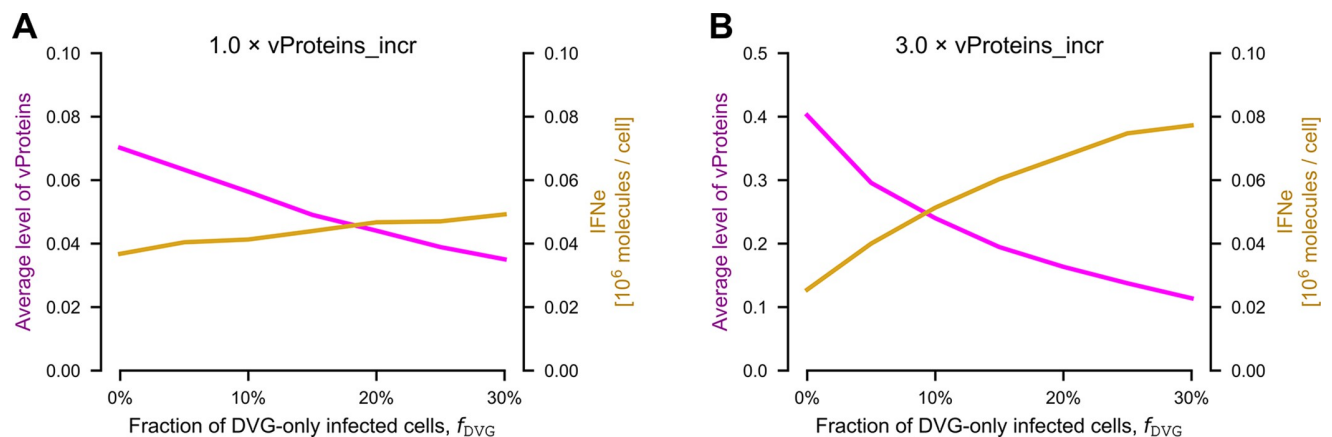


Fig 7. Influence of DVGs on IFN expression and viral progeny (model). A. The average status of vProteins and IFNε as a function of the percentage of cells infected solely by DVGs at 48 h p.i. Initial MOI: 0.01. B. Same as in panel A, but with the vProteins forward rate coefficient (vProteins_incr) increased 3-fold. Note the different range on the left vertical axis.

<https://doi.org/10.1371/journal.ppat.1011597.g007>

helper virions, cannot express viral proteins. In Fig 7A we show a relatively modest increase of IFNε and about a twofold decrease in the average vProteins level as f_{DVG} increases from 0 to 30%. This surprisingly weak influence is a consequence of the fact that for the default model parameters, cells are able to express IFN before viral proteins are produced. However, when the vProteins forward rate is increased threefold, the effect of DVGs on infected cells is much stronger, leading to a fourfold reduction of the level of vProteins (Fig 7B). Overall, this analysis indicates that DVGs can be critical drivers of innate responses, but only in the case of viruses that are capable of quickly terminating the synthesis of IFN.

Discussion

Viral infection, when viewed at the cell-population level, leads to both the multiplication of viral particles and the triggering of the innate immune response. The intuitively expected and experimentally widely observed antagonism between these two processes [29] can be investigated at the single-cell level. Based on cell-population and single-cell experiments involving RSV, we constructed and constrained a stochastic, agent-based model that elucidates this mutual antagonism.

To assemble the model, we confirmed experimentally the importance of the RIG-I/MAVS $\rightarrow$ IRF3 $\rightarrow$ IFN β $\rightarrow$ IFNAR $\rightarrow$ STAT1/2 $\rightarrow$ ISGs axis in antiviral signaling by characterizing the effects of knockouts of its components at the level of (epithelial) cell population. The lack of RIG-I, MAVS, IRF3, IFNAR, STAT1, or STAT2 leads to at least several-fold increase in viral load, in agreement with earlier studies. Seth et al. [39] showed that knockdowns of MAVS and RIG-I reduce IRF3 phosphorylation and promote viral spread (a direct interaction between MAVS and IRF3 in virus-infected cells was demonstrated later by Liu et al. [40]). Wathélet et al. [41] showed that in response to virus infection, IRF3 (as well as IRF7), jointly with ATF-2/c-Jun and NF- κ B, form the enhanceosome to trigger IFN expression. Correspondingly, Hatesuer et al. [42] showed that the lack of both IRF3 and IRF7 resulted in the absence of type I interferons after IAV infection in mice and resulted in a strong increase of mortality. In the case of RSV infection, IFN β $\rightarrow$ IFNAR signaling plays a dominant role, while in the case of IAV, IFN λ $\rightarrow$ IFNLR is of equal importance. As shown by Goritzka et al. [43], IFNAR1-deficient mice displayed increased lung viral load during RSV infection. Correspondingly, Crotta et al. [44] showed that a simultaneous deletion of IFNAR and IFNLR renders mice vulnerable

to IAV infection of the respiratory tract. Both STAT2- and STAT1-deficient mice are more vulnerable to viral infections, however, unsurprisingly, the effect is more pronounced for STAT1-deficient mice as they also cannot respond to type II interferons, reviewed by Meyts & Casanova [45].

The proposed model recapitulates the expected time to production of IFN by early responders, which is in the range of 4–8 h [2,46]. The low fraction of IFN-producing cells is also on par with other estimates indicating that up to a few percent of cells produce interferons, see, e.g., [6,47,48]. As in experiments, the scant sources of IFN, especially at a low MOI, amplify cell population heterogeneity by giving rise to characteristic foci of IFN-activated cells [46].

The model allowed us to investigate correlations between the levels of key proteins in the same cell as well as in adjacent cells, which enabled quantitative comparison with spatial patterns observed in immunostaining images. Both the experimental data and the model indicate that while viral RNA triggers the innate immune response, viral proteins inhibit this response at 3 levels: by blocking activation of IRF3, by blocking the synthesis of IFN, and by inhibiting STATs activation. The first two interactions promote the spread of infection, whereas, as suggested by the model, the last one has a minimal influence on virus spread. The fact that the direct inhibition of STAT phosphorylation by RSV (nonstructural) proteins has little influence on RSV replication in the model suggests that this interaction may be implicated in regulatory processes not included in the model. It is plausible that disabling the production of ISG-encoded proteins redirects ribosomes to the production of viral proteins, enabling efficient generation of viral progeny. Consistent with this hypothesis, RSV has been shown to take over the global control of host cell translation [49]. The innate immune response inhibits viral propagation through paracrine IFN signaling, which activates STAT signaling which, in turn, upregulates expression of ISGs. Once a cell becomes infected, ISGs slow down the synthesis of RSV proteins, giving more time for activation of IRF3 and the synthesis of IFN. However, because the process of RSV protein synthesis is irreversible, at some point, RSV proteins accumulate and terminate both the activity of IRF3 and synthesis of IFN. A consequence of the $v\text{Proteins} \vdash p\text{IRF3}$ and $v\text{Proteins} \vdash \text{IFN}$ inhibitions is that only about 40% of viral proteins-expressing cells have active IRF3 and as little as 10% of such cells produce IFN.

This picture is changed in the presence of DVGs; the model indicates that the fraction of cells infected solely by DVGs can, due to the lack of viral protein expression, serve as a source of IFN and in that way limit virus spread. The effect is modest for default model parameters but significant when faster accumulation of viral proteins is assumed — this implies that DVGs can be critical drivers of interferon-mediated responses, as shown experimentally by Sun et al. [38], but only in the case of viruses that would effectively inhibit interferon synthesis.

In many viral infections, an important factor limiting virus spread is a programmed death of infected cells [50,51]. Although RSV protein F can trigger apoptosis [52], two non-structural proteins of RSV, NS1 and NS2 effectively block apoptosis [53]. This blockage may be only transient, but in our immunostaining images we see that for at least 48 h p.i. the foci of infected regions remain confluent, suggesting an insignificant apoptotic rate. The protective effect may be at least in part attributed to the observed sustained nuclear residence of plausibly anti-apoptotic NF- κ B. Since our model is calibrated for RSV based on our experimental data, it does not include infected cell death. As a consequence, the virus spread can be only slowed down (but not contained) by interferon signaling that mounts an antiviral state in not yet infected, bystander cells. Deaths of infected cells, if occurring during the early onset of infection, would change this picture through the elimination of infected cells before they produce or release new virions, as is the case during some other viral infections [54].

Finally, we note that IRF3 (independently of NF- κ B) induces the expression of several genes—including ISG56/IFIT1, ISG54/IFIT2, and ISG60/IFIT3—in an IFN-independent

manner [55]. IFIT1/2/3 are involved in the inhibition of RSV infection [56], whereas RSV NS1 impinges on the expression of IFIT2/3 [13,57]. These antagonistic interactions are not included in the model. NF- κ B, which is also activated in response to viral RNA, activates the transcription of proinflammatory interleukins IL-6 and IL-8 [58]. Their primary role is the regulation and attraction of immune cells, which are however absent both in our experiments and in our model.

In summary, single-cell experimental data and the proposed stochastic, agent-based model demonstrate the importance of reciprocally inhibitory interactions between the virus and the innate immune response. Because of the irreversible progression of virus replication (not limited by cell death), these interactions do not lead to typical bistability; instead, they result in divergent single-cell trajectories. A single infected cell may either rapidly initiate a cascade of new infections or become a transient source of IFN, allowing neighboring cells to enter an antiviral state and slow down virus spread. As a consequence, when the respiratory epithelium is infected by a small number of virions, heterogeneity of single-cell responses introduced by noise and mutually inhibiting interactions may lead to different outcomes of infection at the organismal level.

Methods

Experimental methods

Cell lines and cell culture. The A549, HeLa (used for RSV proliferation), and HEK-293T (used for the production of lentiviral particles) cell lines were purchased from ATCC. A549 cells were cultured in F-12K basal medium. HeLa and HEK-293T cells were maintained in Dulbecco's modified Eagle's medium (DMEM). Both media were supplemented with 10% FBS and a penicillin–streptomycin antibiotic solution. Cells were cultured under standard conditions (37°C with 5% CO₂) in a Series 8000 WJ incubator (Thermo Fisher Scientific) and kept in monolayers up to 90% confluency.

For STAT1 and STAT2 gene knockouts, a CRISPR lentiviral vector system was used. Commercially available sgRNAs sets were purchased from Applied Biological Materials (catalog numbers, K0002501 and K2300601, respectively). STAT1 knockout was obtained by targeting sequence TGATCCAAGCAAGCATTGGG in exon 9 of STAT1 gene; STAT2 knockout was obtained by targeting sequence AGCTCCCATTTGACCACGGGT in exon 8 of STAT2 gene. Two days after transfection of HEK-293T cells with plasmids encoding Cas9 nuclease (catalog number K002; Abm), lentivirus packaging particles (catalog number LV053; Abm), and mentioned earlier of each single sgRNA, the media with lentiviral particles were collected, enriched with 8 mg/mL Polybrene, and filtered through 0.45-mm syringe filters. Next, each lentiviral supernatant was used to transduce A549 wild-type cells, which were subcultured at low confluence (seeding density of 5×10^4 cells per 30-mm dish). After another 2 days, A549 cells were subjected to selection with 800 mg/mL of G418 for 10 days and seeded into a 96-well plate to obtain single-cell colonies. Clones were validated by Western blotting of p-STAT1 and p-STAT2 in response to IFN β . Additionally, the KO clones selected for further experiments were verified by sequencing.

RIG-I KO cells were purchased from InvivoGen (catalog number a549d-korigi). The A549 IRF3 KO and MAVS KO cell lines were obtained as described by Korwek et al. [7]. The A549 IFNAR1 KO, IFNLR1 KO, IFNAR1 and IFNLR1 double KO cell lines were obtained as described by Czerkies et al. [28].

RSV amplification and isolation. Respiratory syncytial virus strain A2 was purchased from ATCC and amplified in HeLa cells. The cells were seeded into 225-cm² tissue culture flasks (Falcon) and cultured as described above for 2 to 3 days until they reached 90%

confluence. The virus growth medium (DMEM with 2% FBS) and a virus dilution (with a target MOI of ~0.01) were prepared on the day of infection. The culture medium was removed and cells were washed once with phosphate-buffered saline (PBS) and overlaid with 10 mL of the inoculum. The virus was allowed to adsorb to cells for 2 h at 37°C, with occasional stirring. Next, additional virus growth medium was added to a total volume of 40 mL per flask. Infected cells were cultured for 3–6 days at 37°C until the development of cytopathic effects was observed in at least 80% of the cells. Virus-containing culture fluid was then collected and clarified by centrifugation at 3,000 g at 4°C for 20 min. Next, virus particles were precipitated by adding 50% (wt/vol) polyethylene glycol 6000 (PEG 6000) (Sigma-Aldrich) in NT buffer (150 mM NaCl, 50 mM Tris-HCl [pH 7.5]) to a final concentration of 10% and stirring the mixture gently for 90 min at 4°C. Virus was centrifuged at 3,250 g for 20 min at 4°C and re-centrifuged after the removal of the supernatant to remove the remaining fluid. The pellet was suspended in 1 mL 20% sucrose in the NT buffer, aliquoted, and stored at –80°C.

RSV quantification. The virus concentration was quantified using an immunofluorescence protocol. HeLa cells were seeded onto microscopic coverslips and cultured upon reaching 90 to 100% confluence. Serial dilutions of virus samples were prepared in the virus growth medium in a range of 10^{-3} to 10^{-7} . After washing with PBS, cells were overlaid with the diluted virus (for each virus dilution, two coverslips were used). The virus was allowed to adhere for 2 h and occasionally stirred. Afterward, the virus-containing medium was removed, and cells were overlaid with fresh virus growth medium and cultured for 24 h. Next, cells were washed with PBS and fixed with 4% formaldehyde for 20 min at room temperature. Cells were stained using a standard immunofluorescence protocol with an anti-RSV fusion glycoprotein antibody (catalog number ab43812; Abcam). Cells containing stained viral proteins were counted using a Leica TCS SP5 X confocal microscope. The virus concentration was calculated using the following formula: (average number of infected cells)/(dilution factor × volume containing virus added) = infectious particles/mL.

Compounds and stimulation protocols. Human IFNβ 1a was purchased from Thermo Fisher Scientific (catalog number PHC4244) and prepared according to the manufacturer's instructions. For cell stimulation, interferons were further diluted to the desired concentration in F-12K medium supplemented with 2% FBS. The decreased FBS content was used to prevent the inhibition of viral attachment and entry at the second stage of experiments. For interferon-and-virus experiments, the cell culture media were exchanged for interferon-containing or control media at time zero and were not removed until the end of the experiment. Appropriately diluted virus was added in small volumes (less than 10 μl) directly into the wells. The even distribution of the virus across the cell population was aided by the intermittent rocking of the plate for 2 h. For intracellular IFNβ visualization, a brefeldin A solution (catalog number 00-4506-51; Invitrogen) was added 2 h prior to cell fixation.

Western blotting. At the indicated time points, cells were washed twice with PBS, lysed in Laemmli sample buffer containing dithiothreitol (DTT), and boiled for 10 min at 95°C. Equal amounts of each protein sample were separated on 4% to 16% Mini-Protean TGX stain-free precast gels using the Mini-Protean Tetra cell electrophoresis system (Bio-Rad). Upon the completion of electrophoresis, proteins were transferred to a nitrocellulose membrane using wet electrotransfer in the Mini-Protean apparatus (400 mA for 50 min). The membrane was rinsed with TBST (Tris-buffered saline [TBS] containing 0.1% Tween 20; catalog number P7949, Sigma-Aldrich) and blocked for 1 h with 5% BSA–TBS or 5% nonfat dry milk. Subsequently, the membranes were incubated with one of the primary antibodies diluted in a 5% BSA–TBS buffer at 4°C overnight. After thorough washing with TBST, the membranes were incubated with secondary antibodies conjugated to a specific fluorochrome (DyLight 800; Thermo Fisher Scientific) or horseradish peroxidase (HRP-conjugated polyclonal anti-mouse/

anti-rabbit immunoglobulins; Dako) diluted 1:5,000 in 5% non-fat dry milk–TBST for 1 h at room temperature. The chemiluminescence reaction was developed with the Clarity Western ECL system (Bio-Rad). For GAPDH detection, hFAB rhodamine anti-GAPDH primary antibody (Bio-Rad) was used. Specific protein bands were detected using the ChemiDoc MP imaging system (Bio-Rad).

Immunostaining. For staining of intracellular proteins, cells were seeded onto 12-mm round glass coverslips, which were previously washed in 60% ethanol–40% HCl, thoroughly rinsed with water, and sterilized. After stimulation, cells on the coverslips were washed with PBS and immediately fixed with 4% formaldehyde (20 min at room temperature). Cells were then washed 3 times with PBS and incubated with 100% cold methanol for 20 min at -20°C . After washing with PBS, coverslips with cells were blocked and permeabilized for 1.5 h with 5% BSA (Sigma-Aldrich) with 0.3% Triton X-100 (Sigma-Aldrich) in PBS at room temperature. Subsequently, coverslips with cells were incubated with primary antibodies diluted in a blocking solution overnight at 4°C . Cells were then washed 5 times with PBS, and then appropriate secondary antibodies conjugated with fluorescent dyes were added in a blocking solution for 1 h at room temperature. Subsequently, cells were washed with PBS and their nuclei were stained with 200 ng/mL 4',6-diamidino-2-phenylindole (DAPI; Sigma-Aldrich) for 10 min. After a final wash in MilliQ water, coverslips with stained cells were mounted onto microscope slides with a Vectashield Vibrance Antifade Mounting Medium (Vector Laboratories). Multichannel fluorescence images were acquired with a Leica TCS SP5 X confocal microscope.

Antibodies. *Antibodies for Western blotting.* **Primary antibodies:** anti-phospho-STAT1 (Tyr701) (clone 58D6) (catalog number 9167; Cell Signaling Technologies; 1:1000); anti-phospho-STAT2 (Tyr690) (clone D3P2P) (catalog number 88410; Cell Signaling Technologies; 1:1000); anti-phospho-IRF3 (Ser396) (clone 4D4G) (catalog number 4947; Cell Signaling Technologies; 1:1000); anti-IRF3 (clone D6I4C) (catalog number 11904; Cell Signaling Technologies; 1:1000); anti-RIG-I (clone D14G6) (catalog number 3743; Cell Signaling Technologies; 1:1000); anti-STAT1 (catalog number 610116; BD Biosciences; 1:1000), anti-STAT2 (catalog number PAF-ST2; R&D Systems; 1:1000); anti-PKR (clone B-10) (catalog number sc-6282; Santa Cruz Biotechnology; 1:1000); anti-RNase L (clone E-9) (catalog number sc-74405; Santa Cruz Biotechnology; 1:1000); anti-OAS1 (clone F-3) (catalog number sc-374656; Santa Cruz Biotechnology; 1:1000), anti-Respiratory Syncytial Virus (clone 2F7) (catalog number ab43812; Abcam; 1:1000); hFAB Rhodamine anti-GAPDH (catalog number 12004168; Bio-Rad; 1:10,000).

Secondary antibodies: goat anti-rabbit IgG (H+L), DyLight 800 4X PEG (catalog number SA5-35571; Thermo Fisher Scientific; 1:10,000); goat anti-mouse IgG (H+L), DyLight 800 4X PEG (catalog number SA5-35521; Thermo Fisher Scientific; 1:10,000); StarBright Blue 700 goat anti-mouse IgG (catalog number 12004159; Bio-Rad; 1:10,000), rabbit anti-goat immunoglobulins/HRP (catalog number P0449; Agilent; 1:10,000).

Antibodies for immunostaining. **Primary antibodies:** anti-phospho-STAT1 Tyr701 (clone 58D6) (catalog number 9167; Cell Signaling Technologies; 1:1000); anti-IRF-3 (clone D-3) (catalog number sc-376455; Santa Cruz Biotechnology; 1:500), anti-Respiratory Syncytial Virus (catalog number ab20745; Abcam; 1:1000), anti-human interferon beta (catalog number MAB8142; R&D Systems; 1:100).

Secondary antibodies: donkey anti-rabbit IgG (H+L), Alexa Fluor 488 conjugate (catalog number A-21206; Thermo Fisher Scientific; 1:1000); donkey anti-mouse IgG (H+L), Alexa Fluor 555 conjugate (catalog number A-31570; Thermo Fisher Scientific; 1:1000); donkey anti-goat IgG (H+L), Alexa Fluor 633 conjugate (catalog number A-21082; Thermo Fisher Scientific; 1:1000).

Computational methods and modeling

Image quantification. Confocal images obtained from immunostaining were analyzed using our in-house software (<https://pmbm.ippt.pan.pl/software/shuttletracker>). Nuclear regions were detected based on DAPI staining. The nuclei that were partially out of frame or mitotic were excluded from the analysis; outlines of overlapping nuclei were split based on geometric convexity defects when possible. To characterize cells with respect to the levels of RSV proteins, nuclear IRF3 (as a proxy of p-IRF3), accumulated IFN β , and p-STAT1, we used image features that were visually checked and confirmed to adequately capture the immunofluorescence signal in a number of fields of view in several experiments. Specifically, cells were classified as: RSV proteins positive/negative—based on both the mean intensity of pixels of a perinuclear ring (after rejecting $\frac{1}{3}$ of the brightest and $\frac{1}{3}$ of the darkest pixels) and the sum of pixel intensities in the nuclear region and in the sigmoidally weighted nuclear halo around the nucleus in its corresponding Voronoi tile (2 features); p-IRF3 positive/negative—based on both the mean intensity of the nuclear region (after rejecting 10% brightest and 10% darkest pixels) and nuclear dominance (calculated as the difference between the mean nuclear and the mean perinuclear ring intensity normalized by the sum of these mean intensities, all calculated after rejecting 10% brightest and 10% darkest pixels; 2 features); IFN β positive/negative—based on both the mean intensity of the 50% brightest pixels in the perinuclear ring and the sum of pixel intensities in the nuclear region and in the sigmoidally weighted nuclear halo around the nucleus in its corresponding Voronoi tile (2 features); p-STAT1 was quantified as the mean intensity of the nuclear region (after rejecting 10% brightest and 10% darkest pixels; 1 feature, not subjected to thresholding). Thresholds used to binarize the features of RSV proteins, IRF3, and IFN β were selected to correctly identify cells visually identified as RSV proteins, p-IRF3, and IFN β positive/negative in a number of fields of view in several experiments.

To express the extent to which the two distributions are disjoint, we calculated the signed Kolmogorov–Smirnov (sKS) statistic. The absolute value of sKS is that of the standard Kolmogorov–Smirnov statistics, whereas the sign additionally informs about the relative locations of two (empirical) probability densities (shown in histograms as probability density functions, PDFs). We used the notation $\text{sKS}(X | Y)$ to express the sKS statistics for a pair of distributions of X (both continuous in the case of p-STAT1 or both binary otherwise), one obtained for Y -negative cells and the other obtained for Y -positive cells. The sKS statistic attains the values of ± 1 for fully disjointed distributions and 0 for exactly overlapping distributions.

Model justification and choice of parameters. The model structure is justified and model parameters selected based on a series of experiments involving IFN β stimulation and RSV infection of A549 cells, discussed in the following sections. Overall, parameter values were initially constrained based on cell-population experiments (Figs A–E in [S1 Appendix](#)), and then further manually adjusted based on single-cell data (Figs [3C](#), [4BC](#), and [5BC](#)). Finally, we tuned feedback strengths based on IFN β pre-stimulation experiments (Figs D, E in [S1 Appendix](#)) and single-cell data. Although the model parameters are constrained based on ample data collected in Figs A–E in [S1 Appendix](#) and Figs [3–5](#), our choice of parameters may not be unique [[59,60](#)].

When comparing model predictions with experimental data in Figs [3C](#), [4B,4C](#), [5B](#) and [5C](#) we additionally show two types of 95% credibility intervals (95% CrI). First, to estimate the effects of stochastic noise, we run the model 10,000 times for default parameter values on a 50×50 lattice containing $n = 2500$ cells. Second, to estimate the combined effect of stochastic noise and parameter variation, we run the model for log-normally ($\sigma = 0.2$) perturbed parameters, also on a 50×50 lattice. We observe that when two types of noise are taken into account, variability between experimental and numerical replicates is similar.

In Figs A–E in [S1 Appendix](#), overviewed below, we compare population-averaged model trajectories with respective population-based data.

Cascade initiated by RSV infection. In Fig A in [S1 Appendix](#), based on Western blotting and ELISA, we quantified levels of proteins in the cascade initiated by RSV infection at three MOIs (0.01, 0.1 and 1); in Fig B in [S1 Appendix](#), for MOI = 0.1, we analyzed how this cascade is affected by IRF3 knockout. In brief: RSV infection leads to phosphorylation of IRF3 (Fig A panel a in [S1 Appendix](#)), which colocalizes with NF- κ B in cell nuclei (Fig G in [S1 Appendix](#)), triggering synthesis and secretion of IFN β (Fig A panel c in [S1 Appendix](#)). There are several components that form the enhanceosome and are jointly required to initiate *Ifnb1* expression, namely IRF3, NF- κ B, and ATF-2/c-Jun complex [41,61]. As a simplification, in our model we included only IRF3, which localizes to cell nuclei with kinetics comparable to that of NF- κ B (as we show in Fig G in [S1 Appendix](#)). The two transcription factors are well colocalized at 24 h p.i. Secreted IFN β activates STAT1 and STAT2 (jointly represented by the accumulation of pSTATs in the model), which results in the accumulation of three proteins coded by interferon-stimulated genes: RIG-I, PKR, and OAS1 (collectively represented as ISGs in the model), as well as STAT1 and STAT2. In IRF3-deficient cells, we observed no IFN β production (Fig A panel c in [S1 Appendix](#)) and, consequently, no STAT1/2 phosphorylation and no accumulation of ISGs.

Dynamics of STATs and ISGs. The IFN β signaling is key to building an antiviral state in bystander cells, and thus we characterized in detail responses to IFN β in separate experiments. Parameters governing pSTATs kinetics were estimated based on Fig C panels a and c in [S1 Appendix](#). In Fig C panel a in [S1 Appendix](#), we found that in a broad range of IFN β stimulation doses (30 U/ml–1000 U/ml) STAT1 phosphorylation peaks and STAT2 phosphorylation plateaus 0.5 hour after stimulation. STAT2 phosphorylation monotonically increases with stimulation dose, reaching half of the maximal activation for about 100 U/ml, and this value determines the Michaelis–Menten constant in pSTATs heterodimers accumulation. Fig C panel c in [S1 Appendix](#) shows that one hour after IFN β washout STAT1 and STAT2 lose phosphorylation, but can be rapidly re-phosphorylated in response to the next IFN β pulse (which allows estimation of forward and reverse kinetic rate constants for pSTATs accumulation).

The forward and reverse kinetic rate constants for ISGs were estimated based on Fig C panel e in [S1 Appendix](#), in which we observe accumulation of RIG-I, PKR, and OAS1 over 24 hours of IFN β stimulation and their slow degradation over 48 hours after IFN β washout and STATs dephosphorylation.

Comparison of experimental p-STAT1 and p-STAT2 time profiles with pSTATs following from the model (Fig C panels b and d in [S1 Appendix](#)) indicates that the kinetic of p-STAT1 is more complex, but the assumed simplified kinetic of pSTATs which follows the kinetic of p-STAT2 is sufficient to reproduce the kinetics of ISGs (Fig C panel f in [S1 Appendix](#)).

Antiviral effect of IFN signaling. In Fig D panel a in [S1 Appendix](#) we studied the importance of IFN β and IFN λ for attenuating virus spread by knockouts of their respective receptors, IFNAR1 and IFNLR1. We observed that at 48 h p.i. there is significantly more RSV protein F in IFNAR1 KO cells than in WT cells. In turn, IFNLR1 knockout does not influence RSV protein accumulation, as indicated by a comparison of WT and IFNLR1 KO cells as well as of IFNAR1 KO cells and DKO cells (knockout of both IFNAR1 and IFNLR1). This indicates that IFN β is critical for attenuating RSV spread, and accordingly in our RSV model the ‘generic interferon’ can be identified with IFN β .

IFN β attenuates RSV spread by activating STAT1 and STAT2, which dimerize and trigger the expression of ISGs. Accordingly, both in STAT1 KO and STAT2 KO cells we did not observe accumulation of ISGs: RIG-I, PKR, and OAS1, and observed increased accumulation of RSV protein F (Fig D panel c in [S1 Appendix](#)). Stimulation of WT cells for 24 hours before

infection with IFN β results in a decrease of RSV protein F with respect to non-pre-stimulated cells, but this effect is negligible in STAT1 KO and STAT2 KO cells.

The effect of STAT1, STAT2, and interferon receptors knockouts on RSV propagation is reproduced by the model, as shown in Fig D panels b and d in [S1 Appendix](#). When comparing the model with experimental data from Fig D panels a and c in [S1 Appendix](#), we simulated IFNAR1 KO, DKO (with IFNAR1 and IFNLR1 knockouts), STAT1 KO and STAT2 KO cell by prohibiting ISGs from increasing (as in this case the p-STAT1/p-STAT2 dimers cannot be formed), while IFNLR1 KO is treated as WT (i.e., simulated with default parameters). Trajectories of p-STAT1 in STAT2 KO and p-STAT2 in STAT1 KO cells are assumed to be the same as in WT cells (as activation of STATs isoforms is independent), but, obviously, we assumed no increase of p-STAT1 in STAT1 KO cells and no increase of p-STAT2 in STAT2 KO cells.

The inhibitory effect of IFN β prestimulation on RSV propagation is shown using Western blotting (Fig E panel a in [S1 Appendix](#)) and dPCR (Fig E panel c in [S1 Appendix](#)). Generally, the influence of prestimulation is more pronounced for lower MOIs (where the level of RSV RNA and proteins depends on replication cycles inhibited by ISGs accumulated due to IFN β prestimulation). This tendency is reproduced by our model (Fig E panels b and c in [S1 Appendix](#)).

Model implementation. Agents (cells) are arranged on a triangular lattice (in which each node has 6 neighbors), which mimics a monolayer of epithelial cells (in which, based on Voronoi tessellation with geometric centers of cell nuclei used as seeds, each cell has on average 6 neighboring cells). Periodic boundary conditions are assumed along both axes. In all simulations, 100% confluency is assumed, meaning that all lattice nodes are occupied by cells.

For three variables: v_{RNA} , p_{IRF3} , and IFN_i the last state, 3, stands for active species, while the intermediate states {1, 2} are introduced to account for a time delay associated with activation or synthesis. For the remaining three intracellular species: v_{Proteins} , p_{STATs} , ISGs, the intermediate states {1, 2} stand for partial activation (which reflects the assumption of gradual accumulation or activation of these species). Of note, $v_{\text{Proteins}} = 1, 2, 3$ may inhibit the increase of IRF3 and IFN_i , but only $v_{\text{Proteins}} = 3$ renders the cell infectious.

The stochastic internal state of cells is coupled to an external, two-layer interferon field (see [Fig 1B](#)). The field evolves following the diffusion equation, solved using discretization introduced by cell-associated subvolumes. Cells with $\text{IFN}_i = 3$ secrete interferon to the lower layer. Activation of p_{STATs} follows the Michaelis–Menten kinetics, based on the amount of interferon in this layer. The upper layer accounts for the fact that in experimental settings interferon diffusion takes place in 3D (not 2D) space.

Unless specified otherwise, the model is initialized with all stages and interferon layers set to 0. Infected cells are selected randomly with probability dictated by MOI and their v is set to 1. Prestimulation with IFN sets the upper layer to the specified concentration; a wash sets both layers to 0. A typical simulation (100×100 grid, 100% cell confluence, IFN field update every 0.1 min) with default parameters and MOI 0.1 for 24 hours of simulated time takes a few seconds of CPU time on a standard PC.

The model does not account for cell movement, and only infection of neighboring cells is allowed. This is consistent with microscopy images where we observe growing clusters of infected cells (see [Fig 2](#)). The model also does not account for apoptosis of infected cells, as it is known that apoptosis is blocked by RSV [62]. In our experiments we see occasional formation of syncytia (see [Fig 2A](#)), which allows for a direct cell-to-cell transfer of both viral RNA and proteins. Instead of modeling syncytia formation, we have tuned our model to reflect the observed rate of infecting neighboring cells.

To report population-averaged variables, 1000 stochastic simulations were performed on the 100×100 grid, with the exception of Figs [3D](#) and [7](#), and Figs A–E, Fig I, and Fig J panel a

in [S1 Appendix](#), for which 100 stochastic simulations were performed. The estimations of credible intervals ([Fig 3C](#)) were based on 10,000 stochastic simulations.

Comparison of experimental and simulation results. MOI and the time post infection are two parameters that in principle should determine the state of infection. Because of unavoidable differences between experimental replicates in the number of active virus particles actually infecting cells (despite the same MOI), we observe that using the observed fraction of cells expressing RSV proteins in a given time post infection (instead of MOI) allows for more reliable comparisons across experiments, and with model simulations.

Neighboring cells are well defined for simulations (on the model lattice each cell has 6 neighbors). In order to compare simulation results with experimental data, we define neighbors by constructing Voronoi diagrams for microscope images. We identify cell nuclei using DAPI staining and use their centers of mass as Voronoi centers. We then consider cells to be neighbors if their respective Voronoi tiles share an edge. Conveniently, this procedure assigns on average 6 neighbors to each cell, the same as in our model.

In immunostaining images, levels of observed species $IFN\beta$, nuclear IRF3, and RSV proteins are discretized into binary variables (quantification and discretization of immunostaining images are described in the previous subsection Image quantification). We perform a similar binarization in the model when comparing with immunostaining images. In the case of $IFN\alpha$ and $pIRF3$, for which the states $\{1, 2\}$ are introduced solely to account for production delay, only their last state, 3 is considered active. In the case of $vProteins$ intermediate states denote the accumulation of viral proteins, and thus all states $vProteins > 0$ are considered active. In the analysis of experimental data, we leave $p-STAT1$ as a continuous variable, while in the model we use a 4-state discretization. Importantly, this means that for low levels of $p-STAT1$, small differences might be picked up in experimental data but are not accounted for in the model.

To compare Western blots with the model simulation data, for a given simulation protocol we performed 100 simulations on a 100×100 lattice, and then average over all cells and simulations. We assumed that the levels of $p-STAT1$ and $p-STAT2$ are proportional to the average activation level for $pSTATs$, the levels of RIG-I, PKR, and OAS1 are proportional to the average activation level of ISGs, and the level of RSV protein F is proportional to the average activation level of $vProteins$. We also assumed that the level of $p-IRF3$ is proportional to the fraction of cells in which $pIRF3 = 3$ (as in this case states 1 and 2 are introduced solely to account for time delay associated with IRF3 activation). Western blots were first normalized by the corresponding reference GAPDH, then by the maximum value in the series of measurements (from the same blot). Next, we added 0.03 to each value and finally normalized the series by the geometric mean. This procedure reflects the assumption that blots adequately capture fold differences between measurements and that differences between values smaller than 0.03 of the maximum from a given blot are not meaningful. The same normalization procedure (without the reference GAPDH step) is applied to simulation data. Overall, this approach yields a dynamic range that accommodates up to ~30-fold protein level differences.

ELISA measurements ([Fig A panel c in S1 Appendix](#)) are compared with $IFN\alpha$ summed over upper and lower compartments and averaged over cells. dPCR measurements for RSV RNA ([Fig E panel c in S1 Appendix](#)) are compared with the model based on the simplifying assumption that the average level of viral RNA is proportional to the fraction of cells with $vRNA = 3$ (for $vRNA$ states 1 and 2 account solely for time delay).

Supporting information

S1 Appendix. The Appendix PDF file containing Figs A–J and Table A. (PDF)

S1 Dataset. ELISA data for Fig A panel c in [S1 Appendix](#).
(XLSX)

S2 Dataset. Digital PCR data for Fig E panel c in [S1 Appendix](#).
(XLSX)

S3 Dataset. Uncropped Western blot images for Fig A panel a, Fig B panel a, Fig C panel e, Fig D panel c, Fig E panel a in [S1 Appendix](#).
(ZIP)

Author Contributions

Conceptualization: Frederic Grabowski, Marek Kochańczyk, Tomasz Lipniacki.

Data curation: Frederic Grabowski, Marek Kochańczyk.

Formal analysis: Frederic Grabowski.

Funding acquisition: Tomasz Lipniacki.

Investigation: Zbigniew Korwek, Maciej Czerkies.

Methodology: Zbigniew Korwek, Wiktor Prus.

Software: Frederic Grabowski, Marek Kochańczyk.

Writing – original draft: Frederic Grabowski, Marek Kochańczyk, Tomasz Lipniacki.

Writing – review & editing: Frederic Grabowski, Marek Kochańczyk, Zbigniew Korwek, Maciej Czerkies, Wiktor Prus, Tomasz Lipniacki.

References

1. Czerkies M, Korwek Z, Prus W, Kochańczyk M, Jaruszewicz-Błońska J, Tudelska K, et al. Cell fate in antiviral response arises in the crosstalk of IRF, NF- κ B and JAK/STAT pathways. *Nat Commun*. 2018; 9: 493. <https://doi.org/10.1038/s41467-017-02640-8> PMID: 29402958
2. Patil S, Fribourg M, Ge Y, Batish M, Tyagi S, Hayot F, et al. Single-cell analysis shows that paracrine signaling by first responder cells shapes the interferon- β response to viral infection. *Science Signaling*. 2015; 8: ra16–ra16. <https://doi.org/10.1126/scisignal.2005728> PMID: 25670204
3. Rand U, Rinas M, Schwerk J, Nöhren G, Linnes M, Kröger A, et al. Multi-layered stochasticity and paracrine signal propagation shape the type-I interferon response. *Molecular Systems Biology*. 2012; 8: 584–584. <https://doi.org/10.1038/msb.2012.17> PMID: 22617958
4. Schoggins JW. Interferon-Stimulated Genes: What Do They All Do? *Annu Rev Virol*. 2019; 6: 567–584. <https://doi.org/10.1146/annurev-virology-092818-015756> PMID: 31283436
5. Zhao M, Zhang J, Phatnani H, Scheu S, Maniatis T. Stochastic expression of the interferon- β gene. *PLoS Biology*. 2012; 10: e1001249.
6. Drayman N, Patel P, Vistain L, Tay S. HSV-1 single-cell analysis reveals the activation of anti-viral and developmental programs in distinct sub-populations. *eLife*. 2019; 8: e46339. <https://doi.org/10.7554/eLife.46339> PMID: 31090537
7. Korwek Z, Czerkies M, Jaruszewicz-Błońska J, Prus W, Kosiuk I, Kochańczyk M, et al. Non-self RNA rewires IFN β signaling: A mathematical model of the innate immune response. *bioRxiv*; 2023. p. 2022.01.30.478391. <https://doi.org/10.1101/2022.01.30.478391>
8. Sen A, Rothenberg ME, Mukherjee G, Feng N, Kalisky T, Nair N, et al. Innate immune response to homologous rotavirus infection in the small intestinal villous epithelium at single-cell resolution. *Proc Natl Acad Sci USA*. 2012; 109: 20667–20672. <https://doi.org/10.1073/pnas.1212188109> PMID: 23188796
9. García-Sastre A. Ten Strategies of Interferon Evasion by Viruses. *Cell Host Microbe*. 2017; 22: 176–184. <https://doi.org/10.1016/j.chom.2017.07.012> PMID: 28799903
10. Ayllon J, García-Sastre A. The NS1 protein: a multitasking virulence factor. *Curr Top Microbiol Immunol*. 2015; 386: 73–107. https://doi.org/10.1007/82_2014_400 PMID: 25007846

11. Hale BG, Albrecht RA, García-Sastre A. Innate immune evasion strategies of influenza viruses. *Future Microbiol.* 2010; 5: 23–41. <https://doi.org/10.2217/fmb.09.108> PMID: 20020828
12. Barik S. Respiratory syncytial virus mechanisms to interfere with type 1 interferons. *Curr Top Microbiol Immunol.* 2013; 372: 173–191. https://doi.org/10.1007/978-3-642-38919-1_9 PMID: 24362690
13. Pei J, Beri NR, Zou AJ, Hubel P, Dorando HK, Bergant V, et al. Nuclear-localized human respiratory syncytial virus NS1 protein modulates host gene transcription. *Cell Rep.* 2021; 37: 109803. <https://doi.org/10.1016/j.celrep.2021.109803> PMID: 34644581
14. Sedeyn K, Schepens B, Saelens X. Respiratory syncytial virus nonstructural proteins 1 and 2: Exceptional disrupters of innate immune responses. *PLoS Pathog.* 2019; 15: e1007984. <https://doi.org/10.1371/journal.ppat.1007984> PMID: 31622448
15. Rashid F, Xie Z, Suleman M, Shah A, Khan S, Luo S. Roles and functions of SARS-CoV-2 proteins in host immune evasion. *Front Immunol.* 2022; 13: 940756. <https://doi.org/10.3389/fimmu.2022.940756> PMID: 36003396
16. Sa Ribero M, Jouvenet N, Dreux M, Nisole S. Interplay between SARS-CoV-2 and the type I interferon response. *PLoS Pathog.* 2020; 16: e1008737. <https://doi.org/10.1371/journal.ppat.1008737> PMID: 32726355
17. Vignuzzi M, López CB. Defective viral genomes are key drivers of the virus-host interaction. *Nat Microbiol.* 2019; 4: 1075–1087. <https://doi.org/10.1038/s41564-019-0465-y> PMID: 31160826
18. Talemi SR, Höfer T. Antiviral interferon response at single-cell resolution. *Immunol Rev.* 2018; 285: 72–80. <https://doi.org/10.1111/imr.12699> PMID: 30129203
19. Van Eyndhoven LC, Singh A, Tel J. Decoding the dynamics of multilayered stochastic antiviral IFN-I responses. *Trends Immunol.* 2021; 42: 824–839. <https://doi.org/10.1016/j.it.2021.07.004> PMID: 34364820
20. Maier BD, Aguilera LU, Sahle S, Mutz P, Kalra P, Dächert C, et al. Stochastic dynamics of type-I interferon responses. *PLOS Comput Biol.* 2022; 18: e1010623. <https://doi.org/10.1371/journal.pcbi.1010623> PMID: 36269758
21. Gregg RW, Shabnam F, Shoemaker JE. Agent-based modeling reveals benefits of heterogeneous and stochastic cell populations during cGAS-mediated IFN β production. *Bioinformatics.* 2021; 37: 1428–1434. <https://doi.org/10.1093/bioinformatics/btaa969> PMID: 33196784
22. Aponte-Serrano JO, Weaver JJA, Sego TJ, Glazier JA, Shoemaker JE. Multicellular spatial model of RNA virus replication and interferon responses reveals factors controlling plaque growth dynamics. *PLoS Comput Biol.* 2021; 17: e1008874. <https://doi.org/10.1371/journal.pcbi.1008874> PMID: 34695114
23. Sego TJ, Aponte-Serrano JO, Gianlupi JF, Glazier JA. Generation of multicellular spatiotemporal models of population dynamics from ordinary differential equations, with applications in viral infection. *BMC Biol.* 2021; 19: 196. <https://doi.org/10.1186/s12915-021-01115-z> PMID: 34496857
24. Sego TJ, Mochan ED, Ermentrout GB, Glazier JA. A multiscale multicellular spatiotemporal model of local influenza infection and immune response. *J Theor Biol.* 2022; 532: 110918. <https://doi.org/10.1016/j.jtbi.2021.110918> PMID: 34592264
25. Crowe JE. Human Respiratory Viruses. Reference Module in Biomedical Sciences. 2014; B978-0-12-801238-3.02600–3. <https://doi.org/10.1016/B978-0-12-801238-3.02600-3>
26. Falsey AR, Hennessey PA, Formica MA, Cox C, Walsh EE. Respiratory syncytial virus infection in elderly and high-risk adults. *N Engl J Med.* 2005; 352: 1749–1759. <https://doi.org/10.1056/NEJMoa043951> PMID: 15858184
27. Scheltema NM, Gentile A, Lucion F, Nokes DJ, Munywoki PK, Madhi SA, et al. Global respiratory syncytial virus-associated mortality in young children (RSV GOLD): a retrospective case series. *Lancet Glob Health.* 2017; 5: e984–e991. [https://doi.org/10.1016/S2214-109X\(17\)30344-3](https://doi.org/10.1016/S2214-109X(17)30344-3) PMID: 28911764
28. Czerkies M, Kochańczyk M, Korwek Z, Prus W, Lipniacki T. Respiratory syncytial virus protects bystander cells against influenza A virus infection by triggering secretion of type I and type III Interferons. *J Virol.* 2022; 96: e0134122. <https://doi.org/10.1128/jvi.01341-22> PMID: 36326278
29. Mesev EV, LeDesma RA, Ploss A. Decoding type I and III interferon signalling during viral infection. *Nat Microbiol.* 2019; 4: 914–924. <https://doi.org/10.1038/s41564-019-0421-x> PMID: 30936491
30. Ban J, Lee N-R, Lee N-J, Lee JK, Quan F-S, Inn K-S. Human respiratory syncytial virus NS 1 targets TRIM25 to suppress RIG-I ubiquitination and subsequent RIG-I-mediated antiviral signaling. *Viruses.* 2018; 10: 716. <https://doi.org/10.3390/v10120716> PMID: 30558248
31. Boyapalle S, Wong T, Garay J, Teng M, San Juan-Vergara H, Mohapatra S, et al. Respiratory syncytial virus NS1 protein colocalizes with mitochondrial antiviral signaling protein MAVS following infection. *PLoS One.* 2012; 7: e29386. <https://doi.org/10.1371/journal.pone.0029386> PMID: 22383950

32. Ren J, Liu T, Pang L, Li K, Garofalo RP, Casola A, et al. A novel mechanism for the inhibition of interferon regulatory factor-3-dependent gene expression by human respiratory syncytial virus NS1 protein. *J Gen Virol*. 2011; 92: 2153–2159. <https://doi.org/10.1099/vir.0.032987-0> PMID: 21632562
33. Senft AP, Taylor RH, Lei W, Campbell SA, Tipper JL, Martinez MJ, et al. Respiratory syncytial virus impairs macrophage IFN- α /beta- and IFN- γ -stimulated transcription by distinct mechanisms. *Am J Respir Cell Mol Biol*. 2010; 42: 404–414. <https://doi.org/10.1165/rcmb.2008-0229OC> PMID: 19502390
34. Elliott J, Lynch OT, Suessmuth Y, Qian P, Boyd CR, Burrows JF, et al. Respiratory syncytial virus NS1 protein degrades STAT2 by using the elongin-cullin E3 ligase. *J Virol*. 2007; 81: 3428–3436. <https://doi.org/10.1128/JVI.02303-06> PMID: 17251292
35. Lo MS, Brazas RM, Holtzman MJ. Respiratory syncytial virus nonstructural proteins NS1 and NS2 mediate inhibition of Stat2 expression and alpha/beta interferon responsiveness. *J Virol*. 2005; 79: 9315–9319. <https://doi.org/10.1128/JVI.79.14.9315-9319.2005> PMID: 15994826
36. Ramaswamy M, Shi L, Varga SM, Barik S, Behlke MA, Look DC. Respiratory syncytial virus nonstructural protein 2 specifically inhibits type I interferon signal transduction. *Virology*. 2006; 344: 328–339. <https://doi.org/10.1016/j.virol.2005.09.009> PMID: 16216295
37. Snijder B, Sacher R, R  m   P, Damm E-M, Liberali P, Pelkmans L. Population context determines cell-to-cell variability in endocytosis and virus infection. *Nature*. 2009; 461: 520–523. <https://doi.org/10.1038/nature08282> PMID: 19710653
38. Sun Y, Jain D, Koziol-White CJ, Genoyer E, Gilbert M, Tapia K, et al. Immunostimulatory defective viral genomes from respiratory syncytial virus promote a strong innate antiviral response during infection in mice and humans. *PLoS Pathog*. 2015; 11: e1005122. <https://doi.org/10.1371/journal.ppat.1005122> PMID: 26336095
39. Seth RB, Sun L, Ea C-K, Chen ZJ. Identification and characterization of MAVS, a mitochondrial antiviral signaling protein that activates NF- κ B and IRF 3. *Cell*. 2005; 122: 669–682. <https://doi.org/10.1016/j.cell.2005.08.012> PMID: 16125763
40. Liu S, Cai X, Wu J, Cong Q, Chen X, Li T, et al. Phosphorylation of innate immune adaptor proteins MAVS, STING, and TRIF induces IRF3 activation. *Science*. 2015; 347: aaa2630. <https://doi.org/10.1126/science.aaa2630> PMID: 25636800
41. Wathelet MG, Lin CH, Parekh BS, Ronco LV, Howley PM, Maniatis T. Virus infection induces the assembly of coordinately activated transcription factors on the IFN- β enhancer in vivo. *Molecular Cell*. 1998; 1: 507–518.
42. Hatesuer B, Hoang HTT, Riese P, Trittel S, Gerhauser I, Elbahesh H, et al. Deletion of Irf3 and Irf7 genes in mice results in altered interferon pathway activation and granulocyte-dominated inflammatory responses to influenza A infection. *J Innate Immun*. 2017; 9: 145–161. <https://doi.org/10.1159/000450705> PMID: 27811478
43. Goritzka M, Durant LR, Pereira C, Salek-Ardakani S, Openshaw PJM, Johansson C. Alpha/beta interferon receptor signaling amplifies early proinflammatory cytokine production in the lung during respiratory syncytial virus infection. *J Virol*. 2014; 88: 6128–6136. <https://doi.org/10.1128/JVI.00333-14> PMID: 24648449
44. Crotta S, Davidson S, Mahlakoiv T, Desmet CJ, Buckwalter MR, Albert ML, et al. Type I and type III interferons drive redundant amplification loops to induce a transcriptional signature in influenza-infected airway epithelia. *PLoS Pathog*. 2013; 9: e1003773. <https://doi.org/10.1371/journal.ppat.1003773> PMID: 24278020
45. Meyts I, Casanova J-L. Viral infections in humans and mice with genetic deficiencies of the type I IFN response pathway. *Eur J Immunol*. 2021; 51: 1039–1061. <https://doi.org/10.1002/eji.202048793> PMID: 33729549
46. Voigt EA, Swick A, Yin J. Rapid induction and persistence of paracrine-induced cellular antiviral states arrest viral infection spread in A549 cells. *Virology*. 2016; 496: 59–66. <https://doi.org/10.1016/j.virol.2016.05.019> PMID: 27254596
47. Van Eyndhoven LC, Verberne VPG, Bouten CVC, Singh A, Tel J. Transiently heritable fates and quorum sensing drive early IFN-I response dynamics. *Elife*. 2023; 12: e83055. <https://doi.org/10.7554/eLife.83055> PMID: 36629318
48. Wimmers F, Subedi N, van Buuringen N, Heister D, Viv   J, Beeren-Reinieren I, et al. Single-cell analysis reveals that stochasticity and paracrine signaling control interferon- α production by plasmacytoid dendritic cells. *Nat Commun*. 2018; 9: 3317. <https://doi.org/10.1038/s41467-018-05784-3> PMID: 30127440
49. Groskreutz DJ, Babor EC, Monick MM, Varga SM, Hunninghake GW. Respiratory syncytial virus limits alpha subunit of eukaryotic translation initiation factor 2 (eIF2 α) phosphorylation to maintain

- translation and viral replication. *J Biol Chem*. 2010; 285: 24023–24031. <https://doi.org/10.1074/jbc.M109.077321> PMID: 20519500
50. Barber GN. Host defense, viruses and apoptosis. *Cell Death Differ*. 2001; 8: 113–126. <https://doi.org/10.1038/sj.cdd.4400823> PMID: 11313713
51. Orzalli MH, Kagan JC. Apoptosis and Necroptosis as Host Defense Strategies to Prevent Viral Infection. *Trends Cell Biol*. 2017; 27: 800–809. <https://doi.org/10.1016/j.tcb.2017.05.007> PMID: 28642032
52. Eckardt-Michel J, Lorek M, Baxmann D, Grunwald T, Keil GM, Zimmer G. The fusion protein of respiratory syncytial virus triggers p53-dependent apoptosis. *J Virol*. 2008; 82: 3236–3249. <https://doi.org/10.1128/JVI.01887-07> PMID: 18216092
53. Bitko V, Shulyayeva O, Mazumder B, Musiyenko A, Ramaswamy M, Look DC, et al. Nonstructural proteins of respiratory syncytial virus suppress premature apoptosis by an NF-kappaB-dependent, interferon-independent mechanism and facilitate virus growth. *J Virol*. 2007; 81: 1786–1795. <https://doi.org/10.1128/JVI.01420-06> PMID: 17151097
54. Upton JW, Shubina M, Balachandran S. RIPK3-driven cell death during virus infections. *Immunol Rev*. 2017; 277: 90–101. <https://doi.org/10.1111/imr.12539> PMID: 28462524
55. Grandvaux N, Servant MJ, tenOever B, Sen GC, Balachandran S, Barber GN, et al. Transcriptional profiling of interferon regulatory factor 3 target genes: direct involvement in the regulation of interferon-stimulated genes. *J Virol*. 2002; 76: 5532–5539. <https://doi.org/10.1128/jvi.76.11.5532-5539.2002> PMID: 11991981
56. Drori Y, Jacob-Hirsch J, Pando R, Glatman-Freedman A, Friedman N, Mendelson E, et al. Influenza A Virus Inhibits RSV Infection via a Two-Wave Expression of IFIT Proteins. *Viruses*. 2020; 12: 1171. <https://doi.org/10.3390/v12101171> PMID: 33081322
57. Van Royen T, Sedeyn K, Moschonas GD, Toussaint W, Vuylsteke M, Van Haver D, et al. An unexpected encounter: Respiratory syncytial virus nonstructural protein 1 interacts with mediator subunit MED25. *J Virol*. 2022; 96: e0129722. <https://doi.org/10.1128/jvi.01297-22> PMID: 36102648
58. Matsusaka T, Fujikawa K, Nishio Y, Mukaida N, Matsushima K, Kishimoto T, et al. Transcription factors NF-IL6 and NF-kappa B synergistically activate transcription of the inflammatory cytokines, interleukin 6 and interleukin 8. *Proc Natl Acad Sci U S A*. 1993; 90: 10193–10197. <https://doi.org/10.1073/pnas.90.21.10193> PMID: 8234276
59. Grabowski F, Nałęcz-Jawecki P, Lipniacki T. Predictive power of non-identifiable models. *bioRxiv*; 2023. p. 2023.04.07.536025. <https://doi.org/10.1038/s41598-023-37939-8> PMID: 37429934
60. Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP. Universally sloppy parameter sensitivities in systems biology models. *PLoS Computational Biology*. 2007; 3: e189. <https://doi.org/10.1371/journal.pcbi.0030189> PMID: 17922568
61. Panne D, Maniatis T, Harrison SC. An atomic model of the interferon-beta enhanceosome. *Cell*. 2007; 129: 1111–1123. <https://doi.org/10.1016/j.cell.2007.05.019> PMID: 17574024
62. Thomas KW, Monick MM, Staber JM, Yarovinsky T, Carter AB, Hunninghake GW. Respiratory syncytial virus inhibits apoptosis and induces NF-kappa B activity through a phosphatidylinositol 3-kinase-dependent pathway. *J Biol Chem*. 2002; 277: 492–501. <https://doi.org/10.1074/jbc.M108107200> PMID: 11687577

Type III interferons suppress influenza A virus infection independently of STAT activation by triggering cell death

Wiktor Prus¹, Frederic Grabowski¹, Paulina Koza¹, Zbigniew Korwek¹,
Maciej Czerkies¹, Marek Kochańczyk¹, Tomasz Lipniacki^{1,*}

¹ Department of Biosystems and Soft Matter, Institute of Fundamental Technological Research, Polish Academy of Sciences, Warsaw, Poland.

* Corresponding author (e-mail: tlipnia@ippt.pan.pl).

Type III interferons (IFN- λ 1– λ 4) are known to limit influenza infections *in vivo* and are non-redundant to type I interferons (IFN- α and IFN- β). Here, we demonstrated *in vitro* that type III interferons secreted by infected epithelial cells limit infection with influenza A virus (IAV). This effect occurs independently of STAT1 and STAT2 activation. Knockout of the IFN- λ receptor subunit IFNLR1 results in a slight reduction of STAT1/2 phosphorylation during infection with IAV, but leads to a substantially increased fraction of IAV-infected cells and higher levels of viral RNA and protein in the cell population. In contrast, knockout of the IFN- β receptor subunit IFNAR1 results in a pronounced reduction of STAT1/2 phosphorylation but a smaller increase in IAV proliferation. We propose that type III interferons limit the spread of IAV by promoting death of IAV-infected cells and demonstrate that the ratio of dying to infected cells is lower in IFNLR1 KO cells compared to wild-type cells. Overall, our results suggest an additional non-transcriptional role of type III interferons in the control of viral infection.

Keywords: innate immunity / cell signaling / interferons / respiratory viruses / cell death

Key points:

- IFNAR1 deficiency leads to a significant reduction of STAT1/2 activation in cell populations infected with IAV and a small decrease in IAV proliferation.
- IFNLR1 deficiency leads to a small reduction of STAT1/2 activation in cell populations infected with IAV and a significant increase in IAV proliferation.
- IFN- λ controls the proliferation of IAV (but not RSV) independently of STAT1/2 signaling.
- IFN- λ signaling increases the ratio of the dead to the infected cells, likely promoting death in IAV-infected cells.

Introduction

The discovery of type III interferons significantly expanded our awareness of the complexity of anti-viral mechanisms in innate immunity (Kotenko *et al*, 2003; Sheppard *et al*, 2003). This novel cytokine family, consisting of interferons $\lambda 1$, $\lambda 2$, $\lambda 3$, and $\lambda 4$, was initially believed to be mostly supplementary or even redundant to a much better described and understood family of type I interferons (represented importantly by interferons α and β , but also including several others). This was due to similarities between both cytokine families in the mechanisms of their production, recognition, activation of signaling pathways and downstream expression of target genes. Their production is triggered by recognition of pathogen particles by the same cellular pathogen recognition receptors and is dependent on activation of transcription factors NF- κ B and IRF3/7 (Osterlund *et al*, 2007). Type I and type III interferons bind to their cognate receptors, IFNAR1/2 and IFNLR1/IL10R2, respectively, to activate the JAK/STAT pathway, which mediates phosphorylation of STAT1 and STAT2 (collectively referred to as STAT1/2) (Mesev *et al*, 2019). These two transcription factors form a complex with IRF9 and initiate expression of the so-called interferon-stimulated genes (ISGs), which allow the cell to enter a state of increased viral resistance and help activate adaptive immune responses (Schoggins, 2019). These common characteristics of type I and III interferons are generally not shared by type II interferon, IFN- γ , which exhibits a distinct immunomodulatory mode of action and is secreted mostly by specialized immune cells (Schoenborn & Wilson, 2007).

Despite their similarities, accumulating *in vivo* evidence suggests that the roles of type I and III interferons in response to viral infections are partially divergent, with IFN- λ playing a significant role in containing influenza A virus (IAV) infections. In animal studies, Galani *et al*, 2017 demonstrated that IFN- λ suppresses the initial viral spread, ensuring optimal viral clearance, while, in the case of high-dose infections, type I interferons are activated later to enhance antiviral and pro-inflammatory responses. Accordingly, IFNLR1-deficient mice display enhanced type I IFN production, and lethality. Klinkhammer *et al*, 2018 showed that IFN- λ , not type I interferons, conferred long-lasting antiviral protection in the upper airways against influenza virus transmission. A more significant role of type III interferons compared to type I interferons was also observed in mouse models for other viruses, including rotavirus (Pott *et al*, 2011), norovirus (Nice *et al*, 2015).

The non-redundant and significant role of IFN- λ is puzzling. Despite single reports of type III interferons uniquely inducing certain genes in specific cell lines: *Mmp7*, *Serp1b1a*, and *Csprs* in mouse IEC cells by IFN- $\lambda 1$ (Selvakumar *et al*, 2017); or *CXCL10*, *CXCL11*, *IFIT3*, *IFI30*, and *TDRD7* in HECV cells also by IFN- $\lambda 1$ (Caine *et al*, 2019), transcriptome profiling

studies have established that the type I and type III signatures are overlapping, and the ISGs induced by type III IFNs are a subset of those induced by type I IFNs (Lazear *et al*, 2019 and references therein). The major difference is in kinetics and magnitude of expression of genes stimulated by type I and III interferons – strong and rapid in the first group, weaker but more sustained in the second group. In their review, Lazear *et al*, 2019 characterizes type I *versus* type III interferons as having respectively high/low potency, rapid/slower kinetics, inflammatory/less inflammatory, systemic/restricted by anatomic barriers. Observations indicate that the expression of IFNLR is generally limited to epithelial surfaces of the airways, gut, liver, urinary tract and the blood–brain barrier, being especially high in the airway epithelium, which contrasts with a much more ubiquitous expression of IFNAR (Lazear *et al*, 2015; Wack *et al*, 2015). Such localized expression of IFNLR predisposes type III interferons to play a major role in combating infections with respiratory viruses. Thus to some extent the distinct role of type III interferons observed in animal models can be attributed to differences in their expression and the expression of their cognate receptor.

In this study, we show that IFN- λ plays a distinct and significant role in limiting infection with IAV, but not with respiratory syncytial virus (RSV), in respiratory epithelial cells *in vitro*. As expected, we observed that prestimulation with IFN- β leads to higher accumulation of ISGs than prestimulation with IFN- λ 1, resulting in a stronger inhibition of both IAV and RSV proliferation. Upon viral infection, activation of STAT1/2 is dramatically reduced in IFNAR1-deficient but not in IFNLR1-deficient cells, whereas, surprisingly, proliferation of IAV (but not of RSV) is better facilitated by the deficiency of IFNLR1 compared to the deficiency of IFNAR1. Since IFNLR1 has no ligands other than type III interferons (Peterson *et al*, 2019), this indicates that IFN- λ confers protection against IAV through mechanisms beyond JAK/STAT signaling. We observed that the ratio of infected to dead cells is higher in populations of IFNLR1-deficient cells compared to wild-type (WT) cells, indicating that IFN- λ may also play a role in promoting the death of infected cells, thereby limiting viral spread.

Results

Stimulation with IFN- β elicits stronger STAT1/2 response and higher ISG accumulation compared to stimulation with IFN- λ 1

Type III interferons are relatively weak inducers of STAT signaling in A549 cells. As shown in Figs. 1A and 1B, stimulation with IFN- λ 1 (at its saturation concentration of 50 ng/ml) results in lower levels of phosphorylated STAT1/2 (p-STAT1/2) after 1 h, after intermediate times (2 and 6 h), and after longer times (24 and 48 h) compared to stimulation with IFN- β (also at its saturation concentration of 1000 U/ml; saturation concentrations for IFN- λ 1 and IFN- β were

determined in Fig. EV1; all replicates of Western blots used in this study are shown in supplementary Source Data S1). Consequently, stimulation with IFN- λ 1 compared to IFN- β leads to a lower accumulation of STAT1, STAT2, and three important ISG-coded proteins: RIG-I, PKR, and OAS1 (Figs. 1A and 1B). Similarly, mRNA levels are significantly lower after stimulation with IFN- λ 1 compared to IFN- β for 15 out of 16 analyzed ISGs important in the regulation of the innate immune response (Fig. 1C).

Prestimulation with IFN- β inhibits virus proliferation more markedly than prestimulation with IFN- λ 1

In our previous study, we have demonstrated that a 24-hour prestimulation with IFN- β reduces IAV RNA count by approximately 300-fold and RSV RNA count by about 5–10-fold (Czerkies et al. 2022, Fig 2A and 2B therein). Consistent with the weaker activation of ISGs by IFN- λ 1, we have found that a prestimulation with IFN- λ 1 has a lesser effect; it reduces IAV RNA count by about 5–10-fold, while the reduction for RSV RNA is insignificant (Czerkies et al. 2022, Figs. 2A and 2B therein). The effect is similar at multiplicity of infection (MOI) of 0.1 and 1. The difference between IFN- β and IFN- λ 1 is also evident in the number of cells infected with IAV: the prestimulation with IFN- β leads to an about 15-fold reduction in the infected cell count, whereas the prestimulation IFN- λ 1 leads to an about 4-fold reduction (Czerkies et al. 2022, Fig. 4 therein).

Here, we confirm the more substantial effect of prestimulation with IFN- β compared to IFN- λ 1 (for infections with either IAV or RSV) by Western blotting. In the case of an infection with IAV (Figs. 2A and 2B), a 24-hour prestimulation with IFN- β (compared to mock) leads to markedly higher levels of p-STAT1/2, and, consequently, to higher accumulation of STAT1/2 and ISGs observed at 24 and 48 h post infection (p.i.). This results in drastically lower IAV nucleoprotein (NP) levels, below the detection threshold for all combinations of the MOI and time p.i. The effect of IFN- λ 1 prestimulation (compared to mock) is faint — p-STAT1/2 and ISG levels are only slightly elevated, and IAV NP levels are on average about 3-fold lower. In the case of an infection with RSV (Figs. 2C and 2D), p-STAT1/2 levels (at 24 and 48 h p.i.) are similar for non-prestimulated and IFN- λ 1- or IFN- β -prestimulated cells, indicating that a sufficient amount of IFN- β is produced during infection to trigger STAT1/2 phosphorylation. However, the early accumulation of ISGs, observed in Figs. 1A and 1B after IFN- β prestimulation, resulted in about 10-fold reduction of RSV fusion (F) protein levels (compared to non-prestimulated cells), not observed in the case of IFN- λ 1 prestimulation (Figs. 2C and 2D).

In summary, we have shown that (pre)stimulation with IFN- β , compared to IFN- λ 1, leads to a higher level of p-STAT1/2, higher expression of ISGs, and, consequently, when applied before viral infection, more effectively reduces viral spread.

IFNLR1 knockout markedly promotes IAV spread while modestly influencing STAT1/2 activation

In previous experimental protocols, cells responded to a specific interferon provided in prestimulation and, at later stages of the infection, to type I and type III interferons secreted by infected cells. To disentangle the effect of a single type of interferon during viral infection we employed A549 knock-out cell lines: IFNAR1 KO, IFNLR1 KO, and IFNAR1 & IFNLR1 double KO (further referred to as DKO). We found that although IFNLR1 deficiency has a modest impact on p-STAT1/2 levels and ISG accumulation, it significantly increases IAV NP accumulation ($p < 0.001$; approximately 3-fold higher than in WT cells), indicating enhanced virus propagation (Figs. 3A and 3B). The small decrease in p-STAT1/2 levels in the IFNLR1 KO cell population can be explained by a higher fraction of IAV-infected cells, in which (as we demonstrated previously) p-STAT1/2 levels are reduced (Czerkies *et al*, 2022). In contrast, the lack of IFNAR1 results in a high reduction of p-STAT1/2 and STAT1/2 levels as well as ISG accumulation, but does not lead to a significant change in IAV NP levels. Furthermore, when comparing IFNAR1 KO and DKO cells, we observe that the additional loss of IFN- λ signaling leads to an increase in the IAV NP level (Figs. 3A and 3B).

This suggests that IFN- λ slows IAV propagation independently of STAT1/2 signaling. This effect is not observed in the case of infection with RSV, which exhibits a slightly lower accumulation of its F protein in IFNLR1 KO cells compared to WT cells. Given the substantial variability in viral protein levels between experimental replicates, Figs. 3B and 3D present the quantification of eight replicates for IAV and seven replicates for RSV (two from our earlier study, Czerkies *et al*, 2022, and the remaining performed for the current study). Full quantifications, including mock infections and 24 h p.i., can be found in Figs. EV2 and EV3. The increased spread of IAV in IFNLR1 KO cells, compared to WT cells, is also reflected at the IAV RNA level. Digital PCR measurements (Fig. EV4) indicate approximately 3-fold higher accumulation of IAV RNA (for MOI=0.01 and 0.1, 24 to 96 h p.i.), in good agreement with the Western blot results.

Finally, higher levels of IAV mRNA and NP protein observed in IFNLR1 KO cells (but not in IFNAR1 KO cells) with respect to WT cells correlate with a higher fraction of infected cells, shown in Fig. 4A and 4C. We thus showed that IFNLR1 deficiency increases IAV NP, IAV RNA, and the fraction of infected cells (with respect to WT), without substantially changing p-STAT1/2 levels and accumulation of ISG-encoded proteins.

Functions of IFN-λ extend beyond induction of ISGs

As shown in the previous section, during IAV infection, lack of IFNLR1 does not significantly hinder the production of the considered ISG-encoded proteins but leads to significantly higher levels of viral RNA, viral proteins, and a greater fraction of infected cells. To exclude the effect of ISG production resulting from IFNLR1 activation, we performed an immunostaining experiment comparing the fraction of IAV-positive cells in WT and IFNLR1 KO cell populations. We prestimulated the cells with 100 U/ml IFN-β for 24 hours before infection with IAV (MOI = 1) and lysed 24 h p.i. (Fig. 4B and 4D). We used a high MOI and a lower IFN-β concentration, because IFN-β prestimulation substantially inhibits IAV spread. We observed that the fraction of infected cells was about 6-fold higher in IFNLR1 KO cells compared to WT cells. This implies that IFN-λ signaling suppresses IAV spread even more effectively when expression of ISGs is already ensured by IFN-β prestimulation.

In summary, we demonstrated that IFN-λ signaling hinders viral spread during an infection with IAV (but not with RSV), and this effect is independent of its regulation of expression of ISGs. In the following sections, we argue that IFN-λ signaling is linked to cell death (without distinguishing between apoptosis, pyroptosis, or other cell death types), which is crucial for eliminating infected cells.

Measuring an increase in the cell death rate

We postulate that IAV-infected IFNLR1 KO cells are more resistant to cell death compared to WT cells, allowing for faster virus proliferation. The natural inclination is to compare the number of cell deaths in IFNLR1 KO and WT cell populations infected with IAV. However, a change in the cell death rate affects the total number of infected cells, rendering a direct comparison impossible. Instead, the ratio of cell death events to the number of infected cells is a much more reliable indicator of an increased cell death rate. We clarify this by adapting the susceptible, infected, removed (SIR) model (Eqs. 1–3) to represent viral infection of the cell population, interpreting the removed cells (R) as dead. The kinetics are given by the following equations:

$$S' = -\beta IS \quad (1)$$

$$I' = \beta IS - \lambda I \quad (2)$$

$$R' = \lambda I \quad (3)$$

where ' denotes differentiation with respect to time, β denotes the infection rate, and λ denotes the death rate. In Fig. 5, we show time trajectories of observables R , I , and their ratios for varied values of parameters β and λ for an MOI of 0.1 (see Fig. EV5 for MOIs of

0.01 and 0.3). The fraction of dead cells R can increase with λ (at early times) or decrease with λ (at later times), and is generally not monotonic with respect to this parameter (Fig. 5A). This occurs because, at higher λ , the infection progresses more slowly, leading to an overall lower fraction of infected cells (Fig. 5B). However, since the fraction of infected cells (I) is also strongly affected by the infection rate β (Fig. 5C), it cannot be used alone to distinguish between a decrease in mortality and an increase in infectivity. In contrast, the ratio R / I changes strongly and monotonically with λ (Fig. 5D), but weakly and not monotonically with β (Fig. 5E).

In summary, a simultaneous increase in the virus-infected cell fraction (I) and a decrease in the ratio of dead to virus-infected cells (R / I) indicates a decrease in the death rate λ . An important detail is that when experimentally measuring the proportion of infected cells (by immunostaining) or viral proteins or RNA produced by infected cells (by digital PCR or Western blot), we account for live infected cells but also some already dead infected cells. Fortunately, the monotonicity of the ratio R / I with respect to λ implies monotonicity of the ratio $R / (I + \alpha R)$ for any $\alpha \geq 0$. Furthermore, even when $\alpha = 1$, the ratio $R / (I + R)$ still grows substantially with λ (Fig. 5F).

IFN- λ signaling suppresses viral spread by promoting cell death

In Fig. 6A we show that the ratio of cell deaths to accumulated IAV NP is higher in WT cells than in IFNLR1 KO cells, which is equivalent to showing that the ratio of dying to infected cells is higher in the case of WT than IFNLR1 KO cells. The left column of Fig. 6A presents results from Western blots and shows the ratio of IAV NP between IFNLR1 KO and WT cells at 24 h p.i. and 48 h p.i. The amount of IAV NP increases significantly ($p < 10^{-7}$) in IFNLR1 KO cells, consistent with previous experiments (compare with Fig. 3B), despite the high variability observed between experiments. The right column of Fig. 6A shows the ratio of the fraction of cell deaths (measured by FACS) between IFNLR1 KO and WT cells for the same experimental protocol. On average, the ratios are only slightly higher than 1, indicating that the fraction of dead cells in IFNLR1 KO cell populations is comparable to that in WT cell populations. Since the ratio of viral proteins is higher than the ratio of cell deaths, we may conclude that the overall death rate is higher in WT than in IFNLR1 KO cells ($p < 0.01$).

RSV provides a relevant comparison, as it is known that its NS1 protein effectively blocks cell death, promoting viral spread (Bitko *et al*, 2007). To compare cell death rates during infection with IAV and RSV, we measured fractions of cells displaying propidium iodide signal during 24 and 48 hour long infections. Next, the cells were fixed, and fractions of virus positive cells at 24 and 48 hours were measured based on immunostaining images. In Fig.

6B, we show that cell death fractions for RSV and IAV are similar, but the infected cell fraction is much higher for infection with RSV than for infection with IAV. Consequently, this indicates a higher death rate for IAV infections compared to RSV infections ($p < 0.01$). The effect is observed for both WT and IFNLR1 KO cells. This supports our reasoning that an inhibition of cell death can be manifested by an increase of the fraction of infected cells rather than a decrease in the fraction of dead cells, suggesting that IFN- λ signaling suppresses viral spread by promoting cell death.

Discussion

Using WT, IFNLR1 KO, IFNAR1 KO, and IFNLR1–IFNAR1 DKO A549 cells, we demonstrated divergent roles of type I and type III interferons in control of IAV and RSV proliferation. (1) We showed that (pre)stimulation with IFN- β leads to a strong increase of p-STAT1/2 and STAT1/2 levels, high accumulation of ISGs, and, through those mechanisms, a pronounced reduction of proliferation of IAV as well as RSV. Prestimulation with IFN- λ (compared to IFN- β) results in lower levels of p-STAT1/2 and STAT1/2, lower accumulation of ISGs, and only moderately reduces proliferation of IAV, and does not reduce proliferation of RSV. (2) Next, we showed that in IAV- or RSV-infected IFNAR1 KO cells, p-STAT1/2 levels are pronouncedly reduced (with respect to WT cells), while the effect of IFNLR1 knockout on p-STAT1/2 levels is weak. However, IAV proliferates significantly faster in IFNLR1 KO cells than in WT or IFNAR1 KO cells. This indicates that IFN- λ suppresses IAV proliferation independently to STAT signaling. We note the possibility that IFNLR1 inhibits IAV proliferation independently of type III interferons, however mouse studies indicate that type III interferon deficiency has the same consequences as IFNLR1 deficiency during infections with norovirus, reovirus, and influenza virus (Peterson *et al*, 2019). The faster viral propagation in IFNLR1 KO cells is not observed for RSV. (3) Finally, we showed that absence of IFN- λ signaling significantly increases the level of IAV protein and the fraction of infected cells, without an increase in the fraction of dead cells. With the help of a simple mathematical model we interpret this observation as the ability of IFN- λ to promote early death in cells infected with IAV. Additionally, we showed that for infection with IAV, the ratio of dead to infected cells is higher than for infection with RSV. This is in line with earlier observations that NS proteins of RSV very effectively suppress cell death in infected cells, promoting virus proliferation (Bitko *et al*, 2007). This could explain why the effect of IFN- λ on RSV spread appears negligible.

Although divergent roles of type I and type III interferons have been highlighted in numerous studies, reports of a non-transcriptional role, especially of a proapoptotic (or propyrototic) role, of IFN- λ are scarce. In two studies not related to infections, it was shown

that in intestinal cells type III interferons (compared to type I interferons) are a more potent inducer of apoptosis (Li *et al*, 2008) and pyroptosis (preprint: Sposito *et al*, 2022). Major *et al* (Major *et al*, 2020) found that IFN signaling (especially IFN- λ) hampers lung repair by inducing p53 and inhibiting epithelial proliferation and differentiation, which potentially can also promote cell death. Broggi *et al* (Broggi *et al*, 2017), in a study not directly related to infections, showed that IFN- λ regulates functions of neutrophils (known to have high expression of IFNLR) independently of transcription, and that treatment with IFN- λ induced in these cells a reduction in the phosphorylation of AKT at Ser473 and Thr308, which could make these cells more prone to cell death. Earlier, Denisova *et al* (Denisova *et al*, 2014) showed that inhibiting AKT suppresses proliferation of IAV in A549 cells. In the context of viral infections, a straightforward demonstration that IFN- λ increases the cell death rate is challenging because a higher cell death rate does not necessarily result in a greater fraction of dead cells, as it also suppresses viral propagation. We suspect that this may partially explain why this phenomenon was not reported earlier.

Our findings confirm that the role of IFN- β is to promote an antiviral state in not yet infected bystander cells, which increases their resistance to infection and/or slows down virus replication once they are infected. In terms of the SIR model, the effect of IFN- β can be seen as a reduction of the infection rate coefficient β (see Eq. 2). The role of IFN- λ (apart from STAT signaling) would be to promote cell death, and thus removal of infected cells. In the SIR model, this corresponds to an increase of the death rate coefficient λ (see also Eq. 2). In this picture, type III interferons are more universal than type I interferons, being able to both induce some STAT signaling and trigger cell death. Finally, it should be stressed that our data does not exclude the possibility that IFN- λ triggers cell death using ISGs that are accumulated (mostly) due to IFN- β signaling. This is in line with the observation that the effect of IFNLR1 deficiency is the strongest when cells are already prestimulated with IFN- β , and our earlier study (Czerkies *et al*, 2018) which showed that the death rate after poly(I:C) stimulation is much higher in IFN- β -prestimated cells.

Methods

Cell lines and cultures. The A549, HEK-293T, HeLa and MDCK cell lines were purchased from the ATCC. A549 cells were cultured in F-12K basal medium supplemented with 10% fetal bovine serum (FBS) and a penicillin-streptomycin antibiotic solution. MDCK, HeLa cells and HEK-293T cells were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% FBS and a penicillin-streptomycin antibiotic solution. All cell lines were cultured under standard conditions (37°C with 5% CO₂) and kept in monolayers up to 90% confluence. The IFNAR1, IFNLR1, and IFNAR1-IFNLR1 knockouts based on the A549

cell line were developed using the CRISPR lentivirus system (Abm). The detailed procedure for obtaining these cell lines have been described in (Czerkies *et al*, 2022, see Methods section and Supplementary Figure S3 therein).

Virus amplification and isolation. Respiratory syncytial virus strain A2 and influenza A virus H1N1 strain A/PR/8/34 were purchased from the ATCC and amplified in HeLa and MDCK cells, respectively. Cells were seeded into 225-cm² tissue culture flasks (Falcon) and cultured as described above for 2 to 3 days until they reached 90% confluence. On the day of infection, the following virus growth medium was prepared: DMEM with 2% FBS for RSV or basal minimal essential medium (MEM) with 0.3% bovine serum albumin (BSA) and 1 µg/mL of L-1-tosylamido-2-phenylethyl chloromethyl ketone (TPCK-trypsin) for IAV. The dilutions of the virus were prepared in an appropriate medium, with target MOIs of around 0.01 for RSV and 0.05 for IAV. Culture media were removed, and cells were washed once with phosphate-buffered saline (PBS) and overlaid with 10 mL of the inoculum. Virus was allowed to adsorb to cells for 2 h at 37°C, with occasional stirring. Next, additional virus growth medium was added to a total volume of 40 mL per flask. Infected cells were cultured at 37°C until the development of cytopathic effects could be observed in at least 80% of the cells (typically at around 3 days for IAV and 5 days for RSV). Virus-containing culture fluid was then collected and clarified by centrifugation at 3,000×g at 4°C for 20 min. Next, virus particles were precipitated by adding 50% (wt/vol) polyethylene glycol 6000 (PEG 6000) (Sigma-Aldrich) in NT buffer (150 mM NaCl, 50 mM Tris-HCl [pH 7.5]) to a final concentration of 10% and stirring the mixture gently at 4°C for 90 min. Virus was centrifuged at 3,250×g at 4°C for 20 min and recentrifuged after the removal of the supernatant to remove the remaining fluid. The pellet was suspended in 1 mL of NT buffer or in 20% sucrose in NT buffer in the case of RSV, aliquoted, and stored at -80°C.

Virus quantification. The virus concentration in the collected samples was quantified using an immunofluorescence protocol. HeLa or MDCK cells were seeded onto microscopic coverslips and cultured upon reaching 90 to 100% confluence. Serial dilutions of virus samples were made in virus growth medium in a range of 10⁻³ to 10⁻⁶. After washing with PBS, cells were overlaid in duplicates with diluted virus, which was allowed to adhere for 2 h, with occasional stirring. Afterward, the virus-containing medium was removed, and cells were overlaid with fresh virus growth medium and cultured for 16 h (for IAV) or 24 h (for RSV). Next, cells were washed with PBS and fixed with 4% formaldehyde for 20 min at room temperature. Cells were stained using a standard immunofluorescence protocol with anti-RSV antibody (catalog number ab43812; Abcam) or anti-influenza A virus nucleoprotein antibody (clone C43) (catalog number ab128193; Abcam). Cells containing stained viral proteins were counted using a Leica SP5 confocal microscope. The virus concentration was

calculated using the following formula: (average number of infected cells)/(dilution factor × volume containing virus added) = infectious particles/mL. For a given MOI, we observe fewer cells expressing viral proteins in the A549 than in the HeLa and MDCK lines, which are known to be permissive for RSV and IAV infections, respectively.

Compounds and stimulation protocols. Human IFN-β1A and IFN-λ1 were purchased from Thermo Fisher Scientific (catalog numbers PHC4244 and 34-8298-64, respectively) and prepared according to the manufacturer's instructions. For cell stimulation, interferons were further diluted to the desired concentrations in F-12K medium supplemented with 2% FBS. The decreased FBS content was used to prevent the inhibition of viral attachment and entry at the second stage of the experiments.

For interferon–virus experiments, the cell culture media were exchanged for interferon-containing or control media at time zero and were not removed afterward until the end of the experiment. Appropriately diluted virus was added in small volumes (less than 10 μL) directly into the wells. The even distribution of virus across the cell population was aided by the intermittent rocking of the plate for 2 h.

Antibodies. (i) Antibodies for Western blotting. Primary antibodies for Western blotting were anti-phospho-STAT1 (Tyr701) (clone 58D6) (catalog number 9167; Cell Signaling Technologies) (1:1,000), anti-phospho-STAT2 (Tyr690) (clone D3P2P) (catalog number 88410; Cell Signaling Technologies) (1:1,000), anti-RIG-I (clone D14G6) (catalog number 3743; Cell Signaling Technologies) (1:1,000), anti-STAT1 (catalog number 610116; BD Biosciences) (1:1,000), anti-STAT2 (catalog number PAF-ST2; R&D Systems) (1:1,000), anti-PKR (clone B-10) (catalog number sc-6282; Santa Cruz Biotechnology) (1:1,000), anti-OAS1 (clone F-3) (catalog number sc-374656; Santa Cruz Biotechnology) (1:1,000), anti-influenza A virus nucleoprotein (clone C43) (catalog number ab128193; Abcam) (1:1,000), anti-respiratory syncytial virus (clone 2F7) (catalog number ab43812; Abcam) (1:1,000), and hFAB rhodamine anti-glyceraldehyde-3-phosphate dehydrogenase (GAPDH) (catalog number 12004168; Bio-Rad) (1:10,000).

Secondary antibodies for Western blotting were DyLight 800 4× PEG-conjugated goat anti-rabbit IgG(H+L) (catalog number SA5-35571; Thermo Fisher Scientific) (1:10,000), DyLight 800 4× PEG-conjugated goat anti-mouse IgG(H+L) (catalog number SA5-35521; Thermo Fisher Scientific) (1:10,000), StarBright Blue 700-conjugated goat anti-mouse IgG (catalog number 12004159; Bio-Rad) (1:10,000), and horseradish peroxidase (HRP)-conjugated rabbit anti-goat immunoglobulins (catalog number P0449; Agilent) (1:10,000).

(ii) Antibodies for immunostaining. Primary antibodies for immunostaining were anti-phospho-STAT1 (Tyr701) (clone 58D6) (catalog number 9167; Cell Signaling Technologies) (1:1,000), anti-respiratory syncytial virus (catalog number ab20745; Abcam) (1:1,000), anti-influenza A virus (catalog number ab20841; Abcam) (1:1,000), anti-influenza A virus nucleoprotein (clone C43) (catalog number ab128193; Abcam) (1:1,000), and anti-human interferon beta (catalog number MAB8142; R&D Systems).

Secondary antibodies for immunostaining were Alexa Fluor 488-conjugated donkey anti-rabbit IgG(H+L) (catalog number A-21206; Thermo Fisher Scientific) (1:1,000), Alexa Fluor 555-conjugated donkey anti-mouse IgG(H+L) (catalog number A-31570; Thermo Fisher Scientific) (1:1,000), and Alexa Fluor 633-conjugated donkey anti-goat IgG(H+L) (catalog number A-21082; Thermo Fisher Scientific) (1:1,000).

RNA isolation. For gene expression analysis experiments, cells were seeded on 12- or 24-well plates (Falcon) at a density of, respectively, 1.0×10^5 or 1.5×10^5 cells/well. Upon completed stimulation according to chosen protocol, cells were washed once with PBS and submitted to isolation of total RNA using PureLink RNA Mini Kit (Thermo Fisher Scientific), following manufacturer's instructions: cells were harvested and vortexed in Lysis Buffer with 2-mercaptoethanol and then vortexed again with one volume of 70% ethanol. Upon transferring to the spin cartridge, cellular RNA content was bound to the column, washed with appropriate buffers and eluted, all by centrifugation at $12,000 \times g$. Eluted RNA in RNase-free water was used immediately for reverse transcription or stored for later use at -80°C .

Reverse transcription. RNA concentration and quality was determined by measuring UV absorbance of samples diluted 1:100 in 10 mM Tris-HCl (pH 7.5) at 260 and 280 nm, using the Multiskan GO spectrophotometer (Thermo Fisher Scientific). Independently, RIN index for isolated RNA was checked using Agilent 2100 Bioanalyzer (see below). Around 1 μg of RNA was used as a template for reverse transcription, performed using High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific). Following manufacturer's protocol, diluted RNA samples were mixed 1:1 with freshly prepared Master Mix containing RT Buffer, RT Random Primers, dNTP Mix, MultiScribe Reverse Transcriptase and RNase Inhibitor. Reaction was performed in Mastercycle Gradient thermal cycler (Eppendorf) under following conditions: 10 min at 25°C , 120 min at 37°C , and 5 min at 85°C .

Real-Time Quantitative Polymerase Chain Reaction. Reaction was performed on a QuantStudio 12K Flex Real-Time PCR system with Array Card block (Life Technologies). In most cases, 750 ng of cDNA from reverse transcription was mixed with reaction Master Mix and loaded onto TaqMan Array Card containing probes and primers for 24 genes (in 2

replicates), including endogenous reference controls (see Source Data S2). Reaction was conducted using QuantStudio “Standard” protocol, with FAM/ROX chemistry. Upon completion, expression of target genes was analyzed based on absolute Ct quantification or classical $\Delta\Delta C_t$ method with QuantStudio 12K Flex software, normalized against GAPDH gene expression.

Digital PCR. For gene expression experiments, cells were seeded onto 24-well plates at a density of 100,000 cells per well. RNA from virus-infected cells was isolated using a PureLink RNA Mini kit (Thermo Fisher Scientific) according to the manufacturer’s instructions: cells were harvested and vortexed in lysis buffer with 2-mercaptoethanol and then vortexed again with 1 volume of 70% ethanol. Upon transfer to the spin cartridge, the cellular RNA content was bound to the column, washed with the appropriate buffers, and eluted, all by centrifugation at 12,000×g. The eluted RNA in RNase-free water was used immediately for reverse transcription (RT) or stored for later use at –80°C. The concentration and quality of the isolated total RNA were determined by measuring the UV absorbance at 260 nm and 280 nm using a Multiskan Go spectrophotometer (Thermo Fisher Scientific). Approximately 0.5 µg of RNA was used as a template for reverse transcription, performed using a high-capacity cDNA reverse transcription kit (Thermo Fisher Scientific): the diluted RNA samples were mixed 1:1 with freshly prepared master mix containing RT buffer, RT random primers, deoxynucleoside triphosphate (dNTP) mix, MultiScribe reverse transcriptase, and an RNase inhibitor. The reaction was performed in a Mastercycler gradient thermal cycler (Eppendorf) under the following conditions: 10 min at 25°C, 120 min at 37°C, and 5 min at 85°C.

Measurements of the viral RNA copy number were then performed using the QuantStudio 3D system (Thermo Fisher Scientific) and TaqMan Vi99990011_po and Vi99990014_po gene expression assays for the quantification of IAV and RSV RNAs, respectively. Appropriately diluted samples were mixed with the reaction master mix and the chosen assay mix, loaded onto the digital PCR (dPCR) chips in duplicates, and thermocycled using a ProFlex flat-block thermal cycler. Subsequently, chips were analyzed with a QuantStudio 3D digital PCR instrument. The chip quality check and final data analysis were performed using QuantStudio 3D AnalysisSuite software.

Western blotting. Cells at the indicated time points were washed twice with PBS, lysed in Laemmli sample buffer containing dithiothreitol (DTT), and boiled at 95°C for 10 min. Equal amounts of each protein sample were separated on 4 to 16% Mini-Protean TGX stain-free precast gels using the Mini-Protean Tetra cell electrophoresis system (Bio-Rad). Upon the completion of electrophoresis, proteins were transferred to a nitrocellulose membrane using

wet electrotransfer in the Mini-Protean apparatus (400 mA for 50 min). The membrane was rinsed with TBST (Tris-buffered saline [TBS] containing 0.1% Tween 20; Sigma-Aldrich) and blocked for 1 h with 5% BSA–TBS or 5% nonfat dry milk. Subsequently, the membranes were incubated with one of the primary antibodies diluted in a 5% BSA–TBS buffer at 4°C overnight. After thorough washing with TBST, the membranes were incubated with secondary antibodies conjugated to a specific fluorochrome (DyLight 800; Thermo Fisher Scientific) or horseradish peroxidase (HRP-conjugated polyclonal anti-mouse/anti-rabbit immunoglobulins; Dako) diluted 1:5,000 in 5% nonfat dry milk–TBST for 1 h at room temperature. The chemiluminescence reaction was developed with the Clarity Western ECL system (Bio-Rad). For GAPDH detection, hFAB rhodamine anti-GAPDH primary antibody (Bio-Rad) was used. Specific protein bands were detected using the ChemiDoc MP imaging system (Bio-Rad).

Immunostaining. For staining of intracellular proteins, cells were seeded onto 12-mm round glass coverslips, which were previously washed in 60% ethanol–40% HCl, thoroughly rinsed with water, and sterilized. After stimulation with interferon or viral infection at the desired time points, cells on the coverslips were washed with PBS and immediately fixed with 4% formaldehyde (20 min at room temperature). Cells were then washed three times with PBS and incubated with 100% cold methanol for 20 min at –20°C. After washing with PBS, coverslips with cells were blocked and permeabilized for 1.5 h with 5% BSA (Sigma-Aldrich) with 0.3% Triton X-100 (Sigma-Aldrich) in PBS at room temperature. After the removal of the blocking solution, coverslips with cells were incubated with primary antibodies diluted in PBS containing 1% BSA and 0.3% Triton X-100 overnight at 4°C. After the cells were washed five times with PBS, the appropriate secondary antibodies conjugated with fluorescent dyes were added, and the cells were incubated for 1 h at room temperature. Subsequently, the cells were washed, and their nuclei were stained with 200 ng/mL 4',6-diamidino-2-phenylindole (DAPI) (Sigma-Aldrich) for 10 min. After a final wash in MilliQ water, coverslips with stained cells were mounted onto microscope slides with a drop of Mowiol (Sigma-Aldrich).

FACS/cell death assay. Cell death was measured using annexin V and propidium iodide staining method (FITC Annexin V Apoptosis Detection Kit II, BD Bioscience), measuring the amount of phosphatidylserine externalized to the outer layer of cell membrane. For the analysis, cells were seeded on 12-well plates at a density of 100,000 cells/well and treated according to the chosen protocol. Analysis was performed on both the cells attached to the plate surface and dying cells suspended in the medium. Flow cytometry measurement was performed using FACSCalibur (BD) and the CellQuest Pro analysis software. Cells which are both annexin V- and propidium iodide (PI)-negative were considered viable, annexin V-

positive and PI-negative were considered early apoptotic or pyroptotic; PI-positive cells, irrespective of the outcome of annexin V staining, were considered dying or dead.

Live imaging propidium iodide assay. To estimate cell death rates during infection with IAV and RSV, live imaging of PI intracellular accumulation was performed. One day before the experiment, cell nuclei were stained with Hoechst 33342 (Thermo) diluted 1 µg/ml in PBS for 10 minutes at 37°C. After staining, the cells were washed and seeded onto a µClear 96-well microplate at a density of 18,000 cells per well. The following day, the cells were infected with IAV or RSV, and PI was added. Cells were observed for 24 or 48 h, and within these periods fraction of PI positive cells was determined. After 24 or 48 h p.i. cell fixation and immunostaining were performed to estimate the number of infected cells. Imaging was carried out on an Operetta CLS (PerkinElmer) using Harmony software. Images were analyzed using our in-house software, ShuttleTracker (see the Data analysis section).

Image analysis. Confocal images obtained from immunostaining were analyzed using our in-house software (<https://pmbm.ippt.pan.pl/software/shuttletracker>). Nuclear outlines were detected based on DAPI staining. Outlines of nuclei that were partially out of frame or mitotic were excluded from the analysis; outlines of overlapping nuclei were split based on geometric convexity defects when possible. Cells were classified as either virus-negative or virus-positive based on the sum of the intensities of the pixels inside nuclear outlines. Cells were classified as dead/dying based on the intensities of pixels inside nuclear outlines in the PI channel.

Statistical analysis. To assess statistical significance we performed the following tests: Student *t*-test (Fig. 1B, Figs. 2B and 2D, Figs. 3B and 3D, Figs. 4A and 4B, Fig. 6B, Fig. EV4), Wilcoxon signed-rank test (Fig. 1C, Fig. 6B), unpaired Student *t*-test (Fig. 6A), Mann-Whitney *U* test (Fig. 6A). All tests assume two-tailed alternatives.

In most cases (Figs. 1B and 1C, Figs. 2B and 2D, Fig. 4A, Fig. 6A and 6B, Fig. EV4), the statistical tests were performed using experimental replicates obtained for two or three time points pooled together; similarly, in Fig. 4A and Fig. EV4 two MOI values (0.01 and 0.1) are considered jointly. This approach is demonstrated in five cases below, with the remaining cases treated analogously.

In Fig. 1C we compared mRNA expression of ISGs after IFN-β and IFN-λ1 stimulation. To demonstrate that in the period between 6–24 hours, mRNA expression is significantly higher after IFN-β stimulation than after IFN-λ1 stimulation, we performed 3 experimental replicates in which mRNA levels were measured at 3 time points: 6, 10 and 24 h post stimulation; data at each time point was gathered from a separate dish. We then compare these 9 pairs (IFN-

β versus IFN- λ 1) of data points using a single two-tailed Wilcoxon signed-rank test on paired differences. Pooling 3 different time points together is justified by the fact that the studied effect holds for all of them: for all but two ISGs we observed that in all replicates and at all time points mRNA expression after prestimulation with IFN- β is higher than after prestimulation with IFN- λ 1.

In Fig. 1B, we perform a similar comparison by pooling measurements at 2 h with those at 6 h, and 24 h with 48 h in order to compare STAT1/2 and p-STAT1/2 levels after IFN- β and IFN- λ 1 stimulation at an early time point (1 hour), for intermediate times (2–6 hours), and late times (24–48 hours). We performed a parametric two-tailed Student *t*-test on log ratios. In all cases where we report statistical significance, protein levels after IFN- β stimulation are consistently higher than after IFN- λ 1 stimulation (for all replicates and pooled time points).

In Fig. 4A, we pooled together infection experiments with an MOI of 0.1 and 0.01 lasting 24 and 48 h. Again, by performing parametric two-tailed Student *t*-test on log ratios we showed that probing two MOI values and two infection times fraction of cells infected with IAV is higher for IFNLR1 KO than for WT cells.

In Fig. 6A, the left column consists of IAV NP level ratios between IFNLR1 KO and WT cells, where both levels are measured in a single Western blot. The right column consists of ratios between the fraction of dying/dead cells between IFNLR1 KO and WT cells; both fractions are measured in a single FACS experiment. However, ratios from the left column are not paired with ratios from the right column (since the Western blot and FACS measurements were not paired). Thus, we use a (parametric) unpaired Student *t*-test (without assuming equal variances) and a (non-parametric) Mann-Whitney *U* test to compare log ratios from the left column with log ratios from the right column. Data from time points 24 h p.i. and 48 h p.i. are considered jointly.

In Fig. 6B, ratios from the left column (obtained from immunostaining images of fixed cells) and ratios from the right column (obtained from live-cell imaging) are paired, since both types of measurement were performed for the same cell population (for each experimental replicate). For each left–right pair, we can thus compute a ratio-of-ratios. We then perform a (parametric) two-tailed Student *t*-test and a (non-parametric) two-tailed Wilcoxon signed-rank test on log ratio-of-ratios. Data from time points 24 h p.i. and 48 h p.i. and cell lines WT and IFNLR1 KO are considered jointly, since the death rate in IAV is always higher than in RSV (which is equivalent to ratio-of-ratios greater than 1).

541 Data availability

542 All Western blots performed for this study are available in supplementary file Source Data
543 S1. Quantifications of Western blots, RT-PCR, digital PCR, FACS, imaging data and codes
544 to generate figures are available in supplementary file Source Data S2. The raw imaging
545 dataset is available at Zenodo (doi: 10.5281/zenodo.13735878).

546 Author contributions

547 **Conceptualization:** Wiktor Prus, Frederic Grabowski, Paulina Koza, Zbigniew Korwek,
548 Maciej Czerkies, Marek Kochańczyk, Tomasz Lipniacki.

549 **Methodology:** Zbigniew Korwek, Wiktor Prus.

550 **Formal analysis:** Frederic Grabowski.

551 **Investigation:** Wiktor Prus, Paulina Koza, Zbigniew Korwek, Maciej Czerkies.

552 **Data curation:** Frederic Grabowski, Marek Kochańczyk.

553 **Writing – original draft:** Frederic Grabowski, Marek Kochańczyk, Tomasz Lipniacki.

554 **Writing – review & editing:** Wiktor Prus, Frederic Grabowski, Paulina Koza, Zbigniew
555 Korwek, Maciej Czerkies, Marek Kochańczyk, Tomasz Lipniacki.

556 **Visualization:** Frederic Grabowski.

557 **Supervision:** Tomasz Lipniacki, Marek Kochańczyk.

558 **Funding acquisition:** Tomasz Lipniacki.

559 Disclosure and competing interest statement

560 The authors declare no competing interests.

561 Acknowledgements

562 This study was funded by the Norwegian Financial Mechanism GRIEG-1 grant (operated by
563 the National Science Centre, Poland) 2019/34/H/NZ6/00699 (TL) and the National Science
564 Centre Poland grant 2023/51/B/NZ6/02951 (TL). The funders had no role in study design,
565 data collection and analysis, decision to publish, or preparation of the manuscript.

References

- Bitko V, Shulyayeva O, Mazumder B, Musiyenko A, Ramaswamy M, Look DC & Barik S (2007) Nonstructural proteins of respiratory syncytial virus suppress premature apoptosis by an NF-kappaB-dependent, interferon-independent mechanism and facilitate virus growth. *J Virol* 81: 1786–1795
- Broggi A, Tan Y, Granucci F & Zanoni I (2017) IFN-λ suppresses intestinal inflammation by non-translational regulation of neutrophil function. *Nat Immunol* 18: 1084–1093
- Caine EA, Scheaffer SM, Arora N, Zaitsev K, Artyomov MN, Coyne CB, Moley KH & Diamond MS (2019) Interferon lambda protects the female reproductive tract against Zika virus infection. *Nat Commun* 10: 280
- Czerkies M, Kochańczyk M, Korwek Z, Prus W & Lipniacki T (2022) Respiratory syncytial virus protects bystander cells against influenza A virus infection by triggering secretion of type I and type III Interferons. *J Virol* 96: e0134122
- Czerkies M, Korwek Z, Prus W, Kochańczyk M, Jaruszewicz-Błońska J, Tudelska K, Błoński S, Kimmel M, Brasier AR & Lipniacki T (2018) Cell fate in antiviral response arises in the crosstalk of IRF, NF-κB and JAK/STAT pathways. *Nat Commun* 9: 493
- Denisova OV, Söderholm S, Virtanen S, Von Schantz C, Bychkov D, Vashchinkina E, Desloovere J, Tynell J, Ikonen N, Theisen LL, *et al* (2014) Akt Inhibitor MK2206 Prevents Influenza pH1N1 Virus Infection In Vitro. *Antimicrob Agents Chemother* 58: 3689–3696
- Galani IE, Triantafyllia V, Eleminiadou E-E, Koltsida O, Stavropoulos A, Manioudaki M, Thanos D, Doyle SE, Kotenko SV, Thanopoulou K, *et al* (2017) Interferon-λ Mediates Non-redundant Front-Line Antiviral Protection against Influenza Virus Infection without Compromising Host Fitness. *Immunity* 46: 875-890.e6
- Klinkhammer J, Schnepf D, Ye L, Schwaderlapp M, Gad HH, Hartmann R, Garcin D, Mahlaköiv T & Staeheli P (2018) IFN-λ prevents influenza virus spread from the upper airways to the lungs and limits virus transmission. *eLife* 7: e33354
- Kotenko SV, Gallagher G, Baurin VV, Lewis-Antes A, Shen M, Shah NK, Langer JA, Sheikh F, Dickensheets H & Donnelly RP (2003) IFN-lambdas mediate antiviral protection through a distinct class II cytokine receptor complex. *Nat Immunol* 4: 69–77
- Lazear HM, Nice TJ & Diamond MS (2015) Interferon-λ: Immune Functions at Barrier Surfaces and Beyond. *Immunity* 43: 15–28
- Lazear HM, Schoggins JW & Diamond MS (2019) Shared and Distinct Functions of Type I and Type III Interferons. *Immunity* 50: 907–923
- Li W, Lewis-Antes A, Huang J, Balan M & Kotenko SV (2008) Regulation of apoptosis by type III interferons. *Cell Prolif* 41: 960–979
- Major J, Crotta S, Llorian M, McCabe TM, Gad HH, Priestnall SL, Hartmann R & Wack A (2020) Type I and III interferons disrupt lung epithelial repair during recovery from

604 viral infection. *Science* 369: 712–717

605 Mesev EV, LeDesma RA & Ploss A (2019) Decoding type I and III interferon signalling
606 during viral infection. *Nat Microbiol* 4: 914–924

607 Nice TJ, Baldrige MT, McCune BT, Norman JM, Lazear HM, Artyomov M, Diamond MS &
608 Virgin HW (2015) Interferon- λ cures persistent murine norovirus infection in the
609 absence of adaptive immunity. *Science* 347: 269–273

610 Osterlund PI, Pietilä TE, Veckman V, Kotenko SV & Julkunen I (2007) IFN regulatory factor
611 family members differentially regulate the expression of type III IFN (IFN- λ)
612 genes. *J Immunol Baltim Md 1950* 179: 3434–3442

613 Peterson ST, Kennedy EA, Bringle PH, Taylor GM, Urbanek K, Bricker TL, Lee S, Shin H,
614 Dermody TS, Boon ACM, *et al* (2019) Disruption of Type III Interferon (IFN) Genes
615 Ifnl2 and Ifnl3 Recapitulates Loss of the Type III IFN Receptor in the Mucosal
616 Antiviral Response. *J Virol* 93: 10.1128/jvi.01073-19

617 Pott J, Mahlaköiv T, Mordstein M, Duerr CU, Michiels T, Stockinger S, Staeheli P & Hornef
618 MW (2011) IFN- λ determines the intestinal epithelial antiviral host defense. *Proc Natl*
619 *Acad Sci* 108: 7944–7949

620 Schoenborn JR & Wilson CB (2007) Regulation of interferon-gamma during innate and
621 adaptive immune responses. *Adv Immunol* 96: 41–101

622 Schoggins JW (2019) Interferon-Stimulated Genes: What Do They All Do? *Annu Rev Virol* 6:
623 567–584

624 Selvakumar TA, Bhushal S, Kalinke U, Wirth D, Hauser H, Köster M & Hornef MW (2017)
625 Identification of a Predominantly Interferon- λ -Induced Transcriptional Profile in
626 Murine Intestinal Epithelial Cells. *Front Immunol* 8: 1302

627 Sheppard P, Kindsvogel W, Xu W, Henderson K, Schlutsmeyer S, Whitmore TE, Kuestner
628 R, Garrigues U, Birks C, Roraback J, *et al* (2003) IL-28, IL-29 and their class II
629 cytokine receptor IL-28R. *Nat Immunol* 4: 63–68

630 Sposito B, Mambu J, Gwilt KB, Spinelli L, Andreeva N, Galland F, Naquet P, Mitsialis V,
631 Thiagarajah JR, Snapper SB, *et al* (2022) Type III interferons induce pyroptosis in gut
632 epithelial cells and delay tissue restitution upon acute intestinal injury.
633 2022.03.04.482997 doi:10.1101/2022.03.04.482997 [PREPRINT]

634 Wack A, Terczyńska-Dyla E & Hartmann R (2015) Guarding the frontiers: the biology of type
635 III interferons. *Nat Immunol* 16: 802–809

636 Figures

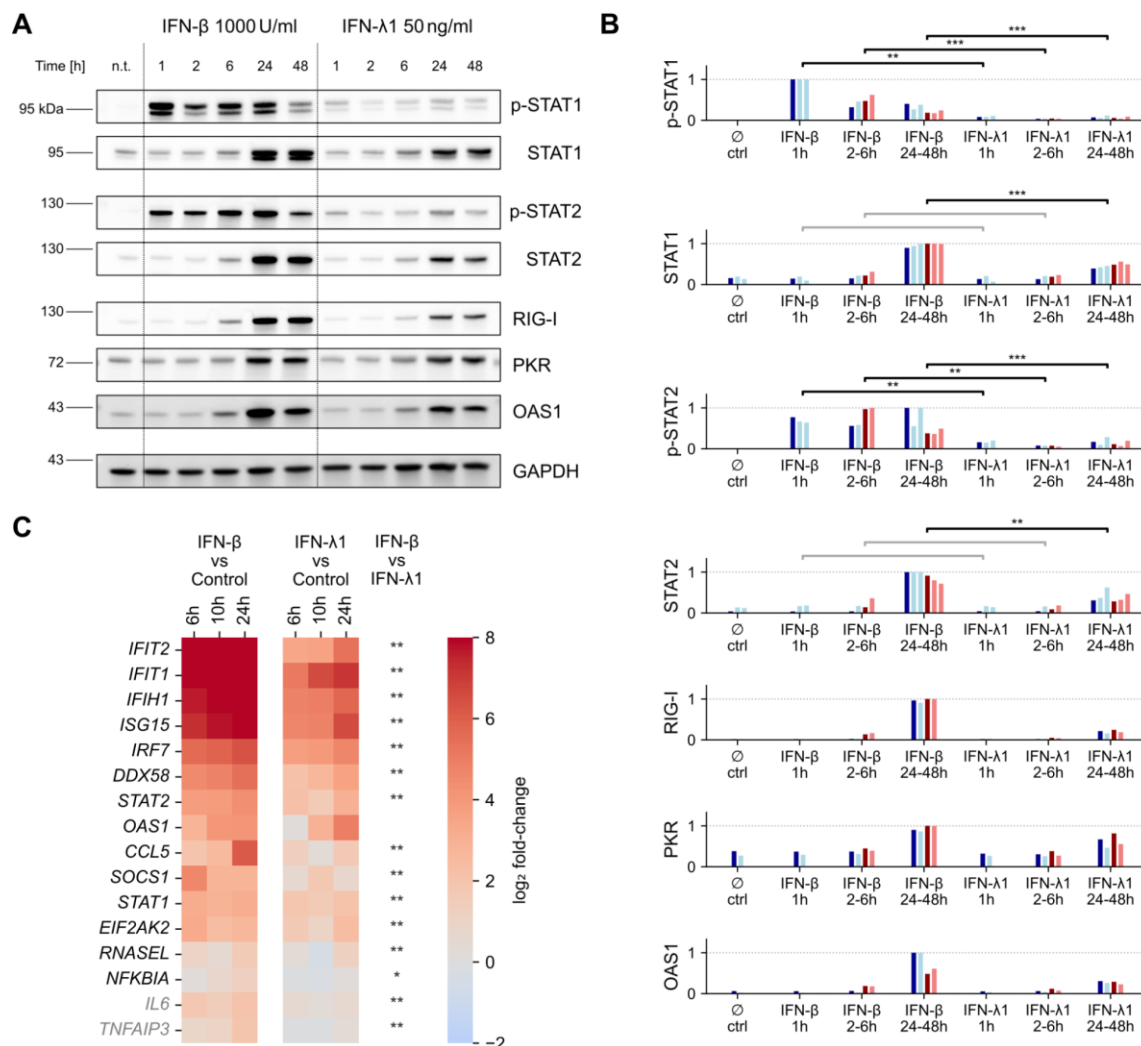
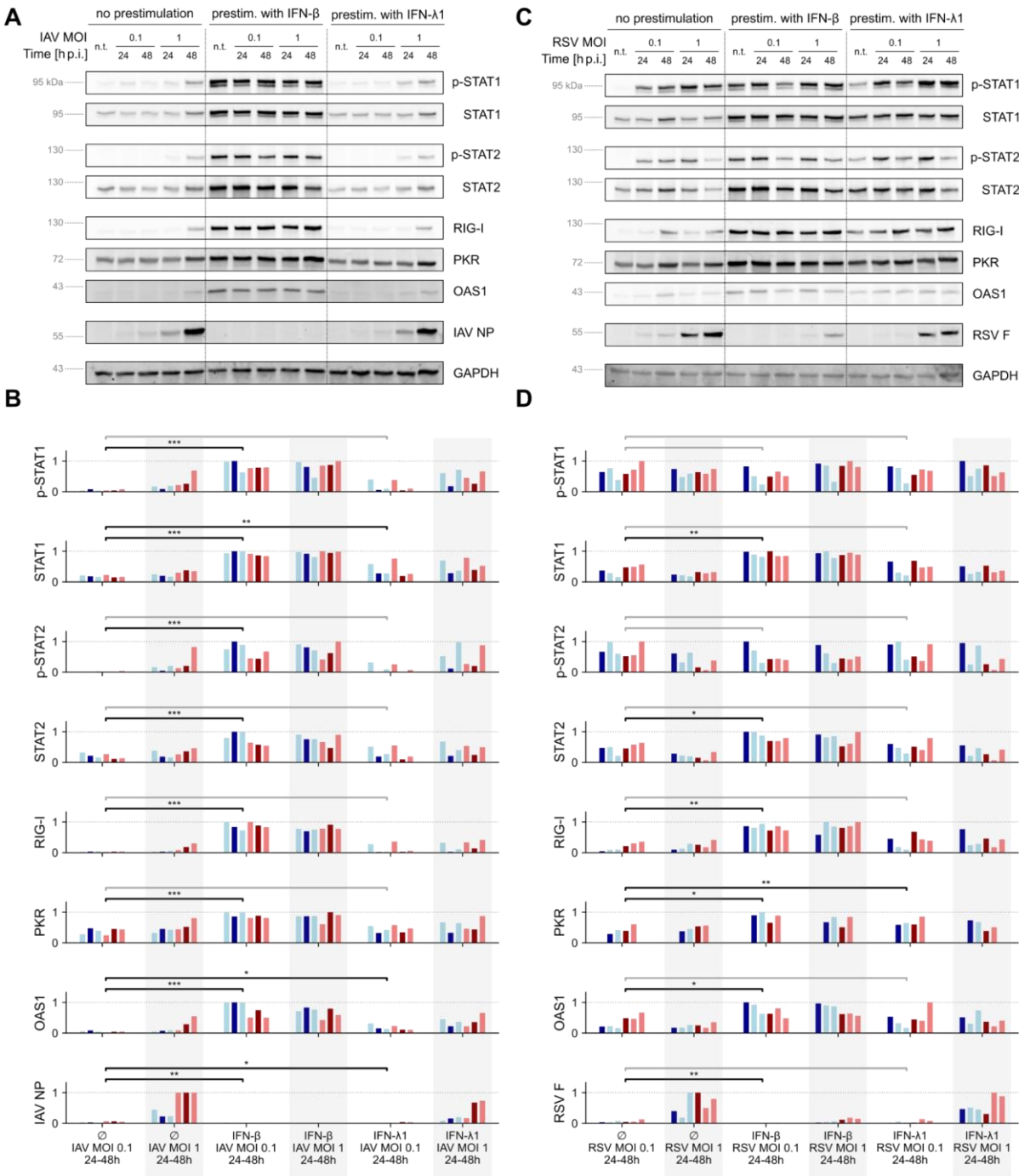


Figure 1. STAT1/2 activation and accumulation of ISGs after stimulation with IFN-β and IFN-λ1.

A A549 cells were stimulated with IFN-β (1000 U/ml) or IFN-λ1 (50 ng/ml) for 1, 2, 6, 24, or 48 h before lysis; p-STAT1, STAT1 phosphorylated at Tyr701; p-STAT2, STAT2 phosphorylated at Tyr690.

B Quantification of two or three experimental replicates (as shown). Blue bars indicate control and time points of 1, 2, and 24 h; red bars indicate 6 and 48 h; darker color indicates the experimental replicate shown in panel A. Each blot quantification is normalized to the maximum. The significances $p < 0.001$ (***) and $p < 0.01$ (**) were determined using a two-tailed Student *t*-test on log ratios.

C RT-PCR of common ISGs as well as *IL6* and *TNFAIP3* after 6, 10, and 24 h stimulation with either IFN-β (1000 U/ml) or IFN-λ1 (50 ng/ml). The averaged log₂ difference (based on 3 experimental replicates) with respect to control is shown. Asterisks indicate that expression of a given gene after stimulation with IFN-β is significantly higher than after IFN-λ1 at $p < 0.05$ (*) or $p < 0.01$ (**); two-tailed Wilcoxon signed-rank test on paired differences; three time points are considered jointly.



655

▲ (continued from the previous page)

656 **Figure 2. Effect of the prestimulation with IFN- β and IFN- λ 1 on STAT1/2**
 657 **activation, expression of ISGs, and virus proliferation.**

658 **A** A549 cells were prestimulated with either IFN- β or IFN- λ 1 (or mock) 24 h before
 659 infection with IAV at MOI of 0.1 or 1 and lysed 24 or 48 h post infection (p.i.). Label
 660 'n.t.' denotes cells not treated with a virus.

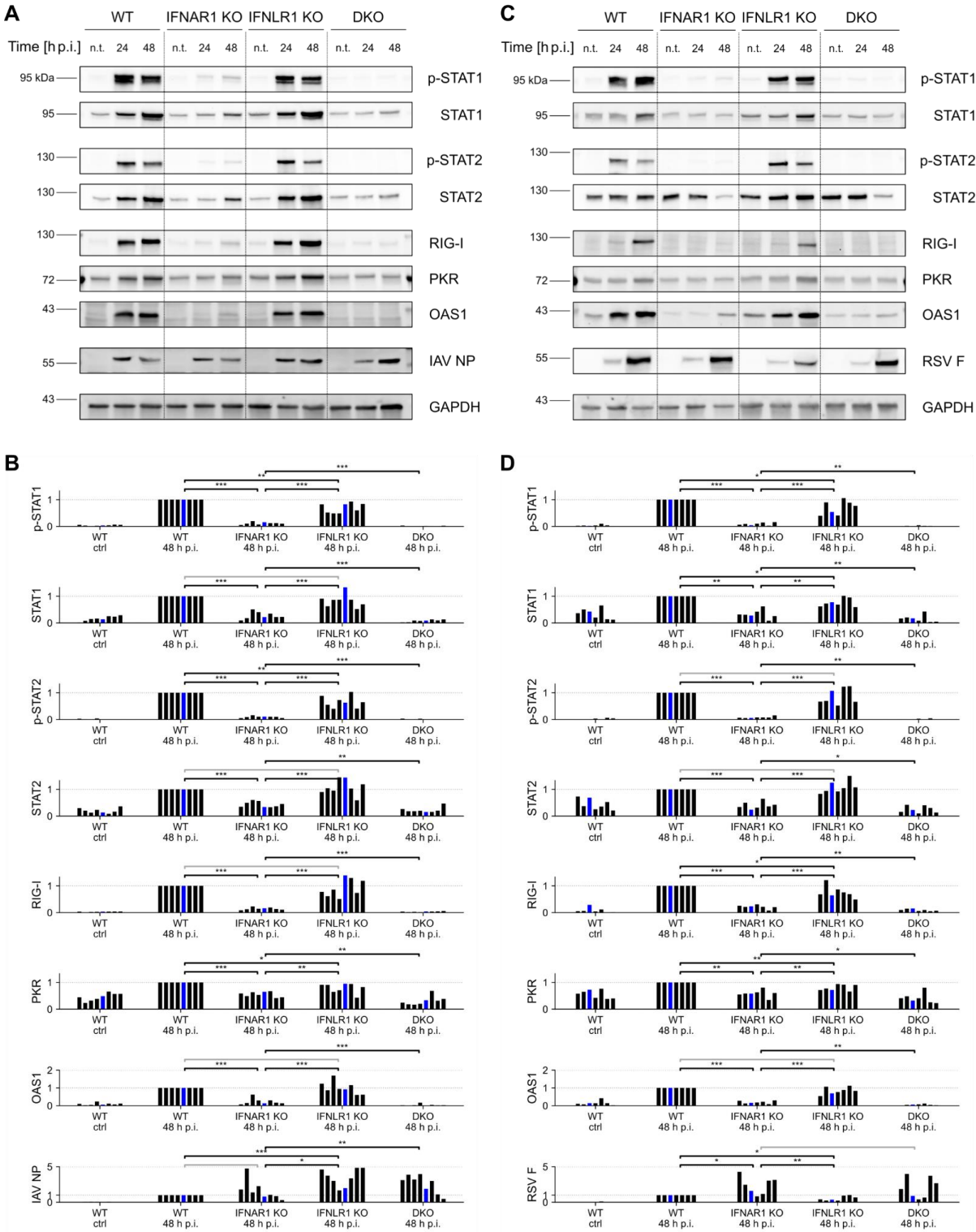
661 **B** Quantification of three replicates of the experiment shown in A. Values for 24 h p.i. and
 662 48 h p.i. are shown in blue and red, respectively. Darker hues correspond to the blot
 663 shown in panel A are shown in blue.

664 **C** A549 cells were prestimulated with IFN- β or IFN- λ 1 (or mock) 24 h before infection
 665 with RSV at an MOI of 0.1 or 1 and lysed 24 or 48 h after the infection. Label 'n.t.'
 666 denotes cells not treated with a virus.

667 **D** Quantification of three replicates of the experiment shown in C. Values for 24 h p.i. and
 668 48 h p.i. are shown in blue and red, respectively. Darker hues correspond to the blot
 669 shown in panel A are shown in blue.

670 In panels A and C, due to all available 15 lanes being used for probes, the marker lane was
 671 omitted. Locations of molecular weight markers were determined based on previous
 672 experiments.

673 In panels B and D, quantification of the bands without infection with IAV or RSV (n.t.) are
 674 omitted. Each blot quantification is normalized to the maximum. The significances $p < 0.001$
 675 (***), $p < 0.01$ (**), $p < 0.05$ (*) were determined using a two-tailed Student t -test on log
 676 ratios.



678

▲ (continued from the previous page)

679 **Figure 3. Effect of IFNLR1 and IFNAR1 knockout on IAV and RSV**
 680 **proliferation and activation of immune response**

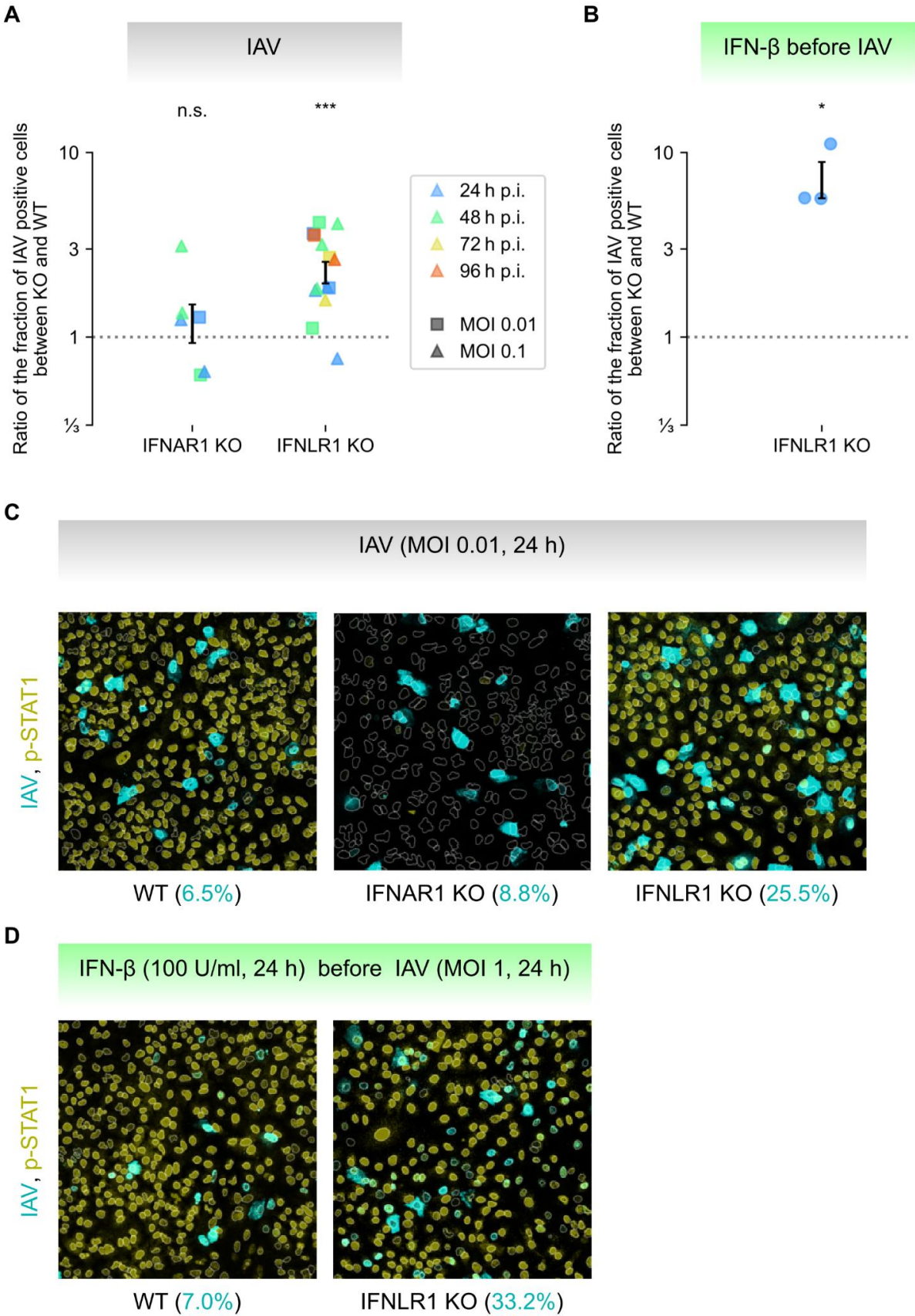
681 **A** A549 WT, IFNAR1 KO, IFNLR1 KO, and IFNAR1 & IFNLR1 double KO (DKO) cells
 682 were infected with IAV at an MOI of 0.1 and lysed 24 or 48 h p.i.

683 **B** Quantification of selected bands from blots in panel A and its seven replicates at 48 h
 684 Quantification of the shown blot is marked in blue.

685 **C** A549 WT, IFNAR1 KO, IFNLR1 KO and DKO cells were infected with RSV at an MOI
 686 of 0.1 and lysed 24 or 48 h p.i.

687 **D** Quantification of selected bands from blots in panel C and its six replicates at 48 h.
 688 Quantification of the shown blot is marked in blue.

689 In panels B and D, $p < 0.001$ (***), $p < 0.01$ (**), and $p < 0.05$ (*) were determined using a
 690 two-tailed Student *t*-test on log ratios. Only selected pairwise differences between WT,
 691 IFNAR1 KO, IFNLR1 KO, and DKO cells were considered. Each blot quantification is
 692 normalized to the value from WT cells at 48 h p.i. The quantification of all time points from
 693 blots in A and C is shown in Figs. EV2 and EV3.



695

▲ (continued from the previous page)

696 Figure 4. Effect of IFNLR and IFNAR knockout on IAV propagation and
697 activation of immune response

698 **A** Fraction of IAV-positive cells in A549 IFNAR1 KO and IFNLR1 KO with respect to the
699 fraction observed for WT. Quantification of three experimental replicates, $p < 0.001$
700 (***) according to the two-tailed Student *t*-test on log ratios for 24 and 48 h p.i. and
701 MOIs of 0.01 and 0.1 considered jointly.

702 **B** Fraction of IAV positive cells in A549 IFNAR1 KO with respect to the fraction observed
703 for WT. Cells were prestimulated with IFN- β at a concentration of 100 U/ml 24 h before
704 infection with IAV at an MOI of 1. Cells were fixed 24 h p.i. Quantification of three
705 experimental replicates, $p < 0.05$ (*), paired two-tailed Student *t*-test on log ratios.

706 **C** Representative fields of view (FOV) showing immunostaining for IAV and p-STAT1 in
707 A549 WT, IFNAR1 KO and IFNLR1 KO cells fixed 24 h p.i. with IAV at an MOI of 0.01.
708 Percentages of infected cells in shown FOVs are given in cyan. All immunostaining
709 images are available at Zenodo (see Data availability).

710 **D** Representative fields of view (FOV) showing immunostaining for IAV and p-STAT1 in
711 A549 WT and IFNAR1 KO cells prestimulated with IFN- β at a concentration of 100
712 U/ml 24 h before infection with IAV at an MOI of 1, and fixed 24 h p.i. Percentages of
713 infected cells in shown FOVs are given in cyan. All immunostaining images are
714 available at Zenodo (see Data availability).

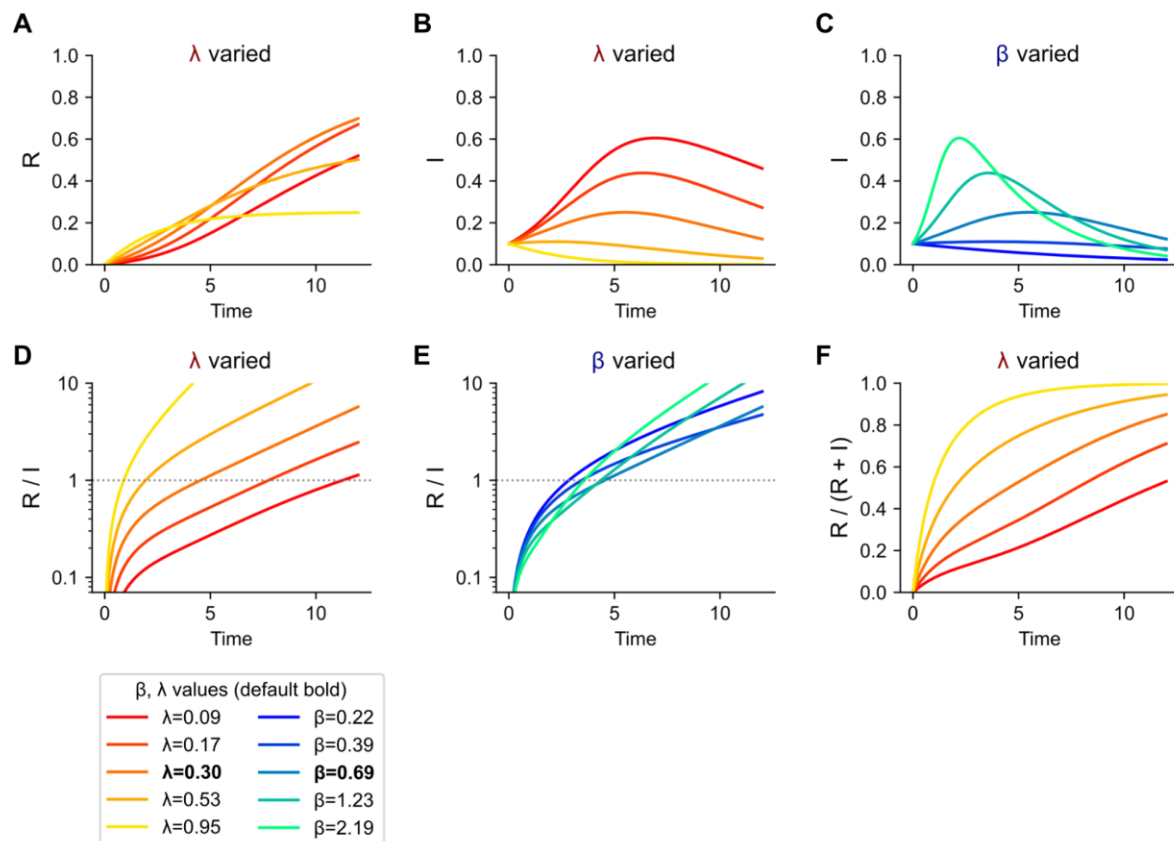


Figure 5. SIR model simulation. Influence of death and infection rates on virus propagation and cell death.

A, B Fraction of dead (A) or infected (B) cells as a function of time and the death rate λ .

C Fraction of infected cells (I) as a function of time and the infection rate β .

D, E Ratio of dead (R) to infected (I) fractions as a function of time and the death rate λ (D) or the infection rate β (E).

F Ratio of dead (R) to joint dead and infected ($R + I$) fractions as a function of time and the death rate λ .

In all panels, $I(0) = 0.1$, corresponding to MOI = 0.1. Unless specified otherwise, $\beta = 0.69$ and $\lambda = 0.3$. The default value of β was selected so that the initial doubling time is 1.

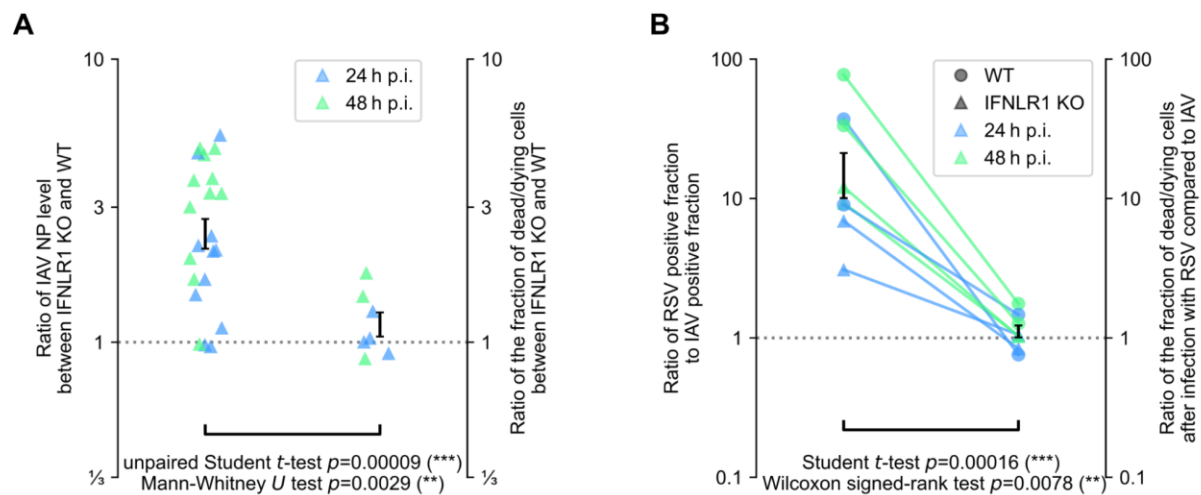
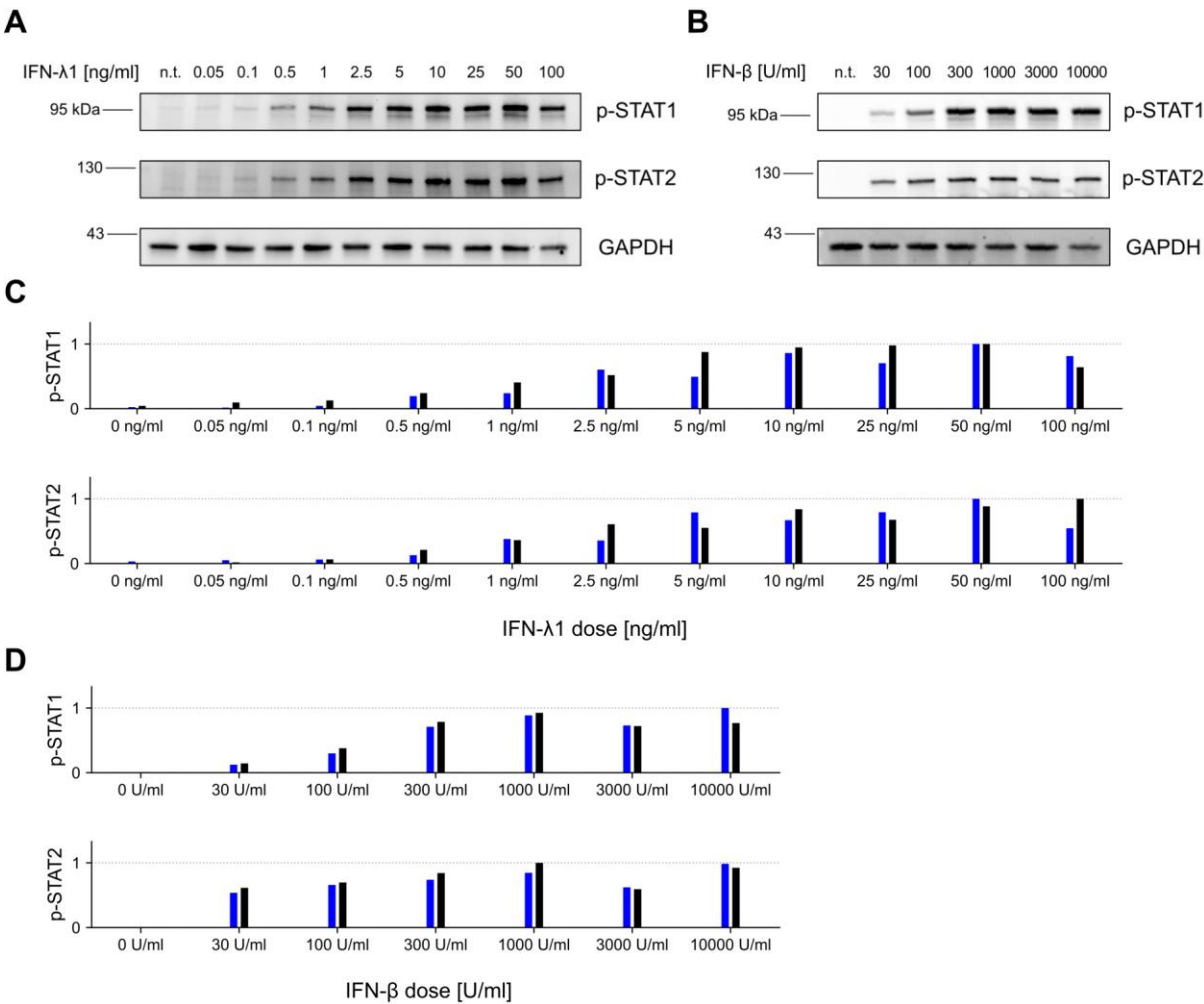


Figure 6. Influence of IFNLR1 deficiency and virus type on infected and dead cell fractions.

- A** Left: ratio of the IAV NP level between IFNLR1 KO and WT cells – calculated based on Western blots, using cells lysed 24 and 48 h p.i. at an MOI of 0.1. Right: ratio of the fraction of dying/dead cells 24 and 48 h p.i. between IFNLR1 KO and WT – calculated based on cell death assay using annexin V and propidium iodide staining (see Methods). Statistical significance was determined by performing an unpaired two-tailed Student *t*-test and a Mann-Whitney *U* test on log ratios (left *versus* right).
- B** Left: ratio of RSV-positive to IAV-positive cell fraction at 24 or 48 h p.i. at an MOI of 0.1 – calculated based on immunostaining images (with staining against respectively RSV and IAV). Right: ratio of the fraction of dead/dying cells after infection with RSV to an analogical fraction after infection with IAV. The dead/dying cells were counted based on the propidium iodide signal during live microscopy experiments lasting 24 or 48 h p.i. The same cells were fixed and immunostained to determine the fraction of virus-infected cells used in the left part of the panel. Statistical significance was determined by performing a two-tailed Student *t*-test and a two-tailed Wilcoxon signed-rank test on log ratio-of-ratios (left / right, see Methods).

744 Expanded View Figures



745

746 Figure EV1. IFN- β and IFN- λ dose-response analysis, determination of
747 saturating concentrations.

748 **A** Cells were stimulated with an array of concentrations of IFN- λ 1 for 30 min before lysis.

749 **B** Cells were stimulated with an array of concentrations of IFN- β for 30 min before lysis.

750 **C** Quantification of 2 experimental replicates, blot shown in panel A is marked in blue.

751 **D** Quantification of 2 experimental replicates, blot shown in panel B is marked in blue.

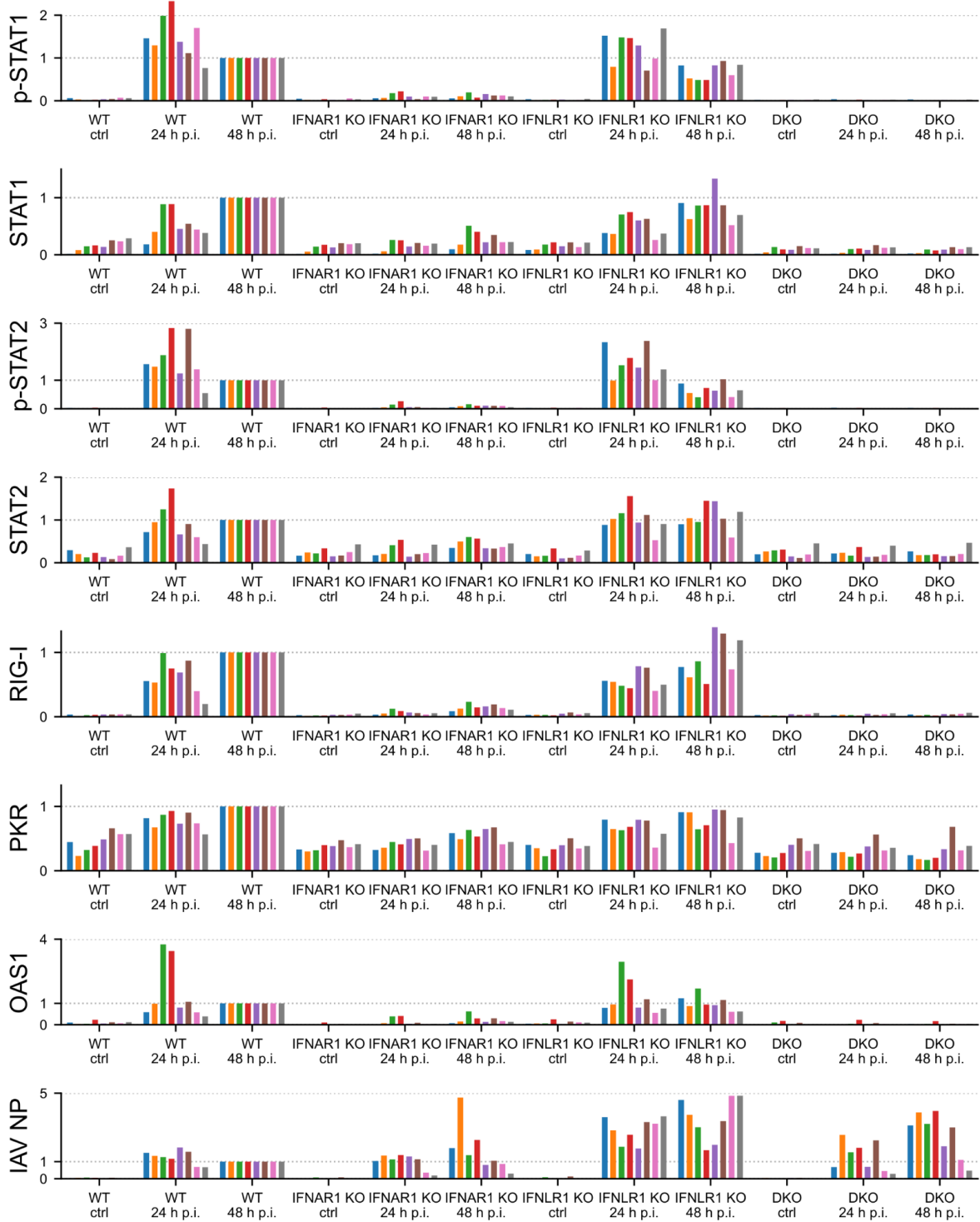


Figure EV2. Effect of IFNLR1 and IFNAR1 knockout on IAV proliferation and activation of the innate immune response – full quantification of experiments shown in Fig. 3A. Colors correspond to independent experimental replicates of experiments shown in Fig. 3A.

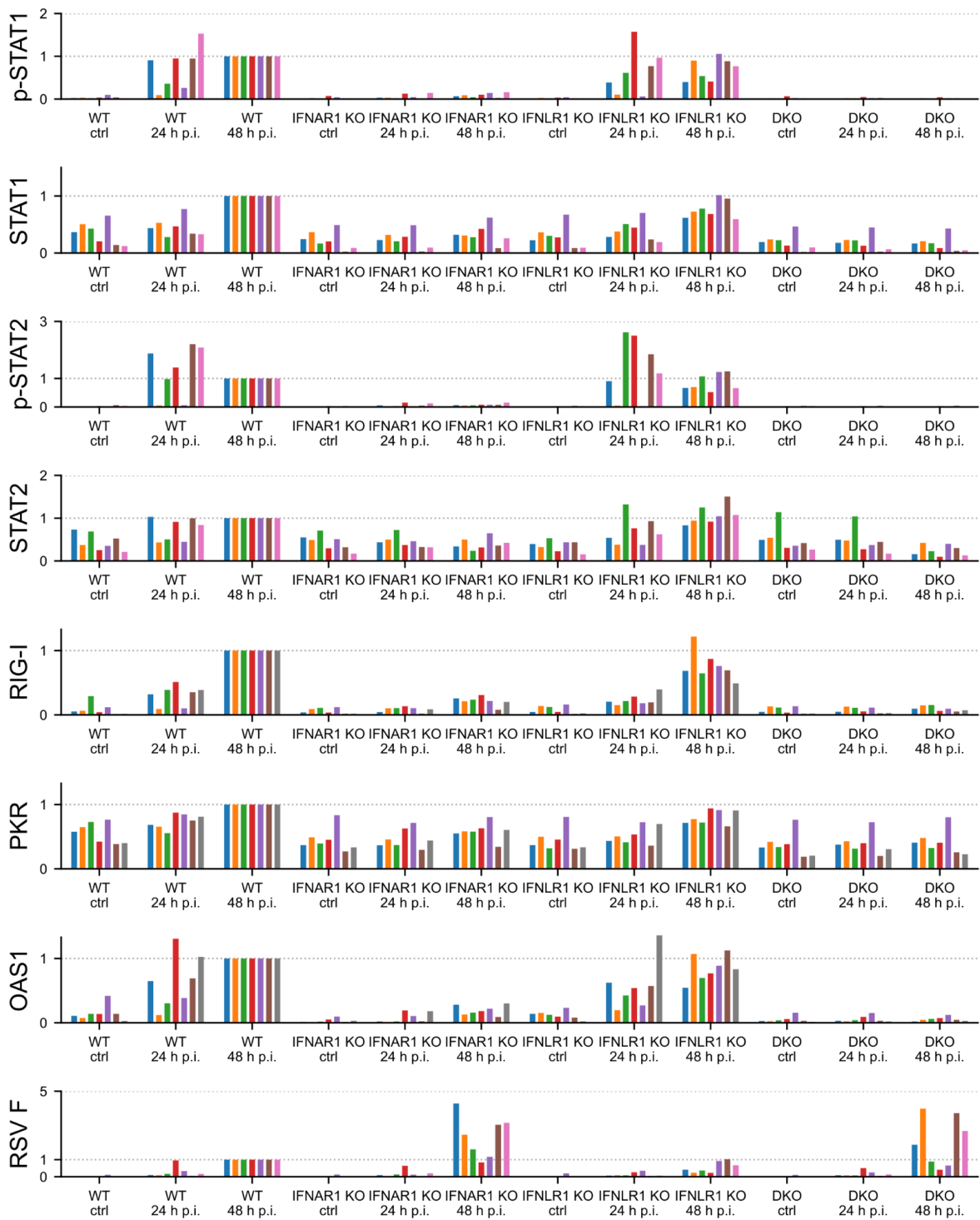
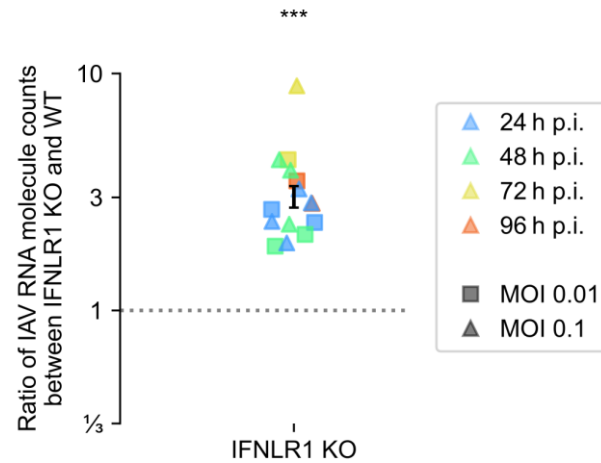


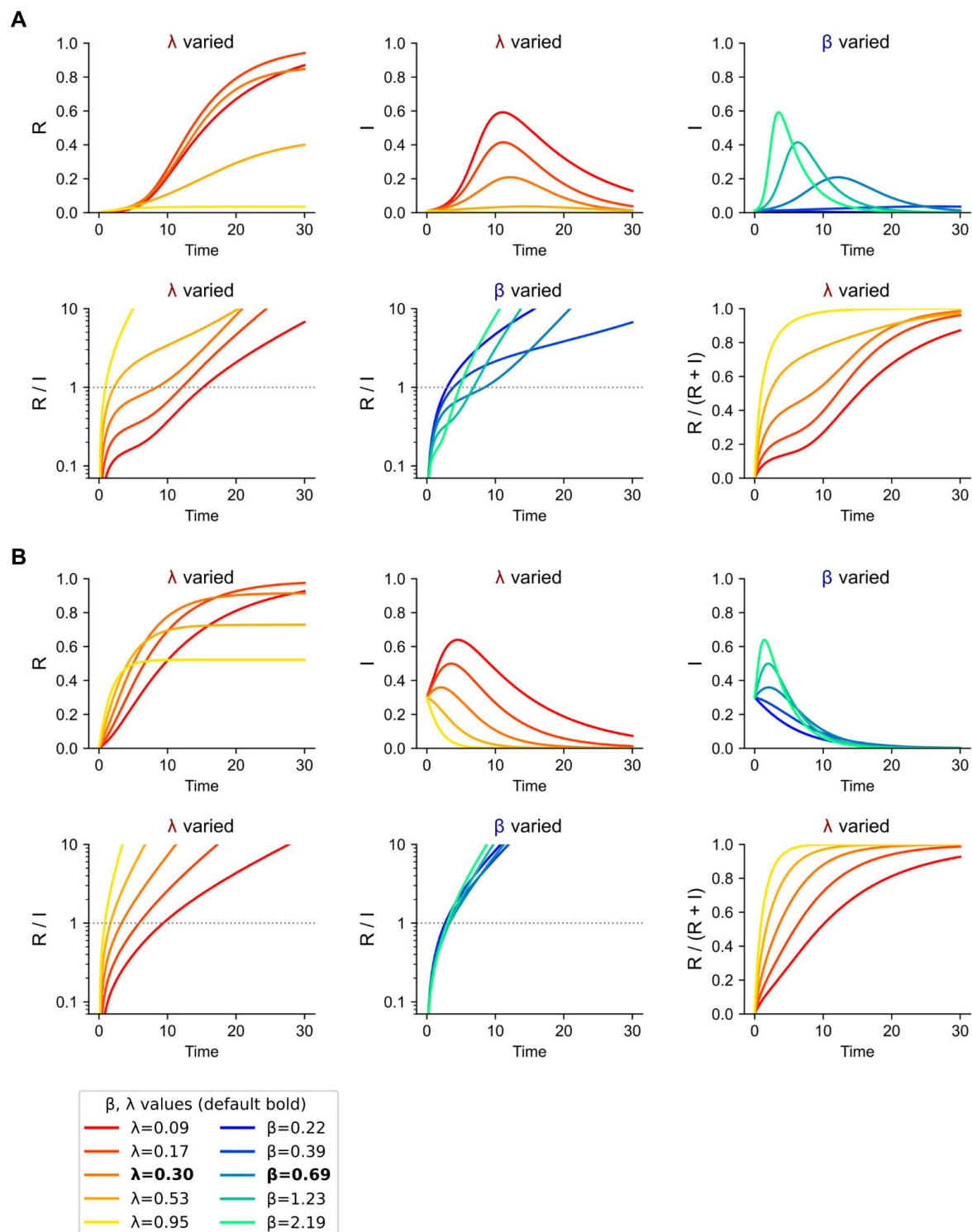
Figure EV3. Effect of IFNLR1 and IFNAR1 knockout on RSV proliferation and activation of the innate immune response – full quantification of experiments shown in Fig. 3B. Colors correspond to independent experimental replicates of experiments shown in Fig. 3B.



762

763 Figure EV4. Effect of IFNLR1 deficiency on IAV RNA expression.

764 Levels of IAV RNA were measured by dPCR for A549 WT and IFNLR1 KO cells at 24, 48,
 765 72 and 96 h p.i. at MOIs of 0.01 and 0.1 (as indicated by the legend). Each square/triangle
 766 corresponds to an experimental replicate. Triple asterisk (***) indicates that IAV RNA
 767 expression is significantly higher in IFNLR1 KO cells than in WT cells ($p < 0.001$, paired two-
 768 tailed Student *t*-test on log ratios)



769

770 Figure EV5. Kinetics of infections in the SIR model. The figure
771 supplements Fig. 5 with data for an MOI of 0.01 (A) and 0.3 (B).

772 **Supplementary Data**

773 Source Data S1 (PDF File). Images of all Western blot replicates.

774 Source Data S2 (Zipped folder). Data quantifications in CSV format (Western blots, dPCR,
775 RT-PCR, FACS, cell counts from immunostaining) and Python codes to generate figures.

DYNAMICS OF COVID-19 PANDEMIC AT CONSTANT AND TIME-DEPENDENT CONTACT RATES

MAREK KOCHAŃCZYK¹, FREDERIC GRABOWSKI² AND
TOMASZ LIPNIACKI^{1,*}

Abstract. We constructed a simple Susceptible–Exposed–Infectious–Removed model of the spread of COVID-19. The model is parametrised only by the average incubation period, τ , and two rate parameters: contact rate, β , and exclusion rate, γ . The rates depend on nontherapeutic interventions and determine the basic reproduction number, $\mathcal{R}_0 = \beta/\gamma$, and, together with τ , the daily multiplication coefficient in the early exponential phase, θ . Initial $\mathcal{R}_0$ determines the reduction of β required to contain the spread of the epidemic. We demonstrate that introduction of a cascade of multiple exposed states enables the model to reproduce the distributions of the incubation period and the serial interval reported by epidemiologists. Using the model, we consider a hypothetical scenario in which β is modulated solely by anticipated changes of social behaviours: first, β decreases in response to a surge of daily new cases, pressuring people to self-isolate, and then, over longer time scale, β increases as people gradually accept the risk. In this scenario, initial abrupt epidemic spread is followed by a plateau and slow regression, which, although economically and socially devastating, grants time to develop and deploy vaccine or at least limit daily cases to a manageable number.

Mathematics Subject Classification. 92D30.

Received March 25, 2020. Accepted April 5, 2020.

1. INTRODUCTION

COVID-19 is a rapidly spreading disease caused by a novel coronavirus, SARS-CoV-2, transmitted between people via respiratory droplets. Typical symptoms are fever, dry cough, and shortness of breath. In severe cases, infection leads to pneumonia and acute respiratory distress syndrome. There is currently no vaccine or specific antiviral treatment. The virus was first reported in Wuhan, Hubei, China, in December 2019. The disease had spread widely over the province of Hubei, but was contained after imposing strict quarantine. Reported peak of the daily confirmed cases of about 4000 was reached on February 3, ten days after the quarantine had been imposed. The number of daily new cases decreased below 100 in early March. Currently (as of March 20, 2020), despite some efforts, nearly exponential growth with more than 20 000 cumulative cases is observed in Italy (nearly 50 000 cumulative cases), Spain, Germany, and the US. The World Health Organization (WHO) declared COVID-19 a pandemic.

Keywords and phrases: basic reproduction number, novel coronavirus.

¹ Department of Biosystems and Soft Matter, Institute of Fundamental Technological Research, Polish Academy of Sciences, 02-106 Warsaw, Poland.

² Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, 02-097 Warsaw, Poland.

* Corresponding author: tlipnia@ippt.pan.pl

© The authors. Published by EDP Sciences, 2020

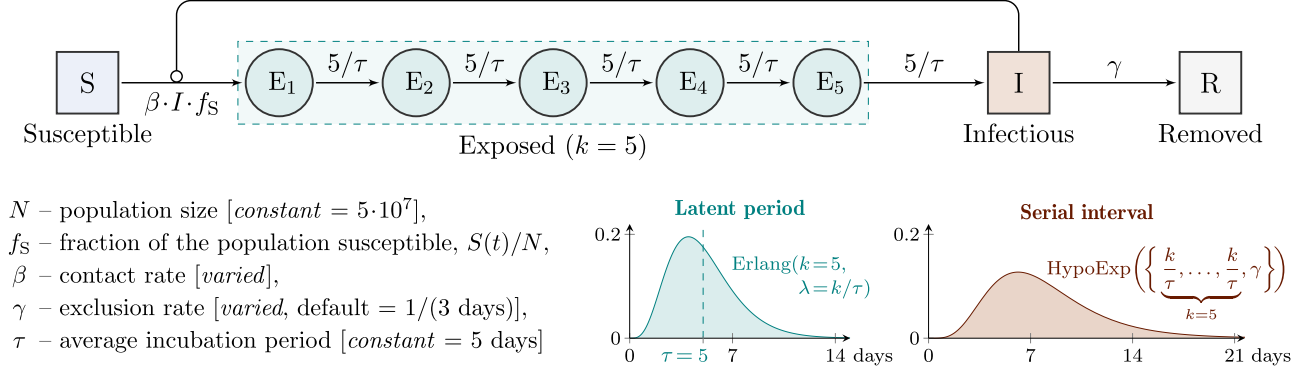


FIGURE 1. Scheme of the model with notation guide and default parameter values. The model is fully characterized by parameters β , γ , and τ . The latent period, assumed to be substitutable with the incubation period, results from the inclusion of $k = 5$ intermediate *exposed* states and follows the Erlang distribution with average $\tau = 5$ days. The serial interval follows the hypoexponential distribution with rates $\Lambda = \{k/\tau, k/\tau, k/\tau, k/\tau, \gamma\}$.

To characterize the dynamics of COVID-19 spread and be able to consider hypothetical long-term scenarios, we developed a simple Susceptible–Exposed–Infectious–Removed model. The model has only 3 parameters: average incubation period, τ , estimated to be approximately 5 days [11, 14], and two parameters that are affected by nonpharmaceutical protective measures: the infectious–susceptible contact rate, β , and the exclusion rate, γ . Isolation (quarantine) reduces β , while testing increases the rate at which infectious individuals are hospitalised or isolated, γ . These two rates determine the basic reproduction number, $\mathcal{R}_0 = \beta/\gamma$. The model is applied to estimate the fold change of β required to contain the epidemic at its early stage. Since initial government-mandated restrictions appear to be insufficient to contain the spread of the disease at its early stage in European countries, we use the model to put forward long-term scenarios, in which we assume time-dependent β that reflects anticipated changes in social behaviour rather than the impact of administratively-imposed control measures.

2. RESULTS

2.1. Model

We consider a variant of the Susceptible–Exposed–Infectious–Removed (SEIR) model, itself an extension of the commonly used SIR model [1, 8]. As in typical SEIR models, we assume that *susceptible* individuals get infected by *infectious* individuals at the rate proportional to the contact rate, β , and in this way become *exposed* (and infected), and later, after the latent period, become *infectious*. Ultimately, at the exclusion rate γ , *infectious* individuals turn into *removed* individuals (which is an umbrella class comprising confirmed cases currently quarantined, handled by the healthcare, recovered with acquired resistance, or the deceased). *Removed* individuals do not infect.

In contrast to most SEIR models, instead of a single *exposed* (infected) state, we have introduced multiple *exposed* states ($E_1, \dots, E_5$ in Fig. 1). With $k = 5$ *exposed* states such representation is equivalent to the assumption that the latent period is distributed according to the Erlang distribution with the “shape” parameter $k = 5$ and the rate parameter $\lambda = k/\tau$, where τ is the mean latent period. Since there are no available estimates of the mean latent period and the disease spreads mainly through airborne droplets disseminated by the coughing or sneezing infected persons, we assume that during the incubation period the infected individuals are effectively noninfectious and thus the mean latent period can be set equal to the estimated average incubation period, $\tau \approx 5$ days [11, 14].

The serial interval, defined as the time between onset of symptoms in a primary case and onset of symptoms in its secondary cases, that can be seen as an emergent property of the model, is distributed according to the hypoexponential distribution resulting from a series of $k = 5$ exponential distributions, each originating from one step of the incubation process, having mean times τ/k , and one exponential distribution, originating from the exclusion process, with mean time $1/\gamma$. At $\tau/k = 1$ day and $1/\gamma = 3$ days, at which the rates of the hypoexponential distribution are $\Lambda = \{1, 1, 1, 1, 1, \frac{1}{3}\}$, the model implicates a distribution of the serial interval that agrees perfectly with the distribution obtained from epidemic data (see Fig. 2B in Ref. [12]).

All model parameters— β , γ , and τ (together with k , as demonstrated further)—are thus constrained by available data. Deterministic model dynamics is described by a system of 8 ordinary differential equations (4.1a)–(4.1e) given and analysed in more detail in Section 4.

2.2. Dynamics of infection spread with respect to all model parameters

The exponential phase of the epidemic growth, during which the number of *removed* (confirmed) individuals grows as $R(t) = R(t=0) \times \exp(\alpha t)$, can be analysed based on an implicit formula (see Sect. 4 for derivation):

$$\frac{\beta}{\gamma + \alpha} = \left(1 + \frac{\alpha \tau}{k}\right)^k, \quad (2.1)$$

which, when the number of intermediate *exposed* steps is very large (formally, when $k \rightarrow \infty$), can be replaced by

$$\frac{\beta}{\gamma + \alpha} = e^{\alpha \tau}. \quad (2.2)$$

In the subsequent analysis we will use the daily multiplication coefficient:

$$\theta := e^{\alpha}, \quad (2.3)$$

that describes the day-to-day increase of the new confirmed cases (for example, when the number of new cases is 25% higher than the number of new cases on the previous day, then $\theta = 1.25$).

In Figure 2 we show the dependence of θ on parameters β , γ , and τ , according to (2.1). Either a decrease of the contact rate, β , or increase of the exclusion rate, γ , result in reduction of θ below 1 and containment of the epidemic. Practically, β is reduced by quarantine or isolation of individuals, which both lower the potential number of contacts per individual per day, while γ is increased by performing more diagnostic tests, which reduces the expected time during which an infected individual may infect susceptible ones. Coefficient θ decreases also with increasing τ (Fig. 2C), but this parameter is not controllable by protective measures.

Equations (2.1) and (2.2) imply that

$$\begin{cases} \alpha > 0 \text{ (equivalently, } \theta > 1) & \text{for } \beta > \gamma, \\ \alpha = 0 \text{ (equivalently, } \theta = 1) & \text{for } \beta = \gamma, \\ \alpha < 0 \text{ (equivalently, } \theta < 1) & \text{for } \beta < \gamma. \end{cases}$$

Further, parameters β and γ determine the expected number of *susceptible* individuals that will be infected by each *infectious* individual, which in the initial phase of the epidemic is expressed by the basic reproduction number:

$$\mathcal{R}_0 := \frac{\beta}{\gamma}. \quad (2.4)$$

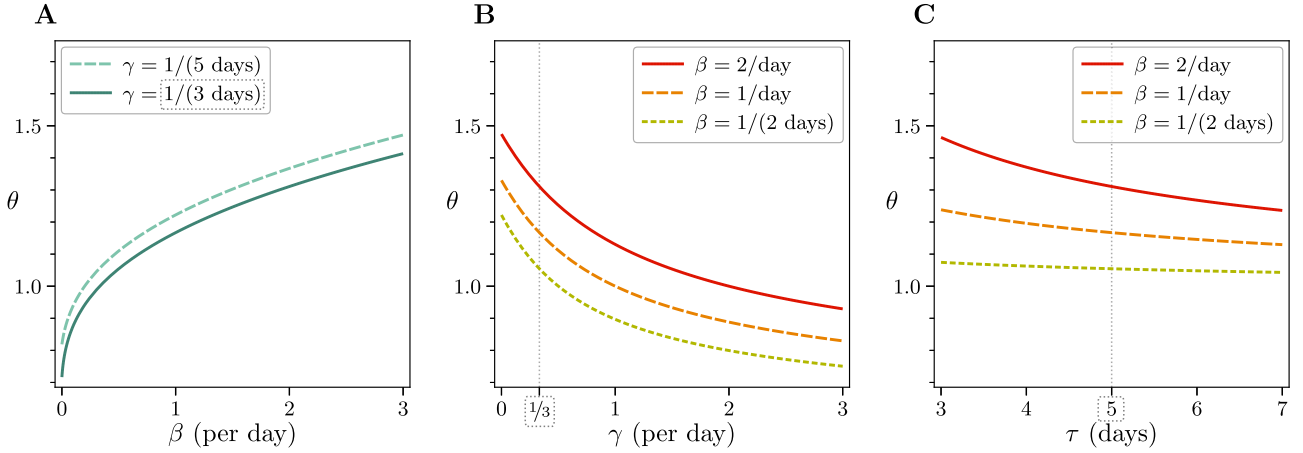


FIGURE 2. Dependence of the daily multiplication coefficient, θ , on all model parameters: (A) contact rate, β , (B) exclusion rate, γ , and (C) average incubation period, τ . Default parameter values are marked with dotted rectangles.

Although the daily multiplication coefficient, θ , is not a function of solely $\mathcal{R}_0$ (Fig. 2), the value of coefficient $\mathcal{R}_0$ is critical for assessment of propagation of the epidemic: $\mathcal{R}_0 > 1$ implies exponential progression, $\mathcal{R}_0 < 1$ implies exponential regression.

2.3. Relation of the daily multiplication coefficient, θ , and the basic reproduction number, $\mathcal{R}_0$

In Figure 3A we show the dependence $\theta(\mathcal{R}_0)$ for two values of γ , $1/(5 \text{ days})$ and $1/(3 \text{ days})$, and in the limit of $\gamma \rightarrow \infty$. The figure should be read as follows. An initial θ that can be determined based on data for a considered country or region corresponds to some value of $\mathcal{R}_0$. Since to terminate the exponential growth phase, $\mathcal{R}_0$ has to be reduced to 1, Figure 3A shows to which extent $\mathcal{R}_0$ should be reduced to contain the epidemic. The reproduction number can be reduced by either reducing the contact rate or increasing the exclusion rate, but the latter is problematic when the number of cases grows rapidly.

Assuming $\gamma = 1/(3 \text{ days})$, which is the default value in the model, the data showing epidemic dynamics in China (Fig. 3B) can be interpreted as follows. Based on the initial exponential phase (trends in one or two weeks marked in Fig. 3B) one can estimate that $\theta = 1.4$ which corresponds to $\mathcal{R}_0 \simeq 8.6$ (Fig. 3A). Thus, to “stabilise” the epidemic, China had to reduce $\mathcal{R}_0$ by a factor of 8.6. However, as indicated by the fall of the number of daily new cases, China reduced θ to approximately $0.89 = \theta'$ (see Sect. 4), which implies $\mathcal{R} = 0.34$ and effective reduction of $\mathcal{R}_0$ about 25 times. We should notice that several authors [9, 19], also based on epidemiological data from China, obtained much lower estimates of $\mathcal{R}_0$; we will come back to this important issue in the Section 2.4.

The ~ 10 -day time lag between introduction of the quarantine and the time of the peak number of daily new cases can be attributed to: (1) the incubation period that causes that infected individuals are registered about $\tau = 5$ days after infection and (2) the fact that the quarantine does not prevent mixing of infectious and susceptible individuals in households, adding next $\tau = 5$ days to onset of the disease among family members. Unfortunately, this implies that even if strict quarantine is imposed immediately in European countries it will have a delayed effect.

In the case of Italy, “soft quarantine” imposed on February 21 in the most affected Italian province, Lombardy, allowed to reduce θ from the initial value of 1.5 (calculated based on the week of February 23–29, 2020) to the current value of $\theta_{\text{last week}} = 1.14$ (calculated based on the last week, March 14–20, 2020). This implies reduction of $\mathcal{R}_0$ from about 12 to 2.6, which is less than 5 times. To just contain the epidemic, that is, reduce $\mathcal{R}_0$ to 1 and stabilise the number of daily cases, Italy should *further* reduce the contact rate more than two-fold. Figure 3A

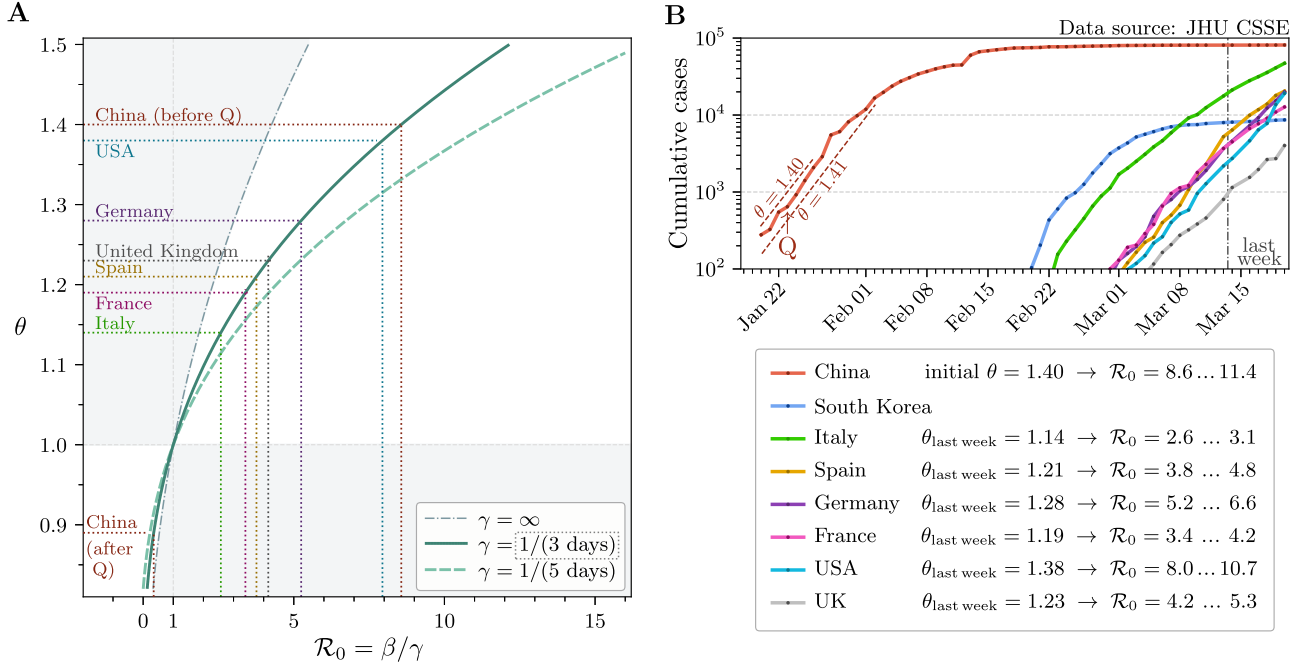


FIGURE 3. Daily multiplication coefficient predicted by the model and estimated from data. (A) Dependence of the daily multiplication coefficient, θ , on the basic reproduction number $R_0 = \beta/\gamma$. Dotted lines guide the eye to relate R_0 and β coming from data for China before and after quarantine ('Q') and for several European countries and the US based on last-week data. (B) Epidemic dynamics in selected countries with high number of cumulative confirmed cases of COVID-19 based on data from Johns Hopkins University (JHU) Center for Systems Science and Engineering (CSSE) [16]. Value of initial θ for China is estimated based on 7 or 14 days of the early exponential phase; current θ is estimated for countries being in the exponential phase based on last-week data (7 days since March 14 to March 20, 2020). Introduction of quarantine in China (January 23) is marked with 'Q'. The range of values of R_0 is computed assuming that $1/\gamma$ is in between 3 and 5 days.

indicates that several other large European countries (Germany, France, Spain, and United Kingdom) as well as the US, each having thousands of cases and last-week $\theta \geq 1.19$ (as of March 20, 2020), to contain the epidemic should reduce the contact rate more than three-fold (in the case of the US, eight-fold). In Figure 3B we give current values of θ and corresponding values of R_0 for the mentioned countries (estimated for $1/\gamma = 3$ days and $1/\gamma = 5$ days).

A much different dynamics can be observed in South Korea, which is exiting the exponential phase (Fig. 3B). Among many preventive countermeasures, until March 20, 2020, South Korea performed 316 000 tests [7], likely substantially increasing the exclusion rate. The example of South Korea is promising as it suggests that, in addition to extensive contact tracing and surveillance, development of easily accessible, fast, and accurate tests may help to contain the spread of the disease.

2.4. Dependence of the basic reproduction number, R_0 , on the number of exposed intermediates, k

Value of R_0 for COVID-19 has been already estimated in several important studies (parameters of which are conveniently summarized on the webpage of an online SEIR model simulator [5]). In particular, Kucharski *et al.* [9] estimated that $R_0 = 2.35$ (95% CI: 1.15–4.77), whereas Tang *et al.* [19] estimated that $R_0 = 6.47$

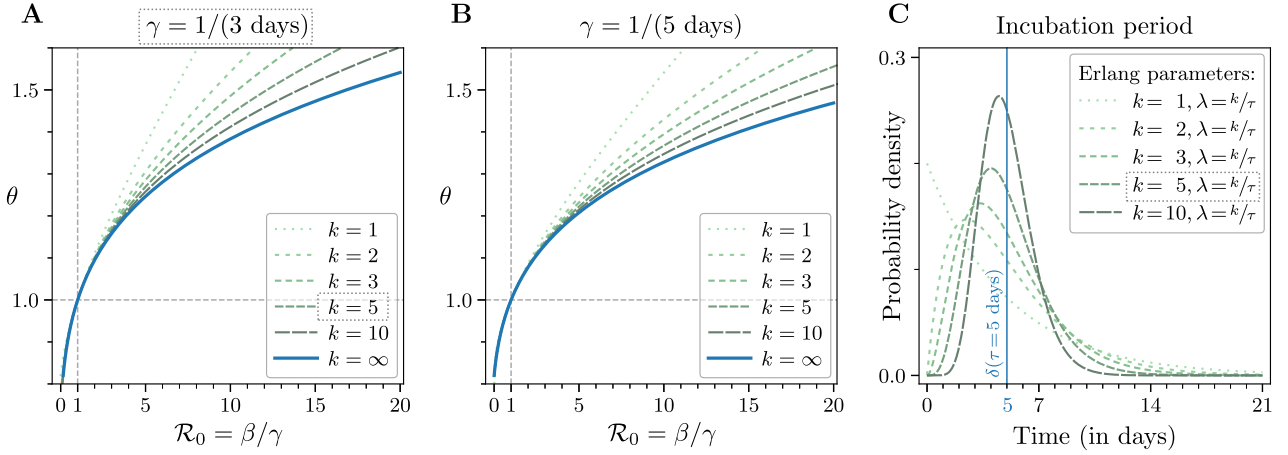


FIGURE 4. (A, B) Dependence of the daily multiplication coefficient, $\theta(\mathcal{R}_0)$, and (C) dependence of the distribution of the incubation period, τ , on the number of intermediate *exposed* states. In the model it is assumed that $k = 5$. In panel C), Erlang distribution with $k = 1$ is equivalent to the exponential distribution with rate parameter $1/\tau$.

(95% CI: 5.71–7.23). Both estimates were based on SEIR-type models. The main difference between our model and these models is the inclusion of a cascade of five intermediate *exposed* states. This is equivalent to assuming that (1) the incubation period is Erlang-distributed with parameters $k = 5$ and $\lambda = k/\tau$, and (2) the serial interval follows a hypoexponential distribution (in agreement with Li *et al.* [12]). Importantly, the simplifying assumption that there is only one intermediate state is equivalent to the assumption that the incubation period follows an exponential distribution. A bit surprisingly, this simplification substantially modifies $\mathcal{R}_0$ estimated from θ . In Figure 4, we show $\theta(\mathcal{R}_0)$ for two values of γ : $1/3$ (default) and $1/5$ day $^{-1}$, calculated assuming that the number of intermediate states is 1, 2, 3, 5, 10, or ∞ (the last case has been determined based on an equivalent model represented by delay differential equations, see Sect. 4). One can observe that for $\theta = 1.4$, the value of $\mathcal{R}_0$ predicted using a model with one intermediate state is more than 1.5 times smaller than $\mathcal{R}_0$ predicted for the model with $k = 5$ intermediate states (5.4 *vs.* 8.6 at $\gamma = 1/3$ day $^{-1}$; of note, the discrepancy would be larger for larger θ ; the ratio of $\mathcal{R}_0$ computed for two different fixed values of k does not depend on γ). As shown in Figure 4C, the exponential distribution of the incubation period with rate $1/\tau$, or even Erlang($k=2, \lambda = k/\tau$), are not supported by data that indicate much narrower distribution with its probability mass separated from zero [11, 14]. In conclusion, we think that our estimate of the basic reproduction number is closer to its true value.

2.5. Hypothetical long-term scenarios

2.5.1. A scenario with constant contact rate, β

The above analysis of the epidemiological situation in several European countries and in the US suggests that the protective measures instituted in these countries failed to contain the epidemic. It is thus worth to consider long-term scenarios in which the cumulative number of *removed* cases, R , becomes comparable to the population size, N , substantially decreasing the susceptible fraction, f_S . In such case, initial exponential growth ceases, the number of daily new cases culminates in a maximum and then drops due to depletion of *susceptible* individuals.

When the reproduction number remains constant, the maximum number of daily new cases is fully determined by initial θ and exceeds 0.3 million for $\theta = 1.05$ and 1 million for $\theta = 1.1$ in a population of $N = 50$ million people (Fig. 5A). It is clear that even the lower of the two numbers exceeds the healthcare capacity of a country

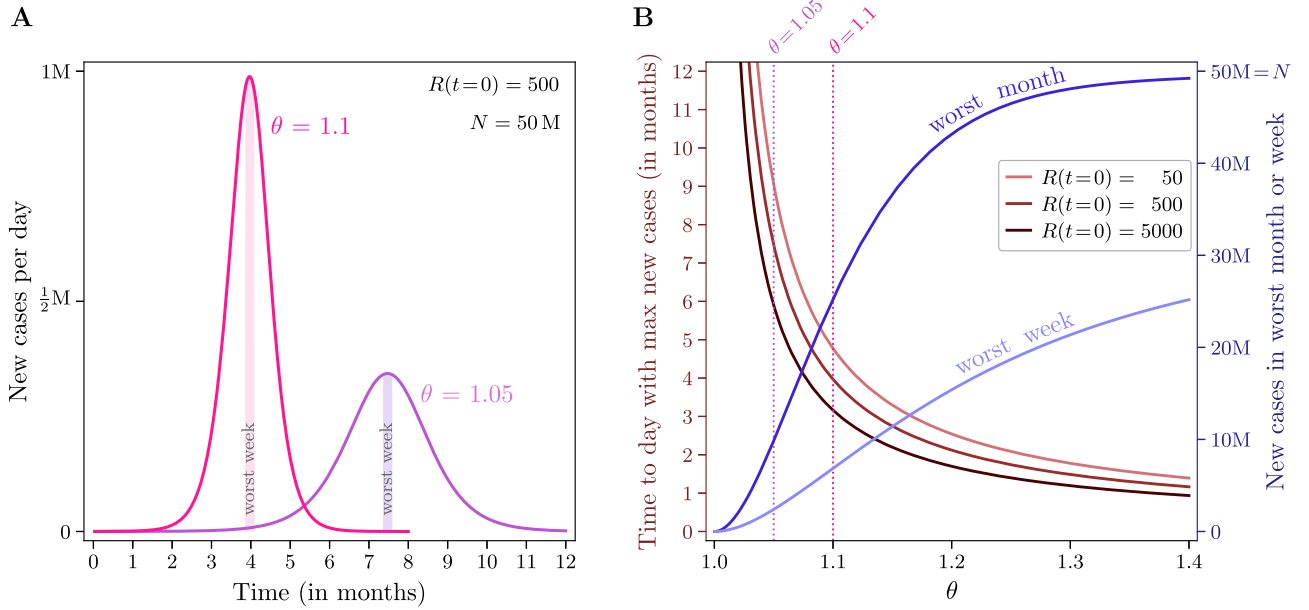


FIGURE 5. Long-term epidemic dynamics at a constant contact rate. (A) Daily new cases for two values of the initial daily multiplication coefficient, θ , and the initial number of *removed* (confirmed) cases, $R(t=0)$, equal 500. (B) Time to the worst day for three initial numbers of *removed* (confirmed) cases and new cases in the worst week and month. In both panels, population size $N = 50$ million, $1/\gamma = 3$ days.

of tens of millions inhabitants such as Italy (~ 60 million), South Korea (~ 50 million), or Poland (~ 40 million). This, in turn, implies higher case fatality rate, which may approach the rate of severe cases (10–15% [10]). Correspondingly, with θ changing from 1.05 to 1.1, time to reach the maximum of daily new cases shortens nearly twice: from 7.5 to 4 months, assuming the initial number of *removed* cases, $R(t=0)$, equal 500. We should notice that achieving $\theta = 1.05$ as well as $\theta = 1.1$ require quarantine; in “non-quarantined” countries, such as currently France, Germany, Spain, United Kingdom, and the US, $\theta \gtrsim 1.2$ (Fig. 3B).

In Figure 5B, we simultaneously show the number of the maximum weekly and monthly new cases and the time to reach the maximum. Parameter $\theta > 1.1$ implies that within one (worst) month more than half of population will be infected, whereas $\theta < 1.1$ require long-term quarantine.

2.5.2. Scenarios with time-dependent contact rate, β

We consider the scenario with the constant reproduction parameter not realistic. Initially, when the number of cases is relatively low, the only way to reduce the contact rate is an administratively-imposed quarantine. However, one could expect that when the number of daily new cases exceeds a threshold, people will begin reducing their contacts voluntarily [2] or taking other protective measures such as careful hand-washing or physical distancing to reduce their number of effective contacts (reviewed in Ref. [4]). To account for this behavioural aspect, we assume that the contact rate is a decreasing function of daily new cases. We propose to assume that

$$\beta = \beta_0 \times \frac{H}{H + (\text{daily new cases})}. \quad (2.5)$$

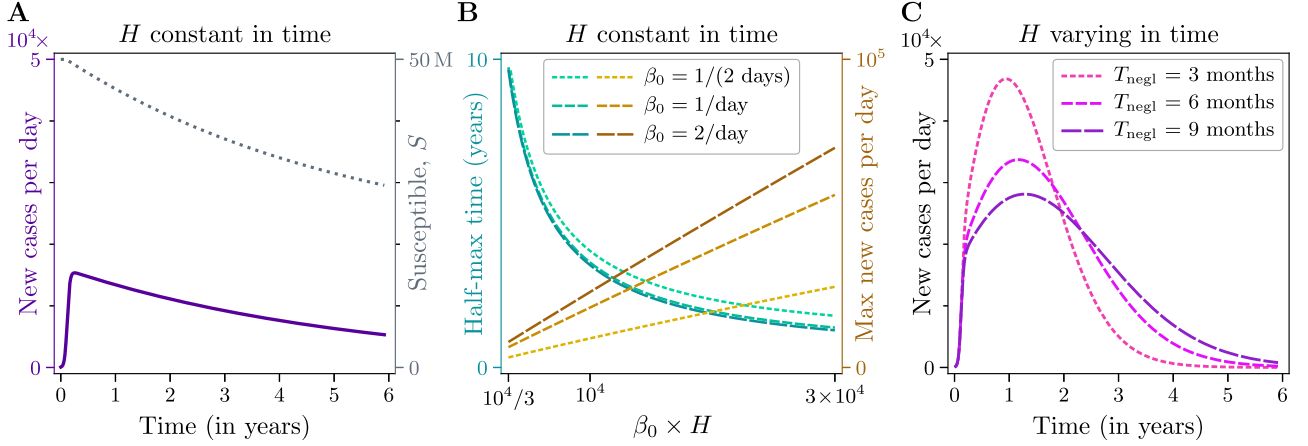


FIGURE 6. Long-term epidemic dynamics with time-dependent contact rate. (A, B) Contact rate reduced in response to the surge in the daily number of cases. (C) Contact rate reduced in response to the surge in the daily number of cases and modulated by the duration of the quarantine. In panels A and C, initial $\theta = 1.14$, which is equal to last-week θ for Italy. In all panels, population size $N = 50$ million and $R(t=0) = 500$.

This simple formula implies, in particular, that when the number of daily new cases equals H , people reduce their contacts twice. The value of parameter H can be culture-dependent. In Figure 6A we show long-term epidemic dynamics assuming $H = 10\,000$. In the considered scenario, the number of daily new cases relatively quickly reaches 15 000 and then decreases slowly (over years) with a decreasing susceptible fraction, f_S . The scenario shown in Figure 6A results in a manageable number of new cases and grants time to develop, produce, and distribute a vaccine. For a given β_0 , the maximum number of daily cases is an increasing function of H , whereas the time in which the number of daily new cases decreases to the half of the maximum value, $t_{\text{max}/2}$, decreases with H ; a bit surprisingly, $t_{\text{max}/2}$ depends mainly on $\beta_0 \times H$ (Fig. 6B).

Finally, we consider a scenario in which a secondary behavioural effect modifies the primary effect discussed above. As before, we assume that people reduce their contacts in response to an increasing risk, but in the course of their (self)quarantine they gradually accept an increasing risk [17]. (As an aside, when protective measures are lifted rapidly, such behaviour can lead to epidemic waves observed during the 1918–1919 A/H1N1 influenza pandemic [6].) Here we propose to model this effect by assuming that

$$H(t) = H_0 \times \frac{t + T_{\text{negl}}}{T_{\text{negl}}}, \quad (2.6)$$

where t is (somewhat arbitrary) time from the beginning of the epidemic and T_{negl} is the time scale on which H (and risk acceptance, or “negligence”) doubles. In Figure 6C we show long-term epidemic dynamics assuming $H = 10\,000$ and three values of T_{negl} : 3, 6, and 9 months. Interestingly, the dynamics is characterized by a relatively broad plateau, which is desirable due to limited capacity of the healthcare system.

3. DISCUSSION

In this study, we proposed a Susceptible–Exposed–Infectious–Removed model to analyse how the dynamics of the COVID-19 spread depends on the contact rate, β , and the exclusion rate, γ —two model parameters that can be modulated by nontherapeutic interventions. The remaining model parameter is the average incubation period, assumed equal 5 days [11, 14]. A decrease of the contact rate, that can be achieved by imposing quarantine, results in a decrease of the daily multiplication coefficient θ that describes growth of the cumulative as well as

the daily new cases in the exponential phase. Coefficient θ can be reduced also by an increase of the exclusion rate that in turn depends on efficiency of the healthcare. Large exclusion rate is achieved by fast identification and isolation of *infectious* individuals. However, when the number of cases is very high, achieving large exclusion rate is nearly impossible; in this case, reduction of the contact rate remains the only option.

In the early phase, before preventive measures were introduced, COVID-19 had been spreading in China and then in Italy with $\theta = 1.4$ and $\theta = 1.5$, respectively. This allowed us to estimate the initial basic reproduction number in these countries: $\mathcal{R}_0 \simeq 9$ and $\mathcal{R}_0 \simeq 12$, respectively. Of note, even without prevention the reproduction number depends on numerous different factors such as sociological context, hygiene, and weather.

Data describing epidemic dynamics in China suggest that very strict quarantine and other preventive efforts imposed in the most affected China province, Hubei, [9, 20] allowed to reduce $\mathcal{R}_0$ about 25-fold (which is on par with the estimated reduction of transmissibility of 97–100% [18]). This allowed the Chinese to reduce coefficient θ from about 1.4 in the exponential growth phase to about $\theta' = 0.89$ in the regression phase. In Italy, however, in response to the mild restrictions, θ was reduced from about 1.5 in the early phase to 1.14 observed currently (March 14–20, 2020), which corresponds to the reduction of $\mathcal{R}_0$ from about 12 to 2.6; thus, further ~ 3 -fold reduction is necessary to contain the exponential epidemic spread.

Higher reduction of $\mathcal{R}_0$ is required in four other considered European countries, France, Germany, Spain, and United Kingdom, with the number of cases exceeding 10 000 and current $\theta \gtrsim 1.2$ (March 14–20, 2020). Such reduction of $\mathcal{R}_0$ would allow only to limit growth of daily new cases, but further four-fold reduction is necessary to bring the epidemic to the regression phase with $\theta' = 0.89$, as was reported in Hubei. Higher $\theta' \in (0.89, 1.0)$ may also allow to eliminate the disease but the required quarantine time would be longer. It appears that most of European countries follow the relatively mild regulations adopted in Italy, which (as of March 20, 2020) proved to be insufficient to contain the epidemic, and, correspondingly, with a week-to-month delay, follow the Italian trajectory of the cumulative daily new cases.

It is important to notice that the current number of *removed* (confirmed) individuals may be much lower than the total number of current *exposed* and *infectious* individuals: $E_1(t) + \dots + E_k(t) + I(t)$. The epidemic growth that is associated with large $\mathcal{R}_0$ implies a proportional large number of not yet confirmed cases that in the next few days will contribute to the number of cumulative confirmed cases. The confirmed fraction $f_{\text{confirmed}}(t) = R(t)/[E_1(t) + \dots + E_k(t) + I(t) + R(t)]$, that remains constant in time in the exponential phase, can be calculated based on (4.3) and (4.4) as $f_{\text{confirmed}} = 1/\mathcal{R}_0$.

Of concern is COVID-19 lethality, which in quantitative terms is usually expressed as the case fatality rate, assessed as the ratio of deaths to confirmed cases. This definition, used by WHO, works well in the steady state when the number of active cases remains constant or when historical data of previous epidemics are analysed. However, in the exponential growth phase, the case fatality rate underestimates the real rate of deaths due to the disease. When the average time between the onset of symptoms or positive test and death is n days, the current cumulative number of deaths should be divided by the cumulative number of cases recorded m days before, which is θ^n times lower than the current number. This observation may to some extent explain puzzling large differences between fatality rates across European countries. In Germany, having the lowest fatality rate (among countries with a high number of cases), identification of people by testing at early stage of the disease and high standard of public healthcare may result in large n , which together with high $\theta = 1.28$ (March 14–20, 2020) may result in a significant underestimation of the case fatality rate. Assuming n between 10 and 14 days, one can find that the case fatality rate may be underestimated by a factor between 1.28^{10} and 1.28^{14} , that is, 12–32 times. Finally, we should notice that in noncontainment scenarios, the expected death toll may not be predicted based on the case fatality rate, because a significant number of cases may be asymptomatic (and some individuals could possibly turn out resistant) [13].

Since current data suggest that in Europe, US, and Iran the epidemic will not be contained in its early phase, we have analysed longer-term scenarios. The model with the constant contact and exclusion rates predicts that even for a relatively small exponential growth phase coefficient, $\theta = 1.05$, the number of daily new cases (in a 50-million population) will exceed 0.3 million, well above the capacity of the healthcare system. Additionally, keeping the coefficient θ at such low level requires stringent limitation of personal contacts, while the time to

the peak of the epidemic would be about 7 months. One of potential consequences of exceeding the healthcare system capacity is the increase of case fatality rate due to the lack of necessary medical staff and equipment.

We think that the more realistic scenarios are these in which the contact rate varies over time. In early phase of the epidemic, the contact rate may be reduced only by forcing people to stay home; in the latter phase, when the number of daily cases exceeds a threshold, people isolate themselves to reduce the risk. In such scenario, the number of daily new cases reaches peak proportional to an assumed “fear” threshold and then slowly decreases due to the decreasing fraction of susceptible individuals. Such scenario seems more realistic and, although devastating for both the economy and social life, grants time to develop and administer vaccine. Historical data on the 1918–1919 A/H1N1 influenza pandemic suggest also that this “fear” threshold may not be constant in time, because people suffering from prolonged quarantine may tend to accept higher risk. When this “negligence” effect is included, one may obtain trajectories for which fast growth is followed by a plateau and then relatively fast decrease of daily cases. A bit surprisingly, this scenario, although not resulting from centrally imposed preventive policies, may be the most plausible non-containment scenario balancing health and economic costs.

4. METHODS

4.1. Model equations

The dynamics of the system depicted in Fig. 1 is governed by the following system of 8 ordinary differential equations:

$$\frac{dS}{dt} = -\beta f_S(t) I(t), \quad (4.1a)$$

$$\frac{dE_1}{dt} = \beta f_S(t) I(t) - \frac{k}{\tau} E_1(t), \quad (4.1b)$$

$$\frac{dE_j}{dt} = \frac{k}{\tau} E_{j-1}(t) - \frac{k}{\tau} E_j(t) \quad \text{for } 2 \leq j \leq k, \quad (4.1c)$$

$$\frac{dI}{dt} = \frac{k}{\tau} E_k(t) - \gamma I(t), \quad (4.1d)$$

$$\frac{dR}{dt} = \gamma I(t), \quad (4.1e)$$

where $k = 5$ is the assumed number of intermediate *exposed* states and $f_S = S(t)/N$ with constant $N = S(t) + E_1(t) + \dots + E_k(t) + I(t) + R(t)$ being the population size (with included deaths).

In the early phase of the epidemic, $f_S \sim 1$. Under the assumption that $f_S = 1$, the systems has the exponential solution:

$$E_j(t) = E_{j0} \exp(\alpha t) \quad \text{for } 1 \leq j \leq k, \quad (4.2a)$$

$$I(t) = I_0 \exp(\alpha t), \quad (4.2b)$$

$$R(t) = R_0 \exp(\alpha t), \quad (4.2c)$$

with R_0 being the initial cumulative number of confirmed cases and

$$I_0(R_0) = R_0 \frac{\alpha}{\gamma}, \quad (4.3a)$$

$$E_{j0}(R_0) = R_0 \frac{\beta}{\gamma} \frac{\alpha \tau}{k} \left(1 + \frac{\alpha \tau}{k}\right)^{-j} \quad \text{for } 1 \leq j \leq k, \quad (4.3b)$$

where α is given implicitly:

$$\beta = (\gamma + \alpha) \left(1 + \frac{\alpha \tau}{k}\right)^k. \quad (4.4)$$

When the population cannot be considered as entirely susceptible, that is, when $N - S(t) \ll N$ is not satisfied, the system becomes essentially nonlinear and has no analytical solutions. This system is solved numerically using Wolfram MATHEMATICA (Wolfram Research, Inc., Champaign, IL, US; notebook is available online [15]). All numerical simulations start from an initial condition in which $E_i(t=0) = E_{i0}$, $I(t=0) = I_0$, $R(t=0) = R_0$, and $S(t=0) = S_0 = N - \sum_{j=1}^k E_{j0} - I_0 - R_0$, where R_0 is the given initial cumulative number of confirmed cases and $E_{i0}(R_0)$ and $I_0(R_0)$ are computed according to 4.3, assuring that the initial condition lies on the exponential trajectory. In the model variant with time-dependent contact rates, daily new cases are computed as $\gamma I(t)$. Simulations of this model variant start from the same initial condition that lies on the exponential trajectory.

It is noteworthy that in the limit of $k \rightarrow \infty$ the considered model converges to the following three-compartment ‘‘SER’’ model with delay τ (equal to the average incubation period):

$$\frac{dS}{dt} = -\beta f_S(t) E(t - \tau), \quad (4.5a)$$

$$\frac{dE}{dt} = \beta f_S(t) E(t - \tau) - \gamma E(t - \tau), \quad (4.5b)$$

$$\frac{dR}{dt} = \gamma E(t - \tau). \quad (4.5c)$$

with $f_S(t) = S(t)/N$ where $N = S(t) + E(t) + R(t) = \text{const}$. The delay differential equations-based formulation (4.5a–4.5c) may admit negative-valued solutions in the epidemic regression phase.

4.2. Estimation of θ

Parameter θ for the initial exponential growth case is estimated based on JHU CSSE data [16] of cumulative cases as:

$$\theta = \left(\frac{\text{cumulative number of cases in day } d_{\text{last}}}{\text{cumulative number of cases in day } d_{\text{first}}} \right)^{\frac{1}{\Delta t}}, \quad (4.6)$$

where for China day d_{first} is January 20 and day d_{last} is January 26 or February 2, 2020, with $\Delta t = d_{\text{last}} - d_{\text{first}} - 1 = 6$ or 13, respectively. For other countries, last-week data are used: d_{first} is March 14 and d_{last} is March 20, 2020; $\Delta t = 6$.

It should be noticed that as long as the epidemic progresses exponentially, coefficient θ estimated based on the increase of cumulative cases $R(t)$ and coefficient θ' estimated from the number of daily new cases, $R(t) - R(t - (1 \text{ day}))$, should be equal (which is obscured by unavoidable fluctuations in daily data). However, when the epidemic ceases, θ calculated based on $R(t)$ reaches 1, while θ' becomes smaller than 1, reflecting an exponential fall of the number of the daily new (as well as the active) cases. Parameter θ' in the epidemic regression phase in China has been estimated as the decay rate of the daily new cases in the time span from February 2 to March 2, 2020.

Note added in the proof. As of April 4, 2020, the nonessential industry lockdown in Italy, Spain, and France, which led to an unprecedented decrease of workplace traffic to 1/3 of pre-epidemic level [3], allowed to terminate the exponential growth phase. These three countries appear to have reached (at least local) maxima of daily new cases.

Acknowledgements. This study was supported by the National Science Centre (Poland) grant 2018/29/B/NZ2/00668.

REFERENCES

- [1] R.M. Anderson and R.M. May, *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, New York (1991).
- [2] V. Capasso and G. Serio, A generalization of the Kermack–McKendrick deterministic epidemic model. *Math. Biosci.* **42** (1978) 43–61.
- [3] COVID-19 Community Mobility Reports. Available at: <https://www.google.com/covid19/mobility/> (2020).
- [4] S. Funk, M. Salathe and V.A.A. Jansen, Modelling the influence of human behaviour on the spread of infectious diseases: a review. *J. R. Soc. Interface* **7** (2010) 1247–1256.
- [5] G. Goh, Epidemic Calculator. Available at: <https://gabgoh.github.io/COVID/index.html> (2020).
- [6] R.J. Hatchett, C.E. Mecher and M. Lipsitch, Public health interventions and epidemic intensity during the 1918 influenza pandemic. *Proc. Natl Acad. Sci. USA* **104** (2007) 7582–7587.
- [7] How many tests for COVID-19 are being performed around the world? Available at: <https://ourworldindata.org/covid-testing> (2020).
- [8] W.O. Kermack and A.G. McKendrick, Contributions to the mathematical theory of epidemics—I. *Proc. R. Soc.* **115A** (1927) 700–721.
- [9] A.J. Kucharski, T.W. Russell, C. Diamond, Y. Liu, J. Edmunds *et al.*, Early dynamics of transmission and control of COVID-19: A mathematical modelling study. To appear in: *Lancet Infect. Dis.* (2020). doi: [10.1016/S1473-3099\(20\)30144-4](https://doi.org/10.1016/S1473-3099(20)30144-4)
- [10] C.-C. Lai, T.-P. Shih, W.-C. Ko, H.-J. Tang and P.-R. Hsueh, Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and coronavirus disease-2019 (COVID-19): the epidemic and the challenges. *Int. J. Antimicrob. Agents* **55** (2020) 105924.
- [11] S.A. Lauer, K.H. Grantz, Q. Bi, F.K. Jones, Q. Zheng *et al.*, The incubation period of 2019-nCoV from publicly reported confirmed cases: estimation and application. Preprint medRxiv doi: [10.1101/2020.02.02.20020016](https://doi.org/10.1101/2020.02.02.20020016) (2020).
- [12] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou *et al.*, Early transmission dynamics in Wuhan, China, of novel Coronavirus infected pneumonia. *N. Engl. J. Med.* **382** (2020) 1199–1207.
- [13] R. Li, S. Pei, B. Chen, Y. Song, T. Zhang *et al.*, Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science* (2020) eabb3221.
- [14] N.M. Linton, T. Kobayashi, Y. Yang, K. Hayashi, A.R. Akhmetzhanov *et al.*, Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data. *J. Clin. Med.* **9** (2020) 538.
- [15] Mathematica notebook. Available at: <http://pmbm.ippt.pan.pl/model/covid19> (2020).
- [16] Novel coronavirus (COVID-19) cases, provided by JHU CSSE. Available at: <https://systems.jhu.edu/research/public-health/ncov> (2020).
- [17] N. Perra, D. Balcan, B. Goncalves and A. Vespignani, Towards a characterization of behavior-disease models. *PLoS One* **6** (2011) e23084.
- [18] J. Riou, A. Hauser, M.J. Counotte and C.L. Althaus, Adjusted age-specific case fatality ratio during the COVID-19 epidemic in Hubei, China, January and February 2020. Preprint medRxiv doi: [10.1101/2020.03.04.20031104](https://doi.org/10.1101/2020.03.04.20031104) (2020).
- [19] B. Tang, X. Wang, Q. Li, N.L. Bragazzi, S. Tang *et al.*, Estimation of the transmission risk of the 2019-nCoV and its implication for public health interventions. *J. Clin. Med.* **9** (2020) 462.
- [20] X. Wang, W. Tian, X. Lv, Y. Shi, X. Zhou *et al.*, Effects of Chinese strategies for controlling the diffusion and deterioration of novel coronavirus-infected pneumonia in China. Preprint medRxiv doi: [10.1101/2020.03.10.20032755](https://doi.org/10.1101/2020.03.10.20032755) (2020).

Research



Cite this article: Kocharczyk M, Grabowski F, Lipniacki T. 2020 Super-spreading events initiated the exponential growth phase of COVID-19 with $\mathcal{R}_0$ higher than initially estimated. *R. Soc. Open Sci.* **7**: 200786.
<http://dx.doi.org/10.1098/rsos.200786>

Received: 12 May 2020

Accepted: 3 September 2020

Subject Category:

Mathematics

Subject Areas:

biomathematics/mathematical modelling

Keywords:

COVID-19, reproduction number

Author for correspondence:

Tomasz Lipniacki

e-mail: tlipnia@ippt.pan.pl

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.5120378>.

Super-spreading events initiated the exponential growth phase of COVID-19 with $\mathcal{R}_0$ higher than initially estimated

Marek Kocharczyk¹, Frederic Grabowski²
and Tomasz Lipniacki¹

¹Department of Biosystems and Soft Matter, Institute of Fundamental Technological Research, Polish Academy of Sciences, 02-106 Warsaw, Poland

²Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, 02-097 Warsaw, Poland

MK, 0000-0003-1215-3920; FG, 0000-0003-4070-9500; TL, 0000-0002-3488-2561

The basic reproduction number $\mathcal{R}_0$ of the coronavirus disease 2019 has been estimated to range between 2 and 4. Here, we used an SEIR model that properly accounts for the distribution of the latent period and, based on empirical estimates of the doubling time in the near-exponential phases of epidemic progression in China, Italy, Spain, France, UK, Germany, Switzerland and New York State, we estimated that $\mathcal{R}_0$ lies in the range 4.7–11.4. We explained this discrepancy by performing stochastic simulations of model dynamics in a population with a small proportion of super-spreaders. The simulations revealed two-phase dynamics, in which an initial phase of relatively slow epidemic progression diverts to a faster phase upon appearance of infectious super-spreaders. Early estimates obtained for this initial phase may suggest lower $\mathcal{R}_0$.

1. Introduction

The basic reproduction number $\mathcal{R}_0$ is a critical parameter characterizing the dynamics of an outbreak of an infectious disease. By definition, $\mathcal{R}_0$ quantifies the expected number of secondary cases generated by an infectious individual in an entirely susceptible population. $\mathcal{R}_0$ may be influenced by natural conditions (such as seasonality) as well as socio-economic factors (such as population density or ingrained societal norms and practices) [1]. Accurate estimation of $\mathcal{R}_0$ is of crucial importance

because it informs the extent of control measures that should be implemented to terminate the spread of an epidemic. Also, $\mathcal{R}_0$ determines the immune proportion f of population that is required to achieve herd immunity, $f = 1 - 1/\mathcal{R}_0$.

A preliminary estimate published by the World Health Organization (WHO) suggested that $\mathcal{R}_0$ of coronavirus disease 2019 (COVID-19) lies in between 1.4 and 2.5 [2]. Later this estimate has been revised to 2–2.5 [3], which is broadly in agreement with numerous other studies that, based on official data from China, implied the range of 2–4 (see Liu *et al.* [4] or Boldog *et al.* [5] for a summary). This range suggests an outbreak of a contagious disease that should be containable by imposition of moderate restrictions on social interactions. Unfortunately, moderate restrictions that were implemented in e.g. Italy or Spain turned out to be insufficient to prevent a surge of daily new cases and, consequently, nationwide quarantines had to be introduced.

We estimated the range of $\mathcal{R}_0$ of COVID-19 based on the doubling times observed in the exponential phases of the epidemic in China, Italy, Spain, France, UK, Germany, Switzerland and New York State. For each of these locations, we used trajectories of both cumulative confirmed cases and deaths [6]. Since our stochastic simulations suggested that the epidemic may have two-phase dynamics—slow (and susceptible to extinction) before any super-spreading events occur, and fast and steadily expanding after the occurrence of super-spreading events—to capture the second phase of the trajectories, we analysed them after a fixed threshold of cases or deaths had been exceeded, in two-week intervals. Both the stochastic simulations and $\mathcal{R}_0$ estimates were obtained within a susceptible–exposed–infected–removed (SEIR) model that correctly reproduces the shape of the latent period distribution and yields a plausible mean generation time. We concluded that the range of $\mathcal{R}_0$ is 4.7–11.4, which is considerably higher than most early estimates. We conjecture that these early estimates were obtained for the first phase of the epidemic in which super-spreading events were absent.

2. Results

2.1. The SEIR model

We used an SEIR model (see Methods for model equations and justification of parameter values) in which:

- we assumed that the latent period is the same as the incubation period and is Erlang-distributed with the shape parameter $m = 6$ and the mean of 5.28 days $= 1/\sigma$ [7];
- we assumed that the infectious period is Erlang-distributed with the shape parameter $n = 1$ (exponentially distributed) or $n = 2$, and the mean of 2.9 days $= 1/\gamma$ [8,9];
- the infection rate coefficient β was determined from σ , γ , m , n and doubling time $\mathcal{T}_d$, which in turn was estimated based on the epidemic data as described in the next subsection, ultimately allowing us to estimate $\mathcal{R}_0 = \beta/\gamma$ as $\mathcal{R}_0(\mathcal{T}_d)$.

The use of the Erlang distributions directly translates to the inclusion of multiple consecutive substates in the SEIR model, meaning that we assumed m ‘exposed’ substates and n ‘infectious’ substates (Erlang distribution is a distribution of a sum of independent exponentially distributed variables of the same mean).

2.2. Estimation of $\mathcal{R}_0$ in the exponential growth phase

First, we estimated the doubling time $\mathcal{T}_d$ within two-week periods beginning on the day in which the number of confirmed (in the SEIR model naming convention, ‘removed’, see Methods) cases exceeded 100 or the number of deaths exceeded 10 in China, six European countries and New York State (figure 1a,b). Values of $\mathcal{T}_d$ that we obtained lie in between $\mathcal{T}_d^{\min} = 1.86$ days (based on cases in New York State) and $\mathcal{T}_d^{\max} = 2.96$ days (based on deaths in Switzerland).

Then, we estimated the range of $\mathcal{R}_0$ as a function of the doubling time $\mathcal{T}_d$ using a formula that takes into account the mean latent and infectious period, $1/\sigma$ and $1/\gamma$, respectively, as well as the shape parameters m and n , see equation (4.8) in Methods. The lower bound has been obtained using the model variant with $n = 2$ (two ‘infectious’ substates), whereas the upper bound results from the model with $n = 1$ (one ‘infectious’ substate), figure 1c. After plugging $\mathcal{T}_d^{\max}$ and $\mathcal{T}_d^{\min}$ in, respectively, the variant of our model with the lower $\mathcal{R}_0(\mathcal{T}_d)$ curve ($n = 2$) and the variant with the higher $\mathcal{R}_0(\mathcal{T}_d)$ curve ($n = 1$), we arrived at the estimated $\mathcal{R}_0$ range of 4.7–11.4. The cases-based doubling time for China, 2.36, is consistent with the value of 2.4 reported by Sanche *et al.* [11], who estimated that $\mathcal{R}_0$

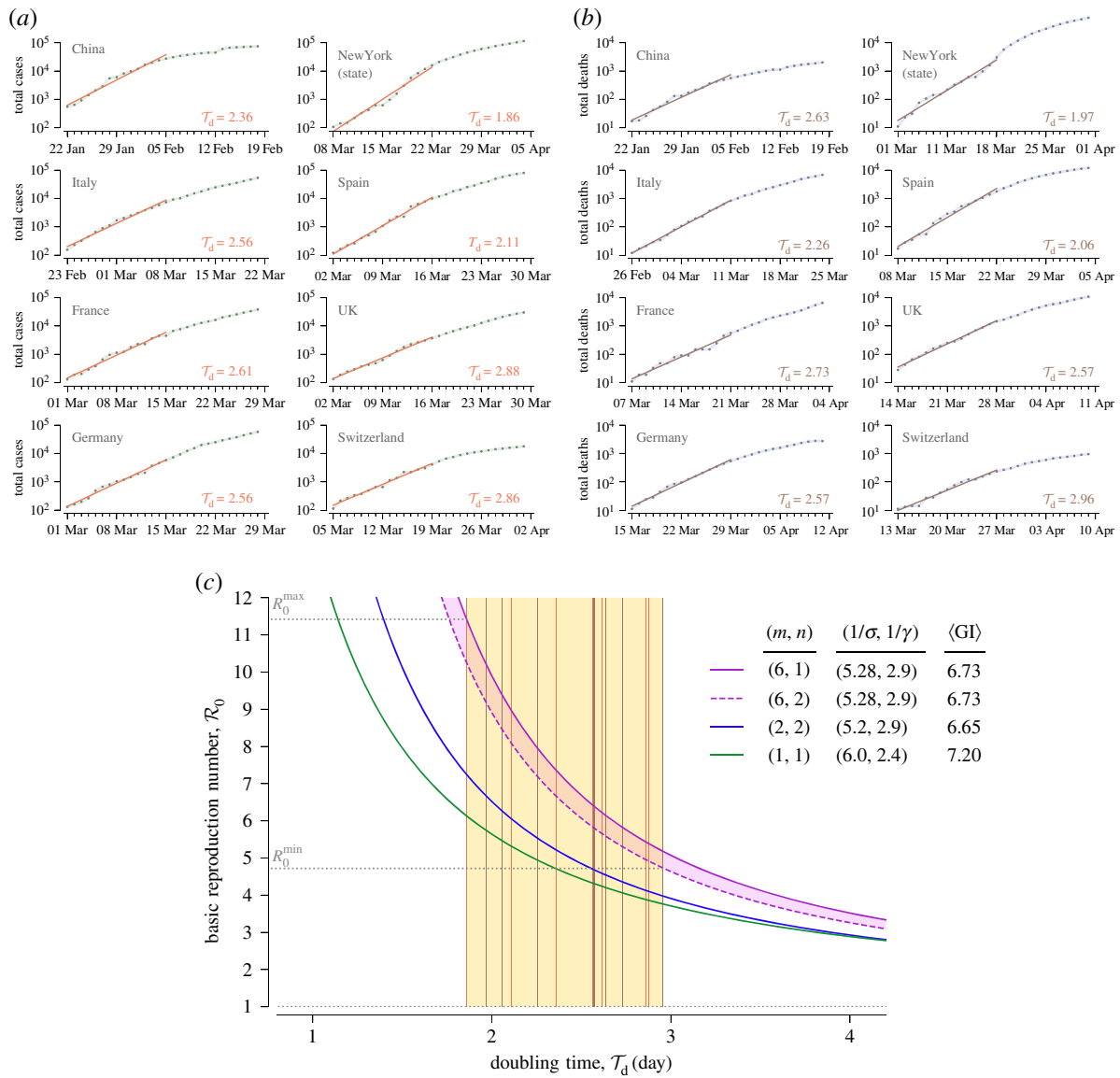


Figure 1. Estimation of the doubling time and the resulting basic reproduction number $\mathcal{R}_0$. (a,b) Estimates of the doubling time $\mathcal{T}_d$ for China, six European countries and New York State using two-week periods beginning (a) when the number of confirmed cases exceeds 100 or (b) when the number of deaths exceeded 10, according to data gathered and made available by Johns Hopkins University [6]. (c) The range of $\mathcal{R}_0$ estimated using two variants of our SEIR model (violet solid and dashed curves) for the range of $\mathcal{T}_d$ estimated in (a) and (b). Vertical lines in the yellow area are $\mathcal{T}_d$ estimates based on the cumulative number cases (orange, from (a)) or the cumulative number of deaths (brown, from (b)). Blue and green solid curves correspond to $\mathcal{R}_0(\mathcal{T}_d)$ according to SEIR models structured and parametrized as in the study of Kucharski *et al.* [9] ($m = 2, n = 2$) and Wu *et al.* [10] ($m = 1, n = 1$).

for China lies in the range 4.7 to 6.6, overlapping with our estimated range for China: 5.6–7.3. The models having one or two ‘exposed’ substates, often used to estimate the value of $\mathcal{R}_0$, substantially underestimated $\mathcal{R}_0$, cf. figure 1c and the articles by Wearing *et al.* [12], Wallinga & Lipsitch [13] and Kořańczyk *et al.* [14].

There are two main reasons why our estimates of the basic reproduction number are higher compared to other published estimates:

- Our SEIR model comprises six ‘exposed’ substates to account for the latent period distribution. As shown in figure 1c, the broader latent period distribution, exponential (i.e. Erlang with $m = 1$), results in lower $\mathcal{R}_0$ estimates than the Erlang with $m = 2$ (at the same remaining model parameters). We characterized sensitivity of $\mathcal{R}_0$ with respect to the mean latent period, $1/\sigma$, in electronic supplementary material, figure S1, while in figure S2 we show that the assumed latent period distribution is in agreement with epidemiological estimates [7,15,16].
- We estimated the doubling time, $\mathcal{T}_d$, from the growth of the number of cumulative cases and cumulative deaths in the two-week-long exponential phases of the epidemic in six locations, obtaining $\mathcal{T}_d$ ranging from 1.86 to 2.96. These values are much lower than the values reported

Table 1. Relation between $\mathcal{T}_d$ model parameters (mean latent period or mean incubation period, $1/\sigma$; mean period of infectiousness, $1/\gamma$; and consequent mean generation interval, $\langle \text{GI} \rangle$), mean serial period, $\langle \text{SI} \rangle$ and $\mathcal{R}_0$. All estimates are based on the epidemic development in Hubei province of China. The unit of all values, except for $\mathcal{R}_0$, is day. Confidence intervals are given in oval brackets; a credible interval is given in square brackets.

$\mathcal{T}_d$	$1/\sigma$	$1/\gamma$	$\langle \text{SI} \rangle$ or $\langle \text{GI} \rangle$	$\mathcal{R}_0$	reference
?	5.2	2.9	6.65 ^a	2.35 (1.15–4.77)	Kucharski <i>et al.</i> [9]
5.2 (4.6–6.1)	6.5	?	7.0 (5.8–8.1)	1.94 (1.83–2.06)	Wu <i>et al.</i> [17]
6.4 [5.8–7.1]	6	2.4 ^b	8.4	2.68 (2.47–2.86)	Wu <i>et al.</i> [10]
7.4	5.2 (4.1–7.0)	?	7.5 (5.3–19)	2.2 (1.4–3.9)	Li <i>et al.</i> [16]

^aThe $\langle \text{GI} \rangle$ value is not given in the article but calculated from the assumed values of $1/\sigma$ and $1/\gamma$ as $\langle \text{GI} \rangle = 1/\sigma + \frac{1}{2}/\gamma$ [18].

^bThe value $1/\gamma$ was obtained by the authors as $\langle \text{SI} \rangle - 1/\sigma$, which is inconsistent with the assumption that the infection occurs in a random time during the period of infectiousness.

in the early influential studies of Wu *et al.* [10,17] and Li *et al.* [16]: 5.2 days, 6.4 days and 7.4 days, correspondingly. In these studies, the basic reproduction number has been estimated to lie in between 1.94 and 2.68. A summary in table 1 shows that the lower $\mathcal{R}_0$ estimates follow from much longer estimates of $\mathcal{T}_d$.

2.3. Impact of super-spreading on $\mathcal{T}_d$ estimation

The discrepancy in $\mathcal{T}_d$ estimation may be potentially attributed to the fact that not all ‘removed’ individuals are registered. In the case when the ratio of registered to ‘removed’ individuals is increasing over time, the true increase of the ‘removed’ cases may be overestimated. We do not rule out this possibility, although we consider it implausible as the expansion of testing capacity in considered countries has been slower than the progression of the outbreak. We rather attribute the discrepancy to the fact that in the early phase, in which the doubling time (growth rate) is estimated based on individual case reports, the consequences of potential super-spreading events (such as football matches, carnival fests, demonstrations, masses or hospital-acquired infections) are negligible due to a low probability of such events when the number of infected individuals is low. In a given region or country, occurrence of first super-spreading events triggers transition to the faster-exponential growth, in which subsequent super-spreading events become statistically significant and may become decisive drivers of the epidemic spread [19]. Based on case reports in China, Sanche *et al.* [11] inferred that the initial epidemic period in Wuhan has been dominated by simple transmission chains. Phylogenetic analyses by Worobey *et al.* [20] revealed that the first cases recorded in USA and Europe did not initiate sustained SARS-CoV-2 transmission networks. In turn, super-spreading events were very likely the main drivers of the epidemic spread in e.g. Italy and Germany, where, in the early exponential phase, spatial heterogeneity of registered cases had been evident [21,22]. In Italy, Spain and France, this explosive phase was followed by a phase of slower growth, during which mass gatherings were forbidden, but quarantine (that finally brought the effective reproduction number below 1) had not been yet introduced.

Motivated by these considerations, we analysed the impact of super-spreading on estimation of $\mathcal{T}_d$ based on stochastic simulations of SEIR model dynamics (see electronic supplementary material, listing S1). Simulations were performed in the perfectly mixed regime according to the Gillespie algorithm [23]. We assumed that a predefined fixed proportion of individuals (equal to 33%, 10%, 3% or 1%) has higher infectiousness and as such is responsible for on average either half of infections (super-spreaders) or two-third of infections (hyper-spreaders). To reproduce these fractions in systems with different assigned proportions of super- or hyper-spreaders, their infectiousness is assumed to be inversely proportional to their ratio in the simulated population. In figure 2, we show dynamics of the epidemic

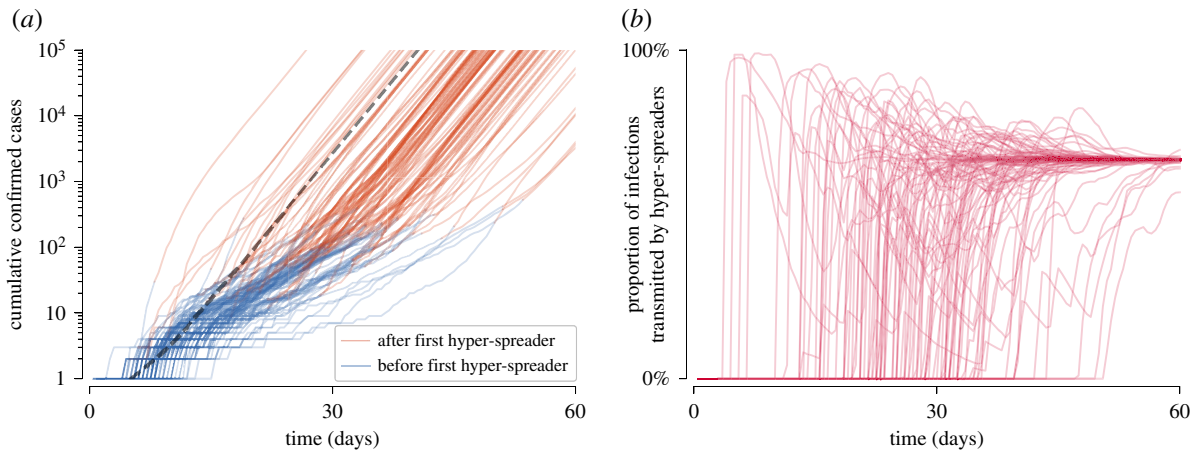


Figure 2. Stochastic epidemic spread in the presence of 1% of hyper-spreaders. (a) Trajectories of confirmed cases (cumulative R in terms of SEIR compartments) resulting from 100 independent stochastic simulations. When the first hyper-spreading event occurs, the colour of the line is changed from blue to brown. Dashed grey line shows a deterministic trajectory. (b) Proportion of infections transmitted by hyper-spreaders among all transmission events over time. Stochastic trajectories stabilize at 66.7%. Trajectories shown in both panels results from the same set of simulations; simulations resulting in outbreak failure were discarded. Model parameters used for simulations in both panels: $(m, n) = (6, 1)$, $(1/\sigma, 1/\gamma) = (5.28 \text{ days}, 2.9 \text{ days})$. Infection rate coefficient of hyper-spreaders was set $\beta_h = 198 \times \beta_n$ (where β_n is the infection rate coefficient for normal spreaders), which assures that in the deterministic limit 66.7% of infections are transmitted by hyper-spreaders. In turn, β_n was set such that the average infection rate coefficient $\beta = 2.97 \times \beta_n$ gives $\mathcal{T}_d = 2$ days (see equation (4.7) in Methods).

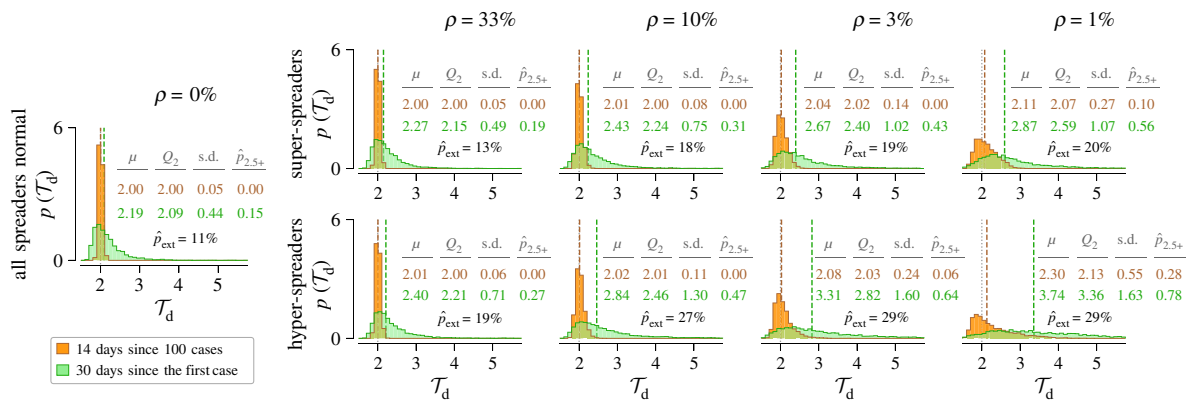


Figure 3. Estimation of the doubling time $\mathcal{T}_d$ based on stochastic simulations of the SEIR model with super- and hyper-spreaders. Histograms show probability density $p(\mathcal{T}_d)$ estimated using the '14 days since 100 cases' method (orange) and the '30 days since the first case' method (green). In each column, ρ denotes a fixed proportion of super-spreaders (top row) or hyper-spreaders (bottom row) in the population. For decreasing proportions of super- and hyper-spreaders (from left, except the shared leftmost panel with $\rho = 0$, to right), their infection rate coefficient β has been reduced to give the same deterministic $\mathcal{T}_d = 2$ days (vertical dotted grey lines). Remaining model parameters: $(m, n) = (6, 1)$; $(1/\sigma, 1/\gamma) = (5.28 \text{ days}, 2.9 \text{ days})$. Each histogram results from 5000 stochastic simulations starting from a single infected normal individual; trajectories resulting in outbreak failure were discarded; fraction of trajectories that resulted in epidemic extinction for given conditions is given as $\hat{p}_{\text{ext}}$. Each distribution is described in terms of its mean (μ), median (Q_2 and vertical dashed lines), standard deviation (s.d.) and the fraction of probability mass for $\mathcal{T}_d > 2.5$ days ($\hat{p}_{2.5+}$).

spread in the presence of 1% of hyper-spreaders to demonstrate that the phase of slower growth is transformed into the faster-exponential growth phase upon the occurrence of hyper-spreading events.

We estimated $\mathcal{T}_d$ in two ways: based on one month of growth of the number of new cases since the first registered case (30 days since the first case) and based on growth of new cases in the two-week period after the number of registered cases exceeds 100 (14 days since 100 cases). As we are interested in the initial phase characterized by exponential growth, we assumed that the susceptible population remains constant. In figure 3, we show histograms of $\mathcal{T}_d$ calculated using either the '14 days since 100 cases' method or the '30 days since the first case' method. One may observe that the histograms calculated using the '30 days since the first case' method are broader than those calculated using the

Table 2. Doubling time ($\mathcal{T}_d$) estimates using the ‘14 days since 100 cases’ or the ‘30 days since the first case’ method. For all locations except China, the data gathered and made available by Johns Hopkins University [6] are used. As for the second method, the calculation either starts from the first case or from two first cases (if no date is provided for the first case). In the case of China (for which Johns Hopkins University [6] early data are not available) data from [24] are used.

location	‘14 days since 100 cases’			‘30 days since the first case’		
	$\mathcal{T}_d$	from	up to	$\mathcal{T}_d$	from	up to
China	2.36	548 cases (22 Jan 2020)	23 707 cases (4 Feb 2020)	4.43	1 case (1 Dec 2019)	37 cases (30 Dec 2019)
Italy	2.56	155 cases (23 Feb 2020)	5883 cases (7 Mar 2020)	3.25	2 cases (31 Jan 2020)	1128 cases (29 Feb 2020)
Spain	2.11	120 cases (2 Mar 2020)	7798 cases (15 Mar 2020)	6.35	1 case (1 Feb 2020)	84 cases (1 Mar 2020)
France	2.61	130 cases (1 Mar 2020)	4496 cases (14 Mar 2020)	12.30	2 cases (24 Jan 2020)	12 cases (22 Feb 2020)
UK	2.88	134 cases (2 Mar 2020)	3077 cases (15 Mar 2020)	7.66	2 cases (31 Jan 2020)	61 cases (29 Feb 2020)
Germany	2.56	130 cases (1 Mar 2020)	4585 cases (14 Mar 2020)	12.56	1 case (27 Jan 2020)	17 cases (25 Feb 2020)
Switzerland	2.86	114 cases (5 Mar 2020)	3028 cases (18 Mar 2020)	2.34	1 case (25 Feb 2020)	10 897 cases (25 Mar 2020)
NY (state)	1.86	106 cases (8 Mar 2020)	11 727 cases (21 Mar 2020)	1.92	1 case (2 Mar 2020)	75 833 case (31 Mar 2020)

‘14 days since 100 cases’ method, and the width of all histograms increases with increasing infectiousness (which is set inversely proportional to ρ). When $\mathcal{T}_d$ is calculated using the ‘14 days since 100 cases’ method, its median value is slightly larger than $\mathcal{T}_d$ in the deterministic model (equal to 2 days); however, when $\mathcal{T}_d$ is calculated using the ‘30 days since the first case’ method, then for high infectiousness of super- and hyper-spreaders (correspondingly, for low ρ) its median value becomes much larger than the deterministic $\mathcal{T}_d$. Using the ‘30 days since the first case’ method for the case of the lowest considered $\rho = 1\%$, when super-spreaders (hyper-spreaders) have their infectiousness about 100 times (200 times) higher than the infectiousness of normal individuals, one obtains median $\mathcal{T}_d$ larger than $\mathcal{T}_d$ obtained in the deterministic model by 29% (67%), while for ‘14 days since 100 cases’ the $\mathcal{T}_d$ overestimation is negligible, 3% (6%). This difference is caused by low probability of appearance of super- or hyper-spreaders in the first weeks of the outbreak.

We note that $\mathcal{T}_d$ estimation for a given country based on available data is equivalent to the analysis of a single stochastic trajectory and that at a very initial stage the epidemic can cease. Probability of extinction is larger when a small fraction of super-spreaders is responsible for a large fraction of cases. In figure 3, we provide extinction probability, $\hat{p}_{\text{ext}}$, which in the extreme case of 1% of hyper-spreaders reaches 29%, whereas without hyper-spreaders (or super-spreaders) is 11%.

The examples shown in figures 2 and 3 are focused on the case in which the $\mathcal{T}_d = 2$ days, which is close to $\mathcal{T}_d$ estimated for Spain and New York State. After removing super-spreaders (assumed to be responsible for 50% of transmissions) the doubling time would be equal to 3.05 days, whereas after removing hyper-spreaders (responsible for 66.7% of transmissions) the doubling time would be equal to 4.24 days. The doubling times in the range 5.2–7.4, obtained by analysing early onsets of the epidemic (Wu *et al.* [10,15] and Li *et al.* [16]), exceed our model prediction obtained after removing 66.7% of transmissions by hyper-spreaders, suggesting that the fraction of transmissions for which hyper-spreaders are responsible can be even larger. Endo *et al.* estimated that 80% of secondary transmissions could have been caused by 10% of infectious individuals [19].

Finally, we compare $\mathcal{T}_d$ estimates obtained for eight considered locations using the ‘14 days since 100 cases’ method (as in figure 1) or the ‘30 days since the first case’ method. As expected, $\mathcal{T}_d$ estimates using ‘30 days since the first case’ method in most cases are larger and more dispersed (in range 1.92–12.6) than

the estimates based on the primary method (1.86–2.88). Results shown in table 2 clearly indicate that the ‘14 days since 100 cases’ method is more reliable. Its disadvantage lies in the fact that for a given location the $\mathcal{T}_d$ estimate is possible when the epidemic is fully developed (see third and fourth columns of table 2). It is, however, important to note that for China the ‘14 days since 100 cases’ method estimate (chosen in our study and giving $\mathcal{R}_0(\mathcal{T}_d)$ in range 5.6–7.3) was possible on 4 February 2020, before the surge of the epidemic in Europe and USA, and more than one month ahead of the first European country-wide lockdown that was imposed in Italy (9 March 2020).

3. Conclusion

Based on epidemic data from China, New York State and six European countries, we have estimated that the basic reproduction number $\mathcal{R}_0$ lies in the range 4.7–11.4 (5.6–7.3 for China), which is higher than most previous estimates [4,5,8]. There are two sources of the discrepancy in $\mathcal{R}_0$ estimation. First, in agreement with data on the incubation period distribution (assumed to be the same as the latent period distribution), we used a model with six ‘exposed’ states, which substantially increases $\mathcal{R}_0$ ($\mathcal{T}_d$) with respect to the models with one or two ‘exposed’ states. Second, we estimated $\mathcal{T}_d$ based on the two-week period of the exponential growth phase beginning on the day in which the number of cumulative registered cases exceeds 100, or when the number of cumulative registered fatalities exceeds 10. Importantly, values of $\mathcal{T}_d$ estimated from the growth of registered cases and from the growth of the registered fatalities led to similar $\mathcal{R}_0$ estimates. This approach, in contrast to estimation of $\mathcal{R}_0$ based on individual case reports, allows to implicitly take into account super-spreading events that substantially shorten $\mathcal{T}_d$. Spatial heterogeneity of the epidemic spread observed in many European countries, including Italy, Spain and Germany, can be associated with larger or smaller super-spreading events that initiated outbreaks in particular regions of these countries. Lack of, or sporadic super-spreading events in first phase of epidemic explains why $\mathcal{T}_d$ estimated using the ‘30 days since the first case’ method gives in most cases larger and more dispersed values than our method of choice: ‘14 days since 100 cases’. This, in turn, suggests that in general the reproduction number calculated based on early epidemic development can be probably underestimated, and thus in the case of future epidemics must be considered with caution.

Our estimates are consistent with current epidemic data in Italy, Spain and France. As of 24 April 2020, these countries managed to terminate the exponential growth phase by means of country-wide quarantine. Current COVID-19 Community Mobility Reports [25] show about 80% reduction of mobility in retail and recreation, transit stations and workplaces in these countries. Together with increased social distancing, this reduction possibly lowered the infection rate β at least fivefold; additionally, massive testing reduced the infectious period, $1/\gamma$. Consequently, we suspect that the reproduction number $R = \beta/\gamma$ was reduced more than fivefold, which brought it to the values somewhat smaller than 1. This suggests that $\mathcal{R}_0$ in these countries could have been larger than 5.

4. Methods

4.1. SEIR model equations and parametrization

The dynamics of our SEIR model is governed by the following system of ordinary differential equations:

$$\frac{dS}{dt} = -\beta I(t) S(t)/N \quad (4.1)$$

$$\frac{dE_1}{dt} = \beta I(t) S(t)/N - m \sigma E_1(t), \quad (4.2)$$

$$\frac{dE_i}{dt} = m \sigma E_{i-1}(t) - m \sigma E_i(t), \quad 2 \leq i \leq m, \quad (4.3)$$

$$\frac{dI_1}{dt} = m \sigma E_m(t) - n \gamma I_1(t), \quad (4.4)$$

$$\frac{dI_j}{dt} = n \gamma I_{j-1}(t) - n \gamma I_j(t), \quad 2 \leq j \leq n, \quad (4.5)$$

$$\frac{dR}{dt} = n \gamma I_n(t), \quad (4.6)$$

where $N = S(t) + E_1(t) + \dots + E_m(t) + I_1(t) + \dots + I_n(t) + R(t)$ is the constant population size, and $I(t) = I_1(t) + \dots + I_n(t)$ is the size of infectious subpopulation. As m is the number of ‘exposed’ substates and n is the

number of ‘infectious’ substates, there are $m + n + 2$ equations in the system. In the early phase of the epidemic, $1 - S(t)/N \ll 1$ and with constant coefficients β , σ and γ the growth of R (as well as E_i and I_j) is exponential.

An important property of a given SEIR model parametrization is its implied distribution of generation interval (GI), the period between subsequent infection events in a transmission chain. While the expected GI is easily computable from model parameters as $\langle \text{GI} \rangle = \sigma^{-1} + \frac{1}{2} \gamma^{-1}$ (the mean period of infectiousness is halved to reflect the assumption that the infection occurs in a random time during the period of infectiousness [17]), it can be hardly estimated based on even detailed epidemiological data. It should be noted that in some sources the formula $\langle \text{GI} \rangle = \sigma^{-1} + \gamma^{-1}$ is used (e.g. [26]), in which it is assumed that the infection occurs at the end of the period of infectiousness, not at a random point of this period. GI may be related to the serial interval, SI, the period between the occurrence of symptoms in the infector and the infectee. Although GI and SI may have different distributions, their means are expected to be equal and thus may be directly compared. Our parametrization implies $\langle \text{GI} \rangle = 6.73$ days, which is consistent with $\langle \text{GI} \rangle$ of the model by Ferguson *et al.* (6.5 days) [27] and the estimates of $\langle \text{SI} \rangle$ by Wu *et al.* (7.0 days) [15], Ma *et al.* (6.8 days) [28] or Bi *et al.* (6.3 days) [29].

The short period of effective infectiousness reflects the assumption that the individuals with confirmed infection are quickly isolated or self-isolated and then cannot infect susceptible individuals. This enabled us to identify the reported increase of confirmed cases with the transfer of the individuals from the (last substate of the) ‘infectious’ compartment to the ‘removed’ compartment of the SEIR model. In addition to the currently diseased individuals that remain isolated, the ‘removed’ compartment contains the recovered (and assumed to be resistant) and the deceased individuals.

We assume the same $1/\gamma = 2.9$ days in all locations and times, being, however, aware that the mean infectious period may shorten over time due to the implementation of protective health-care practices, increased diagnostic capacity, and contact tracing [29]. In turn, the mean latent period, $1/\sigma$, may be considered an intrinsic property of the disease. As the distribution of the latent period is not known, as a simplification, in our model, the distribution of the latent period (time since infection during which an infected individual cannot infect) is assumed to be the same as the distribution of the incubation period (time since infection during which an infected individual has not yet developed symptoms). We demonstrated the influence of $1/\sigma$ on the estimation of $\mathcal{R}_0$ in electronic supplementary material, figure S1.

4.2. Estimation of the doubling time and the basic reproduction number

Growth rates used for estimation of respective doubling times, $\mathcal{T}_d$, were determined by linear regression of the logarithm of the cumulative confirmed cases and cumulative deaths in the exponential phase of the epidemic separately in each of eight considered location. We discarded initial parts of trajectories with less than 100 confirmed cases (or 10 registered fatalities) and used two-week-long periods to strike a balance between: (i) analysis of epidemic progression when stochastic effects associated with individual transmission events, including super-spreading, are relatively small (see stochastic simulation trajectories in electronic supplementary material, figure 2a), and (ii) analysis of the exponential phase of epidemic progression, which is relatively short due to imposition of restrictions. We expect that the trajectories of deaths may be less affected by under-reporting; nevertheless, doubling times obtained from growth rates of cumulative cases and cumulative deaths turn out to be quite consistent.

In the context of our SEIR model, the doubling time $\mathcal{T}_d$ and parameters β , σ , γ , n , m satisfy the relation

$$\beta(\mathcal{T}_d; \sigma, \gamma, m, n) = \frac{\frac{\log 2}{\mathcal{T}_d} \left(\frac{\log 2}{\mathcal{T}_d m \sigma} + 1 \right)^m}{1 - \left(\frac{\log 2}{\mathcal{T}_d n \gamma} + 1 \right)^{-n}} \quad (4.7)$$

that enables calculation of the basic reproduction number using the doubling time $\mathcal{T}_d$ estimated directly from the epidemic data as

$$\mathcal{R}_0(\mathcal{T}_d) = \frac{\beta(\mathcal{T}_d; \sigma, \gamma, m, n)}{\gamma}, \quad (4.8)$$

in accordance with Wearing *et al.* [12] and Wallinga & Lipsitch [13].

Data accessibility. All data used in this theoretical study are referenced.

Authors' contributions. M.K. conceived study, performed model and data analysis, prepared figures and wrote manuscript; F.G. conceived study, performed model analysis and prepared figures; T.L. conceived study and wrote manuscript. All authors gave final approval for publication.

Competing interests. We declare we have no competing interest.

Funding. This study was supported by the National Science Centre (Poland) grant no. 2018/29/B/NZ2/00668.

Acknowledgements. We thank the reviewers whose comments helped us to improve the manuscript.

References

1. Delamater P, Street E, Leslie T, Yang YT, Jacobsen K. 2019 Complexity of the basic reproduction number (R_0). *Emerg. Infect. Dis.* **25**, 1. (doi:10.3201/eid2501.171901)
2. WHO. 2020 Statement on the meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus 2019 (n-CoV) on 23 January 2020. [https://www.who.int/news-room/detail/23-01-2020-statement-on-the-meeting-of-the-international-health-regulations-\(2005\)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-\(2019-ncov\)](https://www.who.int/news-room/detail/23-01-2020-statement-on-the-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov)) (accessed 26 June 2020).
3. WHO. 2020 Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). [https://www.who.int/publications-detail/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-\(covid-19\)](https://www.who.int/publications-detail/report-of-the-who-china-joint-mission-on-coronavirus-disease-2019-(covid-19)) (accessed 26 June 2020).
4. Liu Y, Gayle AA, Wilder-Smith A, Rocklöv J. 2020 The reproductive number of COVID-19 is higher compared to SARS coronavirus. *J. Travel Med.* **27**, taaa021. (doi:10.1093/jtm/taaa021)
5. Boldog P, Tekeli T, Vizi Z, Dénes A, Bartha FA, Röst G. 2020 Risk assessment of novel coronavirus COVID-19 outbreaks outside China. *J. Clin. Med.* **9**, 571. (doi:10.3390/jcm9020571)
6. Dong E, Du H, Gardner L. 2020 An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* **20**, 533–534. (doi:10.1016/S1473-3099(20)30120-1). For GitHub repository see <https://github.com/CSSEGISandData/COVID-19>.
7. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith H, Azman AS, Reich NG, Lessler J. 2020 The incubation period of 2019-nCoV from publicly reported confirmed cases: estimation and application. *medRxiv*. (doi:10.1101/2020.02.02.20020016)
8. Liu T *et al.* 2020 Time-varying transmission dynamics of novel coronavirus pneumonia in China. *bioRxiv*. (doi: 10.1101/2020.01.25.919787)
9. Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, Funk S, Eggo RM. 2020 Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *Lancet Infect. Dis.* **20**, 553–558. (doi:10.1016/S1473-3099(20)30144-4)
10. Wu JT, Leung K, Leung GM. 2020 Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet* **395**, 689–697. (doi:10.1016/S0140-6736(20)30260-9)
11. Sanche S, Lin YT, Xu C, Romero-Severson E, Hengartner N, Ke R. 2020 High contagiousness and rapid spread of severe acute respiratory syndrome coronavirus 2. *Emerg. Infect. Dis.* **26**, 1470–1477. (doi:10.3201/eid2607.200282)
12. Wearing HJ, Rohani P, Keeling MJ. 2005 Appropriate models for the management of infectious diseases. *PLoS Med.* **2**, 0020174. (doi:10.1371/journal.pmed.0020174)
13. Wallinga J, Lipsitch M. 2007 How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc. R. Soc. B* **274**, 599–604. (doi:10.1098/rspb.2006.3754)
14. Kocharczyk M, Grabowski F, Lipniacki T. 2020 Dynamics of COVID-19 pandemic at constant and time-dependent contact rates. *Math. Model. Nat. Phenom.* **15**, 28. (doi:10.1051/mmnp/2020011)
15. Backer JA, Klinkenberg D, Wallinga J. 2020 Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20–28 January 2020. *EuroSurveill* **25**, 2000062. (doi:10.2807/1560-7917.ES.2020.25.5.2000062)
16. Li Q *et al.* 2020 Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *N. Engl. J. Med.* **382**, 1199–1207. (doi:10.1056/NEJMoa2001316)
17. Wu JT, Leung K, Bushman M, Kishore N, Niehus R, de Salazar PM, Cowling BJ, Lipsitch M, Leung GM. 2020 Estimating clinical severity of COVID-19 from the transmission dynamics in Wuhan, China. *Nat. Med.* **26**, 506–510. (doi:10.1038/s41591-020-0822-7)
18. Nelson KE, Williams K. 2014 *Infectious disease epidemiology: theory and practice*, 3rd edn. Burlington, MA: Jones & Bartlett Learning.
19. Endo A, Abbott S, Kucharski A, Funk S. 2020 Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China [version 3; peer review: 2 approved]. *Wellcome Open Res.* **5**, 67. (doi:10.12688/wellcomeopenres.15842.3)
20. Worobey M *et al.* 2020 The emergence of SARS-CoV-2 in Europe and the US. *bioRxiv*. (doi:10.1101/2020.05.21.109322)
21. Cereda D *et al.* 2020 The early phase of the COVID-19 outbreak in Lombardy, Italy. *arXiv*. (<http://arxiv.org/abs/2003.09320>)
22. Mercker M, Betzin U, Wilken D. 2020 What influences COVID-19 infection rates: a statistical approach to identify promising factors applied to infection data from Germany. *medRxiv*. (doi:10.1101/2020.04.14.20064501)
23. Harris LA *et al.* 2016 BioNetGen 2.2: advances in rule-based modeling. *Bioinformatics* **32**, 3366–3368. (doi:10.1093/bioinformatics/btw469)
24. Huang C *et al.* 2020 Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395**, 497–506. (doi:10.1016/S0140-6736(20)30183-5)
25. Google LLC. 2020 Google COVID-19 Community Mobility Reports. <https://www.google.com/covid19/mobility> (accessed 26 June 2020).
26. Lipsitch M *et al.* 2003 Transmission dynamics and control of severe acute respiratory syndrome. *Science* **300**, 1966–1970. (doi:10.1126/science.1086616)
27. Ferguson N, Laydon D, Nedjati Gilani G, Imai N, Ainslie K, Baguelin M. 2020 Report 9. <https://spiral.imperial.ac.uk:8443/handle/10044/1/77482> (accessed 26 March 2020).
28. Ma S, Zhang J, Zeng M, Yun Q, Guo W, Zheng Y, Zhao S, Wang MH, Yang Z. 2020 Epidemiological parameters of coronavirus disease 2019: a pooled analysis of publicly reported individual data of 1155 cases from seven countries. *medRxiv*. (doi:10.1101/2020.03.21.20040329)
29. Bi Q *et al.* 2020 Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. *Lancet Infect. Dis.* **20**, 911–919. (doi:10.1016/S1473-3099(20)30287-5)

Article

SARS-CoV-2 Variant of Concern 202012/01 Has about Twofold Replicative Advantage and Acquires Concerning Mutations

Frederic Grabowski ^{1,†} , Grzegorz Preibisch ², Stanisław Giziński ³ , Marek Kochańczyk ^{1,*}  and Tomasz Lipniacki ^{1,*} 

¹ Institute of Fundamental Technological Research, Polish Academy of Sciences, 02-106 Warsaw, Poland; frederic.grabowski@ippt.pan.pl

² Inter-Faculty Individual Studies in Mathematics and Natural Sciences, The MISMaP College, University of Warsaw, 02-097 Warsaw, Poland; grzegorz.preibisch@gmail.com

³ Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, 02-097 Warsaw, Poland; sg385513@students.mimuw.edu.pl

* Correspondence: mkochan@ippt.pan.pl (M.K.); tlipnia@ippt.pan.pl (T.L.)

† These authors contributed equally to this work.

Abstract: The novel SARS-CoV-2 Variant of Concern (VOC)-202012/01 (also known as B.1.1.7), first collected in United Kingdom on 20 September 2020, is a rapidly growing lineage that in January 2021 constituted 86% of all SARS-CoV-2 genomes sequenced in England. The VOC has been detected in 40 out of 46 countries that reported at least 50 genomes in January 2021. We have estimated that the replicative advantage of the VOC is in the range 1.83–2.18 [95% CI: 1.71–2.40] with respect to the 20A.EU1 variant that dominated in England in November 2020, and in range 1.65–1.72 [95% CI: 1.46–2.04] in Wales, Scotland, Denmark, and USA. As the VOC strain will likely spread globally towards fixation, it is important to monitor its molecular evolution. We have estimated growth rates of expanding mutations acquired by the VOC lineage to find that the L18F substitution in spike has initiated a fast growing VOC substrain. The L18F substitution is of significance because it has been found to compromise binding of neutralizing antibodies. Of concern are immune escape mutations acquired by the VOC: E484K, F490S, S494P (in the receptor binding motif of spike) and Q677H, Q675H (in the proximity of the polybasic cleavage site at the S1/S2 boundary). These mutants may hinder efficiency of existing vaccines and expand in response to the increasing after-infection or vaccine-induced seroprevalence.

Keywords: COVID-19 pandemic; SARS-CoV-2; spike protein; VOC-202012/01; spike L18F; genome sequencing; mutation



Citation: Grabowski, F.; Preibisch, G.; Giziński, S.; Kochańczyk, M.; Lipniacki, T. SARS-CoV-2 Variant of Concern 202012/01 Has about Twofold Replicative Advantage and Acquires Concerning Mutations. *Viruses* **2021**, *13*, 392. <https://doi.org/10.3390/v13030392>

Academic Editor: Allan Brasier

Received: 13 January 2021

Accepted: 24 February 2021

Published: 1 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The earliest genome belonging to the novel SARS-CoV-2 Variant of Concern (VOC)-202012/01, also known as B.1.1.7 lineage, was collected on 20 September 2020, in Kent, UK (GISAID sequence accession ID: EPI_ISL_601443). The lineage, characterized by nine spike protein mutations (deletions: 69–70 HV, 145V; substitutions: N501Y, A570D, D614G, P681H, T716I, S982A, D1118H), started to spread rapidly in mid-October 2020 to constitute in January 2021 86% of all SARS-CoV-2 genomes sequenced in England [1,2]. Spread of the VOC-202012/01 variant, hereafter referred to as the VOC, co-occurred with a rapid surge of cases in December in Kent and Greater London [3,4].

In England, the VOC is currently replacing the recently dominant 20A.EU1 strain, characterized by A222V substitution in spike protein [5]. Both strains are independent substrains of a spike glycoprotein D614G variant that has spread in spring 2020 in England and worldwide, almost reaching fixation [6]. The 20A.EU1 strain started expanding in England in mid-August 2020 and constituted more than 65% of genomes sequenced in England in November 2020 [2]. The VOC strain in the receptor-binding motif (RBM) of

spike shares mutation N501Y with the 501Y.V2 and P.1 strains that are currently rapidly spreading in South Africa and Brazil, respectively [7,8].

Deletion 69H–70V in spike glycoprotein, which is characteristic, but not unique, to the VOC, prevents detection of the spike gene by the dPCR probe used by some laboratories of the English diagnostic system (spike gene target failure, SGTF) [9]. As the VOC has a high multiplicative potential, it has become the most prevalent $\Delta 69-70$ HV strain; consequently, the proportion of the SGTF has been used as a proxy for the prevalence of the VOC genome [9]. Based on the SGTF, the Public Health England agency determined the multiplicative advantage of the VOC in NHS STP areas of England in weeks 44–48 of 2020. As an average for the considered STP areas the authors obtained the ratio of reproduction numbers equal 1.47 [95% CI: 1.34–1.59] [9]. Leung et al., based on GISAID data from the period 22 September–16 November 2020, and a competition transmission model of two viruses, estimated this ratio as 1.75 [95% CrI: 1.70–1.80] [10]. Davies et al., based on data from the COVID-19 Genomics UK (COG-UK) Consortium from October and November 2020 estimated that the VOC is 43–82% [95% CrI: 38–106%] more transmissible than preexisting variants of SARS-CoV-2 [11,12]. Our early estimate of the ratio of reproduction numbers of the VOC to non-VOC strains, based on GISAID data in weeks 43–47 of 2020 in England was 2.24 [95% CI: 2.03–2.48] [13].

In this study, based on GISAID data available on 12 February 2021 (when this manuscript has been revised), we calculated the growth of the ratio of the VOC to the 20A.EU1 genome sequences collected in the period between week 43 and week 51 of 2020 (19 October–20 December 2020) and estimated the replicative advantage of the VOC strain in relation to the 20A.EU1 strain. Then, using the same approach, we estimated the replicative advantage of the 20A.EU1 strain in relation to previous D614G strains and the D614G strain in relation to the D614 strains (i.e., the strains with non-mutated residue 614). In such approach, we analyze the progression of strains with increasing replication advantage: D614G $\rightarrow$ 20A.EU1 $\rightarrow$ the VOC. In contrast to the approach of Public Health England (PHE) agency [9] and Davies et al. [11], we used England-aggregated weekly data. We prefer to use such aggregated data because of the geographically non-uniform emergence of VOC substrains that, as we will show, have diverse replication advantages. Some substrains that emerge locally may have low or no replicative advantage and in the long term may be out-competed and eventually eliminated; however, such transient substrains can considerably contribute to the genome composition in specific NHS STP areas of England.

The VOC strain, as all strains, mutates continuously. Because of its significant replicative advantage, any accrued mutations gain an opportunity to spread (potentially worldwide), depending on their replicative advantage with respect to the bulk VOC strain, or higher ability to infect seroprevalent individuals. We have systematically estimated growth rates of spreading mutations acquired by the VOC lineage to find that spike L18F substitution has likely initiated a substrain of replicative advantage. The L18F mutation is of significance because when recently analyzed in the context of the South African strain 501Y.V2 it has been found to compromise binding of neutralizing antibodies [14,15].

2. Materials and Methods

2.1. Data Sources

In Figures 1–7 we use weekly aggregated data of samples collected in England or South Africa submitted to GISAID until 12 February 2021, and samples collected in Wales, Scotland, Denmark, and the USA and submitted until 17 February 2021. Table 1 is based on GISAID data as of 16 February 2021. In Supplementary Figure S1, we use COG-UK pillar 2 data as of 15 February 2021. All data used for Figures 1–7 and supplementary Figure S1 are collected in a single file, Supplementary Data File S1.

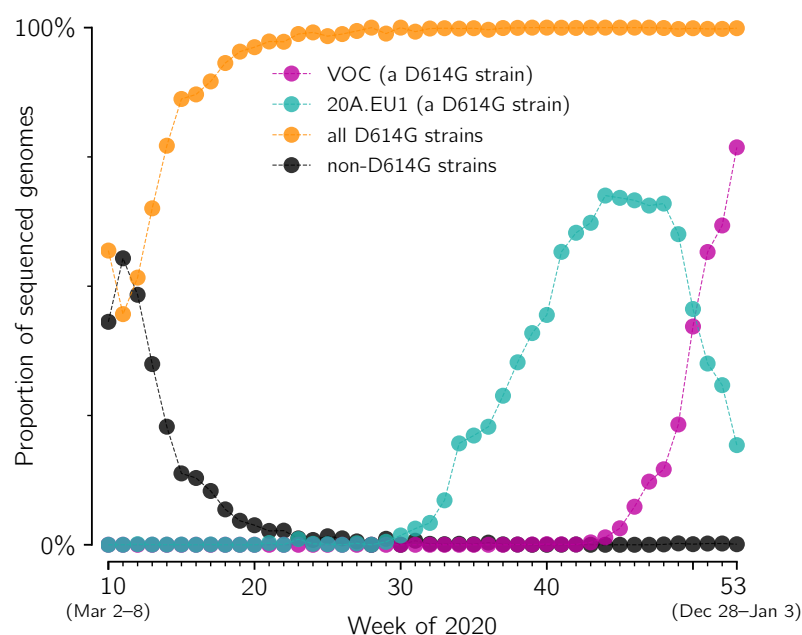


Figure 1. Fixation of D614G strain and competition between its substrains: 20A.EU1 and the VOC in England. The proportion of D614G (orange), 20A.EU1 (cyan), VOC (magenta) and D614 strains (black) in weeks 10–53 of 2020 has been determined based on GISAID data available on 12 February 2021 (provided in Supplementary Data File S1).

2.2. Genome Sequence Analysis

We analyzed mutations in spike gene sequences that had at most 5% letters other than A, C, G, T. Sequences devoid of full daily collection date were excluded from the analysis. The spike gene was localized using EMBOSS stretcher and then re-aligned using EMBOSS needle [16] to the reference sequence: GenBank’s NC_045512.2 from Wuhan or GISAID’s EPI_ISL_601443 from Kent. To minimize ambiguous reporting of mutations variants, indels were left-aligned using an in-house script; insertions and deletions on consecutive residues were collated and considered a single mutation.

2.3. Monte Carlo Estimation of the Confidence Interval for the Growth Rate of the Ratio of Strains

To estimate the 95% credible interval for k subsequent weeks *a priori* we assume that genomes of two compared strains in each week of the considered period follow a respective binomial distribution having a success probability $p = n_x / (n_x + n_y)$, where n_x and n_y are the numbers of two compared strain genomes (e.g., VOC and 20A.EU1). By sampling from k such binomial distributions for the considered time window of k weeks 10^5 times, we obtained 10^5 series of k simulated sequenced genome proportions. We performed fitting to each such series to obtain 10^5 estimates of the weekly growth rate of the ratio of the genome sequences. In all cases but one for L18F analysis (Figure 7) the 95% credible interval obtained using the *a priori* method has been narrower than the 95% confidence interval calculated as $1.96 \times$ (standard error of the slope). In these cases we reported confidence interval, while in L18F case we reported the credible interval.

2.4. Estimation of the Relative Replication Advantage of Viral Strains

To estimate the ratios of replication numbers of strain x and y we estimated the weekly growth of the ratio of the number of their sequenced genomes, v . Then, assuming that both strains have the same average serial interval of 6.73 days [17], we obtained the ratio of their replication numbers $\mathcal{R}_t^x / \mathcal{R}_t^y := v^{6.73/7}$. This estimation of the relative replicative advantage of viral strains does not involve a direct calculation of their reproduction numbers.

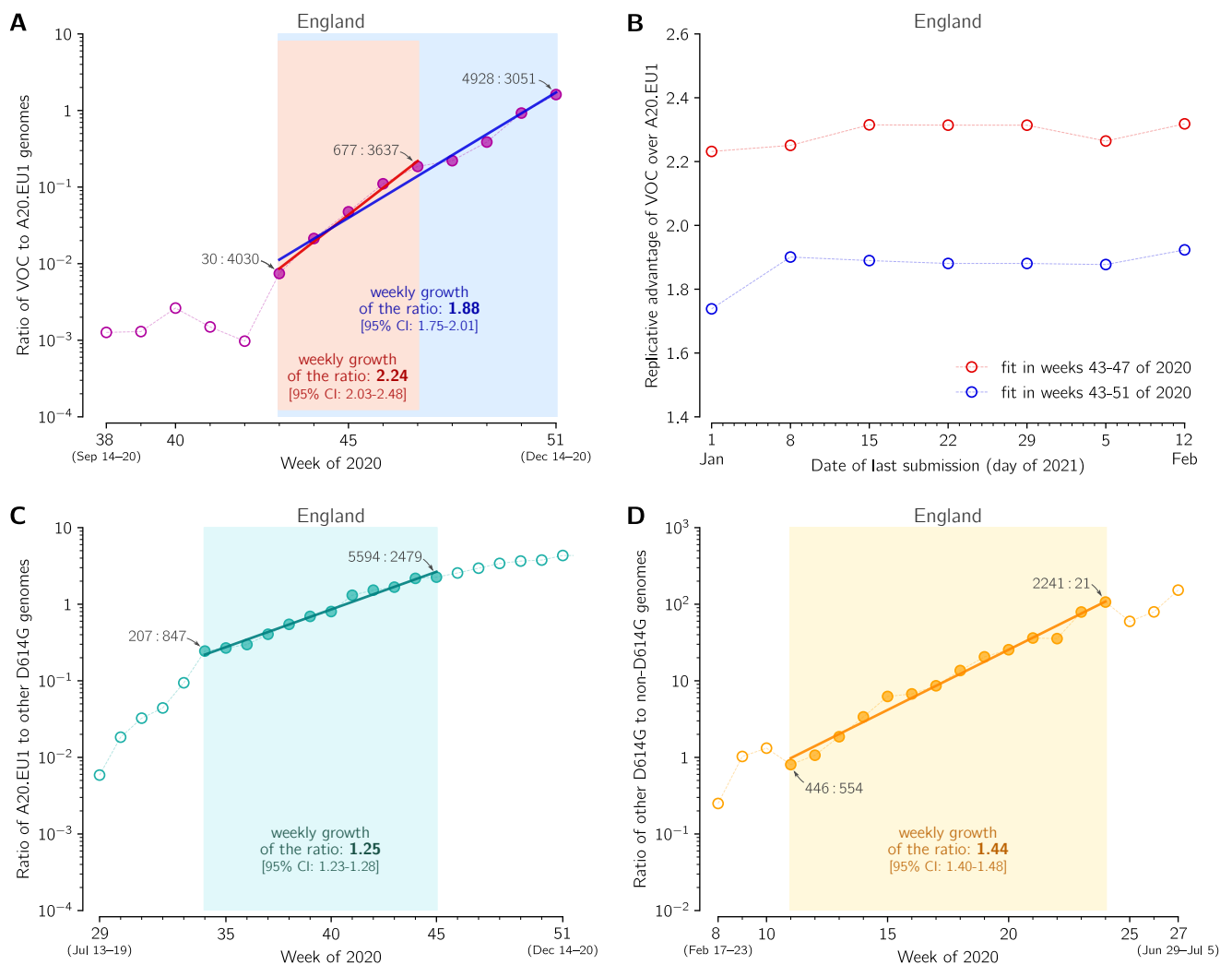


Figure 2. The replicative advantage of the VOC, 20A.EU1, and the D614G strain. **(A)** The ratio of the VOC to 20A.EU1 sequences collected in weeks 38–51 of 2020 in England. The trend line is fitted to data points from weeks 43–51 (blue) and from weeks 43–47 (red). The weekly growth rate is 1.88 [95% CI: 1.75–2.01] for weeks 43–51 and 2.24 [95% CI: 2.03–2.48] for weeks 43–47. **(B)** Stabilization of fits shown in panel A with time-shifting date of the last submission. **(C)** The ratio of the 20A.EU1 to non-20A.EU1 non-VOC D614G genomes collected in weeks 29–51 of 2020 in England. The trend line is fitted to data points from weeks 34–45. The weekly growth rate is 1.25 [95% CI: 1.23–1.28]. **(D)** The ratio of the D614G to D614 genomes collected in weeks 11–27 of 2020 in England. The trend line is fitted to data points from weeks 11–24. The weekly growth rate is 1.44 [95% CI: 1.40–1.48]. Data points for selected weeks are labeled with ratios of the counts of genome sequences. Panels A, C and D are based on GISAID data submitted till 12 February 2021 (provided in Supplementary Data File S1).

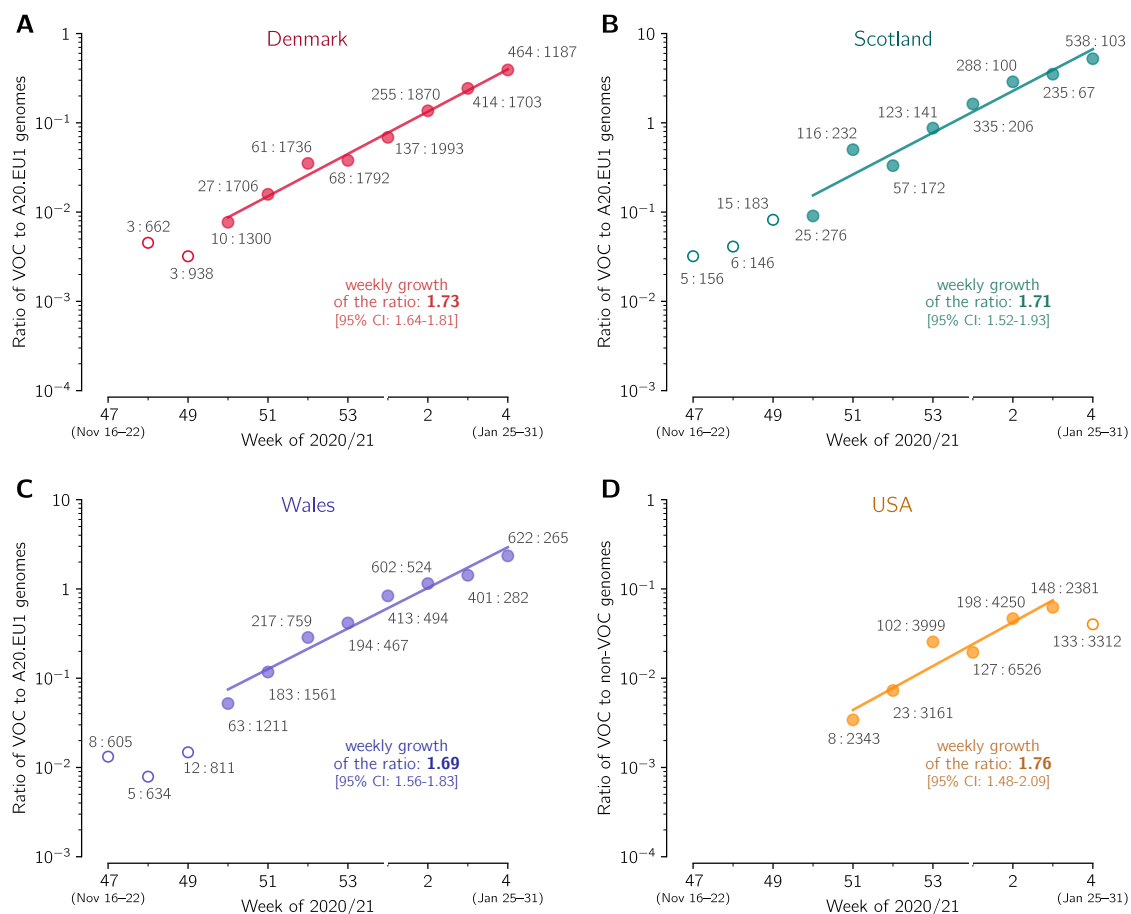


Figure 3. The replicative advantage of the VOC in (A) Denmark, (B) Scotland, (C) Wales, and (D) USA. (A) The ratio of the VOC to the 20A.EU1 sequences collected in Denmark. The trend line is fitted to data points from week 50 of 2020 to week 4 of 2021. (B) The ratio of the VOC to 20A.EU1 sequences collected in Scotland. The trend line is fitted to data points from weeks 50 of 2020 to week 4 of 2021. (C) The ratio of the VOC to 20A.EU1 sequences collected in Wales. The trend line is fitted to data points from week 50 of 2020 to week 4 of 2021. (D) The ratio of the VOC to non-VOC sequences collected in USA. The trend line is fitted to data points from week 51 of 2020 to week 3 of 2021. Data points are labeled with ratios of the counts of genome sequences. The weekly growth rates and respective confidence intervals are given in each panel. All panels are based on GISAID data submitted till 17 February 2021 (provided in Supplementary Data File S1).

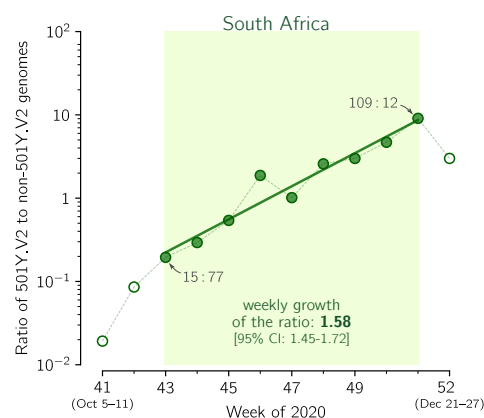


Figure 4. The replicative advantage of the 501Y.V2 strain in South Africa. Shown is the ratio of the 501Y.V2 to the non-501Y.V2 genomes collected in weeks 41–50 of 2020 in South Africa. The trend line is fitted to data points from weeks 43–50. The weekly growth rate is 1.58 [95% CI: 1.45–1.72]. Data points for selected weeks are labeled with ratios of the counts of genome sequences. The figure is based on GISAID data submitted till 12 February 2021 (provided in Supplementary Data File S1).

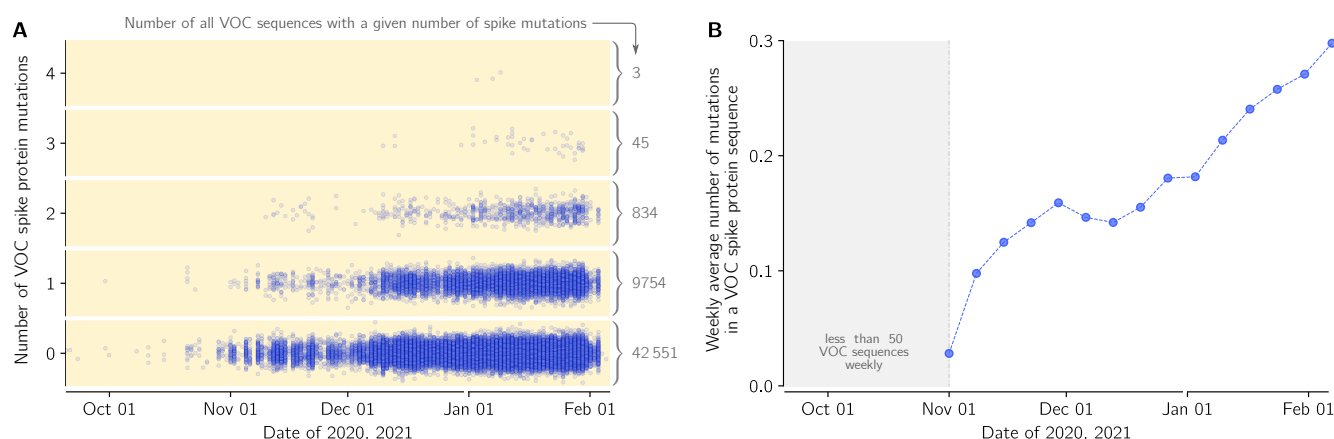


Figure 5. Mutations in spike protein in the VOC lineage in England in relation to the first VOC genome collected in September 20, 2020 (GISAID sequence accession ID: EPI_ISL_601443): (A) Dots denote genomes collected in a given date (horizontal axis) with a respective number of novel (amino acid-level) mutations in relation to the first VOC genome (vertical axis). The number of the genomes with a given number of novel mutations is provided as gray numbers next to a brace. (B) The average number of novel (amino-acid level) mutations in a VOC lineage sequence (weekly average).

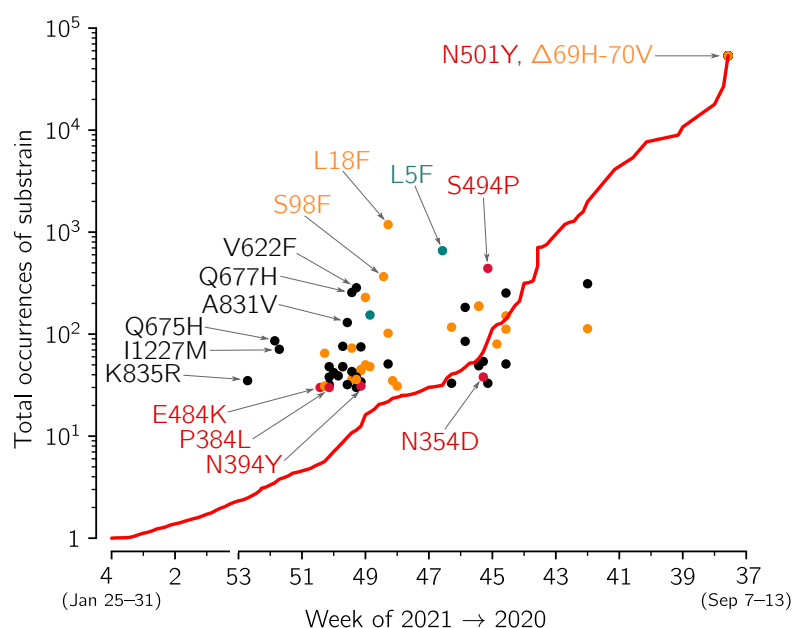


Figure 6. The growth of VOC substrains with respect to the whole VOC strain. The figure is based on spike protein sequences from genomes collected in England and submitted to GISAID till 12 February 2021. The growth of the whole VOC strain (red line) in the time span between t and $t_0 = 12$ February 2021 is calculated as $VOC(t_0)/VOC(t)$, i.e., the ratio of the number of the VOC genomes collected till 12 February 2021 and the number of VOC genomes collected till t (date given on horizontal axis). Disks denote VOC substrains defined by particular mutations (with at least 30 genomes submitted); horizontal axis value for each circle is the date of first collection of a sequence with a given mutation; vertical axis value is the number of genomes collected till 12 February 2021, conferring a given mutation. Red, orange, and green circles denote substrains with leading mutation in respectively in RBD, NTD, and signal peptide; black circles denote mutations in other domains. Selected substrains that may have replicative advantage and all substrains with the leading mutation in RBD are annotated. Detailed data that were analyzed to prepare this figure are provided in Supplementary Data File S1 (sheet ‘Mutations’).

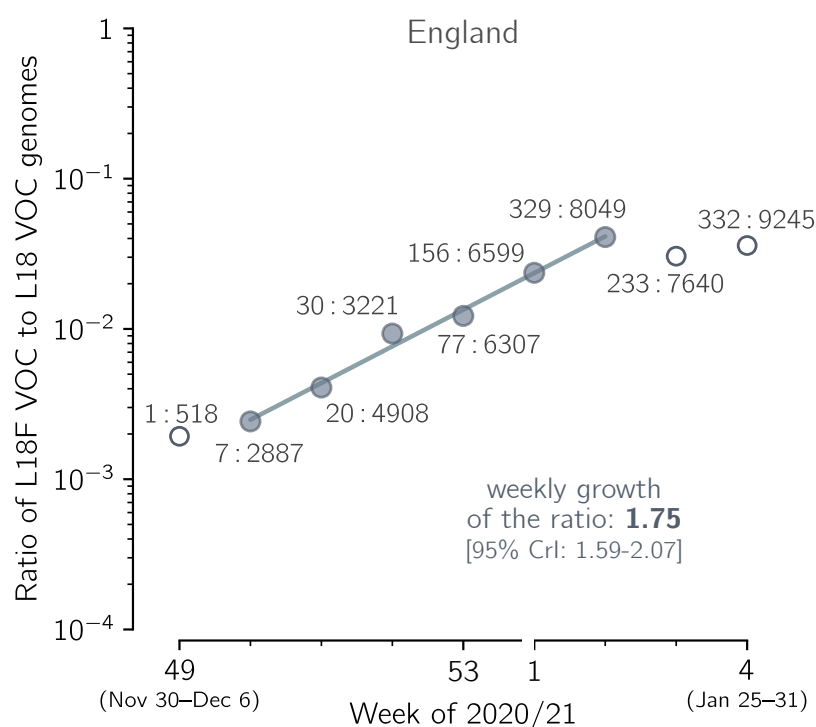


Figure 7. The growth of the spike L18F substrain in relation to the L18 VOC strains. The ratio v_L of the number of VOC genomes conferring spike L18F mutation to the number of non-mutated (L18) VOC genomes collected in the period between week 49 of 2020 and week 3 of 2021 in England. Data aggregated into weeks indicated that v_L changes from 7:2887 = 0.0024 in week 50 of 2020 to 329:8049 = 0.041 in the second week of 2021, i.e., 16.9 times, meaning that v_L increases $16.9^{1/5} = 1.76$ -fold per week. The trend line is fitted to data points shown as filled circles. The weekly growth rate of the ratio is 1.75. The 95% CI calculated as $1.96 \times$ standard error of the slope is [1.66–1.85]; the 95% CrI calculated assuming binomial distribution of substrain genomes is [1.59–2.07]. Data points in week 50 of 2020 and week 2 of 2021 are labeled with ratios of the counts of both types of genome sequences. The figure is based on GISAID data submitted until 12 February 2021 (provided in Supplementary Data File S1).

3. Results

3.1. Evolution of SARS-CoV-2 in England

The first prevailing mutation of SARS-CoV-2 was D614G substitution in spike protein (the first GISAID reported genome, EPI_ISL_913915, was collected on 2 January 2020, in Mexico). This substitution initiated a strain that spread worldwide nearly reaching global fixation [6]. In England, the D614G strain appeared during the spring wave of epidemic in 2020, and in summer it exceeded 98% of all sequenced genomes (Figure 1). Its substrain, 20A.EU1, started expanding in England in week 31 of 2020 and reached its maximum of 68% of all sequenced genomes in week 44 (Figure 1). The VOC strain is also a substrain of the D614G lineage, independent of 20A.EU1, which started expanding in week 43 (before week 43, less than 5 VOC genomes were collected per week), and in week 51 reached 57% of all sequenced genomes. At the same time the proportion of the 20A.EU1 strain dropped to 35% and the proportion of all other genomes dropped to 8%. One can observe that in weeks 44–51 the decrease of the proportion of 20A.EU1 is about twofold (from 68% to 35%), while the decrease of the proportion of remaining strains (that is, strains other than considered 20A.EU1 and VOC) is nearly fourfold (from 31% to 8%). This motivates us to calculate the replicative advantage of the VOC strain in relation to the 20A.EU1 strain instead of all non-VOC strains (the replicative advantage of the VOC over non-VOC strains would depend on the proportion of the 20A.EU1 strain in all non-VOC genomes).

3.2. Replicative Advantage of the VOC Strain over the 20A.EU1 Strain in England

In the eight-week period of weeks 43–51, the ratio q of the VOC to the 20A.EU1 strain increased from $q = 30/4030 \approx 0.0074$ to $q = 4928/3051 \approx 1.62$, that is, 217 times. This implies that q increased $217^{1/8} \approx 1.96$ -fold per week. To estimate the growth of q in a more rigorous way, we fitted the trend line using two fitting windows (Figure 2A). Fitting in weeks 43–47 gives weekly growth $v = 2.24$ [95% CI: 2.03–2.48], whereas fitting in weeks 43–51 gives $v = 1.88$ [95% CI: 1.75–2.01]. These confidence intervals are estimated as $1.96 \times$ standard error of the slope. However, as the estimate for weeks 43–47 is based on only 5 data points and the number of the VOC genomes in the first data point is small ($n = 30$), we additionally estimated the 95% credible interval for weeks 43–47 *a priori* (assuming binomial distributions of VOC and 20A.EU1 strain genomes, see Methods). Using this auxiliary method, we estimated that the 95% CrI is 2.09–2.45, which is somewhat narrower than the 95% CI calculated from the standard error of the slope. This demonstrates that the CI calculated from the standard error of the slope was not a result of an incidental linearity of data points.

To estimate the ratio of reproduction numbers of the VOC and the 20A.EU1 strain, $\mathcal{R}_t^{\text{VOC}}/\mathcal{R}_t^{20\text{A.EU1}}$, we assumed that both strains have the same average serial interval of 6.73 days [17]. Then, $\mathcal{R}_t^{\text{VOC}}/\mathcal{R}_t^{20\text{A.EU1}} = 2.24^{6.73/7} = 2.18$ [95% CI: 1.97–2.40] for the fitting window in between weeks 43 and 47, and $\mathcal{R}_t^{\text{VOC}}/\mathcal{R}_t^{20\text{A.EU1}} = 1.88^{6.73/7} = 1.83$ [95% CI: 1.71–1.96] for fitting window in between weeks 43 and 51. The eight-week window estimate gives a smaller advantage of the VOC than the four-week window. We think that this discrepancy is caused by two factors. As shown in the report of PHE [3] in week 51 of 2020, the VOC almost reached fixation in the Greater London and Kent, being nearly absent in central England. Additionally, more stringent measures implemented in Greater London and Kent limited the growth of absolute numbers of VOC cases. The heterogeneity in ratio q is not important as long as q is low (as it was until week 47); however, when q becomes high in some subregions, growth of q in the whole region decelerates. For this reason, we have not extended the fitting window past week 51. Since we may not rule out a chance that the higher growth of q in the window of 43–47 weeks is an artifact caused by a small number of data points, we conclude that that ratio $\mathcal{R}_t^{\text{VOC}}/\mathcal{R}_t^{20\text{A.EU1}}$ is in the range 1.83–2.18 [95% CI: 1.71–2.40].

We should notice that sequenced genomes are submitted to GISAID with some time delay after sample collection; however, as of 12 February, the data for weeks 44–51 of 2020 appears nearly complete. As shown in Figure 2B, replicative advantage estimates stabilize with the progression of the ‘date of last submission’. We have also verified our fits using the COG-UK genome database. We performed the analysis based on pillar 2 genomes that excludes routine tests of health and care workers and other tests made for particular purposes (pillars 1, 3, and 4) (<https://www.gov.uk/government/publications/coronavirus-covid-19-testing-data-methodology/covid-19-testing-data-methodology-note>, accessed on 12 February 2021). For the COG-UK data we obtained nearly the same ($\pm 3\%$) replicative advantage of the VOC strain in the 43–47 and 43–51 week windows, see Supplementary Figure S1 for fits and values.

In the same way we estimated that the ratio of $\mathcal{R}_t$ of the VOC strain to $\mathcal{R}_t$ of other strains (that are neither VOC nor A20.EU1) is in the range of 2.03–2.47 [95% CI: 1.89–2.77], where the lower bound is the estimate within weeks 43–51 and the higher bound is the estimate for weeks 43–47 of 2020. As may be expected, $\mathcal{R}_t^{\text{VOC}}/\mathcal{R}_t^{\text{other}} > \mathcal{R}_t^{\text{VOC}}/\mathcal{R}_t^{20\text{A.EU1}}$, which reflects the fact that the A20.EU1 strain has a replicative advantage over previously dominant non-VOC D614G strains. We estimated this advantage by fitting a trend line to data points from weeks 34–45 of 2020, because for this period the exponential growth of the ratio of the A20.EU1 strain to other non-VOC D614G strains is observed (Figure 2C). In this period, the ratio of the A20.EU1 to other non-VOC D614G strains grows at a rate of 1.25 [95% CI: 1.23–1.28] per week, which gives $\mathcal{R}_t^{20\text{A.EU1}}/\mathcal{R}_t^{\text{other non-VOC D614G}} = 1.25^{6.73/7} = 1.24$ [95% CI: 1.22–1.27]. Finally, we estimated that the ratio $\mathcal{R}_t^{\text{D614G}}/\mathcal{R}_t^{\text{D614}} = 1.44^{6.73/7} = 1.42$ [95% CI: 1.38–1.45]. Concluding, we showed that the D614G strain that spread worldwide

towards fixation had replicative advantage of 1.42 in relation to D614 strains in England. Its substrain A20.EU1 had replicative advantage of 1.24 over bulk D614G, and reached the proportion of 68% of genomes in England. Currently, A20.EU1 is outcompeted by the VOC that has about two-fold replicative advantage in relation to the A20.EU1 strain.

3.3. Worldwide Spread of the VOC Strain

London serves as a major transportation hub and thus, unsurprisingly, among 46 countries that reported SARS-CoV-2 genomes in January 2021, 40 countries reported a VOC genome from this period. We estimated the ratio of the VOC genomes to all genomes in these countries (Table 1). We found that, in addition to England, in 10 countries the VOC genomes constituted more than half of reported genomes. The data suggest that the strain is spreading globally, even though from countries other than England less than 20% of VOC genomes were reported (as of 16 February 2021).

Table 1. Fraction of VOC genomes in countries that in January 2021 reported more than 50 genomes.

Country	Proportion of VOC Genomes	95% CrI
England	86.1% (43,260/50,240)	85.8–86.4%
Scotland	71.3% (1445/2028)	69.3–73.2%
Ghana	68.9% (62/90)	58.9–77.8%
Turkey	68.6% (155/226)	62.4–74.3%
Slovakia	67.7% (42/62)	56.5–79.0%
Ireland	62.0% (364/587)	58.1–65.9%
Italy	61.0% (169/277)	55.2–66.8%
Nigeria	60.3% (70/116)	51.7–69.0%
N. Ireland	57.9% (183/316)	52.5–63.3%
Wales	53.5% (2125/3971)	52.0–55.1%
Romania	52.0% (26/50)	38.0–66.0%
Spain	49.1% (576/1174)	46.2–52.0%
France	44.7% (691/1545)	42.3–47.2%
Israel	36.9% (113/306)	31.7–42.5%
Australia	34.4% (66/192)	27.6–41.2%
Austria	34.2% (178/520)	30.2–38.3%
Belgium	32.6% (538/1651)	30.4–34.9%
Czechia	32.1% (17/53)	20.8–45.3%
Finland	31.8% (67/211)	25.6–37.9%
Sweden	27.7% (109/394)	23.4–32.2%
Netherlands	21.0% (328/1561)	19.0–23.1%
Singapore	20.6% (30/146)	14.4–27.4%
Portugal	20.0% (127/636)	17.0–23.1%
Germany	16.9% (68/403)	13.4–20.6%
Sri Lanka	14.8% (13/88)	8.0–22.7%
Norway	14.3% (27/189)	9.5–19.6%
S. Korea	13.3% (12/90)	6.7–21.1%
India	12.9% (11/85)	5.9–20.0%
Luxembourg	10.7% (27/253)	7.1–14.6%
Denmark	10.5% (1270/12,137)	9.9–11.0%
Brazil	9.5% (13/137)	5.1–14.6%
N. Macedonia	9.3% (5/54)	1.9–18.5%
Switzerland	8.2% (409/4985)	7.4–9.0%
Iceland	7.7% (5/65)	1.5–15.4%
Canada	7.0% (36/514)	4.9–9.3%
Poland	6.7% (11/165)	3.0–10.9%

Table 1. Cont.

Country	Proportion of VOC Genomes		95% CrI
Mexico	6.5%	(5/77)	1.3–13.0%
Latvia	5.7%	(5/87)	1.2–11.5%
USA	3.7%	(608/16,439)	3.4–4.0%
S. Africa	0.9%	(1/112)	0.0–2.7%
Colombia	0.00%	(0/64)	
Mayotte	0.00%	(0/224)	
Lithuania	0.00%	(0/381)	
Egypt	0.00%	(0/84)	
Kenya	0.00%	(0/75)	
French Guiana	0.00%	(0/98)	

Using the same method as in the previous section we estimated the replicative advantage of the VOC strain in four other countries in which the number of reported VOC genomes permits such analysis. In Denmark, Scotland, and Wales, similarly to England, the A20.EU1 strain constitutes a large share of all genomes in November 2020, respectively: 39%, 64%, and 70%. We thus compared the VOC with the A20.EU1 strain in these countries. In the USA, where we have not found a dominating D614G substrain, we compare the VOC with all non-VOC genomes. In all four countries, we found periods of the exponential growth of the ratio of compared genomes (seven-week-long in Denmark, Scotland, and Wales, and five-week-long in the USA; see linear growth in logarithmic scale in Figure 3). This allowed us to estimate the replicative advantage of the VOC strain in a narrow range: from $1.69^{6.73/7} = 1.65$ for Wales to $1.76^{6.73/7} = 1.72$ for the USA with 95% confidence intervals in the range [1.46–2.04]. We should notice that in all four countries the replicative advantage of the VOC strain has been found smaller than estimated for England.

3.4. Replicative Advantage of the 501Y.V2 Strain in South Africa

In addition to $\Delta 69-70\text{HV}$, mutation N501Y in the RBD of spike is considered as the most important recent mutation [9]. This mutation occurred independently in the South African strain 501Y.V2, where it is accompanied by two other mutations in spike RBD: K417N and E484K [7]. Using the same method as previously we estimated the replicative advantage of the 501Y.V2 strain over other South African strains (Figure 4). In weeks 43–50 of 2020, when an exponential growth is observed, the ratio of 501Y.V2 strain to other strains grows at the weekly rate of 1.58 [95% CI: 1.45–1.72], which gives $\mathcal{R}_t^{501Y.V2} / \mathcal{R}_t^{\text{other}} = 1.58^{6.73/7} = 1.55$ [95% CI: 1.43–1.69]. The noisiness of data is associated with the small number of genomes sequenced in the entire period of weeks 43–50 (only 407 501Y.V2 sequences and 376 non-501Y.V2 sequences collected in this period were submitted to GISAID).

3.5. Emergence of Mutations in VOC Genomes in England

Because of its high replicative advantage, the VOC strain will likely become globally dominant, possibly reaching fixation. It is thus crucial to track mutations that arise in this strain, that could further increase its replicative advantage. We thus performed sequence analysis of the spike gene in all genomes from England submitted to GISAID till 12 February 2021 (Supplementary Data File S1, sheet ‘Mutations’). There are 2232 different mutations in genomes collected in England, including 1213 different mutations in VOC genomes with 697 ‘confirmed mutations’ found in more than one submitted VOC genome. Majority of these mutations (535 out of 697) have arisen also in non-VOC strains. This suggests that the mutational space is to a large extent already explored, however the nine mutations characterizing the spike protein of VOC may increase or decrease the replicative advantages of recurrent mutations or allow for propagation of novel mutations. In Figure 5A we show accumulation of mutations in VOC spike in time; within 53,185 analyzed VOC genomes, 20% have at least one mutation, while 1.7% has two mutations or more. In relation to the

first collected VOC lineage genome (GISAID accession ID: EPI_ISL_601443), the VOC genomes collected in the end of January have on average 0.3 mutation in their spike protein (Figure 5B).

In Figure 6 we analyze the replicative advantage of 60 VOC substrains for which there were at least 30 genomes submitted to GISAID from England. Each substrain is marked by disk on the (first collection date, total occurrences) plane. The VOC substrains that from time of its first collection grow faster than the VOC strain on average are above the red line that shows the average growth of the VOC strain sequences. Unsurprisingly, the majority of substrains that exceed 30 submitted genomes fall into this category. This analysis should be taken with great caution, as the total number of substrains is high, some of may grow faster in the considered time window just by chance without having any replicative advantage. Nevertheless, one can use this approach to screen for further analysis the substrains that can potentially have a replicative advantage. The most prevalent mutation, defining a fast growing variant is the L18F substitution (1186 genomes) in the N-terminal domain (NTD) of spike protein. The second most prevalent is L5F (658 genomes) localized in the signal peptide of spike. This mutation (at a highly homoplastic position that may be a sequencing artifact [18]) was found abundant also in non-VOC genomes. The third most prevalent mutation, that is also by far the most prevalent RBM as well as in whole RBD mutation, is S494P (441 genomes). The other two fast growing “sibling” mutations in VOC — Q677H (256 genomes) and Q675H (86 genomes) — are present in the the proximity of the polybasic cleavage site (residues 682–685) at the S1/S2 boundary influencing RBD–ACE2 binding [19]. Mutations at residue Q677 (either Q677H or Q677P) were found in several independent lineages spreading over the autumn of 2020 and into the winter of 2021 in the USA [5].

3.6. Growth of the L18F Substrain

The first occurrence of the spike L18F substitution has been reported in a VOC strain genome collected on 4 December 2020 (GISAID ID: EPI_ISL_720875). As of 12 February 2021, as much as 1186 spike L18F VOC genomes have been reported in England. Of note, in Autumn 2020, that is, before the VOC lineage has become the dominant strain, the L18F substitution was a ubiquitous mutation in England. Till 12 February 2021, most of the L18F non-VOC genomes in England (97.6%, 25,655 out of 26,280) were found within the 20A.EU1 strain. The fraction of spike L18F mutation in the expanding 20A.EU1 strain was slowly increasing from 35% (1332 out of 3799) in September, 43% (5658 out of 13,046) in October, till 52% (8917 out of 17,470) in November 2020, which may suggest that this mutation was beneficial for the 20A.EU1 strain.

In Figure 7 we show the exponential growth of the L18F VOC substrain in England in the five-week period of 7 December 2020–17 January 2021, in relation to the VOC genomes non-mutated at residue 18, denoted L18. In the considered period, the ratio of the L18F to the L18 genomes increased with the fitted weekly growth rate of 1.75, which gives $\mathcal{R}_t^{L18F}/\mathcal{R}_t^{L18} = 1.75^{6.73/7} = 1.72$ [95% CrI: 1.57–2.02]. This credible interval is calculated assuming a binomial distribution of the number of the L18F and L18 VOC genomes in each week (see Methods). The confidence interval calculated from $1.96\times$ standard error of the slope, [1.63–1.81], is narrower, which means that the nearly perfect co-linearity of six data points is somewhat coincidental, and the binomial distribution-based credible interval is the proper estimate. This analysis suggests an advantage of the L18F VOC substrain in relation to the remaining VOC genomes, but since it is based on short period, it must be taken with caution. The advantage of L18F substrain is supported by data from Wales, UK, where the L18F VOC genomes constituted 17% (390 out of 2301) of all VOC genomes reported in January, substantially more than in the same period in England, 3.0% (1333 out of 43,700). The number of genomes in Wales is however too small to perform an analysis analogous to that in Figure 7.

3.7. VOC Strain Mutations in Spike Receptor-Binding Domain

Of particular concern are the VOC strain mutations occurring in the receptor-binding domain (RBD, residues 333–527), especially mutations in the receptor-binding motif (RBM, residues 438–506). These mutations may potentially lead to immune escape mutants, resulting in reinfection of convalescent individuals and aggravation of the efficacy of current vaccines. Propagation of such mutations is facilitated by high replicative advantage of the VOC strain and potential selection due to the increasing number of convalescent or immunized individuals. The VOC-202012/01 strain spike RBM mutations of special concern are substitutions E484K and S494P.

E484K. A first genome has been collected on 17 December 2020 (GISAID ID: EPI_ISL_782148), and there were 30 genomes reported from the England up till 12 February 2021. The same mutation has occurred in the fast expanding South African and Brazilian (Manaus) strains that share with the VOC substitution N501Y and additionally have a mutation of residue 417: either K417N (South African strain 501Y.V2) [7] or K417T (Manaus strain P.1) [8]. It was suggested that E484K may compromise binding of class 2 neutralizing antibodies, while the A501V mutation interferes with binding of class 1 antibodies. The P.1 strain led to the surge of infections in Manaus in December 2020 despite high seroprevalence of the population (a study of blood donors indicated that 76% [95% CI: 67–98%] of the population in Manaus had been infected with SARS-CoV-2 by October 2020 [20]).

S494P. A first genome has been collected on 12 November 2020 (GISAID ID: EPI_ISL_741039), and there were 441 genomes reported from England up till 12 February 2021. In an *in silico* study, this substitution has been found to increase complementarity between the RBD and ACE2 [21]. This mutation has been also characterized as an escape mutation by Koenig et al. [22], who also distinguished five additional “escape” residues in the RBM: G447, Y449, L452, F490, G496, and six outside the RBM but within the RBD: Y369, S371, T376, F377, K378, R403. Among these residues, until 12 February 2021, substitution F490S (first collected on 13 December 2020, GISAID accession ID: EPI_ISL_736026) was reported in the highest number of genomes (28 genomes in England).

4. Discussion

The mutations of SARS-CoV-2 that substantially increase replicative advantage of emerging strains will likely become dominant, either locally in countries or continents, or worldwide. Substitution D614G in spike protein (first GISAID reported genome, EPI_ISL_913915, was collected on 2 January 2020, in Mexico) initiated a strain with replicative advantage over D614 strains estimated based on data from England as 1.42 [95% CI: 1.38–1.45]. The D614G strain has spread worldwide nearly reaching fixation; it was present in more than 99% of genomes collected worldwide in January 2021. The 20A.EU1 strain, a substrain of D614G that harbors A222V mutation in spike, emerged in Spain in early summer, 2020, spread over Europe, becoming the dominating strain (more than half of sequenced genomes) in several countries (Spain, England, Scotland, Wales, Ireland and Italy) in November, 2020, but was nearly absent outside of Europe. Based on GISAID data in England we estimated its replicative advantage over other non-VOC D614G English strains as 1.24 [95% CI: 1.22–1.27]. The VOC strain started spreading in England in October 2020, outcompeting the A20.EU1 strain, and reached about 80% of genomes in England in January 2021. We have estimated that its replicative advantage over the A20.EU1 strain is in the range 1.83–2.18. The lower bound was obtained by fit in the eight-week-long period of weeks 43–51 of 2020, when the ratio of the VOC to the A20.EU1 strain genomes increased 217 times, from 0.0074 to 1.62, whereas the upper bound was obtained in the four-week-long period of weeks 43–47 of 2020. We think that the slower growth in the period of weeks 47–51 is a consequence of (1) the fact that in Kent, Greater London, and their vicinity, the VOC strain almost reached fixation and (2) the fact that in these regions more stringent measures were implemented to suppress rapid growth of cases.

We also estimated the replicative advantage of the VOC strain in relation to the 20A.EU1 strain in Denmark, Scotland, and Wales, and in relation to bulk non-VOC strains in USA. We found values in range from 1.69 for Wales to 1.76 for USA, with 95% confidence intervals in range [1.46–2.04], that are smaller than in England. One possible explanation is that the VOC strain is able to infect seroprevalent or exposed individuals. Such ability would increase its replicative advantage in a population, in which the fraction of seroprevalent individuals is large [23]. This mechanism would also help to explain the higher replicative advantage of the VOC strain observed in weeks 43–47 of 2020, when the VOC strain was gaining prevalence in London area [3]. The strain P.1, found in Brazil, that shares with VOC RBM mutation N501Y, caused recently the second (higher) wave of deaths in Manaus despite high seroprevalence of the population [8].

In addition to double deletion $\Delta 69-70$ HV, substitution N501Y in the RBD of spike is considered the most important VOC mutation [9]. This mutation occurred independently in the South African strain 501Y.V2. We estimated that 501Y.V2 has replicative advantage over other South African strains equal 1.55 [95% CI: 1.43–1.69]. The replicative advantage of 501Y.V2 strains supports the conjecture that mutation N501Y increases infectiousness of SARS-CoV-2 by increasing the affinity of spike RBD to the angiotensin-converting enzyme 2 (ACE2) [24].

Both our estimates suggests that the replicative advantage of the VOC strain is higher than early estimate, 1.47 [95% CI: 1.34–1.59] [9]. Also Davies et al. [11] estimated that the VOC strain has 43–82% [95% CrI: 38–106%] higher transmissibility. In these studies, the authors estimated the replicative advantage of the VOC strain separately for each region and then averaged over analyzed regions. We think that our approach, based on weekly and England-averaged data, gives a more accurate estimate. This is because the VOC strain (as well as other strains) evolves and some substrains have a lower or even no replicative advantage and will become extinct in the course of evolution, and finally a substrain with the highest replicative advantage will dominate. Substrains with a small replicative advantage may contribute to the VOC to non-VOC replication ratio averaged over NHS STP areas of England but only marginally influence the expansion of the VOC strain globally. By using the aggregated approach we estimate the replicative advantage of the dominating substrain(s). To convert the weekly growth of the ratio of genomes to the ratio of respective reproduction numbers we assumed that both strains have the same mean serial interval equal 6.73 days. This calculation is approximate as the serial interval is not a number but follows a hypoexponential distribution [17]. Additionally, although no current data indicate this, it may happen that the faster spread of VOC strain partially results from a shorter generation time.

The VOC strain, because of its high replicative advantage, is likely to become globally dominant, possibly reaching fixation. It will continue to evolve so it is crucial to track which mutations already present in this strain can further increase its replicative advantage. We thus performed detailed sequence analysis of the spike protein identifying, as of 12 February 2021, 1213 different mutations in VOC genomes collected in England with 697 of them found in more than one submitted VOC genome. By systematic analysis of the propagation of VOC substrains we found that substrain(s) conferring L18F substitution is/are the most abundant and rapidly growing VOC substrain(s).

The data collected in the five-week period of 7 December 2020–17 January 2021, in England suggest a replicative advantage of the L18F substrain in relation to the remaining VOC strains. The estimate of the replicative advantage of 1.72 [95% CrI: 1.57–2.02] is based on a relatively short time interval so it must be taken with caution. Importantly, the L18F mutation has also expanded in the South African strain 501Y.V2 defined by three spike mutations K417N, E484K, N501Y (thus sharing with the VOC strain spike mutation N501Y). Among the 501Y.V2 genomes collected after 1 December 2020, the L18F substrain constitutes 41% genomes (127 out of 309), according to GISAID as of 12 February 2021). In Brazil, in strain P.1 defined by three spike mutations K417T, E484K, N501Y (differing from the South African strain 501Y.V2 by substitution K417T instead of K417N), mutation L18F

has been found in 93% of genomes (69 out of 74) collected after 1 December 2020. This data suggests a replicative advantage of L18F substrains within the VOC, 501Y.V2, and P.1 strains, in, respectively, England, South Africa, and Brazil. This replicative advantage of the VOC L18F substrain must be considered with caution until the mechanism promoting faster spread of strains containing L18F substitution is elucidated. Leucine 18 lies in the N-terminal domain (NTD), that has not been typically considered as a target for neutralizing antibodies. However, there is a growing number of studies showing that the NTD is targeted by antibodies and that NTD deletion 69H–70V (characterizing the VOC strain) compromises binding of antibodies [25–27]. With respect to L18F, an in vitro study by Cele et al. shows that an African variant L18F, D80A, D215G, K417N, E484K, N501Y, D614G, A701V propagates much faster than a variant without L18F mutation in the presence of plasma antibodies collected from donors infected in the first wave of epidemic in South Africa (June–August 2020) [28]. Correspondingly, McCallum et al. showed that L18F substitution compromises binding of neutralizing antibodies [15]. Findings by Cele et al. and McCallum et al., together with the increase of L18F variants in 501Y.V2, P.1, and VOC strains, suggests that the replicative advantage of L18F mutants can be partly associated with their ability to infect seroprevalent individuals, and thus depend on the fraction of seroprevalent individuals in given territory. In turn, growth of strains with mutations in escape residues L18 and S494 on the VOC strain suggests an increasing selection pressure resulting from the growth of the seroprevalent fraction of the population of England. This trend can be enhanced by the ongoing English vaccination program, in which the relatively large time span between the first and second dose can be a contributing factor.

In summary, we have shown that the new VOC strain has about twofold replicative advantage over the 20A.EU1 strain of SARS-CoV-2, that was dominating in England in November 2020. The strain has already spread across the world and will likely spread further towards fixation. It was present in 40 out of 46 countries that reported at least 50 viral genomes in January, 2021. Spread of the faster-replicating VOC-202012/01 strain may hinder the efforts to contain the COVID-19 epidemics prior to mass vaccinations. As the global spread of the VOC strain is very likely, it is important to monitor mutations of this strain, with particular attention to mutations interfering with immune response including the fast spreading NTD mutation L18F, and RBM mutations E484K, F490S, and S494P that may decrease the efficacy of currently available vaccines.

Supplementary Materials: The following are available online at <https://www.mdpi.com/1999-4915/13/3/392/s1>: Figure S1: The replicative advantage of the VOC strain estimated based on COG-UK pillar 2 data; Data File S1: All datasets that were extracted from public data sources and used for analysis in this study.

Author Contributions: Conceptualization, F.G., M.K. and T.L.; Data curation, F.G., G.P., S.G. and M.K.; Investigation, F.G. and M.K.; Software, G.P. and S.G.; Supervision, M.K. and T.L.; Visualization, F.G., G.P., S.G. and M.K.; Writing—original draft, T.L.; Writing—review & editing, M.K. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the Norwegian Financial Mechanism GRIEG-1 grant 2019/34/H/NZ6/00699 (operated by the National Science Centre Poland). The funding agency had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found at <https://www.gisaid.org> (available after registration; accessed on 12 and 17 February 2021, see Methods) and <https://www.cogconsortium.uk> (accessed on 15 February 2021), and at <https://www.ecdc.europa.eu/en/publications-data/data-national-14-day-notification-rate-covid-19> and <https://coronavirus.data.gov.uk/details/cases> (accessed on 17 February 2021). We also provide retrieved and preprocessed data in a single supplementary file, Data File S1.

Acknowledgments: We are very grateful to the GISAID Initiative and all its data contributors, i.e., the Authors from the originating laboratories responsible for obtaining the specimens and the submitting laboratories where genetic sequence data were generated and shared via the GISAID Initiative, on which this research is based.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Rambaut, A.; Loman, N.; Pybus, O. Preliminary Genomic Characterisation of an Emergent SARS-CoV-2 Lineage in the UK Defined by a Novel Set of Spike Mutations. Available online: <https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563> (accessed on 9 January 2021).
- Shu, Y.; McCauley, J. GISAID: Global initiative on sharing all influenza data—From vision to reality. *Eurosurveillance* **2017**, *22*, 30494. [CrossRef] [PubMed]
- Public Health England. Investigation of Novel SARS-CoV-2 Variant: 202012/01. Technical Briefing 2. Available online: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/949639/Technical_Briefing_VOC202012-2_Briefing_2_FINAL.pdf (accessed on 9 December 2020).
- Greater London Authority. Coronavirus (COVID-19) Cases. Available online: <https://data.london.gov.uk/dataset/coronavirus-covid-19-cases> (accessed on 9 January 2021).
- Hodcroft, E.B.; Zuber, M.; Nadeau, S.; Crawford, K.H.D.; Bloom, J.D.; Veeler, D.; Vaughan, T.G.; Comas, I.; Candelas, F.G.; Stadler, T.; et al. Emergence and spread of a SARS-CoV-2 variant through Europe in the summer of 2020. *medRxiv* **2020**. [CrossRef]
- Yurkovetskiy, L.; Wang, X.; Pascal, K.E.; Tomkins-Tinch, C.; Nyalile, T.P.; Wang, Y.; Baum, A.; Diehl, W.E.; Dauphin, A.; Carbone, C.; et al. Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant. *Cell* **2020**, *183*, 739–751. [CrossRef] [PubMed]
- Tegally, H.; Wilkinson, E.; Giovanetti, M.; Iranzadeh, A.; Fonseca, V.; Giandhari, J.; Doolabh, D.; Pillay, S.; San, E.J.; Msomi, N.; et al. Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *medRxiv* **2020**. [CrossRef]
- Sabino, E.C.; Buss, L.F.; Carvalho, M.P.S.; Crispim, M.A.E.; Fraiji, N.A.; Pereira, R.H.M.; Parag, K.V.; da Silva Peixoto, P.; Kraemer, M.U.G.; Oikawa, M.K.; et al. Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence. *Lancet* **2021**, *397*, 452–455. [CrossRef]
- Public Health England. Investigation of Novel SARS-CoV-2 Variant: 202012/01. Technical Briefing 1. Available online: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/947048/Technical_Briefing_VOC_SH_NJL2_SH2.pdf (accessed on 9 December 2020).
- Leung, K.; Shum, M.H.; Leung, G.M.; Lam, T.T.; Wu, J.T. Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020. *Eurosurveillance* **2021**, *26*, 2002106. [CrossRef] [PubMed]
- Davies, N.G.; Abbott, S.; Barnard, R.C.; Jarvis, C.I.; Kucharski, A.J.; Munday, J.; Pearson, C.A.B.; Russell, T.W.; Tully, D.C.; Washburne, A.D.; et al. Estimated transmissibility and severity of novel SARS-CoV-2 Variant of Concern 202012/01 in England. *medRxiv* **2021**. [CrossRef]
- COVID-19 Genomics UK Consortium. Available online: <https://www.cogconsortium.uk> (accessed on 12 February 2021).
- Grabowski, F.; Preibisch, G.; Kochańczyk, M.; Lipniacki, T. SARS-CoV-2 Variant Under Investigation 202012/01 has more than twofold replicative advantage. *medRxiv* **2021**. [CrossRef]
- Wibmer, C.K.; Ayres, F.; Hermanus, T.; Madzivhandila, M.; Kgagudi, P.; Lambson, B.E.; Vermeulen, M.; van den Berg, K.; Rossouw, T.; Boswell, M.; et al. SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. *BioRxiv* **2021**. [CrossRef]
- McCallum, M.; Marco, A.D.; Lempp, F.; Tortorici, M.A.; Pinto, D.; Walls, A.C.; Beltramello, M.; Chen, A.; Liu, Z.; Zatta, F.; et al. N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. *BioRxiv* **2021**. [CrossRef]
- Rice, P.; Longden, I.; Bleasby, A. EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.* **2000**, *16*, 276–277. [CrossRef]
- Kochańczyk, M.; Grabowski, F.; Lipniacki, T. Super-spreading events initiated the exponential growth phase of COVID-19 with R_0 higher than initially estimated. *R. Soc. Open Sci.* **2020**, *7*, 200786. [CrossRef]
- De Maio, N.; Borges, R.; Walker, C.; Weilguny, L.; Slodkiewicz, G.; Goldman, N. Issues with SARS-CoV-2 Sequencing Data. Available online: <https://virological.org/t/issues-with-sars-cov-2-sequencing-data> (accessed on 12 February 2021).
- Qiao, B.; Olvera de la Cruz, M. Enhanced Binding of SARS-CoV-2 Spike Protein to Receptor by Distal Polybasic Cleavage Sites. *ACS Nano* **2020**, *14*, 10616–10623. [CrossRef]
- Buss, L.F.; Prete, C.A.J.; Abraham, C.M.M.; Mendrone, A.J.; Salomon, T.; de Almeida-Neto, C.; França, R.F.O.; Belotti, M.C.; Carvalho, M.P.S.S.; Costa, A.G.; et al. Three-quarters attack rate of SARS-CoV-2 in the Brazilian Amazon during a largely unmitigated epidemic. *Science* **2021**, *371*, 288–292. [CrossRef]
- Chakraborty, S. Evolutionary and structural analysis elucidates mutations on SARS-CoV2 spike protein with altered human ACE2 binding affinity. *Biochem. Biophys. Res. Commun.* **2021**, *534*, 374–380. [CrossRef]
- Koenig, P.A.; Das, H.; Liu, H.; Kümmerer, B.M.; Gohr, F.N.; Jenster, L.M.; Schiffelers, L.D.J.; Tesfamariam, Y.M.; Uchima, M.; Wuerth, J.D.; et al. Structure-guided multivalent nanobodies block SARS-CoV-2 infection and suppress mutational escape. *Science* **2021**, *371*, eabe6230. [CrossRef] [PubMed]
- Chen, S.; Flegg, J.A.; White, L.J.; Aguas, R. Levels of SARS-CoV-2 population exposure are considerably higher than suggested by seroprevalence surveys. *medRxiv* **2021**. [CrossRef]

24. Fratev, F. The N501Y and K417N mutations in the spike protein of SARS-CoV-2 alter the interactions with both hACE2 and human derived antibody: A Free energy of perturbation study. *BioRxiv* **2020**. [[CrossRef](#)]
25. Collier, D.A.; Marco, A.D.; Ferreira, I.A.; Meng, B.; Datir, R.; Walls, A.C.; Bassi, J.; Pinto, D.; Fregni, C.S.; Bianchi, S.; et al. SARS-CoV-2 B.1.1.7 escape from mRNA vaccine-elicited neutralizing antibodies. *MedRxiv* **2021**. [[CrossRef](#)]
26. Kemp, S.A.; Collier, D.; Datir, R.; Ferreira, I.; Gayed, S.; Jahun, A.; Hosmillo, M.; Rees-Spear, C.; Mlcochova, P.; Lumb, I.U.; et al. Neutralising antibodies in Spike mediated SARS-CoV-2 adaptation. *MedRxiv* **2020**. [[CrossRef](#)]
27. Voss, W.N.; Hou, Y.J.; Johnson, N.V.; Kim, J.E.; Delidakis, G.; Horton, A.P.; Bartzoka, E.; Paresi, C.J.; Tanno, Y.; Abbasi, S.A.; et al. Prevalent, protective, and convergent IgG recognition of SARS-CoV-2 non-RBD spike epitopes in COVID-19 convalescent plasma. *BioRxiv* **2020**. [[CrossRef](#)]
28. Cele, S.; Gazy, I.; Jackson, L.; Hwa, S.H.; Tegally, H.; Lustig, G.; Giandhari, J.; Pillay, S.; Wilkinson, E.; Naidoo, Y.; et al. Escape of SARS-CoV-2 501Y.V2 variants from neutralization by convalescent plasma. *MedRxiv* **2021**. [[CrossRef](#)]

The Spread of SARS-CoV-2 Variant Omicron with a Doubling Time of 2.0–3.3 Days Can Be Explained by Immune Evasion

Frederic Grabowski, Marek Kościński  and Tomasz Lipniacki * 

Institute of Fundamental Technological Research, Polish Academy of Sciences, 02-106 Warsaw, Poland; frederic.grabowski@ippt.pan.pl (F.G.); mkochan@ippt.pan.pl (M.K.)

* Correspondence: tlipnia@ippt.pan.pl

Abstract: Omicron, the novel highly mutated SARS-CoV-2 Variant of Concern (VOC, Pango lineage B.1.1.529) was first collected in early November 2021 in South Africa. By the end of November 2021, it had spread and approached fixation in South Africa, and had been detected on all continents. We analyzed the exponential growth of Omicron over four-week periods in the two most populated of South Africa's provinces, Gauteng and KwaZulu-Natal, arriving at the doubling time estimates of, respectively, 3.3 days (95% CI: 3.2–3.4 days) and 2.7 days (95% CI: 2.3–3.3 days). Similar or even shorter doubling times were observed in other locations: Australia (3.0 days), New York State (2.5 days), UK (2.4 days), and Denmark (2.0 days). Log-linear regression suggests that the spread began in Gauteng around 11 October 2021; however, due to presumable stochasticity in the initial spread, this estimate can be inaccurate. Phylogenetics-based analysis indicates that the Omicron strain started to diverge between 6 October and 29 October 2021. We estimated that the weekly growth of the ratio of Omicron to Delta is in the range of 7.2–10.2, considerably higher than the growth of the ratio of Delta to Alpha (estimated to be in the range of 2.5–4.2), and Alpha to pre-existing strains (estimated to be in the range of 1.8–2.7). High relative growth does not necessarily imply higher Omicron infectivity. A two-strain SEIR model suggests that the growth advantage of Omicron may stem from immune evasion, which permits this VOC to infect both recovered and fully vaccinated individuals. As we demonstrated within the model, immune evasion is more concerning than increased transmissibility, because it can facilitate larger epidemic outbreaks.

Keywords: COVID-19 pandemic; SARS-CoV-2; Omicron variant; genome sequencing; mutation



Citation: Grabowski, F.; Kościński, M.; Lipniacki, T. The Spread of SARS-CoV-2 Variant Omicron with a Doubling Time of 2.0–3.3 Days Can Be Explained by Immune Evasion. *Viruses* **2022**, *14*, 294. <https://doi.org/10.3390/v14020294>

Academic Editor: Craig E. Cameron

Received: 10 December 2021

Accepted: 28 January 2022

Published: 30 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Omicron, the novel SARS-CoV-2 Variant of Concern (VOC, Pango lineage B.1.1.529, Nextstrain clade identifier 21K) was first collected in South Africa on 2 November 2021, (GISAID [1] sequence accession ID: EPI_ISL_8182767). Compared to the original SARS-CoV-2 virus, Omicron carries 30 amino acid non-synonymous substitutions, three small deletions, and one small insertion in the spike protein [2]. Altogether, Omicron has 51 amino-acid level mutations, and its closest known sibling has 15 mutations (GISAID sequence accession ID: EPI_ISL_622806) with only nine common mutations, implying a distance of 42 mutations from the last common ancestor (based on the phylogenetic tree generated by Nextstrain [3]). The collection date of the sibling genome, 13 September 2020, suggests more than a year of evolution in an isolated niche, possibly in an immunocompromised host, but more data is necessary to rule out or confirm the existence of hidden branches (see [4] for discussion). The lineage started spreading rapidly in South Africa's Gauteng province in November 2021, approaching fixation in the whole of South Africa by the end of that month and causing abrupt epidemic outbreaks across South Africa, then Europe, and finally other continents. In all these locations, Omicron outcompeted Delta VOC (lineage B.1.617.2), which in October 2021 accounted for more than 99% of genomes sequenced in

Europe, North America, and Oceania, more than 90% in Asia and South America, and nearly 90% in Africa.

The rapid spread of Omicron both in South Africa, having a widespread infection-induced seroprevalence [5], and in Western European countries, where a high proportion of the population is vaccinated [6], suggests immune evasion that may be linked to the high number of mutations in viral spike glycoprotein, the major target of neutralizing antibodies [7]. This is confirmed by a growing number of (i) *in vitro* studies showing Omicron resistance to humoral immunity provided by vaccine- or pre-existing variant infection-induced antibodies, as well as epidemiological studies indicating (ii) significantly reduced vaccine effectiveness against infection with Omicron and (iii) higher chance of reinfections with Omicron compared to Delta.

- (i) For two mRNA-based vaccines, BNT162b2 (Pfizer) and mRNA-1273 (Moderna), Liu et al. demonstrated a >21-fold and >8.6-fold decrease (Omicron versus D614G) in ID50 (infectious dose), respectively. For two vector vaccines, Ad26.COV2.S (Johnson & Johnson) and ChAdOx1 (AstraZeneca), all samples obtained from patients without a previous history of SARS-CoV-2 infection were below the level of detection against Omicron [8]. After three homologous mRNA vaccinations, the average ID50 drop was 6.5-fold [8]. Planas et al. showed that sera from either BNT162b2 or ChAdOx1 vaccine recipients (sampled 5 months after complete vaccination) barely inhibited Omicron. Sera from COVID-19 convalescent patients (collected 6 or 12 months post symptoms) displayed low-to-absent neutralizing activity against Omicron, whereas administration of a booster dose of BNT162b2 as well as vaccination of previously infected individuals generated an anti-Omicron neutralizing response, but with titers 5–31-fold lower than against Delta [9]. Omicron VOC was found to be 5.3–7.4-fold less sensitive than Beta VOC when assayed with serum samples obtained from individuals inoculated with 2 mRNA-1273 doses [10]. A meta-analysis of 24 studies showed a decrease in the neutralization titer (not significantly different between different vaccines) compared to the ancestral virus for the previous four VOCs: Alpha (1.6-fold), Gamma (3.5-fold), Delta (3.9-fold), and Beta (8.8-fold) [11]. This loss of neutralization activity is not as substantial as in the case of Omicron [8]. These findings are in line with another study that shows barely detectable serum neutralizing activity against Omicron after two mRNA vaccination doses (and still much lower neutralizing activity after the “booster” dose in relation to wild-type virus as well as the Delta VOC) [12].
- (ii) Andrews et al. showed a decrease in vaccine effectiveness against symptomatic infection by Omicron with respect to Delta [13]. Half a year after two-dose ChAdOx1 vaccination, the effectiveness was 42% against Delta, with no effect observed against Omicron starting 15 weeks after the second ChAdOx1 vaccination. In the case of BNT162b2, the protection 15 weeks after vaccination was 63% against Delta and 34–37% against Omicron. The BNT162b2 booster increases protection to above 93% against Delta and 75% against Omicron [13]. A report from the UK Health Security Agency confirms these results and additionally indicates that the mRNA “booster” effect against Omicron, but not against Delta, wanes rapidly in time to about 40% 10 weeks post “booster” dose [14]. These findings are in line with a report from the MRC Centre for Global Infectious Disease Analysis indicating a significantly increased risk of an Omicron case compared to Delta for those with vaccine status AZ 2+weeks post Dose 2 (PD2), Pfizer 2+w PD2, AZ 2+w post Dose 3 (PD3) and PF 2+w PD3 vaccine states with hazard ratios of 1.86 (95% CI: 1.67–2.08), 2.68 (95% CI: 2.54–2.83), 4.32 (95% CI: 3.84–4.85), and 4.07 (95% CI: 3.66–4.51), respectively [15].
- (iii) The same report indicates that Omicron is associated with a 5.41 (95% CI: 4.87–6.00)-fold higher risk of reinfection compared with Delta [15].

Although all the evidence is based on a limited number of cases and may be influenced by population-level biases, one can expect that the hazard ratio of Omicron versus Delta infection is in the range 2–5 and depends principally on specific vaccines, the proportion of the population vaccinated by the “booster” dose, and resistance after recovery from COVID-19.

In this work, we estimated the Omicron variant doubling time in South Africa based on the daily number of new COVID-19 cases in the two most populated South African provinces, Gauteng and KwaZulu-Natal [16]. Further, we repeated this estimation for Australia, the UK, Denmark, and New York State, four locations that differ with respect to their COVID-19 epidemic histories and proportions of COVID-19 vaccine manufacturers. In all these locations, new COVID-19 cases are expected to be relatively reliably reported and to have their viral genomes screened by sequencing. Based on data for the three latter locations, we showed that the weekly growth of the ratio of Omicron to Delta cases significantly exceeds previous weekly growths of the ratio of Delta to Alpha cases, and the ratio of Alpha cases to cases caused by pre-existing strains. To rule out hidden spread of the Omicron variant, which could potentially influence the above estimates, we used an Omicron phylogenetic tree from Nextstrain to estimate the strain divergence date. Finally, we employed a two-strain mathematical model to demonstrate that observed rapid outbreaks of Omicron strain can be explained solely by immune evasion, which expands the pool of individuals susceptible to infection.

2. Materials and Methods

2.1. Genomic Sequence-Based Analysis

All data used in this study were retrieved as of 9 January 2022. The GISAID [1] data include genomes with submission dates earlier than 6 January 2022.

In Figure 1 the weekly cases of Omicron, Delta, and pre-existing strains were estimated based on GISAID data and the cumulative number of COVID-19 cases in Gauteng and KwaZulu-Natal (DSFSI at the University of Pretoria, <https://github.com/dsfsi/covid19za/tree/master/data>, accessed on 9 January 2022). For Gauteng this dataset was amended by changing the number of cases from 8099 to 605 for 23 November 2021 using information from the South African National Institute for Communicable Diseases (NICD, <https://www.nicd.ac.za/latest-confirmed-cases-of-covid-19-in-south-africa-23-november-2021>, accessed on 9 January 2022). According to NICD, the difference results from a retrospective addition of 7494 antigen test results. The dataset used for Figure 1A,C is provided as Supplementary Table S1.

Mobility in Gauteng and KwaZulu-Natal (Figure 1B) was assessed based on the COVID-19 Community Mobility Reports from Google (<https://www.google.com/covid19/mobility>, accessed on 9 January 2022). Weekly averages were calculated based on workdays.

The Omicron strain divergence date and mutation accumulation rate (Figure 1D) were determined by Poisson regression, assuming that the mean number of mutations grows linearly with time. The Omicron phylogenetic tree (with last update on 3 January 2022) was retrieved from Nextstrain [4]; sample collection dates of genome sequences selected by Nextstrain were retrieved from GISAID. The resulting dataset is provided as Supplementary Table S2.

Two lists of GISAID IDs and corresponding acknowledgments for all genomes collected from 19 April 2021 until 2 January 2022 in Gauteng and KwaZulu-Natal are provided as Supplementary Tables S3 and S4.

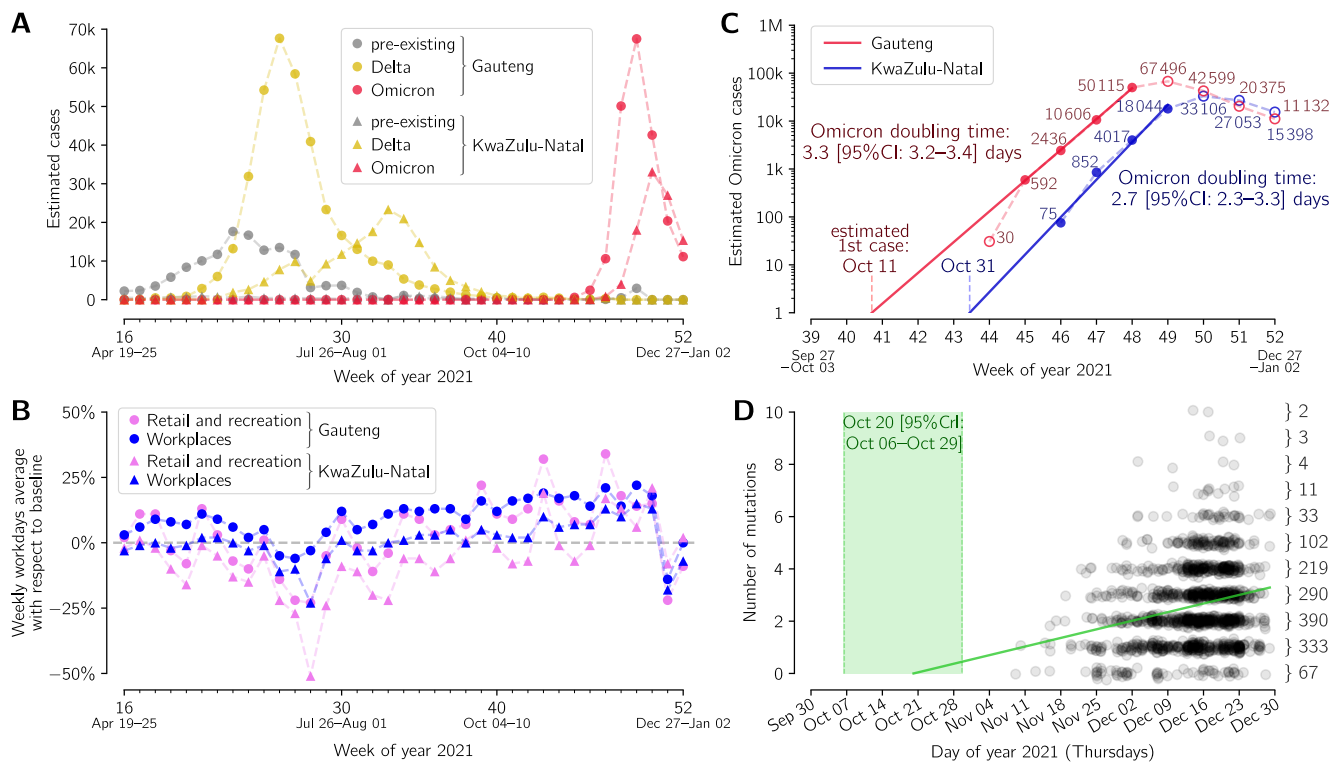


Figure 1. Growth and divergence of the Omicron strain. **(A)** Weekly aggregated cases of Omicron, Delta and other variants in two South African provinces, Gauteng and KwaZulu-Natal. **(B)** Weekly averaged workday mobility in Gauteng (filled circles) and KwaZulu-Natal (filled triangles) in workplaces (blue) and retail and recreation centers (pink). **(C)** Exponential growth of the Omicron strain in weeks 45–48 in 2021 (8 November–5 December) in Gauteng and in weeks 46–49 in KwaZulu-Natal. **(D)** Accumulation of mutations by the Omicron strain worldwide based on the Nextstrain phylogenetic tree. The green line shows the mutation accumulation trend determined by the linear regression assuming Poisson distribution of the number of mutations at a given time. The 95% credible interval of time is marked in light green. The dataset for panels **(A,B)** is provided as Supplementary Table S1; the phylogenetic tree (with dates) is provided as Supplementary Table S2.

In Figures 2 and 3 the weekly cases of Omicron, Delta, and pre-existing strains were estimated based on sequence data from GISAID and case data aggregated by Johns Hopkins University [17]. The datasets used for Figures 2 and 3 are provided as Supplementary Table S5.

The Omicron doubling time (Figures 1C, 2B,E and 3B,E) was estimated by log-linear regression of the (estimated) number of Omicron cases in the exponential phase of its growth. Daily new cases were aggregated by week. An appropriate four-week period was used in each location except Australia, for which a six-week period was used.

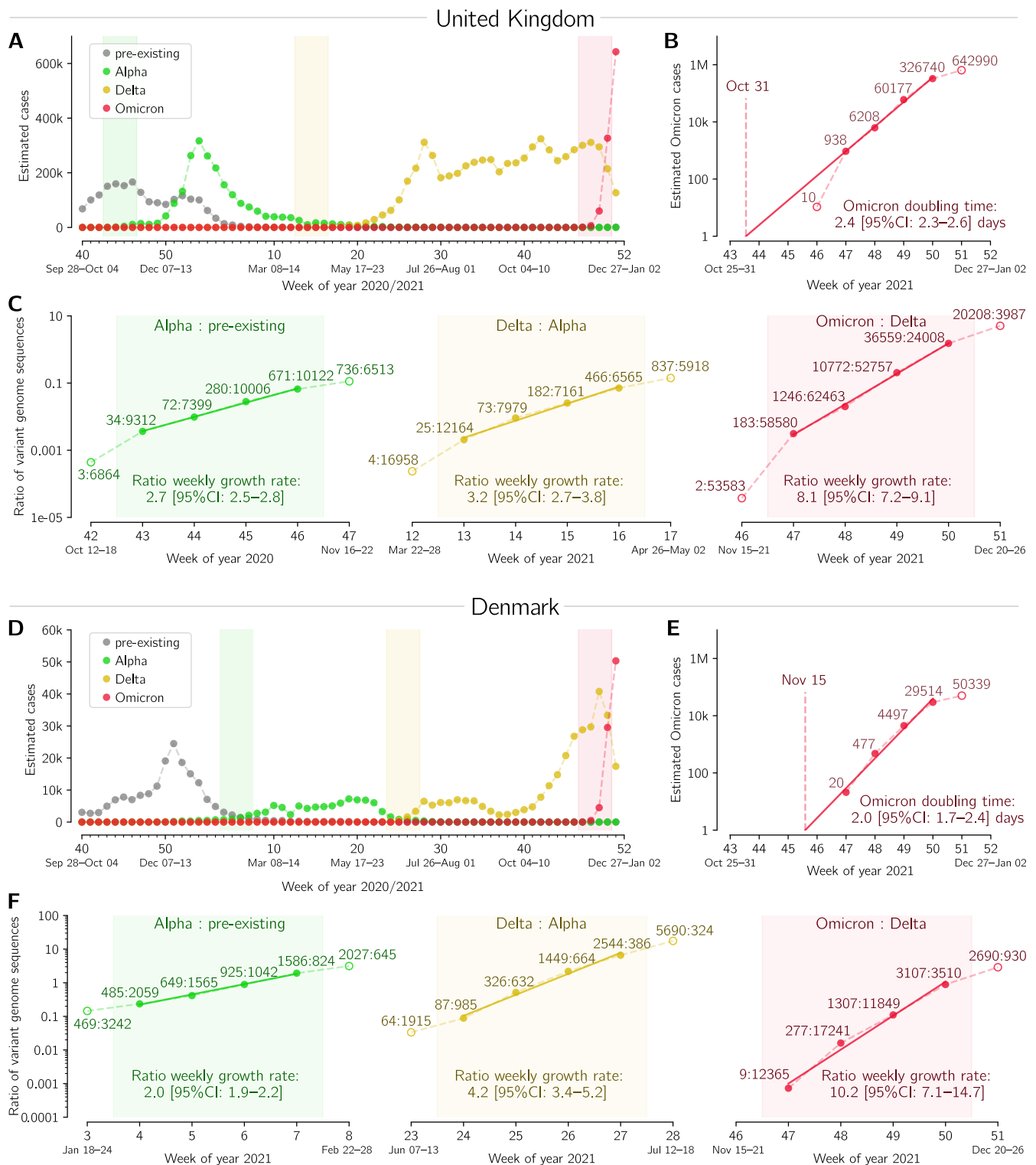


Figure 2. Succession of SARS-CoV-2 strains in the UK and in Denmark. **(A,D)** Estimated number of weekly cases infected with a particular strain over the weeks of 2020 and 2021. **(B,E)** Estimated number of weekly Omicron cases and the doubling time estimate based on log–linear regression in four whole-week periods (filled circles). **(C,F)** Ratios of weekly cases of an emergent strain to the previously dominant strain, and the estimate of ratios’ growth rates based on log–linear regression in four whole-week periods (subpanels correspond to shaded regions in respective panels **(A,D)**). A dataset for this figure is provided as Supplementary Table S5.

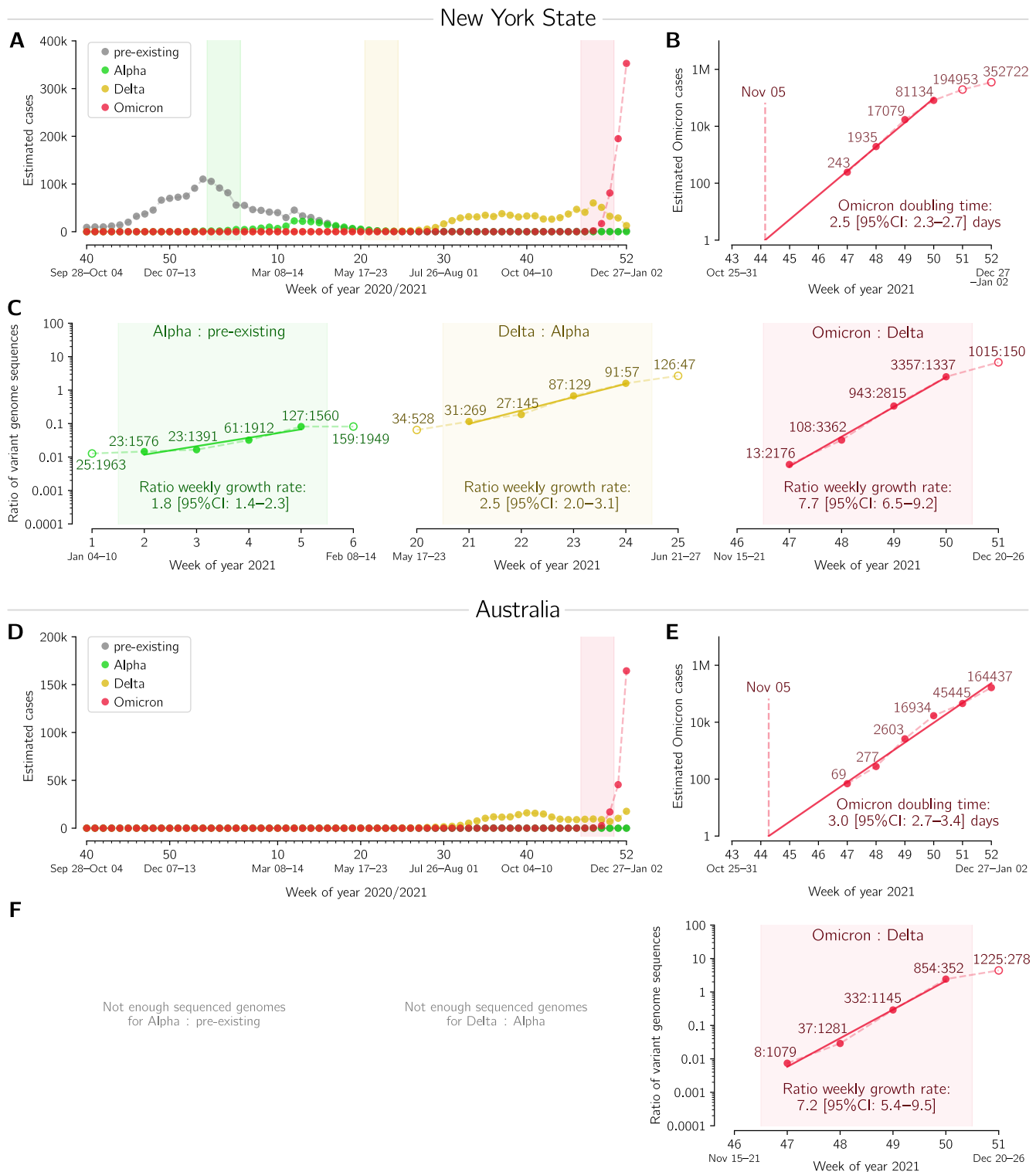


Figure 3. Succession of SARS-CoV-2 strains in New York State and in Australia. (A,D) Estimated number of weekly cases infected with a particular strain over the weeks of 2020 and 2021. (B,E) Estimated number of weekly Omicron cases and the doubling time estimate based on log-linear regression in four (Panel (B)) and six (Panel (E)) whole-week periods (filled circles). (C,F) Ratios of weekly cases of an emergent strain to the previously dominant strain, and the estimate of ratios' growth rates based on log-linear regression in four whole-week periods (subpanels correspond to shaded regions in respective panels (A,D)). A dataset for this figure is provided as Supplementary Table S5.

The weekly (multiplicative) growth rates of the ratios of an emerging to a pre-existing strain (Figures 2C,F and 3C,F) were also estimated by log-linear regression. To enable comparison between emerging strains, we select four-week periods of the fastest growth, so that in at least two of those weeks the emerging strain has lower counts than the pre-existing one. This additional criterion ensures that the analyzed period captures the emergence of the new strain. We use the reporting convention in which the weekly (multiplicative) growth rate of two implies doubling of the strain-to-strain ratio every week.

2.2. Mathematical Modeling

The mathematical model has been formulated as an extended susceptible–exposed–infectious–recovered (SEIR) model amended with a vaccinated (V) compartment. We assumed that the latent period was the same as the incubation period and was Erlang-distributed with the shape parameter $m = 6$ (which in the model structure is reflected by the inclusion of six exposed subcompartments) and the mean of $1/\sigma = 3$ days. The average period of infectiousness is $1/\gamma = 3$ days (such a short period reflects the assumption that the individuals with confirmed infection are quickly isolated and then cannot infect susceptible individuals). The recovered individuals become susceptible at the rate of $\rho = 1/\text{year}$. Susceptible individuals are vaccinated at the rate of $\nu = 2/\text{year}$, and their vaccine-induced immunity wanes at the rate ρ (same as the rate of transition from a recovered to a susceptible compartment).

We consider two model variants—see Figure 4A,B. In Model A, there is a single pool of individuals susceptible to both Delta and Omicron, whereas in Model B there are two additional compartments of individuals that are susceptible to either only Delta or only Omicron, and that are fed with the post-Omicron recovered or the post-Delta recovered, respectively, at the rate of $\pi = 1.5/\text{year}$. From these two additional susceptible compartments there are transitions to the compartment of individuals susceptible to both Delta and Omicron (at the rate ρ). In this way we account for only partial overlap in reciprocal post-infection protection. In Model A, transmissibility of Omicron is 4-fold higher than that of Delta, whereas in Model B the transmissibility of Omicron and Delta is the same but, due to the specific choice of transition parameters, the aggregated pool of individuals susceptible to Omicron is four-fold higher than that of individuals susceptible to Delta in the steady state before the appearance of Omicron. Model B is symmetrical with respect to both strains; however, by including an additional transition from the vaccinated compartment to the compartment of individuals susceptible only to Omicron (at the rate π), we account for faster waning of post-vaccination immunity to Omicron.

Model dynamics are governed by a system of ordinary differential equations (18 ODEs in Model A, 20 ODEs in Model B). The ODEs may be unambiguously derived, assuming mass-action kinetics, based on model schemes in Figure 4A,B, and parametrized with kinetic rates given in Figure 4C. On the first day shown in Figure 4D–G, when the system is in equilibrium with Delta, just one individual (in a population of 10^6 individuals) is exposed to Omicron. As emphasized further in Results, the initial exponential phase of the Omicron outbreak is nearly identical within both models.

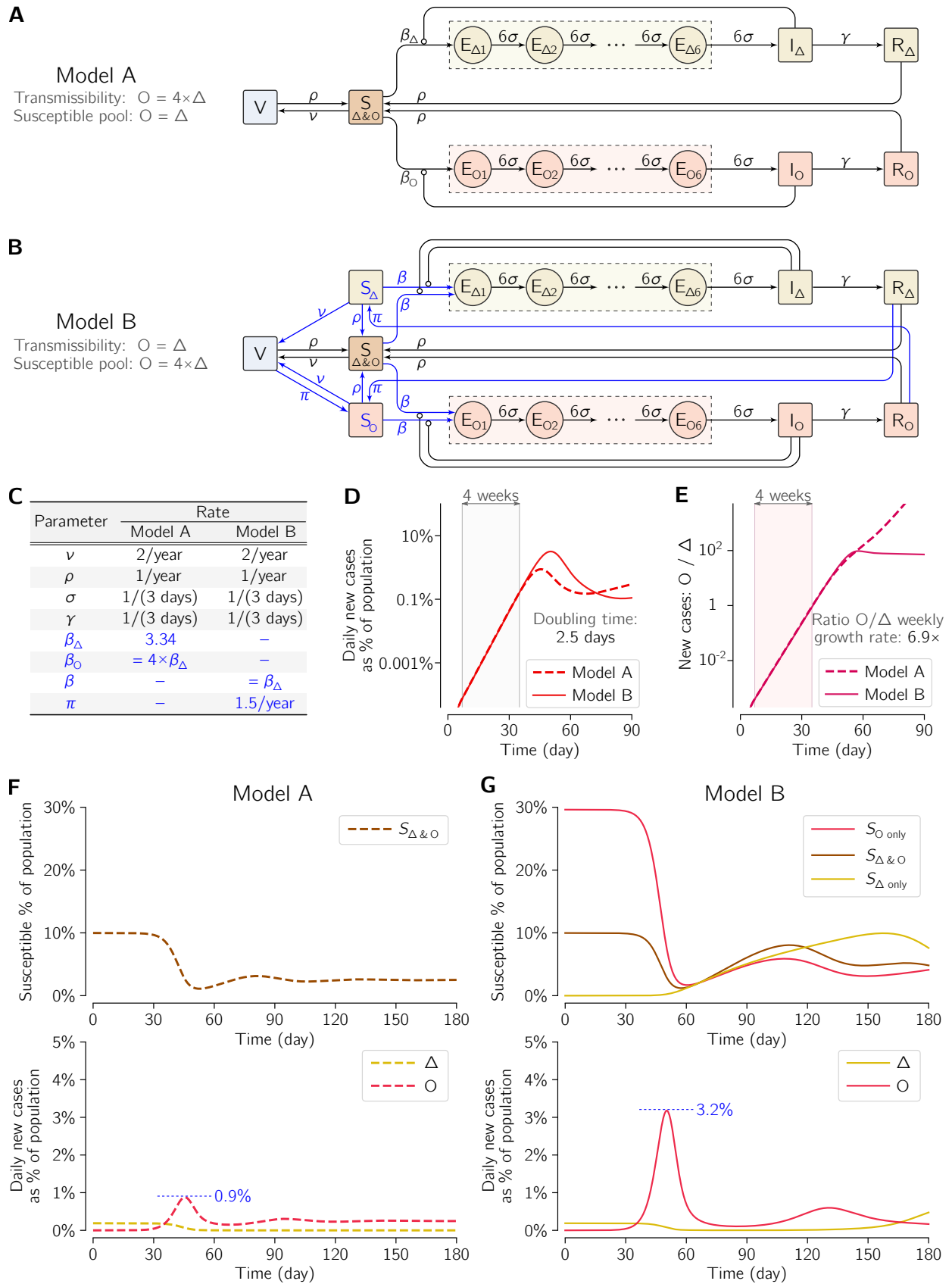


Figure 4. Two COVID-19 two-strain SEIR models with vaccination. (A) Scheme of Model A, in which the transmissibility of Omicron (O) is four-fold higher than that of Delta (Δ) and both strains have a

common pool of susceptible individuals (10% of simulated population in the pre-Omicron steady state). (B) Scheme of Model B, in which the transmissibility of Omicron (O) and Delta (Δ) is the same but the aggregated pool of individuals susceptible to Omicron is four-fold higher than to Delta (40% vs. 10% of simulated population in the pre-Omicron steady state). Essential modifications with respect to Model A are shown in blue. (C) Values of rate parameters of both models. Model variant-specific parameters are blue. (D) Initial dynamics of Omicron dynamics in both models shows similar growth and identical doubling time in the 4-week time window in the initial exponential phase of the epidemic outbreak, but not in later time points. (E) Ratio of Omicron to Delta new daily cases and its growth rate in both models. (F) Dynamics of the outbreak of Omicron infections in Model A (G).

3. Results

3.1. Divergence and Growth of Omicron Strain in South Africa

The Delta VOC became the dominant variant in Gauteng and KwaZulu-Natal in June 2021, causing an epidemic wave that peaked at the beginning of July 2021 in Gauteng and the end of August 2021 in KwaZulu-Natal, Figure 1A. The number of Omicron (or other variant) cases were estimated by multiplying the weekly number of total COVID-19 confirmed cases by the proportion of Omicron (or other variant) genomes among all collected genomes in a given week (see Supplementary Table S1). Between July and October 2021 in Gauteng and in September and October in KwaZulu-Natal, the weekly number of COVID-19 cases was decreasing despite no significant reduction in population mobility at workplaces and retail and recreation centers (Figure 1B). The emergence of the Omicron variant has caused recent rapid epidemic outbreaks in both provinces considered (Figure 1A). In Figure 1C we show the exponential growth of Omicron variant cases in weeks 45–48 in 2021 (8 November–5 December) in Gauteng and in weeks 46–49 in KwaZulu-Natal. This method, in contrast to analyzing only the proportion of new strain genomes [18], enabled us to follow Omicron growth after its fixation in weeks 47–48 in 2021. The Omicron doubling time, estimated based on the log–linear regression of the number of weekly cases in the four-week periods, is equal to 3.3 days (95% CI: 3.2–3.4 days) in Gauteng, and 2.7 days (95% CI: 2.7–3.3 days) in KwaZulu-Natal.

The log–linear regression suggests that the exponential growth of Omicron started in Gauteng around October 11, 2021; however, the initial epidemic growth is highly stochastic and may be heavily disturbed by superspreaders in the cascade of infections [19]. The profile of mutation accumulation in Figure 1D indicates that the Omicron strain started diverging between 6 October and 29 October 2021 (95% CrI), at an average mutation accumulation rate equal 0.33/week (95% CrI: 0.26–0.40 per week). This is lower than the average (global) SARS-CoV-2 mutation accumulation rate equal to approximately 0.45/week (based on the Nextstrain [3] estimate as of 15 January 2021 [20]). An assumption of a higher mutation rate would yield a later divergence date.

3.2. Succession of SARS-CoV-2 Variants of Concern

In Figure 2 we analyze dynamics of Omicron-driven COVID-19 outbreaks in the UK and Denmark, and compare them with outbreaks caused therein by two previous VOCs, Alpha and Delta. In both these countries, the Alpha variant outcompeted pre-existing strains, approaching fixation, and then it was outcompeted by the Delta variant, which also approached fixation; in December 2021, Omicron became the dominant strain (Figure 2A,D). We estimate the doubling time of Omicron to be 2.4 and 2.0 days in the UK and in Denmark, respectively, based on its nearly exponential growth in the 4-week period from 22 November to 19 December 2021, Figure 2B,E. The short doubling time of Omicron is on par with its rapid gain of dominance over the Delta VOC. In the 4-week period considered, the Omicron:Delta ratio increases exponentially with a weekly growth rate of 8.1 in the UK and 10.2 in Denmark. This is faster than the earlier growth of the Delta:Alpha ratio, estimated to be 3.2 in the UK and 4.2 in Denmark, and much faster than

the growth of Alpha with respect to pre-existing strains, which was 2.7 in the UK and 2.0 in Denmark. All data used in Figure 2 are provided in Supplementary Table S5.

In Figure 3 we analogously analyze Omicron-driven outbreaks in New York State and Australia, and again compare them with outbreaks caused by two previous VOCs, Alpha and Delta. In New York State, the Alpha variant arrived during the end phase of the second wave, and consequently caused a noticeable but relatively modest rise in cases. It was then outcompeted by Delta, which, causing the third epidemic wave, quickly reached fixation. In Australia, due to stringent lockdowns and strict border rules, there were relatively few COVID-19 cases before the Delta variant. In both regions, Omicron became the dominant variant within one month after its first detection (Figure 2A,D). We estimate the doubling time of Omicron to be 2.5 and 3.0 days in New York State and Australia, respectively (Figure 2B,E). In the 4-week period considered, the Omicron/Delta ratio was found to grow exponentially with a weekly rate 7.7 in New York State and 7.2 in Australia. This is significantly higher than the Delta/Alpha and Alpha/pre-existing strains growth rates in the New York State, estimated to be 2.5 and 1.8, respectively. All data used in Figure 3 are provided in Supplementary Table S5.

3.3. Two-Strain Mathematical Model

To corroborate whether the observed rapid surge of Omicron cases that displace Delta may be attributed to immune evasion, as widely suggested by the references cited in the Introduction, we analyzed two variants of a mathematical model of the COVID-19 pandemic (Figure 4). In Model A (Figure 4A), transmissibility of Omicron is four-fold higher than that of Delta and both strains share a common pool of susceptible individuals. In Model B (Figure 4B), transmissibility of Omicron and Delta is identical but, in the steady state before the appearance of Omicron, the aggregated pool of individuals susceptible to Omicron is four-fold higher than that of individuals susceptible to Delta. The models were structured and parametrized (Figure 4C) such that the doubling time of Omicron cases in both models is equal to 2.5 days (Figure 4D), which lies roughly in the middle of the range of doubling times observed in six geographical locations considered previously. Moreover, the ratio of new Omicron to new Delta cases is accordant in both models and equal to 6.9 days (Figure 4E), only slightly lower than in the four locations for which this ratio was determined.

Despite having non-distinguishable initial exponential phases, in later time points the two models exhibit divergent trajectories that differ markedly in terms of the peak of the number of daily new Omicron cases. According to the model, with increased Omicron transmissibility (Model A), the maximum number of new daily cases is below 1% of the population (Figure 4F). In the model with the increased pool of individuals susceptible to Omicron (Model B), the initial exponential growth phase lasts longer, which contributes to delaying and, most importantly, elevating this maximum to above 3% of population (Figure 4G).

4. Discussion

We have demonstrated the exponential growth of the Omicron strain in the South African provinces of Gauteng and KwaZulu-Natal in the four-weeks period starting, respectively, on November 8 and 15, 2021, with the doubling times equal to 3.3 days (95% CI: 3.2–3.4 days) and 2.7 days (95% CI: 2.3–3.3 days). Based on the mutation accumulation profile, we found that the Omicron strain started diverging between 6 October and 29 October 2021, which agrees with the date suggested by the log-linear regression of the number of weekly cases in the first affected province of Gauteng, 11 October 2021. Notably, an unnoticed spread before October 2021 would imply that the strain growth rate is lower than that estimated based on the exponential growth rate in the analyzed four-week period in Gauteng. Before the Omicron outbreaks, the Delta variant was the dominant strain in Gauteng and KwaZulu-Natal, and in September and October the COVID-19 epidemic was receding without significant mobility reduction, suggesting that the population of

these provinces might have reached a transient herd immunity to the Delta variant. The population-level immunity has been apparently overcome by the Omicron variant.

The potential of Omicron to create rapid outbreaks was confirmed by analyzing its spread in the UK, Denmark, New York State, and Australia. In these locations the doubling time of Omicron was in the range of 2.0–3.0 days. Additionally, based on the relatively high number of sequenced genomes sampled in these locations, we estimated the weekly growth of the ratio of Omicron to Delta to be in the range of 7.2–10.2, considerably higher than the growth of the ratio of Delta to Alpha (estimated to be in the range of 2.5–4.2), and Alpha to pre-existing strains (estimated to be in the range of 1.8–2.7). These findings are in line with the observed ability of Omicron to infect vaccinated and recovered individuals, which endows it with a natural advantage over Delta [15]. Notably, the Alpha outbreak took place in winter 2020/2021, when the proportion of vaccinated individuals was very low. In turn, the reduction of vaccine effectiveness for Delta in reference to Alpha [11,21] was substantially lower than in the case of Omicron in relation to Delta [15]. This suggests that, in contrast to Omicron, the Alpha and Delta variants become transiently dominant mainly because of their higher infectivity and not due to significant immune evasion.

In the six locations considered, the Omicron doubling time was found to be in the range of 2.0–3.3 days, which is comparable to the doubling times during the first COVID-19 pandemic outbreaks in spring 2020. For that time, doubling time was estimated to lie between 1.86 and 2.88 days for China, Italy, France, Germany, Spain, UK, Switzerland, and New York State [19]. Here, the weekly growth of the ratio of Omicron to Delta was found in the range of 7.2–10.2, considerably higher than the previous ratios of Alpha and Delta at the times when they were gaining dominance. These findings strongly suggest that Omicron will outcompete Delta and become (transiently) the dominant strain.

Omicron accumulated more than 30 mutations in its spike protein, with 15 substitutions in the receptor binding domain (RBD, residues 319–541) alone [2]. Many of these RBD mutations are thought to decrease potency of neutralizing antibodies [22], which is in agreement with growing evidence that Omicron has a several-fold increased ability to infect both vaccinated and recovered individuals, as discussed in the Introduction.

Our mathematical model-based analysis of COVID-19 dynamics clearly demonstrates two points relevant to the initial phase of the outbreak. First, if only rough epidemiological data are analyzed, in the first weeks of the outbreak immune evasion may be indistinguishable from increased transmissibility. Second, being able to distinguish between the two scenarios based on tangential evidence supporting immune evasion is of crucial importance for predicting the impact of a new strain on longer term epidemic dynamics. Immune evasion is more concerning than increased transmissibility, because dodging protection provided by vaccination or infection with prior variant(s) renders a significant share of the population susceptible to an emerging variant, promoting larger outbreaks. These outbreaks may be hard to curb by lockdowns due to increasing lockdown fatigue, but in the case of Omicron plausibly will not result in proportionally high death toll, as suggested by early estimates of Omicron-associated mortality [5,23].

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/v14020294/s1>. Supplementary Table S1: Data used for Figure 1A,C; Supplementary Table S2: containing phylogenetic tree with dates; Supplementary Table S3: list of GISAID IDs and acknowledgments for genomes collected in Gauteng since 19 April 2021 till 2 January 2022; Supplementary Table S4: list of GISAID IDs and acknowledgments for genomes collected in KwaZulu-Natal since 19 April 2021 till 2 January 2022; Supplementary Table S5: Data used for Figures 2 and 3.

Author Contributions: Conceptualization, F.G., M.K. and T.L.; data curation, F.G.; software, F.G.; investigation, F.G. and M.K.; supervision, M.K. and T.L.; visualization, F.G. and M.K.; writing—original draft preparation, T.L.; writing—review and editing, M.K. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the National Science Centre Poland OPUS, grant 2018/29/B/NZ2/00668 and by the Norwegian Financial Mechanism GRIEG-1 grant 2019/34/H/NZ6/00699 (operated by the National Science Centre Poland). The funding agency had no role in the study design, data collection, analysis, decision to publish, or preparation of the manuscript.

Data Availability Statement: We analyzed publicly available genome datasets retrieved from GISAID (<https://www.gisaid.org>, accessed on 9 January 2022) and Community Mobility Reports available from Google (<https://www.google.com/covid19/mobility/>, accessed on 9 January 2022). Cases data for Gauteng and KwaZulu-Natal can be found in daily reports from the National Institute for Communicable Diseases (<https://www.nicd.ac.za>, accessed on 9 January 2022) as aggregated by the University of Pretoria (<https://github.com/dsfsi/covid19za>, accessed on 9 January 2022). Cases for all other regions can be found in data aggregated by Johns Hopkins University (<https://github.com/CSSEGISandData/COVID-19>, accessed on 9 January 2022). Mutation counts for Omicron sequences can be found on Nextstrain (<https://nextstrain.org/groups/nehrlab/ncov/21K-diversity>, accessed on 9 January 2022).

Acknowledgments: We are very grateful to the GISAID Initiative and all its data contributors, i.e., the authors from the originating laboratories responsible for obtaining the specimens and the submitting laboratories where genetic sequence data were generated and shared via the GISAID Initiative, on which this research is based.

Conflicts of Interest: The authors have no competing interest.

References

- Shu, Y.; McCauley, J. GISAID: Global initiative on sharing all influenza data—from vision to reality. *Eurosurveillance* **2017**, *22*, 30494. [CrossRef] [PubMed]
- Implications of the Further Emergence and Spread of the SARS-CoV-2 B.1.1.529 Variant of Concern (Omicron) for the EU/EEA—First Update. Available online: <https://www.ecdc.europa.eu/en/publications-data/covid-19-threat-assessment-spread-omicron-first-update> (accessed on 2 December 2021).
- Hadfield, J.; Megill, C.; Bell, S.M.; Huddleston, J.; Potter, B.; Callender, C.; Sagulenko, P.; Bedford, T.; Neher, R.A. Nextstrain: Real-time tracking of pathogen evolution. *Bioinformatics* **2018**, *24*, 4121–4123. [CrossRef] [PubMed]
- Kupferschmidt, K. Where did ‘weird’ Omicron come from? *Science* **2021**, *374*, 1179. [CrossRef] [PubMed]
- Madhi, S.A. South African Population Immunity and Severe Covid-19 with Omicron Variant. *medRxiv* **2021**. [CrossRef]
- Ritchie, H.; Mathieu, E.; Rodés-Guirao, L.; Appel, C.; Giattino, C.; Ortiz-Ospina, E.; Hasell, J.; Macdonald, B.; Beltekian, D.; Roser, M. Coronavirus Pandemic (COVID-19). 2020. Available online: <https://ourworldindata.org/coronavirus> (accessed on 15 January 2022).
- Greaney, A.J.; Loes, A.N.; Crawford, K.H.D.; Starr, T.N.; Malone, K.D.; Chu, H.Y.; Bloom, J.D. Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell Host Microbe* **2022**, *29*, 463–476.e6. [CrossRef] [PubMed]
- Liu, L.; Iketani, S.; Guo, Y.; Chan, J.F.; Wang, M.; Liu, L.; Luo, Y.; Chu, H.; Huang, Y.; Nair, M.S.; et al. Striking Antibody Evasion Manifested by the Omicron Variant of SARS-CoV-2. *Nature* **2021**, 1–8. [CrossRef] [PubMed]
- Planas, D.; Saunders, N.; Maes, P.; Benhassine, F.G.; Planchais, C.; Porrot, F.; Staropoli, I.; Lemoine, F.; Pere, H.; Veyer, D.; et al. Considerable escape of SARS-CoV-2 variant Omicron to antibody neutralization. *bioRxiv* **2021**. [CrossRef]
- Doria-Rose, N.A.; Shen, X.; Schmidt, S.D.; O’Dell, S.; McDanal, C.; Feng, W.; Tong, J.; Eaton, A.; Maglinao, M.; Tang, H.; et al. Booster of mRNA-1273 Strengthens SARS-CoV-2 Omicron Neutralization. *medRxiv* **2021**. [CrossRef]
- Cromer, D.; Steain, M.; Reynaldi, A.; Schlub, T.E.; Wheatley, A.K.; Juno, J.A.; Kent, S.J.; Triccas, J.A.; Khoury, D.S.; Davenport, M.P. Neutralising antibody titres as predictors of protection against SARS-CoV-2 variants and the impact of boosting: A meta-analysis. *Lancet Microbe* **2021**, *3*, e52–e61. [CrossRef]
- Garcia-Beltran, W.F.; Denis, K.J.S.; Hoelzemer, A.; Lam, E.C.; Nitido, A.D.; Sheehan, M.L.; Berrios, C.; Ofoman, O.; Chang, C.C.; Hauser, B.M.; et al. mRNA-based COVID-19 vaccine boosters induce neutralizing immunity against SARS-CoV-2 Omicron variant. *Cell* **2022**. [CrossRef] [PubMed]
- Andrews, N.; Stowe, J.; Kirsebom, F.; Toffa, S.; Rickeard, T.; Gallagher, E.; Gower, C.; Kall, M.; Groves, N.; O’Connell, A.M.; et al. Effectiveness of COVID-19 vaccines against the Omicron (B.1.1.529) variant of concern. *medRxiv* **2021**. [CrossRef]
- SARS-CoV-2 Variants of Concern and Variants under Investigation in England Technical Briefing 33. Available online: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1043807/technical-briefing-33.pdf (accessed on 15 January 2022).
- Report 49: Growth, Population Distribution and Immune Escape of Omicron in England. Available online: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1041911/S1449_COVID19_Report_49.pdf (accessed on 15 January 2022).

16. Marivate, V.; Combrin, H.M. Use of Available Data to Inform The COVID-19 Outbreak in South Africa: A Case Study. *Data Sci. J.* **2020**, *19*, 1–7. [[CrossRef](#)]
17. Dong, E.; Du, H.; Gardner, L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* **2020**, *20*, 533–534. [[CrossRef](#)]
18. Grabowski, F.; Preibisch, G.; Giziński, S.; Kocharczyk, M.; Lipniacki, T. SARS-CoV-2 Variant of Concern 202012/01 Has about Twofold Replicative Advantage and Acquires Concerning Mutations. *Viruses* **2021**, *13*, 392. [[CrossRef](#)] [[PubMed](#)]
19. Kocharczyk, M.; Grabowski, F.; Lipniacki, T. Super-spreading events initiated the exponential growth phase of COVID-19 with R_0 higher than initially estimated. *R. Soc. Open Sci.* **2020**, *7*, 200786. [[CrossRef](#)] [[PubMed](#)]
20. Nextstrain Clock. Available online: <https://nextstrain.org/ncov/gisaid/global?l=clock> (accessed on 15 January 2021).
21. Bernal, J.L.; Andrews, N.; Gower, C.; Gallagher, E.; Simmons, R.; Thelwall, S.; Julia, S.; Elise, T.; Natalie, G.; Gavin, D.; et al. Effectiveness of Covid-19 Vaccines against the B.1.617.2 (Delta) Variant. *N. Engl. J. Med.* **2021**, *385*, 585–594. [[CrossRef](#)] [[PubMed](#)]
22. Callaway, E.; Ledford, H. How bad is Omicron? What scientists know so far. *Nature* **2021**, *600*, 197–199. [[CrossRef](#)] [[PubMed](#)]
23. Lewnard, J.A.; Hong, V.X.; Patel, M.M.; Kahn, R.; Lipsitch, M.; Tartof, S.Y. Clinical outcomes among patients infected with Omicron (B.1.1.529) SARS-CoV-2 variant in southern California. *medRxiv* **2022**. [[CrossRef](#)]

Research



Cite this article: Grabowski F, Czyż P, Kocharczyk M, Lipniacki T. 2019 Limits to the rate of information transmission through the MAPK pathway. *J. R. Soc. Interface* **16**: 20180792.
<http://dx.doi.org/10.1098/rsif.2018.0792>

Received: 24 October 2018

Accepted: 13 February 2019

Subject Category:

Life Sciences – Mathematics interface

Subject Areas:

computational biology, systems biology

Keywords:

cellular signal transduction, pulsatile stimulation, pulse – interval transcoding, bandwidth, representation problem

Authors for correspondence:

Marek Kocharczyk

e-mail: mkochan@ippt.pan.pl

Tomasz Lipniacki

e-mail: tlipnia@ippt.pan.pl

[†]These authors contributed equally to the study.

Electronic supplementary material is available online at <https://dx.doi.org/10.6084/m9.figshare.c.4407791>.

Limits to the rate of information transmission through the MAPK pathway

Frederic Grabowski^{1,†}, Paweł Czyż^{2,†}, Marek Kocharczyk³
and Tomasz Lipniacki³

¹Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, Warsaw, Poland

²Mathematical, Physical and Life Sciences Division, University of Oxford, Oxford, UK

³Institute of Fundamental Technological Research, Polish Academy of Sciences, Warsaw, Poland

id FG, 0000-0003-4070-9500; PC, 0000-0002-6262-0614; MK, 0000-0003-1215-3920; TL, 0000-0002-3488-2561

Two important signalling pathways of NF-κB and ERK transmit merely 1 bit of information about the level of extracellular stimulation. It is thus unclear how such systems can coordinate complex cell responses to external cues. We analyse information transmission in the MAPK/ERK pathway that converts both constant and pulsatile EGF stimulation into pulses of ERK activity. Based on an experimentally verified computational model, we demonstrate that, when input consists of sequences of EGF pulses, transmitted information increases nearly linearly with time. Thus, pulse-interval transcoding allows more information to be relayed than the amplitude–amplitude transcoding considered previously for the ERK and NF-κB pathways. Moreover, the information channel capacity C , or simply bitrate, is not limited by the bandwidth $B = 1/\tau$, where $\tau \approx 1$ h is the relaxation time. Specifically, when the input is provided in the form of sequences of short binary EGF pulses separated by intervals that are multiples of τ/n (but not shorter than τ), then for $n = 2$, $C \approx 1.39$ bit h^{-1} ; and for $n = 4$, $C \approx 1.86$ bit h^{-1} . The capability to respond to random sequences of EGF pulses enables cells to propagate spontaneous ERK activity waves across tissue.

1. Introduction

Cells secrete and recognize chemical signals in order to coordinate their actions at a supercellular level. In addition to chemical identity, quantifiable features of the signal such as amplitude or duration are usually relevant for eliciting a proportionate physiological cell response. The amplitude of an input signal can be translated into the amplitude or duration of activity of an effector protein such as a transcription factor. It has been shown for the nuclear factor kappa–light chain enhancer of activated B cells (NF-κB) system that the response in a population of cells is proportional to the product of signal amplitude and its duration [1]. Some signalling pathways that exhibit oscillatory behaviour transcode the input amplitude to the frequency of effector pulses [2]. Such an ability has been recently demonstrated in the mitogen-activated protein kinase (MAPK) pathway, where, in response to extracellular epidermal growth factor (EGF) (input) of constant amplitude, extracellular signal-regulated kinase (ERK) (effector) exhibits activity pulses with constant amplitude but frequency and pulse duration determined by concentrations of EGF [3]. In the MAPK system, the amplitude-to-frequency transcoding is enabled by a particular topology of feedback loops and associated time scales [4].

Molecular interpretation of pulsatile (or more complex) temporal codes of effector molecules is an area of active research [5–7]; however, less attention has been devoted to the characterization of responses of signalling systems challenged with time-varying inputs. It has been demonstrated that both the NF-κB and the MAPK/ERK pathways respond in a pulsating manner to periodic pulses of cytokines, tumour necrosis factor (TNF) and EGF, respectively [8,9]. Experimental and computational analysis of responses to pulsating inputs has shown that the

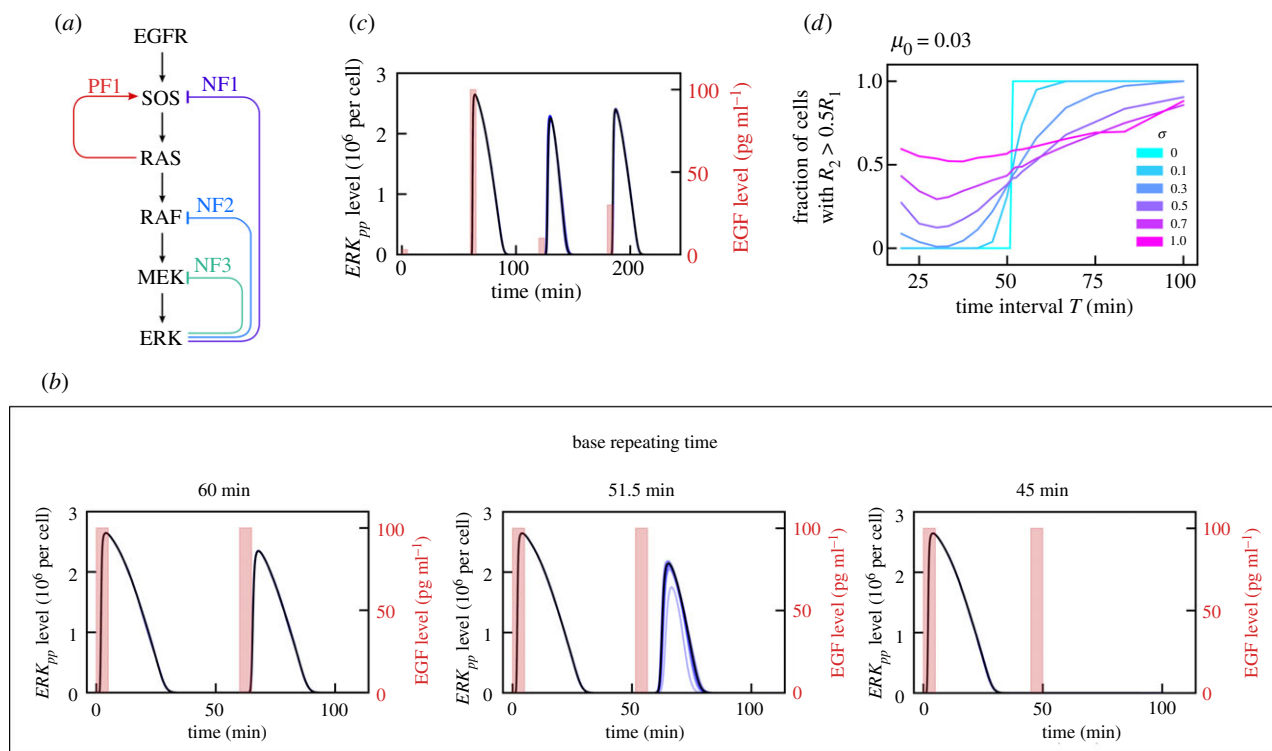


Figure 1. The MAPK pathway and analysis of ERK responses to repeated EGF pulses. (a) Schematic diagram of the MAPK pathway. As analysed in [4], PF1 and NF1 lead to relaxation oscillations, whereas NF2 and NF3 shape the time profile of the pulse of active ERK (ERK_{pp}). (b) $ERK_{pp}(t)$ profile in response to two 5 min long 100 pg ml^{-1} EGF pulses separated by $T = 45, 51.5, 60$ min. Black bold line: deterministic model trajectory; thin blue lines: 10 stochastic trajectories (visible only for $T = 51.5$ min). (c) $ERK_{pp}(t)$ profile in response to four EGF pulses repeated every 60 min with successive amplitudes 3, 100, 10, 30 pg ml^{-1} . (d) Fraction of simulated cells for which $R_2 > 0.5R_1$, as a function of the time interval T between EGF pulses. The fraction is calculated based on at least 500 independent numerical simulations (number of simulations for T close to τ was up to 5000) performed assuming lognormal distributions $\log N(\mu_i, \sigma^2)$ of protein levels (and first-order reaction coefficients) with mean μ_i equal to the default value of a given variable, and six different values of σ . EGFR, epidermal growth factor receptor; MEK, mitogen-activated protein kinase kinase; NF, negative feedback; PF, positive feedback; SOS, son of sevenless. (Online version in colour.)

ability of the NF- κ B system to respond to a series of TNF pulses is inherently limited by the relaxation time τ associated with oscillation-generating negative feedback [10,11]. Information transmission through both the NF- κ B and the MAPK/ERK pathways has been analysed within the generic framework of information theory [12,13], revealing that these pathways are able to transmit about 1 bit of information about a constant stimulus concentration [14–17].

In this work, we present a computational analysis of information relay within an experimentally verified model of the MAPK/ERK pathway [4], calibrated on MCF10A cells [3], where a fast positive feedback loop (involving RAS and SOS) is nested within a slow negative feedback loop (involving ERK and SOS; see figure 1a). Such network topology allows both to transcode constant EGF stimulation to periodic ERK activity pulses and to transcode analogous EGF pulses to nearly digital pulses of ERK activity. Based on the model, we approach the following representation problem: What is the signal transcoding strategy that allows the maximum amount of information to be transmitted through a given transmission channel in a given time span or to achieve the highest information transmission rate in the long term? We propose to code input information in the form of sequences of short digital EGF pulses. Then, we estimate the maximum transmission rate with respect to the noise strength. In particular, we demonstrate that when input information is encoded in intervals between subsequent EGF pulses, then the information channel capacity, computed as the maximum mutual information (MI) rate, may exceed channel bandwidth, $B = 1/\tau$. By theoretical and numerical analysis of transmission of binary EGF sequences with varying inter-pulse intervals we demonstrate

that the MAPK/ERK pathway is able to transmit more than 1 bit per hour.

2. Methods

2.1. Parameters and simulations

To perform our analyses, we employed the computational model described in [4], which has been amended with extrinsic noise and perturbed according to stimulation protocols, characterized by the following parameters:

- pulse duration: 5 min, pulse amplitude: 0 or 100 pg ml^{-1} (unless otherwise specified);
- (square of) the second parameter of the lognormal distribution describing additive noise: $\sigma_0 = 1.0$;
- (square of) the second parameter of the lognormal distribution describing cell-specific noise: $\sigma = 0, 0.1, 0.3, 0.5, 0.7, 1.0$;
- basic term to determine the first parameter of the lognormal distribution describing cell-specific noise: $\mu_0 = \{0, 0.01, 0.03, 0.1, 0.3\}$;
- number of pulses: 6 or 8, inter-pulse interval: $T = 15, 20, 30, 40, 50, 60, 70, 80$ min;
- number of simulations for a given input sequence of pulses: $M = 1000$.

All parameters of the MAPK model are provided in electronic supplementary material, table S1. Abundances of molecules of each pathway component were drawn from lognormal distributions $\log N(\mu_i, \sigma)$ with median μ_i equal to a default value for this component. Also, all pseudo-first-order reaction parameters, which are proportional to the level of an implicit enzyme, were independently drawn from a lognormal distribution. These randomly set parameters of the MAPK pathway are written in

green in electronic supplementary material, table S1. For each considered combination of inter-pulse interval T , noise strength σ and specific input sequence of pulses, we generated M cell-specific random sets of parameters (as described above) and simulated model dynamics of each cell deterministically using an adaptive ODE solver embedded in BioNetGen [18]. To simulate the effect of the additive noise of different strengths, we post-processed the obtained M trajectories of $ERK_{pp}(t)$.

2.2. Mutual information calculation and maximization

Estimation of the MI defined in equation (3.2) has been performed as proposed by Kraskov *et al.* [19]. In this approach (see [19], eqn (8)), effectively the output probability distributions, Y_i , and marginal distribution, Y , are estimated based on the distance to the k th nearest neighbour and on the number of points within this distance. For all estimates described in this article, we have set the number of nearest neighbours $k = 15$, and assumed that distances in input space X are much larger than those in Y space (which implies that in eqn (8) from [19] we set $n_x = M$). We sampled multi-dimensional output distributions in $M = 1000$ d -dimensional points obtained by running M simulations, each with parameters independently drawn from lognormal distributions as described in §2.1. The number of dimensions d was 2, 3, 4, 5, 6 or 8, equal to the number of pulses (or pseudopulses) of the considered sequences. The number of distinct binary sequences was 2^d . From each sequence we obtained d numbers quantifying the ERK response to d subsequent pulses; see equation (3.1). We implemented the method of Kraskov *et al.* [19] such that for each input category, x , the counts of nearest neighbours from all other categories, $x' \neq x$, are weighted according to current $p(x')$, enabling iterative constrained maximization of the estimated MI over the complete set of input probabilities. To obtain estimates of the maximum MI (channel capacity) and $p(x)$ associated with inputs, we employed a stochastic gradient optimizer, Adam [20], available in TensorFlow [21]. Our implementation of the channel capacity estimator is provided in the form of a Python module called CCE (see <http://pmbm.ippt.pan.pl/software/cce>). The accuracy of our implementation of the Kraskov algorithm is checked within unit tests, where we calculate MI numerically based on eight three-dimensional Gaussian distributions either (1) centred along an S-shaped curve or (2) centred in vertices of the cube. In both cases the distributions are overlapping so MI is about $1.8 < \log_2(8) = 3$ bit. These nearly accurate values can be then compared with the values obtained using the Kraskov algorithm based on M points drawn at random from each distribution. The Kraskov algorithm overestimates MI by about 2% for $M = 1000$ and 1.5% for $M = 5000$; in the second case, 3.5% for $M = 1000$ and 2.5% for $M = 5000$. Accuracy is nearly the same in the case when all eight input probabilities are equal (test_use_case_1.py) and in the case when input probabilities are varied to maximize MI (test_use_case_3.py).

3. Results

3.1. ERK responses to the repeated EGF pulses of different frequencies and amplitudes

The ability of a cell to respond to EGF pulses depends on their amplitude, duration and pulsing frequency. We focus on relatively short pulses of 5 min. In figure 1b we show trajectories resulting from one deterministic (black bold line) and 10 stochastic (thin blue lines) simulations of the MAPK model in response to two EGF pulses, both of amplitude 100 pg ml^{-1} , separated by three different time intervals T of 45, 51.5 and 60 min. Let $R_i(T)$ denote the integrated ERK response after the i th EGF pulse:

$$R_i(T) = \int_{iT}^{(i+1)T} ERK_{pp}(t) dt. \quad (3.1)$$

Since the ERK_{pp} pulse lasts about 30 min, for $T > 30$ min, $R_1(T) = \text{const} = : R_{\max}$ with the numerically calculated $R_{\max} = 2.65 \times 10^9 \text{ molecule} \times \text{s}$. Let us consider the integrated ERK responses to the first and the second EGF pulse. For $T = 45$ min, the response to the second pulse, R_2 , is negligible with respect to R_1 , $R_2(45)/R_1(45) \ll 1$, while for $T = 51.5$ min, $R_2(51.5)/R_1(51.5) = 0.5$. For $T = 60$ min, $R_2 = 2.00 \times 10^9 \text{ molecule} \times \text{s}$ is comparable to R_1 . This indicates a threshold $\tau = 51.5$ min above (below) which the system can (cannot) be activated in response to the next pulse. It is of note that this threshold time (relaxation time) τ is very close to the minimal period of oscillations under constant EGF stimulation observed experimentally [3].

In figure 1c, we analyse the ERK activity time profile in response to EGF pulses repeated every 60 min, with successive amplitudes 3, 100, 10, 30 pg ml^{-1} . While there is no response to the 3 pg ml^{-1} pulse, the responses to the remaining three pulses have comparable amplitudes, even though the 10 pg ml^{-1} pulse follows the much stronger 100 pg ml^{-1} pulse, which has an inhibitory effect. Overall, the results presented in figure 1b,c show that the MAPK system exhibits nearly all-or-nothing ERK responses to the pulsatile stimulation. We thus expect that, for such stimulation, information is not coded by the level of EGF, but by the presence or absence of an EGF pulse that exceeds some threshold. Consequently, in the further analysis we restrict ourselves to 5 min long EGF pulses of amplitude 100 pg ml^{-1} .

We notice that the stochastic trajectories in figure 1b,c are nearly indistinguishable from the deterministic one except for trajectories with $T = 51.5$ min in figure 1b. This is a consequence of a large number of reacting molecules at each step of the pathway, which makes the intrinsic noise of signal processing negligible. When comparing model predictions with experiment we found that, to reproduce experimentally observed heterogeneity of single cell responses, one has to include extrinsic noise, that is, assume that the levels of the MAPK pathway components vary among cells [4]. Since intrinsic noise was found to be negligible, henceforth we will use a deterministic approximation (ODEs). In order to analyse how extrinsic noise influences the transmitted information, we will consider two noise sources, as follows.

- Cell-specific noise associated with uncertain levels of pathway components (in individual simulated cells) and with pseudo-first-order rates of dephosphorylation reactions (by implicit phosphatases). We will assume that these variables follow the lognormal distributions $\log N(\mu_i, \sigma^2)$ with median μ_i equal to the default value of an i th parameter.
- Additive noise associated with ERK activation mediated by pathways that are not included in the model. For the additive noise we will also assume a lognormal distribution $\log N(\mu^*, \sigma_0^2)$ with $\sigma_0 = 1$ and $\mu^* = \mu_0 \times R_{\max} \times T/(60 \text{ min})$. For most of the analysis we take $\mu_0 = 0.03$, which is equivalent to the assumption that about 3% of ERK activity results from the additive noise.

The parameters σ and μ_0 that characterize the strength of the cell-specific and the additive noise will be varied. The MAPK channel information capacity (or simply bitrate) is controlled by the maximal frequency of EGF pulses that can induce ERK activation. For figure 1d we perform similar simulations to those for figure 1b, but accounting for the extrinsic noise. We

compute the fraction of cells for which the response to the second pulse is at least half of the response to the first pulse, as a function of the time interval T between EGF pulses, for several values of σ (keeping $\mu_0 = 0.03$). For $\sigma = 0$, all cells exhibit significant responses to the second pulse when $T > \tau$ and none of them exhibit such responses when $T < \tau$. For $\sigma = 0.1$ and $\sigma = 0.3$ the significant responses to the second EGF pulse are reached for T substantially larger than τ ; additionally, even for $T < \tau$, a fraction of cells exhibit significant responses to the second EGF pulse. For larger noise, $\sigma \geq 0.5$, regardless of T , there are cells that exhibit a significant response to the second pulse and cells that fail to respond. In summary, based on the above analysis, we may expect that, for zero or small noise, EGF pulses separated by intervals $T > \tau$ will lead to a significant ERK activation, whereas, for larger noise, EGF pulses will be missed in some cells even for $T > \tau$, but some other cells will show responses for $T < \tau$.

3.2. ERK activation in response to sequences of three EGF pulses

In figure 2 we show responses to eight sequences of EGF pulses. We consider three base repeating times T_i (time intervals between repeated pulses) of 60, 30 or 20 min. In the following convention the sequence '101' for $T = T_1 = 60$ min denotes a sequence of three EGF pulses of successive amplitudes: 1 (in dimensional units $1 \times 100 \text{ pg ml}^{-1}$) at $t = 0$ min, 0 (a 'pseudo-pulse') at $t = 60$ min, and 1 at $t = 120$ min. For $T = T_2 = 30$ min, the same sequence denotes pulses of amplitude 1 at $t = 0$ min and $t = 60$ min, and a pseudo-pulse at $t = 30$ min.

For illustration purposes, times T_i are chosen so that $T_1 > \tau$, $T_2 < \tau < 2T_2$, and $2T_3 < \tau < 3T_3$. For each of the eight sequences and each T_i , we perform 30 simulations (each corresponding to a single cell) with $\sigma = 0.3$, $\mu_0 = 0.03$; the resulting trajectories of $\text{ERK}_{pp}(t)$ are shown in figure 2. Since the noise is moderate, one can expect that for $T = 60$ min most cells will respond to all EGF pulses (by looking closely one can spot that only one and two cells show a very weak response to the second pulse in sequences '110' and '111', respectively). One can thus say that for $T = 60$ min almost all EGF sequences are transmitted properly.

For $T = 30$ min, cells are not able to properly transmit EGF sequences '011', '110' and '111'. Sequence '011' leads only to a single ERK_{pp} pulse—in essence producing the same response as sequence '010'. We will say that these two input EGF sequences lead to a '010' ERK_{pp} response. In some simplification, the distinction between '0' and '1' of the ERK_{pp} response is based on integrated ERK_{pp} . When the integral exceeds some threshold, the ERK_{pp} response is interpreted as '1'. The threshold however is not predefined, but rather set by likelihood. Above the threshold the probability that the input was '1' is higher than the probability that the input was '0'. EGF sequence '110' is interpreted as '100', whereas EGF sequence '111' is interpreted as '101'. In the last case—because the second pulse does not elicit ERK activation—the third pulse can be transmitted. Overall, for $T = 30$ min, one can distinguish two EGF sequences, '000' and '100', that are accurately transmitted (i.e. produce '000' and '100' ERK_{pp} response, respectively), and three groups {'010', '011'}, {'100', '110'} and {'101', '111'} (grouped in bold frames in figure 2) that are transcoded, respectively, to ERK_{pp} sequences '010', '100' and '101'. The element of the group which is accurately transmitted (i.e. here '010', '100' and '101') will be

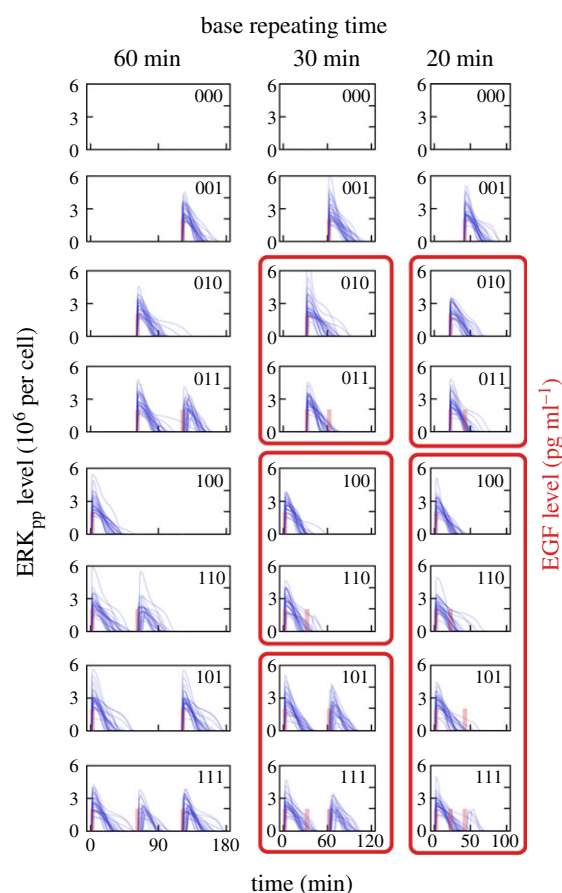


Figure 2. Responses to eight possible series of three pulses having amplitude of either '0' ('pseudo-pulse') or '1' (100 pg ml^{-1} in dimensional units). For each base repeating time T_i (60, 30 or 20 min) and each pulse sequence, 30 independent numerical simulations were performed with $\sigma = 0.3$, $\mu_0 = 0.03$. For $T = 30$ min and $T = 20$ min, the sequences transmitted to the same output sequence are grouped and encircled by bold frames. The first element of each group is accurately transmitted and will be referred to as a group representative. (Online version in colour.)

referred to as the group representative. For $T = 20$ min, the analysis is analogous: there are also two sequences that are accurately transmitted, but only two groups—one containing two EGF sequences, the other containing four EGF sequences. Because an ERK_{pp} pulse at the first position inhibits signal transmission for about 50 min, EGF sequences '100', '110', '101' and '111' are all transcribed to '100'.

Transmitted information is limited from above by $\log_2(K)$, where K is the number of distinct output sequences. In our case, $K = 8$ for $T = 60$ min, $K = 5$ for $T = 30$ min and $K = 4$ for $T = 20$ min. Thus, transmitted information (or MI) is limited from above by, respectively, $\log_2(8) = 3$, $\log_2(5) = 2.32$ and $\log_2(4) = 2$. The estimated bitrate, that is, $\text{MI}/(3T_i)$, equals, respectively, 1 bit h^{-1} , 1.55 bit h^{-1} and 2 bit h^{-1} . As we can see, the bitrate is highest when the base repeating time is 20 min, which is only a fraction of the relaxation time τ . We will see that this observation remains true also for longer sequences and that this simple theoretical estimates agree well with numerical values obtained for small noise.

3.3. Quantitative estimation of transmitted information

In order to estimate the bitrate numerically we will calculate the MI between input and output sequences. The inputs $x \in X$ are EGF binary sequences of length L . We have performed simulations for $L = 8$ and $L = 6$. For each input sequence we

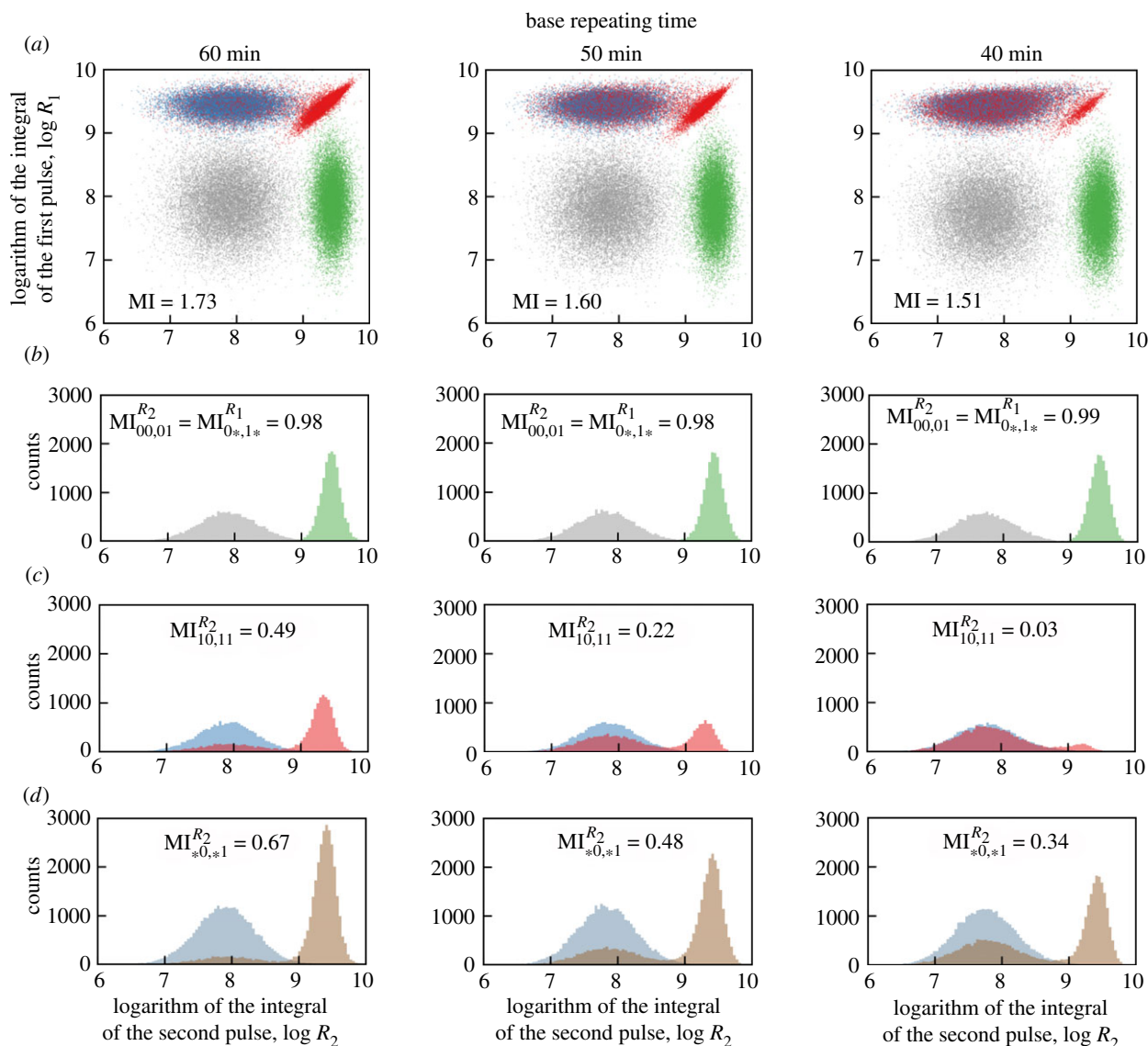


Figure 3. ERK responses to four binary sequences: 00, 01, 10 and 11 for three base repeating times. (a) Scatter plots of R_1, R_2 for all sequences (grey 00, green 01, blue 10, red 11), based on 16 000 simulations for each sequence. (b) Histograms of R_2 when the first pulse has amplitude 0 (grey 00, green 01). (c) Histograms of R_2 when the first pulse has amplitude 1 (blue 10, red 11). (d) Histograms of R_2 independently of the first pulse amplitude (blue *0, brown *1). In all scatter plots, MI denotes the mutual information calculated for two-dimensional distributions; in all histograms, MI denotes the mutual information calculated based on two shown one-dimensional distributions.

performed $M = 1000$ simulations obtaining M response vectors $y = (R_1, \dots, R_L) \in Y$, where R_i are defined as in equation (3.1) with the additive noise included. These simulations probe a (continuous) output probability distribution Y . The analysis of short sequences ($L < 6$) is based on truncated simulation data obtained for $L = 6$. In this way we gained larger statistics; for example, for $L = 2$ (figure 3), we have 16 000 simulations for each of four binary input sequences. In our case the domain of input is discrete, so MI can be expressed as

$$MI(X; Y) = \sum_{x \in X} p(x) \int_{y \in Y} p(y|x) \log \left(\frac{p(y|x)}{p(y)} \right) dy. \quad (3.2)$$

To calculate MI, one has to assume or determine input sequence probabilities $p(x)$. We proceed as follows:

- (1) assume that all input sequences have equal probabilities, or
- (2) assume that all group representatives have equal probabilities, while the other sequences have zero probabilities, or
- (3) calculate input probabilities that maximize MI.

3.4. Numerical estimation of the transmitted information in sequences of two pulses

In figure 3 we analyse the ability of the system to resolve binary sequences '00', '01', '10' and '11' for three base repeating times $T_i = 60, 50$ and 40 min, for moderate noise ($\sigma = 0.3$, $\mu_0 = 0.03$). In figure 3a we present scatter plots in the (R_1, R_2) -plane showing responses to the four binary sequences with corresponding MI values. The scatter plot for $T = 40$ min shows that for this time interval the system is essentially able to transmit three input sequences, '00', '01' and '10', while the sequence '11' produces almost the same output as sequence '10'. Consequently, $MI = 1.51 < \log_2(3) \approx 1.58$. For $T = 50$ min and $T = 60$ min, the system distinguishes '10' and '11' sequences to a certain extent and MI values (equal, respectively, to 1.60 and 1.73) exceed $\log_2(3)$ but are still substantially lower than the upper bound $MI_{\max} = \log_2(4) = 2$.

We further investigate the system's inability to transmit four sequences by detailed analysis of responses to the second pulse. When the amplitude of the first pulse is 0,

Table 1. Information transmitted in $L = 4$ pulses with intervals $T = 60, 30, 20, 15$ min. Comparison of numerical and theoretical estimates for small noise, $\sigma = 0.1$ and $\mu_0 = 0.03$.

T (min)	theoretical MI value		numerical MI value, $\sigma = 0.1$ and $\mu_0 = 0.03$			maximum and minimum values of group probabilities together with the group representative ^a
	equal probabilities of input sequences	maximized $\log_2(K)$	numerically maximized	equal probabilities of all group representatives	equal probabilities of all input sequences	
60	$\log_2(16) = 4$	$\log_2(16) = 4$	3.93	3.93	3.93	'0001': 0.065; '0110': 0.060
30	$4 - 9/8 = 2.875$	$\log_2(8) = 3$	2.96	2.96	2.82	'0010': 0.129; '1010': 0.120
20	$4 - 13/8 = 2.375$	$\log_2(6) = 2.58$	2.55	2.55	2.33	'0100': 0.171; '1001': 0.161
15	$4 - 17/8 = 1.875$	$\log_2(5) = 2.32$	2.32	2.32	1.89	'1000': 0.202; '0000': 0.195

^aData obtained via numerical MI maximization.

then, regardless of T_i , the system almost perfectly distinguishes between amplitudes 0 and 1 of the second pulse (figure 3b). The value $MI_{00,01}^{R_2}$ (given on each histogram in figure 3b) is calculated between the inputs $X = \{ '00', '01' \}$ (assuming that both sequences are equiprobable) and the output to the second pulse, R_2 . For each T_i , $MI_{00,01}^{R_2} \geq 0.98$ ($MI_{00,01}^{R_2} = 1$ means that both inputs are perfectly distinguishable). In the case when the amplitude of the first pulse is 1, the ability to distinguish between amplitudes 0 and 1 of the second pulse is much lower (figure 3c): for $T = 60$ min, $MI_{10,11}^{R_2} = 0.49$; for $T = 50$ min, $MI_{10,11}^{R_2} = 0.22$; and for $T = 40$ min, $MI_{10,11}^{R_2} = 0.03$; the nearly zero value for $T = 40$ min is a consequence of the fact that the system cannot respond to two EGF pulses with a time interval of 40 min. In figure 3d we show histograms of the marginal probability distributions of responses to the second pulse. The $MI_{*,0,*1}^{R_2}$ values shown in these histograms measure the ability to distinguish between amplitudes 0 and 1 of the second pulse when the first pulse is unknown.

Let us notice that two-dimensional MI is greater than the sum of unidimensional MIs, $MI_{0*,1*}^{R_1} + MI_{*,0,*1}^{R_2}$. Here $MI_{0*,1*}^{R_1} = MI_{00,01}^{R_2}$, because the response to the second pulse in the case when the first pulse has amplitude 0 is identical to the response to the first pulse—regardless of the amplitude of the second pulse. This implies that information that can be inferred from two pulses together is higher than information inferred from the first and the second pulse separately (i.e. when the interpretation of the second pulse ignores information inferred from the first pulse). This indicates that a receiver with memory, which can record and analyse the full response to an EGF sequence, can infer more information than a memoryless receiver that processes each pulse individually.

3.5. Theoretical versus numerical estimation of information carried by pulse sequences

To estimate MI for small noise, one can express it in terms of entropy and conditional entropy:

$$MI(Y; X) = H(Y) - H(Y|X) = H(X) - H(X|Y), \quad (3.3)$$

where $H(X)$ and $H(Y)$ are the entropies of input and output, $H(Y|X)$ is the conditional entropy of the output given the input, while $H(X|Y)$ is the conditional entropy of the input given the output. First, we note that this formula implies

that MI is limited by the input entropy $H(Y)$ as well as the output entropy $H(X)$. In the simplest case of $T > \tau$, all EGF sequences are properly transmitted (transcoded) to ERK sequences, and thus $H(Y|X) = H(X|Y) = 0$ (which means that the knowledge of input is sufficient to deduce output and, conversely, knowledge of output is sufficient to infer input). Thus, for $T > \tau$ we have $MI(Y; X) = H(Y) = H(X) = L$ (where, recall, L is the length of binary input sequences). In this case, the bitrate is equal to the base repeating frequency $\Phi = 1/T$. In the case of $T < \tau$, $MI(Y; X) = H(X)$ can be reached when one is confined to input sequences that can be unambiguously transmitted, that is, to group representatives (as defined in figure 2).

Let us now consider the case of $L = 4$, $T = 30$ min. The input sequences form eight groups: $\{ '0000' \}$, $\{ '0001' \}$, $\{ '0010', '0011' \}$, $\{ '0100', '0110' \}$, $\{ '0101', '0111' \}$, $\{ '1000', '1100' \}$, $\{ '1001', '1101' \}$ and $\{ '1010', '1110', '1011', '1111' \}$. In each group, input sequences are indistinguishable, so $K = 8$. The maximum MI is $\log_2(8) = 3$ and is achieved when probabilities of all group representatives (first sequence in each set) are equal to $1/8$. The same maximal MI value is achieved when in each group the sum of all probabilities is equal to $1/8$. In this case, one also learns 3 bits about input as knowing the output sequence allows one to infer from which of the eight equiprobable groups the input signal originates. Unsurprisingly, by maximizing MI numerically in the case of small noise ($\sigma = 0.1$, $\mu_0 = 0.03$) we obtain almost equal probabilities of each group of input sequences (table 1, last column).

Let us notice that by assuming that all 16 input sequences have equal probabilities, we obtain a somewhat lower value of $MI(Y; X) = H(X) - H(X|Y)$. Now, $H(X) = \log_2(16) = 4$, while $H(X|Y) = (2/16)\log_2(1) + (5/8)\log_2(2) + (1/4)\log_2(4) = 9/8$, which gives $MI(Y; X) = 23/8 < 3$. The calculations for $T = 20$ min and $T = 15$ min are analogous, with the difference that for $T = 20$ min there are $K = 6$ groups, and for $T = 15$ min there are $K = 5$ groups. Also, in the two latter cases the maximized $MI = \log_2(K)$ is larger than MI calculated under the assumption that all input sequences are equiprobable; moreover, the difference is larger than for $T = 30$ min. These theoretical values are compared with numerical estimates for small noise (table 1). One can see that both the maximized MI and the MI calculated by assuming that all sequences are equiprobable agree well with theoretical

estimates. Additionally, one may notice that the numerically maximized MI perfectly agrees with the value calculated by assuming that all group representatives have equal probabilities and the other sequences have zero probability.

3.6. Bitrate estimation in the limit of infinitely long sequences

To estimate the upper limit of MI for small noise for a given base repeating time T , one has to calculate the number of sequences of length L that can be accurately transmitted. For $T = 30$ min, the accurately transmitted sequences are those without '11' subsequences. It is straightforward to show that the numbers n_L of such sequences follow the Fibonacci sequence. Let n_L^0 and n_L^1 denote the number of such sequences of length L ending in, respectively, '0' and '1'. In the first case the sequence can be extended by adding either '0' or '1' at its end; in the second case it can be extended only by adding '0'. Thus, we have $n_{L+1}^0 = n_L^0 + n_L^1$, and $n_{L+1}^1 = n_L^0$. Next, $n_{L+2}^0 = n_{L+1}^0 + n_{L+1}^1 = n_{L+1}$ and $n_{L+2}^1 = n_L^0 + n_L^1 = n_L$, hence $n_{L+2} = n_{L+1} + n_L$ (i.e. $n_{L+1} = n_L + n_{L-1}$). Analogously for $T = 20$ min (or any T satisfying $2T < \tau < 3T$), we may notice that the accurately transmitted sequences are those without both '11' and '101' subsequences. It is also straightforward to show that the number of such sequences satisfies a recurrence relation similar to the Fibonacci sequence, i.e. $n_{L+1} = n_L + n_{L-2}$. Finally, for any T satisfying $kT < \tau < (k+1)T$, the number of accurately transmitted sequences satisfies $n_{L+1} = n_L + n_{L-k}$. In the limit of large L , these recurrence relations can be solved: $n_L(k) = a_{k,0} \times a_k^L$, where a_k satisfies $a_k^{k+1} = a_k^k + 1$, and, for $k \leq 10$, $a_{k,0}$ are constants of order 1, i.e. $a_{1,0} \approx 1.17$, $a_{2,0} \approx 1.03$, $a_{3,0} \approx 0.94$, $a_{10,0} \approx 0.74$.

MI that can be transmitted in a sequence of length L is thus $\text{MI}(k, L) = \log_2(a_{k,0}) + L \times \log_2(a_k)$, therefore the bitrate $C(k, L) = (k+1) \times \text{MI}(k, L)/L$ and, in the limit of $L \rightarrow \infty$, $C(k, L) = C(k) = (k+1) \times \log_2(a_k)$. By straightforward calculation one can find that for $T = 30$ min the bitrate is $C(2) \approx 1.39 \text{ bit h}^{-1}$; for $T = 20$ min, $C(3) \approx 1.65$; and for $T = 15$ min, $C(4) \approx 1.86 \text{ bit h}^{-1}$. $C(k)$ is the monotonically increasing function of k , and $C(k)/\log_2(k) \rightarrow 1$ for $k \rightarrow \infty$.

In figure 4a, for inter-pulse intervals $T = 60, 30, 20$ and 15 min we compare bitrates that were predicted theoretically with those calculated based on numerical simulations. Despite a moderate magnitude of noise ($\sigma = 0.3$, $\mu_0 = 0.03$), there is good agreement between theoretical low-noise prediction and numerical simulations both for not maximized MI (equal input probabilities) and maximized MI. This suggests that the theoretical asymptotic bitrate value, $C(k)$, serves as a good approximation even for low and moderate noise, $\sigma \leq 0.3$ and $\mu_0 \leq 0.03$.

Bitrates increase monotonically with repeating frequency $\Phi = 1/T$. The maximized bitrate is higher than the bitrate calculated for equal input sequence probabilities. The difference is most significant for the smallest $T = 15$ min, when the ratio of sequences that can be properly transmitted to all sequences is low. However, even for equal input probabilities the bitrate for $T = 15$ min is much larger than for $T = 60$ min.

3.7. Influence of noise on bitrate

Noise can substantially decrease the amount of transmitted information. As shown in figure 1d, noise reduces a system's ability to respond to the second EGF pulse, especially when

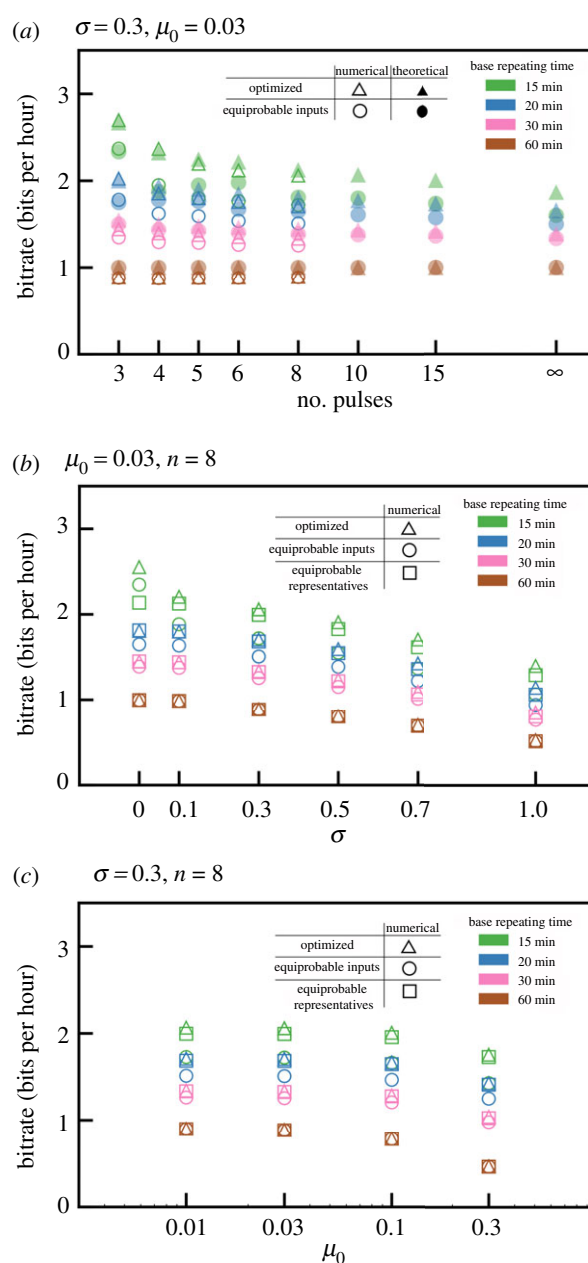


Figure 4. (a) Information transmitted per hour for moderate noise for different carrying frequencies $\phi = 1/T$, where $T = 15, 20, 30, 60$ min. MI was calculated as follows: empty triangles—numerically optimized; solid triangles—optimized, i.e. $\log_2(K)$; empty circles—numerically assuming equal probabilities of all input sequences; solid circles—theoretically assuming equal probabilities of all input sequences. (b) Influence of noise on transmitted information. (c) Influence of additive noise on transmitted information. In (b) and (c) MI was calculated numerically as follows: triangles—maximized; circles—assuming equal probabilities of all group representatives; squares—assuming equal probabilities of all input sequences.

the time span between the pulses is comparable to τ . In figure 4b we show bitrate with respect to σ , with additive noise kept constant, $\mu_0 = 0.03$. As expected, bitrates for all considered T (60, 30, 20 and 15 min) decrease with σ . Interestingly, the ratio of the optimized bitrate for $T = 15$ min to the bitrate for $T = 60$ min increases with noise, which implies that coding using short base repeating time is less sensitive to noise. This is because, in the optimized coding using the short repeating times, the two (non-pseudo) pulses are less frequently separated by a 60 min interval than for the base repeating time

$T = 60$ min. For $T = 60$ min, subsequences '10' and '11' are equiprobable, and the latter is sensitive to noise (figure 1d). Let us notice that even for very high noise parameter $\sigma = 1.0$, for $T = 15$ min, $C > 1 \text{ bit h}^{-1}$ (i.e. C exceeds the bitrate for $T = 60$ min in the limit of small noise).

Small noise $\sigma = 0.1$ influences bitrate only for $T = 15$ min but more detailed analysis suggests that for $T = 15$ min the high bitrate for $\sigma = 0.0$ can be considered an artefact. We found that in this case the system can distinguish between sequences '10' and '11', because the second EGF pulse increases the ERK activity tail. This effect is small and becomes invisible even for small noise ($\sigma \geq 0.1$). The analysis shown in figure 4b is continued in electronic supplementary material, figure S1, where we consider the base repeating times from 30 to 80 min. For $\sigma \geq 0.3$, bitrate increases monotonically with Φ ; only for small noise, $\sigma \leq 0.1$, is the bitrate for $T = 60$ min somewhat higher than for $T = 50$ min. This effect is clear in the context of figure 1d, which shows that for small noise the pathway transmits pulses separated by 60 min intervals but not by 50 min intervals.

In figure 4c we repeat the analysis shown in figure 4b, but now analysing the effect of the additive noise, keeping $\sigma = 0.3$. This effect is very weak up to $\mu_0 = 0.1$ (recall that for $\mu_0 = 0.1$ the amplitude of the additive noise is equal to 10% of the response amplitude). For $\mu_0 = 0.3$, the relative reduction of bitrate is significant, and the relative reduction again is the strongest for $T = 60$ min. The analysis is continued in electronic supplementary material, figure S2 for the base repeating times from 30 to 80 min and shows that, regardless of the amplitude of additive noise, bitrate increases monotonically with Φ .

In summary, the analysis shown in figure 4b,c and electronic supplementary material, figures S1 and S2 shows that, as expected, noise reduces bitrate. Generically, bitrate increases monotonically with $\Phi = 1/T$, for T between 15 min and 80 min. Moreover, this increase is steeper for higher noise. In other words, high-frequency coding is less sensitive to noise than low-frequency coding.

4. Discussion

Two important regulatory pathways of NF- κ B and ERK were found to transmit merely 1 bit of information about the level of extracellular stimulation [14,16], which roughly means that these pathways only relay information about the absence or presence of a stimulus. Information about the stimulation dose is increased only slightly when output is measured in more than one time point [16,22]. These results suggest that, in the considered pathways, information is encoded not in the input amplitude but in the temporal profile of, respectively, TNF and EGF, which are transcribed to the temporal profiles of nuclear NF- κ B or active ERK. The question about optimal information coding can be formulated as the following representation problem: What is the input signal coding that allows the highest information transmission rate to be achieved in long time? It has been investigated previously whether the amplitude or the frequency coding can transduce more information from receptors to transcription factors [23]. Here, we proposed considering information coded in sequences of EGF pulses and, for such coding, we theoretically estimated from below the maximum transmission rate.

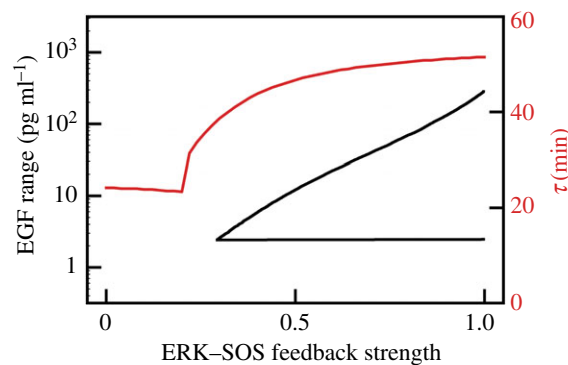


Figure 5. The range of EGF for which ERK_{pp} exhibits oscillations in response to constant EGF stimulation (black, lower line) and influence of the ERK-to-SOS feedback strength on threshold time τ (red, upper line). (Online version in colour.)

The considered model predicts that nearly all EGF inputs are converted to pulses of ERK activity [4]. Relaxation oscillations arise in response to constant EGF stimulation. The oscillation period is a non-monotonic function of the EGF level, attaining minima for the moderate stimulation. The ratio of the ON phase to the OFF phases increases monotonically with the EGF dose; at high dose, oscillations are replaced by constant response [4]. Analogous pulses of EGF are converted to nearly digital pulses of ERK activity. Conversion of various inputs to nearly digital ERK activity pulses may suggest that pulses are used as symbols in intracellular communication. This motivated us to quantify MI between sequences of short EGF pulses and corresponding sequences of ERK activity. Based on this we estimated from below the EGF-to-ERK-channel information capacity, or simply the bit rate at which information can be transmitted.

We showed that high repeating frequency coding allows to substantially exceed bandwidth limit equal $1 \text{ bit}/\tau$, where, recall, $\tau \approx 1 \text{ h}$ is the threshold (or relaxation) time. Specifically, we considered sequences of EGF pulses of '1' or '0' amplitude separated by time intervals T that are multiples of hour/n , and showed that, in the case when intervals between pulses of amplitude '1' (true pulses) are not smaller than 1 h, $C \approx 1.39 \text{ bit h}^{-1}$ for $n = 2$, and $C \approx 1.86 \text{ bit h}^{-1}$ for $n = 4$. Because ERK activity pulses are nearly digital, the considered encoding is not very sensitive to noise. Moreover, the high repeating frequency coding ($n > 1$) is less sensitive to noise than coding with $n = 1$. The high repeating frequency coding is equivalent to coding by pulse/spike intervals, proposed many years ago for neural information processing [24]. Neuronal spikes are separated by intervals of at least an order of magnitude longer than spike duration [25]. In this case, information is more efficiently encoded by time intervals between subsequent '1' (that are not shorter than τ and resolved with accuracy τ/n) than by sequences of '0' and '1' separated by τ .

The bitrate is inversely proportional to threshold time, which for our model is $\tau = 51.5 \text{ min}$. It should be noted that the value of τ may be cell line dependent. According to the model, τ decreases with the strength of the negative feedback from ERK to SOS (figure 5). The abrupt decrease of the τ from 40 min to 25 min is observed when the strength of the feedback is close to one-quarter of its nominal value. At this value, ERK_{pp} does not exhibit oscillations in response to the constant EGF stimulation (figure 5). Therefore, the model suggests that there is a trade-off between short relaxation

time and the ability of the system to translate constant EGF stimulation into series of pulses. In rat PC-12 cells, Ryu *et al.* [9] observed pronounced oscillations in response to 3 min long pulses with a period of 23 min, and also smaller oscillations (that are not arising in our model) in response to the higher frequency pulses. As expected from our analysis, these cells do not exhibit oscillations in response to a constant EGF stimulation. Responses exhibited by the PC-12 cell line can be reproduced by the considered model by removing the negative feedback from ERK to SOS and the positive feedback between SOS and RAS. This may suggest that signalling at the level of SOS is different in the PC-12 cell line from that in the human MCF-10A epithelial cells studied by us [4].

Because the negative feedback loop emanating from ERK targets SOS, a protein at the very beginning of the MAPK cascade, ERK activity oscillations are associated with oscillations of all components between SOS and MEK. Thus, the same or a higher amount of information is reached by pathway components preceding ERK. These proteins control different cellular functions. For example RAF-1 interacts with ROK α , which is implicated in the reorganization of cytoskeletal filaments [26] and with MST2/LATS pathway controlling apoptosis [27].

Characteristic spatio-temporal profiles of cytokines such as TNF and EGF have not been fully characterized. There is however growing evidence that these profiles are bursty and spatially localized rather than slowly fluctuating and spatially uniform. In the context of NF- κ B, recently Bagnall *et al.* [28] showed that TNF propagation is of short range and NF- κ B signalling is limited to distances of a few cell diameters from the neighbouring tissue-resident macrophages. Capuani *et al.* [29] demonstrated that at physiological EGF concentrations (which do not exceed 1 ng ml⁻¹) only a fraction of EGF receptors are activated, which suggests that EGF diffusion is of short range, with EGF being removed from the extracellular space by the receptors. This allows for the existence of spatially localized ERK activity waves that may arise spontaneously [30] or in response to wounding [31], and coordinate motility of cells in the tissue. Hiratsuka *et al.* [31] observed *in vivo* ERK activation waves propagating from the wound edge. Under steady-state conditions, the epidermis occasionally revealed bursts of ERK activation patterns where ERK activation is propagated from cell to cell in a radial manner. ERK activity waves induce collective cell migration in the opposite direction [32], promoting wound healing or sealing a hole in the

epithelium after individual cell apoptosis. These waves, although not yet fully characterized, are dependent on EGF receptor ligands because inhibition of EGF receptors completely blocks ERK activity propagation [27], which in turn suppresses collective cell migration [32]. Propagation of ERK activity waves requires timely digital ERK activation in subsequent layers (or lines in epithelium) of cells. To maintain directionality cells behind the wavefront have to remain inhibited until the signal propagates to the next layer. The ability to transmit spontaneous signals requires a single cell to accurately respond to random sequences of EGF pulses, which is equivalent to transmitting information at sufficiently high bit rates.

Regulatory pathways can be perceived as noisy information-transmitting channels, but can also be considered as decision-making modules that employ feedbacks and other nonlinear regulatory elements to convert input into one of several predefined outputs. MAPK/ERK and NF- κ B pathways convert both constant and pulsating cytokine stimulation into pulses of ERK or NF- κ B activity. Here, we consider an optimal representation problem and investigate which EGF time profiles can transmit the maximum amount of information to ERK. Motivated by the growing evidence that physiological EGF stimulation is short ranged and bursty, we focus on sequences of short EGF pulses. We found that information can be transmitted with the rate exceeding the classical bandwidth limit of about 1 bit h⁻¹ in the case when inter-pulse intervals are used to code information. The timely activation of ERK in neighbouring cells allows for propagation of waves that coordinate collective cell migration during wound healing.

Data accessibility. All data supporting the presented results are included in the article or electronic supplementary material. The source code of our Python module for estimation of information channel capacity is available at <http://pmbm.ippt.pan.pl/software/cce>.

Authors' contributions. F.G. and P.C. prepared the computational models and performed the analysis; M.K. prepared the computational models and drafted the manuscript; T.L. conceived and coordinated the study and drafted the manuscript. All authors gave final approval for publication.

Competing interests. We have no competing interests.

Funding. This work has been supported by the National Science Centre (NCN grant no. 2014/14/M/NZ6/00537) and the Vienna Science and Technology Fund (WWTF grant no. MA14-049).

Acknowledgements. We thank the Institute of Fundamental Technological Research for providing computational resources (Grafen).

References

- Kellogg RA, Tian C, Lipniacki T, Quake SR, Tay S. 2015 Digital signaling decouples activation probability and population heterogeneity. *Elife* **4**, e08931. (doi:10.7554/eLife.08931)
- Hao N, O'Shea EK. 2011 Signal-dependent dynamics of transcription factor translocation controls gene expression. *Nat. Struct. Mol. Biol.* **19**, 31–39. (doi:10.1038/nsmb.2192)
- Albeck JG, Mills GB, Brugge JS. 2013 Frequency-modulated pulses of ERK activity transmit quantitative proliferation signals. *Mol. Cell* **49**, 249–261. (doi:10.1016/j.molcel.2012.11.002)
- Kochańczyk M, Kocieniewski P, Kozłowska E, Jaruszewicz-Błońska J, Sparta B, Pargett M, Albeck JG, Hlavacek WS, Lipniacki T. 2017 Relaxation oscillations and hierarchy of feedbacks in MAPK signaling. *Sci. Rep.* **7**, 38244. (doi:10.1038/srep38244)
- Behar M, Hoffmann A. 2010 Understanding the temporal codes of intra-cellular signals. *Curr. Opin. Genet. Dev.* **20**, 684–693. (doi:10.1016/j.gde.2010.09.007)
- Levine JH, Lin Y, Elowitz MB. 2013 Functional roles of pulsing in genetic circuits. *Science* **342**, 1193–1200. (doi:10.1126/science.1239999)
- Purvis JE, Lahav G. 2013 Encoding and decoding cellular information through signaling dynamics. *Cell* **152**, 945–956. (doi:10.1016/j.cell.2013.02.005)
- Tay S, Hughey JJ, Lee TK, Lipniacki T, Quake SR, Covert MW. 2010 Single-cell NF- κ B dynamics reveal digital activation and analogue information processing. *Nature* **466**, 267–271. (doi:10.1038/nature09145)
- Ryu H *et al.* 2015 Frequency modulation of ERK activation dynamics rewires cell fate. *Mol. Syst. Biol.* **11**, 838. (doi:10.15252/msb.20156458)

10. Ashall L *et al.* 2009 Pulsatile stimulation determines timing and specificity of NF- κ B-dependent transcription. *Science* **324**, 242–246. (doi:10.1126/science.1164860)
11. Korwek Z, Tudelska K, Nałęcz-Jawecki P, Czerkies M, Prus W, Markiewicz J, Kochančzyk M, Lipniacki T. 2016 Importins promote high-frequency NF- κ B oscillations increasing information channel capacity. *Biol. Direct* **11**, 61. (doi:10.1186/s13062-016-0164-z)
12. Shannon CE. 1948 A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423. (doi:10.1002/j.1538-7305.1948.tb01338.x)
13. Shannon CE. 1948 A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 623–656. (doi:10.1002/j.1538-7305.1948.tb00917.x)
14. Cheong R, Rhee A, Wang CJ, Nemenman I, Levchenko A. 2011 Information transduction capacity of noisy biochemical signaling networks. *Science* **334**, 354–358. (doi:10.1126/science.1204553)
15. Uda S, Saito TH, Kudo T, Kokaji T, Tsuchiya T, Kubota H, Komori Y, Ozaki Y, Kuroda S. 2013 Robustness and compensation of information transmission of signaling pathways. *Science* **341**, 558–561. (doi:10.1126/science.1234511)
16. Selimkhanov J, Taylor B, Yao J, Pilko A, Albeck J, Hoffmann A, Tsimring L, Wollman R. 2014 Accurate information transmission through dynamic biochemical signaling networks. *Science* **346**, 1370–1373. (doi:10.1126/science.1254933)
17. Tudelska K *et al.* 2017 Information processing in the NF- κ B pathway. *Sci. Rep.* **7**, 15926. (doi:10.1038/s41598-017-16166-y)
18. Faeder JR, Blinov ML, Hlavacek WS. 2009 Rule-based modeling of biochemical systems with BioNetGen. In *Systems biology* (ed. IV Maly), pp. 113–167. Totowa, NJ: Humana Press.
19. Kraskov A, Stögbauer H, Grassberger P. 2004 Estimating mutual information. *Phys. Rev. E* **69**, 066138.
20. Kingma DP, Ba J. 2014 Adam: a method for stochastic optimization. In *Proceedings of the 3rd International Conference on Learning Representations (ICLR), San Diego, CA, USA, May 7–9, 2015*.
21. Abadi M *et al.* 2015 *TensorFlow: large-scale machine learning on heterogeneous systems*. See <https://www.tensorflow.org/>.
22. Potter GD, Byrd TA, Mugler A, Sun B. 2017 Dynamic sampling and information encoding in biochemical networks. *Biophys. J.* **112**, 795–804. (doi:10.1016/j.bpj.2016.12.045)
23. Hansen AS, O'Shea EK. 2015 Limits on information transduction through amplitude and frequency regulation of transcription factor activity. *Elife* **4**, e06559. (doi:10.7554/eLife.06559)
24. MacKay DM, McCulloch WS. 1952 The limiting information capacity of a neuronal link. *Bull. Math. Biophys.* **14**, 127–135. (doi:10.1007/BF02477711)
25. Cessac B, Paugam-Moisy H, Vieville T. 2010 Overview of facts and issues about neural coding by spikes. *J. Physiol. Paris* **104**, 5–18. (doi:10.1016/j.jphysparis.2009.11.002)
26. Varga A, Ehrenreiter K, Aschenbrenner B, Kocieniewski P, Kochančzyk M, Lipniacki T, Baccarini M. 2017 RAF1/BRAF dimerization integrates the signal from RAS to ERK and ROK α . *Sci. Signal.* **10**, eaai8482. (doi:10.1126/scisignal.aai8482)
27. Romano D, Nguyen LK, Matallanas D, Halasz M, Doherty C, Kholodenko BN, Kolch W. 2014 Protein interaction switches coordinate Raf-1 and MST2/Hippo signalling. *Nat. Cell Biol.* **16**, 673–684. (doi:10.1038/ncb2986)
28. Bagnall J *et al.* 2018 Quantitative analysis of competitive cytokine signaling predicts tissue thresholds for the propagation of macrophage activation. *Sci. Signal.* **11**, eaaf3998. (doi:10.1126/scisignal.aaf3998)
29. Capuani F, Conte A, Argenzio E, Marchetti L, Priami C, Polo S, Di Fiore PP, Sigismund S, Ciliberto A. 2015 Quantitative analysis reveals how EGFR activation and downregulation are coupled in normal but not in cancer cells. *Nat. Commun.* **6**, 7999. (doi:10.1038/ncomms8999)
30. Aoki K, Kumagai Y, Sakurai A, Komatsu N, Fujita Y, Shionyu C, Matsuda M. 2013 Stochastic ERK activation induced by noise and cell-to-cell propagation regulates cell density-dependent proliferation. *Mol. Cell* **52**, 529–540. (doi:10.1016/j.molcel.2013.09.015)
31. Hiratsuka T, Fujita Y, Naoki H, Aoki K, Kamioka Y, Matsuda M. 2015 Intercellular propagation of extracellular signal-regulated kinase activation revealed by *in vivo* imaging of mouse skin. *Elife* **4**, e05178. (doi:10.7554/eLife.05178)
32. Aoki K, Kondo Y, Naoki H, Hiratsuka T, Itoh RE, Matsuda M. 2017 Propagating wave of ERK activation orients collective cell migration. *Dev. Cell* **43**, 305–317.e5. (doi:10.1016/j.devcel.2017.10.016)



OPEN Predictive power of non-identifiable models

Frederic Grabowski, Paweł Nałęcz-Jawecki & Tomasz Lipniacki✉

Resolving practical non-identifiability of computational models typically requires either additional data or non-algorithmic model reduction, which frequently results in models containing parameters lacking direct interpretation. Here, instead of reducing models, we explore an alternative, Bayesian approach, and quantify the predictive power of non-identifiable models. We considered an example biochemical signalling cascade model as well as its mechanical analogue. For these models, we demonstrated that by measuring a single variable in response to a properly chosen stimulation protocol, the dimensionality of the parameter space is reduced, which allows for predicting the measured variable's trajectory in response to different stimulation protocols even if all model parameters remain unidentified. Moreover, one can predict how such a trajectory will transform in the case of a multiplicative change of an arbitrary model parameter. Successive measurements of remaining variables further reduce the dimensionality of the parameter space and enable new predictions. We analysed potential pitfalls of the proposed approach that can arise when the investigated model is oversimplified, incorrect, or when the training protocol is inadequate. The main advantage of the suggested iterative approach is that the predictive power of the model can be assessed and practically utilised at each step.

Non-identifiable or sloppy models are ubiquitous in systems biology^{1,2}. Non-identifiability, in practical terms, means that parameters obtained through model fitting cannot be trusted: either because arbitrarily large changes of some parameters can be fully compensated by changes of other parameters (structural non-identifiability) or because plausible parameters may differ widely due to noise in measurements (practical non-identifiability). A closely related property is sloppiness. In sloppy models, some combinations of parameters define stiff directions, along which model responses change rapidly, whereas the remaining combinations define sloppy directions, along which model responses change insignificantly. It is important to note that non-identifiability or sloppiness is a property of a model together with a dataset, not just the model itself. Additional data can make the model (more) identifiable³ and can reduce the number of sloppy directions.

Non-identifiability and sloppiness have a lot in common, however the first is perceived as a model disadvantage that should be resolved, whereas the second is perceived as a ubiquitous property of complex models we have to live with². Typically, structural non-identifiability is relatively easy to resolve by nondimensionalisation and reparametrization. Practical non-identifiability is a harder problem, usually requiring additional experiments (which can be expensive or time-consuming), or manual model reduction^{4,5}. An identifiable model that reflects underlying biochemistry can be seen as the ultimate goal of systems biology. However, model reduction, performed to reach identifiability, may lead to composite parameters that lack straightforward biochemical interpretation^{6–8}. Additionally, a reduced model may not be able to accommodate new data containing measurements of previously omitted variables. An alternative approach to non-identifiable, sloppy models is to identify stiff directions, and analyse system behaviour along these directions, which is an effective way of reducing the dimensionality of the investigated parameter space, without changing the model⁹.

In this study, we focus on practically non-identifiable models, but instead of reducing them to reach identifiability, we train them on limited datasets and investigate their predictive power after such training. It was observed earlier by Brown et al.¹⁰, Cedersund¹¹, van Mourik et al.¹², Monsalve-Bravo et al.⁹ and others that sloppy, non-identifiable models may lead to informative predictions; nevertheless the standard solution is to make models identifiable before predictions are made³, as reviewed in Refs.^{4,13}. In this paper, we analysed the predictive power of non-identifiable models in more detail in the context of systems biology and investigated how successive inclusion of model variables in training influences predictions¹⁴. We view modelling as the following iterative procedure: perform experiment → train model on gathered data → assess its predictive power, if required perform additional experiments → train model again, narrowing down its plausible parameter space and thus increasing its predictive power. This approach stems from the fact that in (systems) biology additional

Institute of Fundamental Technological Research, Polish Academy of Sciences, Warsaw, Poland. ✉email: tlipnia@ippt.pan.pl

measurements are usually relatively expensive and/or time-consuming, so researchers try to constrain the model by performing a minimal necessary set of experiments^{15,16}. This procedure increases the number of stiff directions and ultimately allows one to predict the investigated variables with high accuracy. To illustrate the procedure outlined above we considered a biochemical signalling cascade with a negative feedback loop—a motif ubiquitous in regulatory pathways (e.g. MAPK^{17,18}, NF- κ B^{19,20}, p53^{21,22}). As an auxiliary example, we analysed a simple damped mechanical oscillator consisting of three masses with one of them subject to forcing.

We investigated the trained model's ability to predict the system responses to different temporal stimulation protocols as well as responses that arise when one of the parameters is multiplied or divided by a factor of 10. The latter corresponds to perturbations that can be introduced by inhibitors or mutations causing protein hyperactivity, while the change of stimulation protocol can be interpreted as a modification of the therapy schedule. In the last decade, investigation of responses to complex temporal stimulations was enabled by optogenetic receptors which allow arbitrarily complex stimulation protocols^{23,24}. We demonstrated that training a relaxed model, with three hypothetical negative feedback loops, leads to equally good predictions as training the nominal model, and correctly suggests the location of the nominal feedback loop. In contrast, training oversimplified or incorrect models leads to inaccurate predictions with (misleadingly) narrow prediction bands. This suggests the use of a more relaxed model structure when the exact network wiring is unknown or uncertain.

Model training relies on the exploration of the space of plausible parameters, i.e. parameters giving predictions within an expected error bound of the measurements. There are several methods to explore the parameter space (e.g. Fisher information matrix^{16,25}, profile likelihood^{3,4}, Bayesian methods^{26–28}, approximate Bayesian computation²⁹, nested sampling³⁰). We chose the Bayesian approach and sampled the parameter space by performing Markov Chain Monte Carlo (MCMC) simulations using the Metropolis–Hastings algorithm³¹. This method allows us to start from broad, non-informative priors and explore the posterior parameter distributions confined by smaller or larger datasets, and consequently obtain straightforward probabilistic predictions. To avoid over-focusing on the details of the Bayesian approach, we will refer to posterior parameter samples as the set of plausible parameters, and to the process of running the MCMC simulation and obtaining the samples as training.

Results

Sequential training of the signalling cascade model. We considered a four-step signalling cascade with negative feedback(s) resembling the MAPK cascade (RAS $\rightarrow$ RAF $\rightarrow$ MEK $\rightarrow$ ERK), outlined in Fig. 1a, with corresponding equations given in Fig. 1b. The cascade is activated by a time-dependent signal $S(t)$, which depends on the stimulation protocol chosen. Nominal model parameters (used for generating training data) and

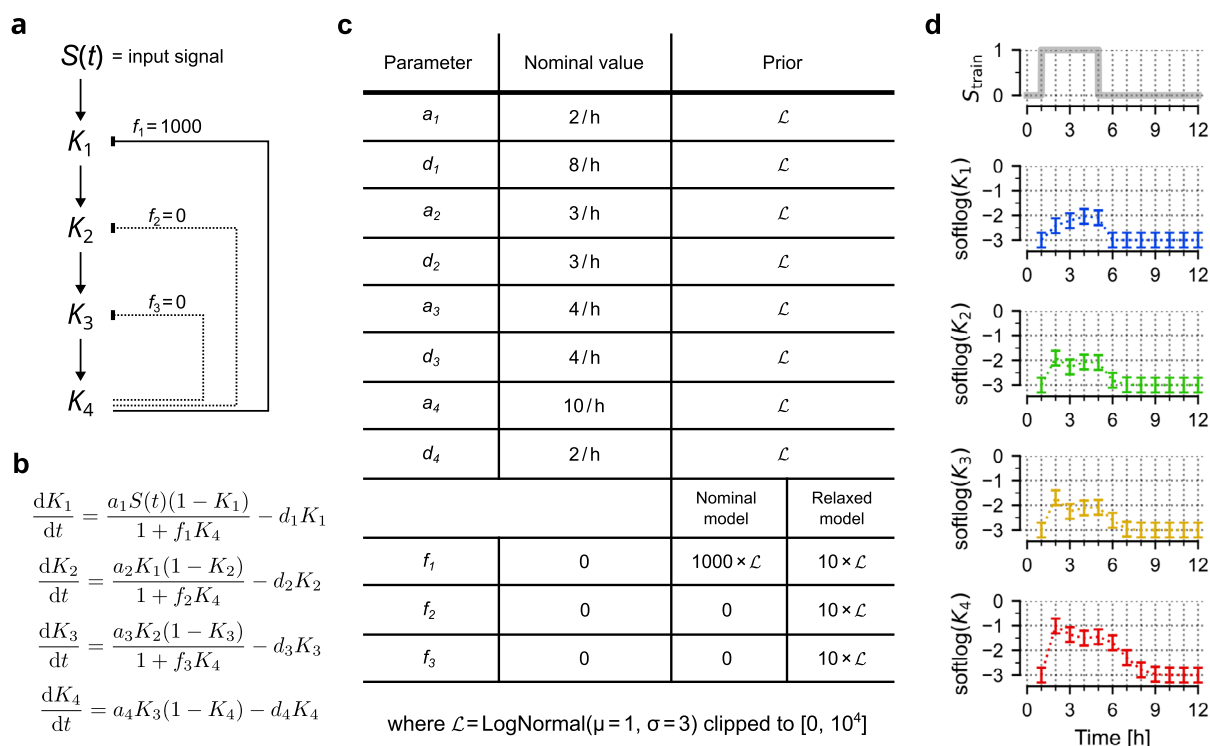


Figure 1. Signalling cascade model. **(a)** Scheme of the 4-level biochemical signalling cascade model activated by a time-dependent signal $S(t)$. Strengths of feedbacks f_2 and f_3 are set to zero in the nominal model but are allowed to assume positive values in the ‘relaxed’ model. **(b)** Model equations. **(c)** Model parameters used for generating training data and prior distributions used for training. **(d)** Simulated trajectories of K_1, K_2, K_3 , and K_4 in response to S_{train} used for model training; softlog(x) = log(0.001 + x). Error bars show measurement errors assumed for the generation of training data.

assumed lognormal parameter priors used during model training are given in Fig. 1c. We note that, by default, the strengths of feedbacks to the second and third step of the cascade are zero ($f_2 = f_3 = 0$)—later we investigate a ‘relaxed’ model in which f_2 and f_3 are allowed to be positive. The training dataset was generated by simulating trajectories of the four model variables, K_1 , K_2 , K_3 , and K_4 using an ‘on–off’ stimulation protocol and randomly perturbing them in order to mimic experimental errors (Fig. 1d). The error bars are plotted at measurement time points (every hour) and show the magnitude of the assumed experimental errors. We used three such generated measurement replicates for training (see “Methods” for details). The training dataset was used to train two models: the ‘nominal’ model for which we assume $f_2 = f_3 = 0$, and a ‘relaxed’ model (in the next section) in which we allow f_2 and f_3 to assume positive values, also with a lognormal prior (Fig. 1c).

First, we trained the nominal model using only the trajectory of the last variable, K_4 . We observe that the resulting model accurately predicts the trajectory of K_4 in response to two very different stimulation protocols (Fig. 2a,b), but fails to predict the trajectories of any of the remaining variables. For these three variables, the 80% prediction bands are very broad, implying that the trained model contains little information about these variables. Next, the training dataset was expanded to additionally include the trajectory of K_2 , and we observe that also the trajectory of this variable can be reliably predicted (Fig. 2c). Finally, the complete training dataset containing trajectories of all four variables resulted in a “well trained” model, which predicts the trajectories of all variables (Fig. 2d).

By investigating the space of plausible parameters (i.e., posterior samples) we may notice that even for the well-trained model the marginal probability distributions show that all 9 parameters may vary by about an order of magnitude, and when training is performed only on K_4 (or K_2 and K_4) parameters may vary by about two orders of magnitude (Fig. 2e). This shows that the model trained only on the last variable of the cascade, K_4 , can accurately predict this variable’s trajectories in response to different stimulation protocols, even though all model parameters remain uncertain, suggesting sloppiness of the trained model¹⁰. To better understand this feature we numerically analysed the dimensionality of the plausible parameter sets (see “Methods” section for details) obtained after training the model on (i) K_4 , (ii) K_2 and K_4 , and (iii) all four variables (Fig. 2f). Since we are interested in multiplicative changes of parameters, we performed a principal component analysis (PCA) on sets of (natural) logarithms of plausible parameters⁹. Exponents of the square roots of the principal values (variances) λ_i can be interpreted as principal multiplicative deviations of parameter values, $\delta_i = \exp(\sqrt{\lambda_i})$. The prior parameter distributions are LogNormal($\mu = 1$, $\sigma = 3$) clipped to $[0, 10^4]$, which gives $\delta_{i,\text{prior}} \approx \exp(3) \approx 20$ (for all i), implying that “on average” all prior parameter values can vary 20-fold from the median. After training on all variables, the largest principal multiplicative deviation is $\delta_i \approx 12$, however, the 4 smallest δ_i are much lower, below 1.5 (note that $\delta_i = 1$ implies no variation along the i -th principal component), indicating that these directions are stiff. This implies that the dimensionality of the plausible parameter space was effectively reduced by 4 due to training. When training is performed only on K_4 , dimensionality is reduced by 1 (only the smallest δ_i is below 1.5), and when training uses K_2 and K_4 , dimensionality is reduced by 2 (the two smallest δ_i are below 1.5). Therefore, in agreement with intuition, model training reduces the dimensionality of the plausible parameter space. In our example, measurement of n variables reduces dimensionality by n . Importantly, even reduction from 9 to 8 dimensions allows for reliable predictions of a single variable of interest that was measured to train the model.

The principal components (eigenvectors) can be used to gain additional insight. Each stiff direction defines a combination of parameters that is well-constrained by data used for training (Fig. S1a). The single stiff direction that emerges after training on K_4 has four positive coordinates corresponding to activation parameters (a_1 – a_4) and five negative coordinates corresponding to deactivation parameters (d_1 – d_4) and the negative feedback (f_1). The same pattern appears when training includes two, or all four variables; in the emerging stiff directions, coordinates corresponding to activation and deactivation have opposite signs. Stiff directions span a stiff subspace, and this subspace can be more intuitively visualised when a linear combination of stiff vectors minimising the number of dominant coordinates is chosen (Fig. S1b). One can now readily see that even after training on all four model variables, only the ratios of activation/deactivation parameters can be determined, not these parameters themselves.

Finally, we investigated whether the trained model is able to predict responses to test protocols that arise after a multiplicative change of a given model parameter. Such a change is equivalent to a system perturbation that can be introduced experimentally by knockdown or overexpression of a reacting protein. To test this, after training the model on K_4 , we modified each set of plausible parameters by multiplying or dividing a given parameter by a factor of 10. As shown in Fig. 3, for each of the 9 parameters such a modification allows to correctly predict, with relatively narrow prediction bands, the response of the nominal model with the corresponding parameter change.

Training a relaxed model. Frequently, experiments are performed not only to constrain the parameters of a known model but to deduce the structure of the model itself. To explore this scenario, in Fig. 4 we introduced a ‘relaxed’ model, in which we allowed strengths of the three negative feedbacks shown in Fig. 1a to assume positive values during training (assuming prior lognormal distributions, as shown in Fig. 1c). This increases the number of free parameters from 9 to 11. Again, this relaxed model was trained on in silico-generated data, based on the nominal model with one feedback. Regardless of whether only K_4 or all variables are included in the training dataset, the relaxed model performs equally well as the nominal model on test protocols (compare Fig. 4a versus Figs. 2a, 4b versus Fig. 2d). We note that the marginal distributions of the three feedback parameters remain broad (Fig. 4c). This means that, even without knowing strengths of individual feedbacks, the model can give reasonable predictions. Nevertheless, these distributions indicate that the (nominal) feedback f_1 is dominant.

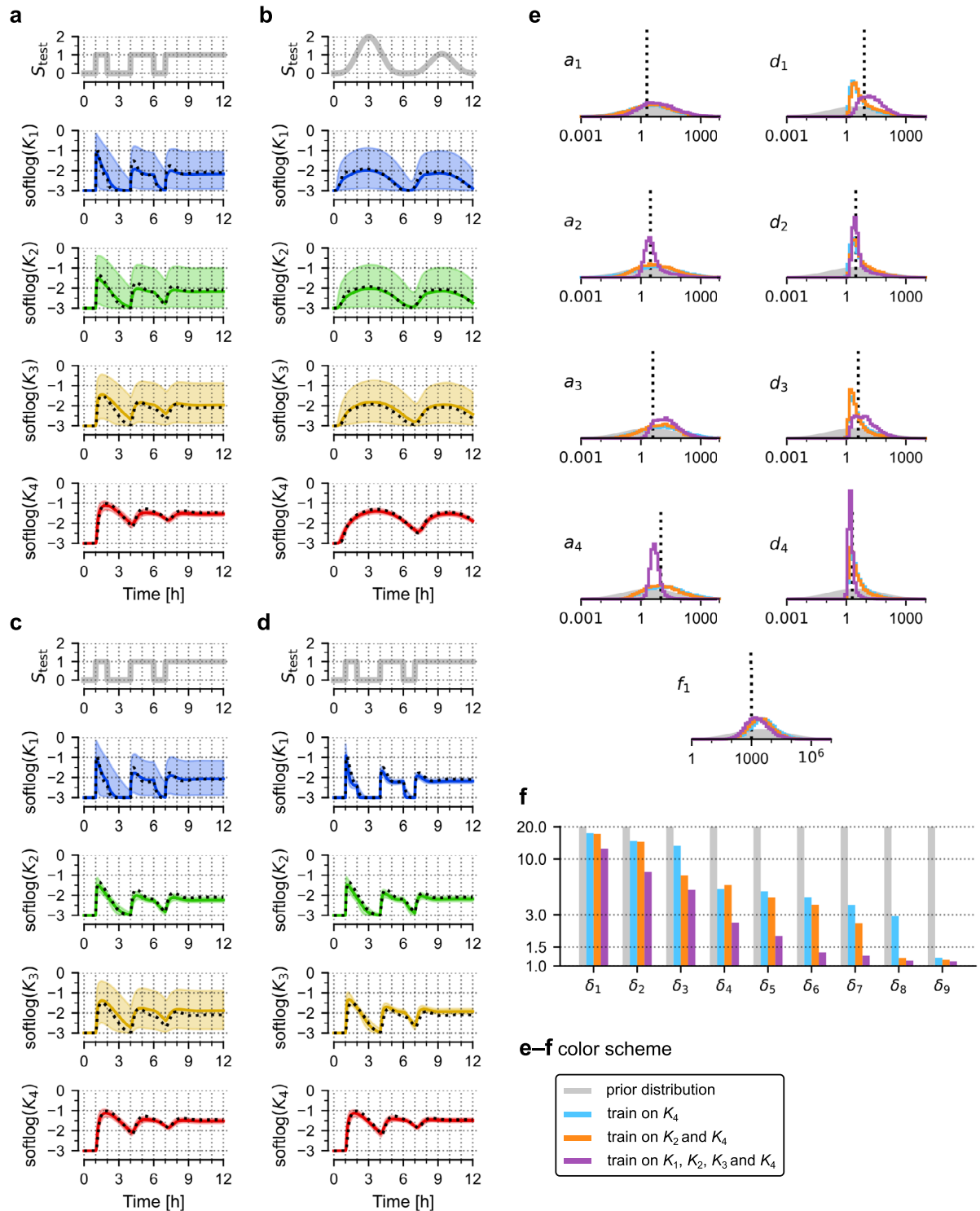


Figure 2. Training the nominal model. **(a,b)** Prediction of model responses to two test signals, S_{test} , after training on the trajectory of K_4 (generated in response to the training protocol, S_{train} , shown in Fig. 1d). Black dotted lines show trajectories of the nominal model, coloured lines show (point-wise) medians of predictions, contours show 80% prediction bands; $\text{softlog}(x) = \log(0.001 + x)$. **(c)** Prediction of model responses after training on trajectories K_2 and K_4 . **(d)** Prediction of model responses after training on trajectories of all four model variables, K_1, K_2, K_3 and K_4 . **(e)** Histograms obtained from 10,000 samples generated from the posterior distribution of model parameters after training on K_4 (blue); K_2 and K_4 (orange); K_1, K_2, K_3 and K_4 (purple). Grey contours show prior distributions. **(f)** Dimensionality reduction of the parameter space by successive inclusion of model variable trajectories into the training set. Bar heights show principal multiplicative deviations $\delta_i = \exp(\sqrt{\lambda_i})$, where λ_i are principal values (variances) of log-parameters.

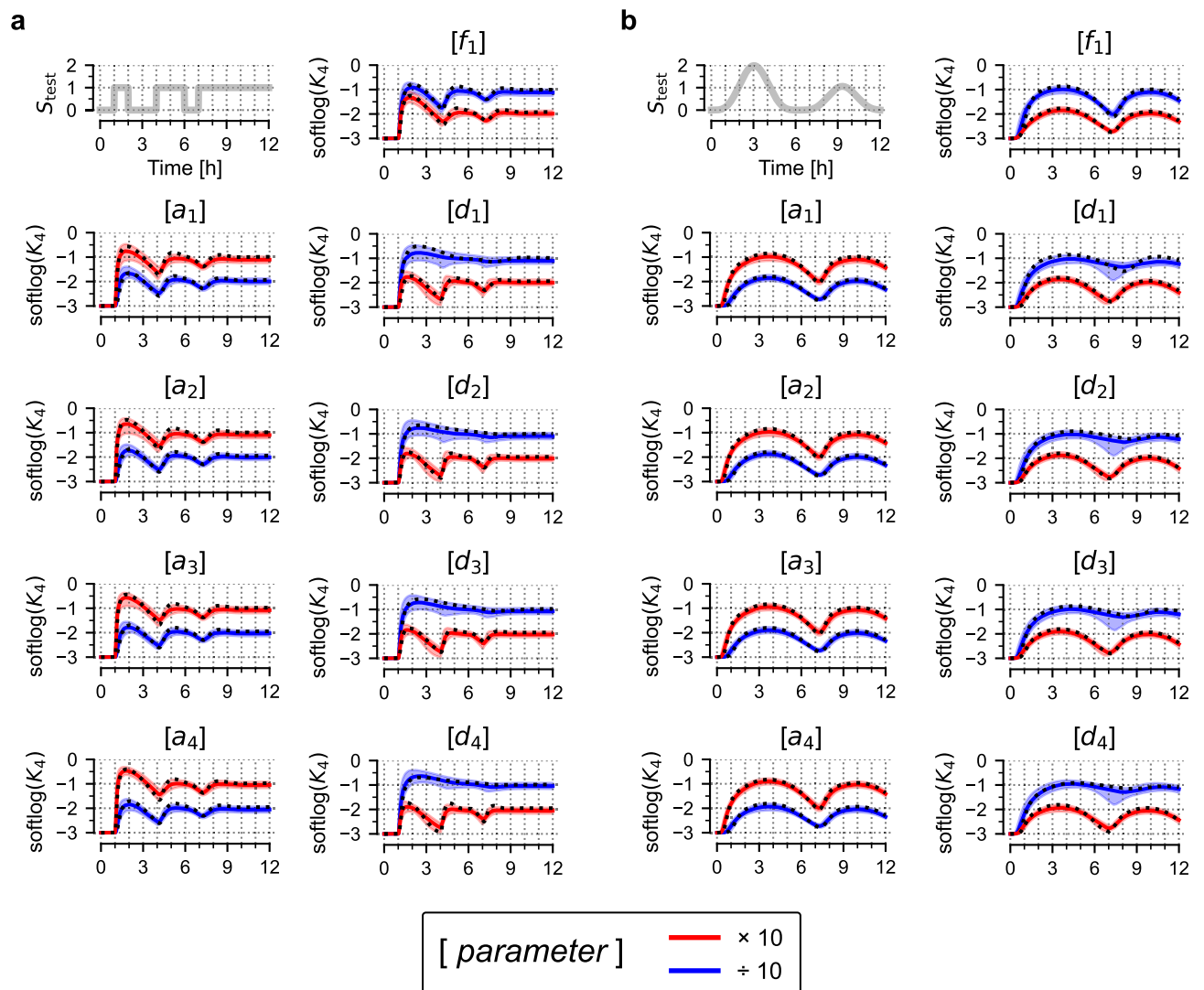


Figure 3. Predicted responses after changing model parameters. **(a,b)** Prediction of model responses to two test signals, S_{test} (shown in each panel top left), after training on the trajectory of K_4 (generated in response to the training protocol, S_{train} , shown in Fig. 1d). In each chart, one of the parameters (shown in square brackets) was multiplied (red) or divided (blue) by a factor of 10. Black dotted lines show trajectories of the nominal model, coloured lines show (point-wise) medians of predictions, contours show 80% prediction bands; $\text{softlog}(x) = \log(0.001 + x)$.

The above example suggests that if there exists uncertainty about model structure, it is advisable to train a more relaxed model, which can still result in satisfactory predictions, and eventually, after gathering more data, can provide a hint about true model structure.

Training a simplified model. Given insufficient knowledge about the cascade structure and a small data-set, it might be tempting to train a simplified model. Here, we give an example that showcases potential pitfalls of this approach. As previously, the training set was a perturbed trajectory of K_4 generated by the nominal model—however, we attempted to train a model that contains only two signalling steps, in which K_1 directly influences K_4 (formally in the equation for K_4 we replace K_3 by K_1 , and equations for K_2 and K_3 are omitted). Training the simplified model on the ‘on–off’ protocol shows some discrepancy between the original and simplified model trajectories (Fig. 5a, note the logarithmic scale of the plot). When the trained model is verified on the test stimulation protocol, the observed discrepancy is more visible, meaning that the simplified model is not able to make accurate predictions (Fig. 5b). Crucially, the 80% prediction band is relatively narrow, similar to that obtained for the correct model (Fig. 5b versus Fig. 2a). Thus the trained simplified model is “sure” (even if marginal parameter distributions of 3 out of 5 model parameters remain very broad—Fig. 5c) about its inaccurate predictions.

In certain cases where approximate predictions are still satisfactory, the simplified model can be a reasonable solution. However, as demonstrated above, one should be very careful when interpreting parameters and making predictions, since some conclusions might be an artefact of using a wrong model to compute posterior probabilities. Finally, in contrast to the relaxed model, the simplified model cannot be improved by additional data.

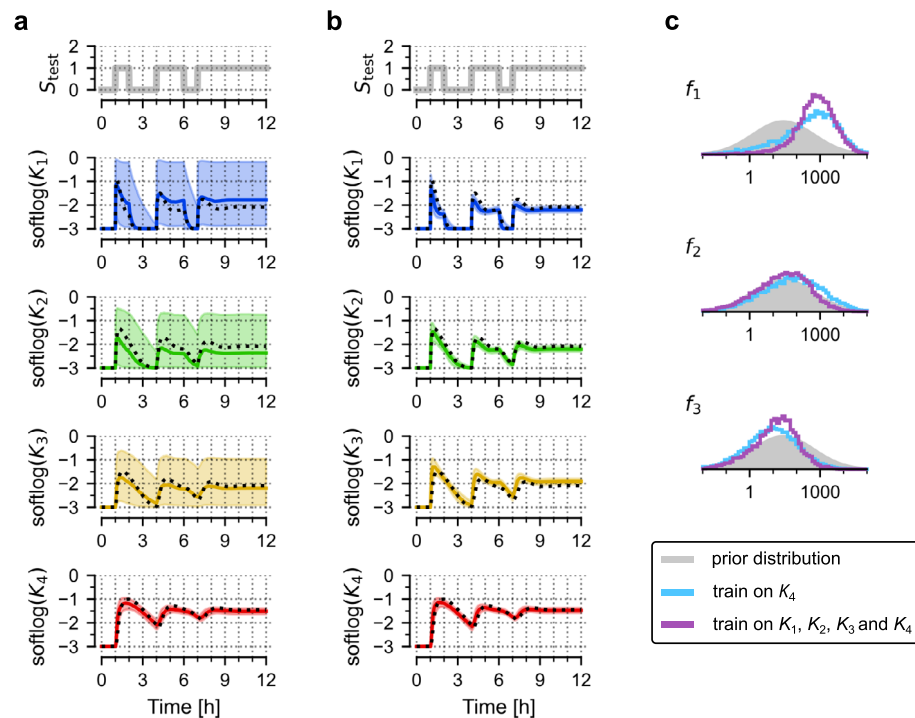


Figure 4. Training a relaxed model. (a) Prediction of model responses to test signal, S_{test} , after training on the trajectory of K_4 (generated in response to the training protocol, S_{train} , shown in Fig. 1d). Black dotted lines show trajectories of the nominal model, coloured lines show (point-wise) medians of predictions, contours show 80% prediction bands; $\text{softlog}(x) = \log(0.001 + x)$. (b) Prediction of model responses after training on trajectories of all four model variables. (c) Histograms were obtained from 10,000 samples generated from the posterior distribution of model parameters after training on K_4 (blue) or all model variables (purple). Grey contours show prior distributions.

Analysis of potential pitfalls. In Fig. S2 we investigate training based on an inadequate stimulation protocol. The protocol's 'on' phase lasts only 1 h (compared to 4 h in Fig. 2), and as a result, the model predictions have a broad 80% prediction band. The prediction band for K_4 is broad when training includes K_4 only, and remains broad even when training includes all four model variables. Higher uncertainty of trained model trajectories is associated with larger principal multiplicative deviations δ_p , compared to values obtained for training shown in Fig. 1d. This example highlights the importance of designing a good stimulation protocol for training. If our goal is to obtain accurate predictions of K_4 only, training on K_4 with the help of an appropriate protocol (Fig. 2a) gives much better results than training on all variables employing a suboptimal protocol (Fig. S2c). This pitfall is relatively simple to resolve—one has to adjust the training protocol so that it allows capturing characteristic temporal responses of the system. Whether the 1-h long 'on' phase is sufficient or not depends on the nominal model parameters.

In Fig. S3 we show an example where the prior model is incorrect, i.e. has no negative feedback at all. We obtain discrepancies between measurement and model, which could be overlooked during training if, for example, the measurement points were not distributed densely enough. From the trained wrong model we also obtain wrong predictions with very narrow prediction bands, as in the case of the simplified model discussed earlier. The key message from Fig. 5 and Fig. S3 is that a simplified or wrong model can make inaccurate or wrong predictions with narrow prediction bands, and thus the width of prediction bands contains no information about the accuracy of the model. Narrow prediction bands imply that obtained constraints on parameters suffice for unique predictions, but these predictions are correct only under the condition that the model structure is also correct.

Mechanical analogue. We performed a similar analysis of a (distant) mechanical analogue of the signalling pathway, a damped harmonic oscillator consisting of three masses connected by two springs (Fig. 6). The system is parametrized by 8 free parameters. For training, we used a protocol in which one of the masses is pushed with constant force for 10 s, which results in a displacement of all masses. Next, we used the trajectory of the opposite mass (or of all masses) to train the model. In the test protocol, a sinusoidal driving force was applied to the same mass as during training. Similarly to the signalling cascade model, we found that training on a single mass trajectory suffices to predict its trajectory (for a limited time) in response to a very different forcing, and training on all three masses allows for very accurate predictions of all trajectories, despite the fact that marginal distributions of model parameters remain broad.

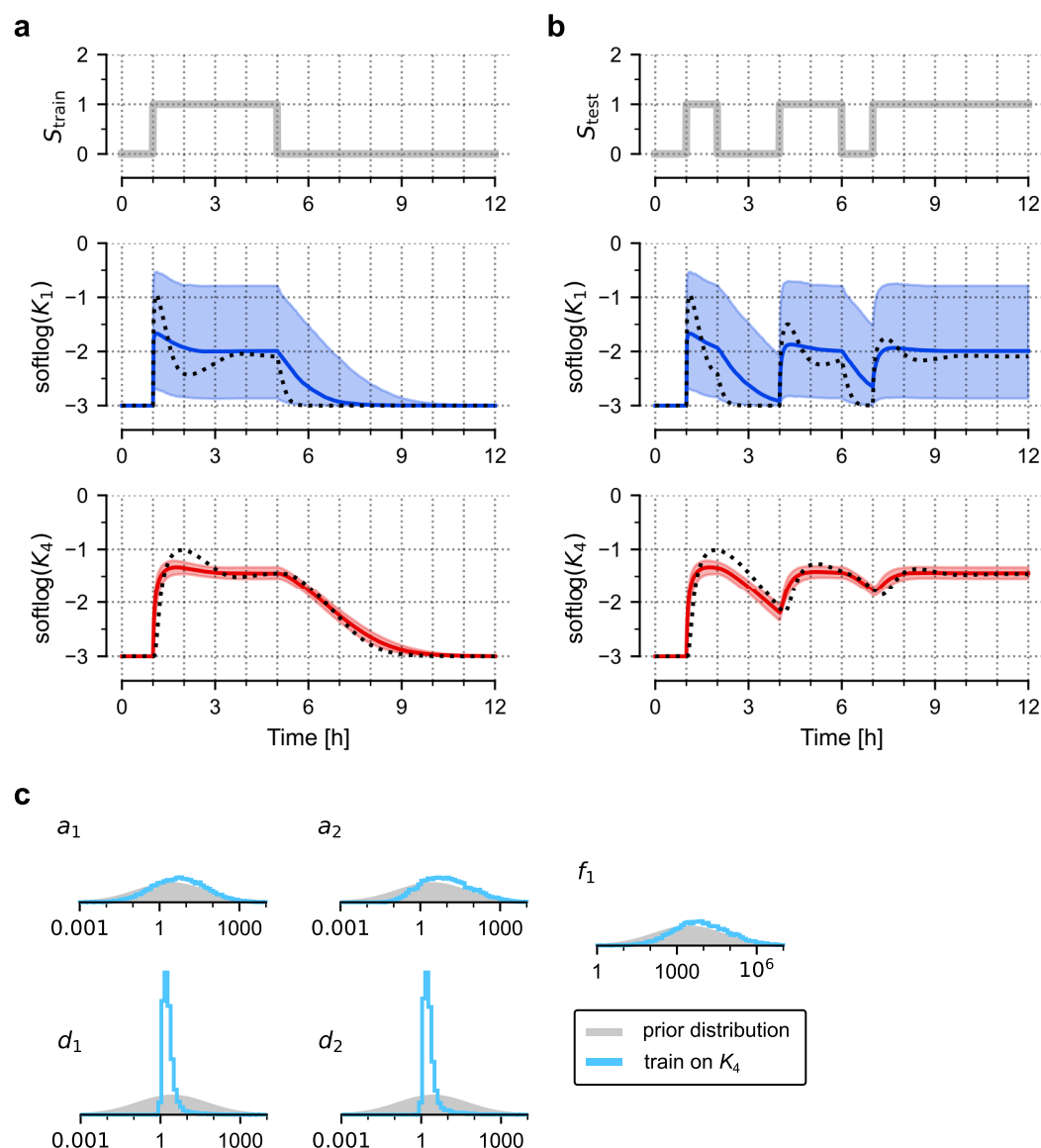


Figure 5. Training a simplified model. **(a)** Training a simplified (2-step) model on the nominal (4-step) model trajectory of K_4 . Only variables present in the simplified model are shown. Black dotted lines show trajectories of the nominal model, coloured lines show (point-wise) medians of the simplified model, contours show 80% prediction bands; $\text{softlog}(x) = \log(0.001 + x)$. **(b)** Prediction of the trained simplified model responses to the test signal, S_{test} . Same graphical convention as in **(a)**. **(c)** Histograms were obtained from 10,000 samples generated from the posterior distribution of simplified model parameters. Grey contours show prior distributions.

Discussion

The majority of systems biology models, including mechanistic models of intensely studied regulatory networks, are non-identifiable (or sloppy). Current techniques and algorithms allow practitioners to automatically resolve structural non-identifiability problems, however, resolving practical non-identifiability is harder, and frequently requires additional data or substantial model reduction. We see the model reduction problem as follows: regulatory networks have built-in signalling cascades that transmit signals between key nodes. In many cases, researchers are aware of components (typically kinases) of these signalling cascades, but measuring the kinetics of all these components is not feasible. It is thus tempting to skip these intermediate steps, but this solution is problematic since these intermediate steps introduce time delays that can be important for whole-network behaviour, such as oscillations. In the case of the discussed cascade $S \rightarrow K_1 \rightarrow K_2 \rightarrow K_3 \rightarrow K_4$ triggered in $t=0$ from state $K_1 = K_2 = K_3 = K_4 = 0$ by the step signal $S(t)$, the first, second and third time derivatives of K_4 (at $t=0$) are zero. When the cascade is reduced to $S \rightarrow K_1 \rightarrow K_4$, the second derivative of K_4 becomes positive, which noticeably changes the system's response, as shown in Fig. 5. To maintain observed dynamics (in particular oscillations) one can introduce nonlinearities, but as a consequence, the resulting model misses some components and has nonlinear terms not supported by underlying biochemistry^{32,33}.

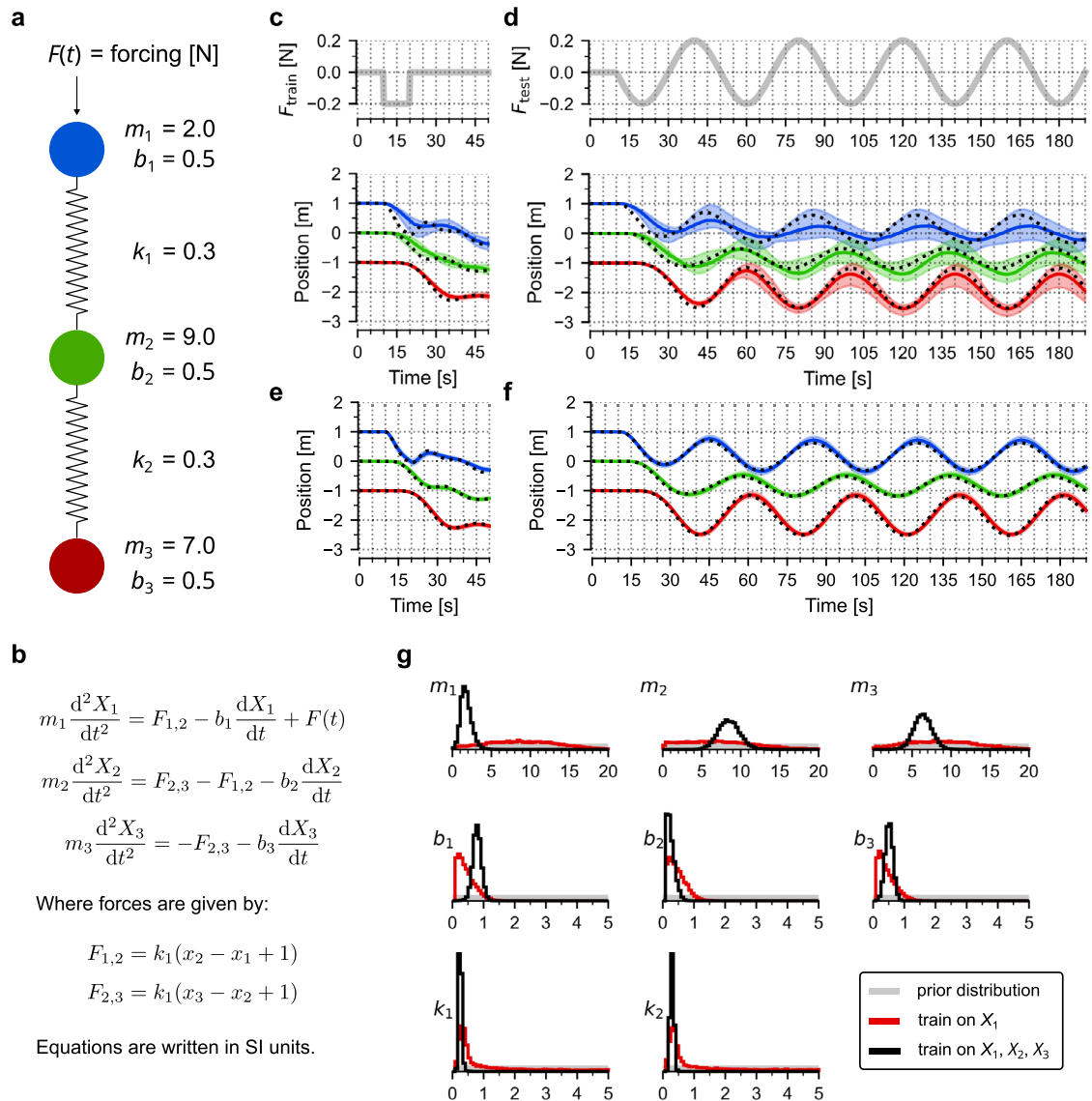


Figure 6. Three-body harmonic oscillator. **(a)** Scheme of the model. All eight parameters (masses, m_i ; damping coefficients b_i ; spring constants k_i) to be estimated are given together with their nominal values. Before the time-dependent force $F(t)$ is applied, the system remains in equilibrium. **(b)** Model equations. **(c,d)** Training **(c)** and prediction **(d)**, with only the trajectory of the red mass included in the training. As in the signalling cascade model, for training, we use three independently generated measurement replicates with Gaussian noise (sigma $\sigma_{\text{error}} = 0.3$) added to the deterministic trajectory (black dotted line). Measurement points are at full seconds. Coloured lines show (point-wise) medians of predictions, contours show 80% prediction bands. **(e,f)** Training **(e)** and prediction **(f)**, with trajectories of all masses included in the training, same convention as in **(c,d)**. **(g)** Histograms were obtained from 10,000 samples generated from the posterior distribution of model parameters after training on red mass trajectory (red line) and after training on all masses trajectories (black line). Grey contours show (uniform) prior distributions.

In response to the outlined problems with model reduction, we studied an alternative approach based on training the full mechanistic model on the available dataset and assessing its predictive power². We showed that a very limited dataset, insufficient to constrain any model parameters, can still allow for satisfactory predictions^{10,11}. In particular, we demonstrated that by training the signalling cascade model on the trajectory of the last component of the cascade (obtained in response to some properly chosen training protocol), one can unambiguously predict its trajectory in response to other protocols, as well as parameter changes. This prediction is possible despite the fact that trajectories of all upstream variables remain unknown.

Investigation of the plausible parameter set allows for additional insight. In the considered example we observe that when training is performed on trajectories of n variables, the ‘effective’ dimensionality of the plausible parameter set is reduced by n , and this reduction suffices to predict trajectories of the measured variables in response to other protocols. As discussed by Machta et al.³⁴, the compression of the parameter space identifies stiff directions corresponding to composite parameters controlling observed variables. By picking a suitable basis for

the stiff space, we confirmed our intuition that ratios of activation/deactivation parameters govern the dynamics of the cascade. Since stiff directions usually do not align with model parameters (as it is in our case), Gutenkunst et al. argued that in sloppy models, measuring system dynamics is an ineffective way of deriving parameter values, and conversely, measuring parameter values is an ineffective way of predicting system dynamics². In fact, even after training on trajectories of all model variables, most model parameters remained unidentified.

Model training can be generally understood as exploring the parameter space, in order to find plausible parameters, i.e. parameters leading to model trajectories in agreement with available data. In this context, many exploration methods are available. Locally, information about the uncertainty of parameters can be calculated using Fisher's information matrix. For models with asymmetric but still unimodal loss landscapes, the (semi-local) profile likelihood method can be used. For arbitrary loss landscapes, a Bayesian approach together with MCMC exploration is convenient, as

1. it allows for the inclusion of prior knowledge; if no such knowledge exists, training can start from some broad uninformative prior,
2. it allows for the analysis of how much a model has learnt by including a given dataset; this can be expressed by the reduction of the 'effective' dimensionality of the plausible parameter space, or entropy of the distribution of plausible parameters,
3. it allows us to assess the predictive power of the trained model, by sampling the posterior parameter distribution and computing prediction bands for variable trajectories.

The main disadvantage of MCMC compared to local (or semi-local) methods is that it can become computationally expensive, especially for high-dimensional parameter spaces²⁸. The computational cost is typically lower if exploration starts from a relatively narrow (informative) prior, or a more complete dataset is available, which results in a narrower posterior distribution³⁵. A promising method of accelerating MCMC exploration is based on reducing expensive model evaluations by local approximation of the likelihood^{36,37}.

In summary, the advantage of the studied approach to non-identifiability is twofold.

1. It does not require pre-emptive model reduction to avoid unmeasured variables in order to obtain predictions. Certainty of the predictions obtained from the full model can be determined by predictive bands, and if these predictions are insufficiently narrow, the model can be further trained using additional measurements of the variables of interest. Notably, these predictions are more reliable than predictions obtained from simplified models. While predictions from the full model may have wide predictive bands, the predictions from simplified models can be misleading (inaccurate with narrow bands) due to introduced simplifications.
2. If at a later time, measurements of previously unobserved variables are performed, no modification of the model structure is necessary. In contrast, reduced models may not be able to accommodate new data, simply because some components were omitted during reduction.

We hope that this approach will allow for easier collaboration, as it allows researchers to jointly work on the same model, improving it by performing new experiments².

Identifiable models or theories stem from physics, where they showed their superiority over other approaches. Whole branches of physics are based on a small number of fundamental theories. In systems biology, the situation is different: there are a lot of not-so-important models, typically having a large number of parameters that can neither be (feasibly) derived from more fundamental theories nor effectively measured. Currently, a very intensively developed approach to complex systems is based on neural networks. In this approach, one gets little to no insight into how the system works but can get very good predictions. In this sense, Bayesian methods are a middle ground between classical, identifiable models and black box models such as neural networks. The Bayesian approach studied here works well in our case because we have a good prior on the model structure, even if the prior on model parameters is broad and uninformative. This approach can be impractical for very complex systems for which neural networks are typically used, but is able to both give useful predictions and obtain some insight into the underlying regulatory system. Even if the explanatory power of the trained non-identifiable model is limited by the uncertainty of its parameters, its predictive power can make it useful (see also discussion in Refs.^{2,38}).

We think that the discussed approach, based on model training on some selected variables and giving only partial insight into the system's behaviour, may reflect true uncertainties in underlying biology^{39,40}. Considering a regulatory network one can identify core nodes. Because of their importance, the corresponding variables are frequently measured, and for the same reason one can expect relatively good replicability of these measurements. The intermediate pathway components are less tightly constrained and may exhibit higher variability⁴¹. Daniels et al. argued that cell-to-cell variability in protein levels might be beneficial for robustness and evolvability³⁹. Variability of intermediate proteins implies that (for very practical reasons) these variables are not (frequently) measured; or even if they are measured, results are (understandably) not reported due to insufficient replicability. One can thus hypothesise, that in the case of intensively investigated regulatory systems, the remaining uncertainty of model parameters and of trajectories of intermediate components may reflect their true variability, at least at the single cell level. One may hope that quantification of the variability of intermediate variables, and correlation between them will help to elucidate existing compensatory properties ensuring robustness of the whole regulatory network.

Methods

Computational model simulations. Each model is represented by a set of ODEs and an input signal $S(t)$, which are solved using an explicit Runge–Kutta (4,5) formula developed by Dormand-Prince. Model simulations are used first to generate in silico measurements, and then to estimate parameter distributions based on these measurements. A measurement trajectory is evaluated at full hours (or at full seconds in the case of the mechanistic analogue model shown in Fig. 6) by adding to the ODE solution Gaussian noise (with sigma $\sigma_{\text{error}} = 0.3$) that emulates measurement error. Parameter estimation (model training) is based on 3 independent measurement trajectories.

In the case of biochemical cascade models, ODE trajectory $x(t)$ is log-transformed to $y(t) = \log_{10}(0.001 + x)$ before adding measurement noise.

Parameter estimation based on MCMC simulations. For each of the considered models we assume some prior parameter distribution $P(\theta)$ and use an MCMC algorithm to draw samples from the posterior parameter distribution $P(\theta | M)$, given in silico generated measurements M .

To estimate $P(\theta | M)$, we first calculate the likelihood of θ given a single measurement $L_{k,t,r}(\theta) = P(m_{k,t,r} | \theta)$ (where k is the measured variable, t is the time of measurement, and $r = 1, 2, 3$ is the measurement replicate) by simulating variable trajectories $y_k(t)$ for the parameter set θ . Then

$$L_{k,t,r}(\theta) = \frac{1}{\sigma_{\text{error}} \sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma_{\text{error}}^2}(y_k(t) - m_{k,t,r})^2\right),$$

where σ_{error} is the measurement error (assumed 0.3 for all our models). The total likelihood given all measurements of variable k is thus:

$$L_k(\theta) = \prod_{t,r} L_{k,t,r}(\theta),$$

and the posterior distribution $P(\theta | M)$ is proportional to

$$f(\theta) = P(\theta)P(M|\theta) = P(\theta) \prod_k L_k(\theta),$$

where the product is over variables k selected for model training. Then $f(\theta)$ is used in MCMC simulations to draw samples from $P(\theta | M)$.

MCMC simulations are performed using the Metropolis–Hastings algorithm with Gaussian jumping distribution. We use 8 parallel chains, each with 1.5 million steps from which the first 0.25 million are removed (burn-in). This gives altogether 10 million parameter samples from which we retain every thousandth sample, i.e. 10,000 samples for model simulations. We compared the between- and within-chain variance estimates using the $\hat{r}$ -hat convergence statistics to find $\hat{r}$ -hat < 1.01, suggesting a sufficient length of chains. The effective sample size of the final sample was found to be at least 1000.

For the biochemical cascade models, we assume wide log-normal priors and thus use log-parameters during MCMC sampling. After this reparameterization, we use a spherically symmetric Gaussian jumping distribution with independent variables and $\sigma = 0.2$ (in the log scale). For the mechanical model, we use an elliptical Gaussian jumping distribution with $\sigma = 1.0$ for masses, $\sigma = 0.2$ for damping coefficients, and $\sigma = 0.4$ for stiffness coefficients. For the final fit (trajectories of all three masses observed) we decrease step size by a factor of 4.

To minimise biases we select wide, non-informative priors, however, it is worth noting that in practical applications informative priors can significantly increase model performance.

In summary, our statistical model can be itemised as follows:

1. $\theta \sim \text{Prior distribution}$.
2. $x(t)$ = Model ODEs solution for parameters θ and initial condition $x(0) = x_0$.
3. $y(t) = \text{softlog}(x(t)) := \log(0.001 + x(t))$ (only for signalling cascade).
4. $m_{k,t,r} \sim \mathcal{N}(y_k(t), \sigma_{\text{error}})$.

where θ are the model parameters, σ_{error} is the measurement error and $m_{k,t,r}$ is the r -th measurement of variable k at time t , drawn from normal distribution $\mathcal{N}(y_k(t), \sigma_{\text{error}})$. Initial conditions for the ODE are kept constant and excluded from estimation. The *softlog* normalisation step is omitted for spring models. For all our models $\sigma_{\text{error}} = 0.3$.

Principal component analysis. In Fig. 2F, Fig. S1 and Fig. S2D we use principal component analysis (PCA) to quantitatively describe the dimensionality reduction of the plausible parameter set. This approach works if the analysed set of points is “sufficiently flat”. Points on a half circle would have two big PC values, even though the space is clearly one-dimensional. In such a case dimensionality can be investigated locally, as proposed by Little et al.⁴². In the half-circle example, this local PCA analysis would result in one small and one big PC value (for a sufficiently small spatial scale). We have verified that in our case local estimates result in higher PC values than the global ones; we thus conclude that our point clouds are sufficiently flat to justify using global PCA analysis.

Implementation. We implemented our models in Python using *jax*⁴³ and *diffax*⁴⁴ (for ODE integration), and the MCMC algorithm using *blackjax*⁴⁵.

Data availability

No experimental datasets were generated or analysed during the current study. All numerical results were obtained using code available at <https://github.com/grfrederic/identifiability>.

Code availability

Our source code is available at <https://github.com/grfrederic/identifiability>.

Received: 7 April 2023; Accepted: 29 June 2023

Published online: 10 July 2023

References

1. Brown, K. S. & Sethna, J. P. Statistical mechanical approaches to models with many poorly known parameters. *Phys. Rev. E* **68**, 021904 (2003).
2. Gutenkunst, R. N. *et al.* Universally sloppy parameter sensitivities in systems biology models. *PLoS Comput. Biol.* **3**, e189 (2007).
3. Raue, A. *et al.* Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics* **25**, 1923–1929 (2009).
4. Wieland, F.-G., Hauber, A. L., Rosenblatt, M., Tönsing, C. & Timmer, J. On structural and practical identifiability. *Curr. Opin. Syst. Biol.* **25**, 60–69 (2021).
5. Jaruszewicz-Błońska, J., Kosiuk, I., Prus, W. & Lipniacki, T. A plausible identifiable model of the canonical NF- κ B signaling pathway. *PLoS ONE* **18**, e0286416 (2023).
6. Roosa, K. & Chowell, G. Assessing parameter identifiability in compartmental dynamic models using a computational approach: Application to infectious disease transmission models. *Theor. Biol. Med. Model.* **16**, 1 (2019).
7. Chowell, G. Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts. *Infect. Dis. Model.* **2**, 379–398 (2017).
8. Transtrum, M. K. & Qiu, P. Model reduction by manifold boundaries. *Phys. Rev. Lett.* **113**, 098701 (2014).
9. Monsalve-Bravo, G. M. *et al.* Analysis of sloppiness in model simulations: Unveiling parameter uncertainty when mathematical models are fitted to data. *Sci. Adv.* **8**, 5952 (2022).
10. Brown, K. S. *et al.* The statistical mechanics of complex signaling networks: Nerve growth factor signaling. *Phys. Biol.* **1**, 184 (2004).
11. Cedersund, G. Conclusions via unique predictions obtained despite unidentifiability—New definitions and a general method. *FEBS J.* **279**, 3513–3527 (2012).
12. van Mourik, S., ter Braak, C., Stigter, H. & Molenaar, J. Prediction uncertainty assessment of a systems biology model requires a sample of the full probability distribution of its parameters. *PeerJ* **2**, e433 (2014).
13. Villaverde, A. F. & Banga, J. R. Reverse engineering and identification in systems biology: Strategies, perspectives and challenges. *J. R. Soc. Interface* **11**, 20130505 (2014).
14. Casey, F. P. *et al.* Optimal experimental design in an epidermal growth factor receptor signalling and down-regulation model. *IET Syst. Biol.* **1**, 190–202 (2007).
15. Liepe, J., Filippi, S., Komorowski, M. & Stumpf, M. P. H. Maximizing the information content of experiments in systems biology. *PLoS Comput. Biol.* **9**, e1002888 (2013).
16. Fox, Z. R. & Munsky, B. The finite state projection based Fisher information matrix approach to estimate information and optimize single-cell experiments. *PLoS Comput. Biol.* **15**, e1006365 (2019).
17. Lake, D., Corrêa, S. A. L. & Müller, J. Negative feedback regulation of the ERK1/2 MAPK pathway. *Cell Mol. Life Sci.* **73**, 4397–4413 (2016).
18. Kochańczyk, M. *et al.* Relaxation oscillations and hierarchy of feedbacks in MAPK signaling. *Sci. Rep.* **7**, 38244 (2017).
19. Hoffmann, A., Levchenko, A., Scott, M. L. & Baltimore, D. The I κ B-NF- κ B signaling module: Temporal control and selective gene activation. *Science* **298**, 1241–1245 (2002).
20. Lipniacki, T., Paszek, P., Brasier, A. R., Luxon, B. & Kimmel, M. Mathematical model of NF- κ B regulatory module. *J. Theor. Biol.* **228**, 195–215 (2004).
21. Harris, S. L. & Levine, A. J. The p53 pathway: Positive and negative feedback loops. *Oncogene* **24**, 2899–2908 (2005).
22. Hat, B., Kochańczyk, M., Bogdał, M. N. & Lipniacki, T. Feedbacks, bifurcations, and cell fate decision-making in the p53 system. *PLoS Comput. Biol.* **12**, e1004787 (2016).
23. Toettcher, J. E., Weiner, O. D. & Lim, W. A. Using optogenetics to interrogate the dynamic control of signal transmission by the Ras/Erk module. *Cell* **155**, 1422–1434 (2013).
24. Nałęcz-Jawecki, P. *et al.* The MAPK/ERK channel capacity exceeds 6 bit/hour. *PLoS Comput. Biol.* **19**, e1011155 (2023).
25. Jeong, J. E. & Qiu, P. Quantifying the relative importance of experimental data points in parameter estimation. *BMC Syst. Biol.* **12**, 103 (2018).
26. Valderrama-Bahamóndez, G. I. & Fröhlich, H. MCMC techniques for parameter estimation of ODE based models in systems biology. *Front. Appl. Math. Stat.* **5**, 55 (2019).
27. Linden, N. J., Kramer, B. & Rangamani, P. Bayesian parameter estimation for dynamical models in systems biology. *PLoS Comput. Biol.* **18**, e1010651 (2022).
28. Hamon, J., Jennings, P. & Bois, F. Y. Systems biology modeling of omics data: Effect of cyclosporine a on the Nrf2 pathway in human renal cells. *BMC Syst. Biol.* **8**, 76 (2014).
29. Tavaré, S., Balding, D. J., Griffiths, R. C. & Donnelly, P. Inferring coalescence times from DNA sequence data. *Genetics* **145**, 505 (1997).
30. Mikelson, J. & Khammash, M. Likelihood-free nested sampling for parameter inference of biochemical reaction networks. *PLoS Comput. Biol.* **16**, e1008264 (2020).
31. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087–1092 (1953).
32. Geva-Zatorsky, N. *et al.* Oscillations and variability in the p53 system. *Mol. Syst. Biol.* **2**, 20060033 (2006).
33. Krishna, S., Jensen, M. H. & Sneppen, K. Minimal model of spiky oscillations in NF- κ B signaling. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 10840–10845 (2006).
34. Machta, B. B., Chachra, R., Transtrum, M. K. & Sethna, J. P. Parameter space compression underlies emergent theories and predictive models. *Science* **342**, 604–607 (2013).
35. Raue, A., Kreutz, C., Theis, F. J. & Timmer, J. Joining forces of Bayesian and frequentist methodology: A study for inference in the presence of non-identifiability. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **371**, 20110544 (2013).
36. Fielding, M., Nott, D. J. & Liang, S.-Y. Efficient MCMC schemes for computationally expensive posterior distributions. *Technometrics* **53**, 16–28 (2011).
37. Conrad, P. R., Marzouk, Y. M., Pillai, N. S. & Smith, A. Accelerating asymptotically exact MCMC for computationally intensive models via local approximations. *J. Am. Stat. Assoc.* **111**, 1591–1607 (2016).

38. Cedersund, G. & Roll, J. Systems biology: Model based evaluation and comparison of potential explanations for given biological data. *FEBS J.* **276**, 903–922 (2009).
39. Daniels, B. C., Chen, Y.-J., Sethna, J. P., Gutenkunst, R. N. & Myers, C. R. Sloppiness, robustness, and evolvability in systems biology. *Curr. Opin. Biotechnol.* **19**, 389–395 (2008).
40. Sturm, O. E. *et al.* The mammalian MAPK/ERK pathway exhibits properties of a negative feedback amplifier. *Sci. Signal* **3**, 90 (2010).
41. Sigal, A. *et al.* Variability and memory of protein levels in human cells. *Nature* **444**, 643–646 (2006).
42. Little, A., Maggioni, M. & Rosasco, L. *Multiscale Geometric Methods for Estimating Intrinsic Dimension* (2011).
43. Bradbury, J. *et al.* *JAX: Composable Transformations of Python+NumPy Programs* (2018).
44. Kidger, P. *On Neural Differential Equations* (University of Oxford, 2021).
45. Lao, J. & Louf, R. *Blackjax: A Sampling Library for JAX* (2020).

Acknowledgements

This study was funded by National Science Centre Poland Grants 2019/34/H/NZ6/00699 (Norwegian Financial Mechanism) and 2019/35/B/NZ2/03898. The authors thank Dr. Marek Kočańczyk for the discussions.

Author contributions

F.G., P.N.J. and T.L. designed the study, F.G. wrote numerical codes, performed simulations and prepared figures. F.G. and T.L. with P.N.J.'s assistance wrote the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-37939-8>.

Correspondence and requests for materials should be addressed to T.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

Beyond Normal: On the Evaluation of Mutual Information Estimators

Paweł Czyż^{*1,2} Frederic Grabowski^{*3}

Julia E. Vogt^{4,5} Niko Beerenwinkel^{†1,5} Alexander Marx^{†2,4}

¹Department of Biosystems Science and Engineering, ETH Zurich ²ETH AI Center, ETH Zurich

³Institute of Fundamental Technological Research, Polish Academy of Sciences

⁴Department of Computer Science, ETH Zurich ⁵SIB Swiss Institute of Bioinformatics

Abstract

Mutual information is a general statistical dependency measure which has found applications in representation learning, causality, domain generalization and computational biology. However, mutual information estimators are typically evaluated on simple families of probability distributions, namely multivariate normal distribution and selected distributions with one-dimensional random variables. In this paper, we show how to construct a diverse family of distributions with known ground-truth mutual information and propose a language-independent benchmarking platform for mutual information estimators. We discuss the general applicability and limitations of classical and neural estimators in settings involving high dimensions, sparse interactions, long-tailed distributions, and high mutual information. Finally, we provide guidelines for practitioners on how to select appropriate estimator adapted to the difficulty of problem considered and issues one needs to consider when applying an estimator to a new data set.

1 Introduction

Estimating the strength of a non-linear dependence between two continuous random variables (r.v.) lies at the core of machine learning. Mutual information (MI) lends itself naturally to this task, due to its desirable properties, such as invariance to homeomorphisms and the data processing inequality. It finds applications in domain generalization [Li et al., 2022, Ragonesi et al., 2021], representation learning [Belghazi et al., 2018, Oord et al., 2018], causality [Solo, 2008, Kurutach et al., 2018], physics [Keys et al., 2015], systems biology [Selimkhanov et al., 2014, Grabowski et al., 2019, Uda, 2020], and epidemiology [Young et al., 2023].

Over the last decades, the estimation of MI has been extensively studied, and estimators have been developed, ranging from classical approaches based on histogram density [Pizer et al., 1987], kernel density estimation [Moon et al., 1995] to k -nearest neighbor [Kozachenko and Leonenko, 1987, Kraskov et al., 2004] and neural estimators [Belghazi et al., 2018, Oord et al., 2018, Song and Ermon, 2020]. However, despite the progress that has been achieved in this area, not much attention has been focused on systematically benchmarking these approaches.

Typically, new estimators are evaluated assuming multivariate normal distributions for which MI is analytically tractable [Darbellay and Vajda, 1999, Kraskov et al., 2004, Suzuki, 2016]. Sometimes, simple transformations are applied to the data [Khan et al., 2007, Gao et al., 2015], moderately high-dimensional settings are considered [Lord et al., 2018, Lu and Peltonen, 2020], or strong dependencies are evaluated [Gao et al., 2015]. Beyond that, Song and Ermon [2020] study the

^{*}Equal contribution [†]Joint supervision

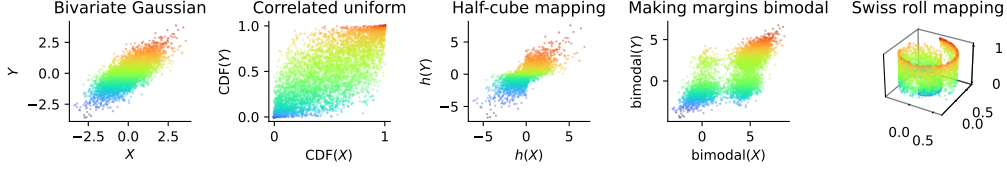


Figure 1: Visualisations of selected proposed distributions. Two correlated Gaussian variables X and Y (1) can be transformed via the Gaussian CDF into correlated uniform variables (2), into a long-tailed distribution via the “half-cube” mapping $t \mapsto t\sqrt{|t|}$ (3), into a multi-modal distribution (4) or embedded in the three-dimensional space via the composition of the Swiss roll mapping and Gaussian CDF (5). Color on the plot corresponds to the original Y variable.

self-consistency (additivity under independence and the data processing inequality) of neural MI estimators in the context of representation learning from image data.

Our Contributions In this work we show a method of developing expressive distributions with known ground-truth mutual information (Sec. 2), propose forty benchmark tasks and systematically study the properties of commonly used estimators, including representatives based on kernel or histogram density estimation, k NN estimation, and neural network-based estimators (Sec. 3). We address selected difficulties one can encounter when estimating mutual information (Sec. 4), such as sparsity of interactions, long-tailed distributions, invariance, and high mutual information. Finally, we provide recommendations for practitioners on how to choose a suitable estimator for particular problems (Sec. 6). Our benchmark is designed so that it is simple to add new tasks and estimators. It also allows for cross-language comparisons — in our work we compare estimators implemented in Python, R and Julia. All of our experiments are fully reproducible by running Snakemake workflows [Mölder et al., 2021]. Accompanying code is available at <http://github.com/cbg-ethz/bmi>.

Overall, our *key findings* from the benchmark can be summarized as follows:

- Testing on multivariate normal distributions gives a biased and overly optimistic view of estimator performance. In this setting, canonical correlation analysis (CCA), a model-based approach, emerges as the most effective method — even if the model is slightly misspecified.
- Compared to classical estimators, neural estimators excel in high-dimensional settings, capturing sparse interactions across multiple dimensions.
- The popular KSG estimator [Kraskov et al., 2004] is accurate in low- and moderate-dimensional settings, but its performance suffers on problems involving high-dimensions or sparse interactions.
- Although MI is invariant to a wide range of transformations (Theorem 2.1), numerical estimates, even with large sample sizes, are not. Thus, MI may not be suitable when metric invariance to specific transformation families is required.
- Multivariate Student distributions pose an important and hard challenge to many mutual information estimators. This effect is partially, but not fully, attributed to their long tails.

2 Mutual Information: Estimation and Invariance

We start by recalling the definition of MI, then discuss the problem of estimating it, and introduce the estimators used in the benchmark.

Consider two r.v. X and Y with domains $\mathcal{X}$ respectively $\mathcal{Y}$, joint probability measure P_{XY} and marginal probability measures P_X and P_Y , respectively. If P_{XY} is absolutely continuous with respect to $P_X \otimes P_Y$ (e.g., X and Y are finite r.v.), MI is equal to the following Kullback–Leibler divergence²:

$$\mathbf{I}(X; Y) = \mathbf{D}_{\text{KL}}(P_{XY} \parallel P_X \otimes P_Y) = \int \log f \, dP_{XY},$$

where $f = dP_{XY}/d(P_X \otimes P_Y)$ is the Radon–Nikodym derivative. If the absolute continuity does not hold, then $\mathbf{I}(X; Y) = +\infty$ [Pinsker and Feinstein, 1964, Theorem 2.1.2]. If $\mathcal{X}$ and $\mathcal{Y}$ are Euclidean spaces and measures P_{XY} , P_X and P_Y have probability density functions (PDFs) with respect to the Lebesgue measure, then the Kullback–Leibler divergence can be written in terms of the PDFs.

²We use the natural logarithm, meaning that the mutual information is measured in *nats*.

However, we will later consider distributions which do not have PDFs with respect to the Lebesgue measure and the general Radon–Nikodym derivative must be used (see “Swiss roll” in Fig. 1).

In almost all applications, P_{XY} and $P_X \otimes P_Y$ are not known and one needs to estimate MI from a finite sample from P_{XY} , i.e., a realization $((x_1, y_1), \dots, (x_N, y_N)) \in (\mathcal{X} \times \mathcal{Y})^N$ of N i.i.d. r.v. $(X_1, Y_1), \dots, (X_N, Y_N)$ distributed according to the joint distribution P_{XY} . As a MI estimator we will denote a family of measurable mappings, indexed by the sample size N and spaces $\mathcal{X}$ and $\mathcal{Y}$ (in most applications, Euclidean spaces of various dimensionalities): $e_N^{\mathcal{X}\mathcal{Y}}: (\mathcal{X} \times \mathcal{Y})^N \rightarrow \mathbb{R}$. Composing this mapping with the r.v. representing the whole data set we obtain a real-valued r.v. $E_N^{\mathcal{X}\mathcal{Y}}$. For a given P_{XY} we can consider the distribution of $E_N^{\mathcal{X}\mathcal{Y}}$. It is often summarized by its first two moments, resulting in the bias and the variance of the estimator. Understanding the bias of an estimator however requires the knowledge of ground-truth MI, so the estimators are typically tested on the families of jointly multivariate normal or uniform distributions, with varying N and ground-truth $\mathbf{I}(X; Y)$ [Khan et al., 2007, Lord et al., 2018, Holmes and Nemenman, 2019, Song and Ermon, 2020].

Constructing Expressive Distributions and Benchmarking Besides considering different types of distributions such as (multivariate) normal, uniform and Student, our benchmark is based on the invariance of MI to the family of chosen mappings. Generally MI is not invariant to arbitrary transformations, as due to the data processing inequality $\mathbf{I}(f(X); g(Y)) \leq \mathbf{I}(X; Y)$. However, under some circumstances MI is preserved, which enables us to create more expressive distributions by transforming the variables.

Theorem 2.1. *Let $\mathcal{X}$, $\mathcal{X}'$, $\mathcal{Y}$ and $\mathcal{Y}'$ be standard Borel spaces (e.g., smooth manifolds with their Borel σ -algebras) and $f: \mathcal{X} \rightarrow \mathcal{X}'$ and $g: \mathcal{Y} \rightarrow \mathcal{Y}'$ be continuous injective mappings. Then, for every $\mathcal{X}$ -valued r.v. X and $\mathcal{Y}$ -valued r.v. Y it holds that $\mathbf{I}(X; Y) = \mathbf{I}(f(X); g(Y))$.*

This theorem has been studied before by Kraskov et al. [2004, Appendix], who consider diffeomorphisms and by Polyanskiy and Wu [2022, Th. 3.7], who assume measurable injective functions with measurable left inverses. For completeness we provide a proof which covers continuous injective mappings in Appendix A. In particular, each of f and g can be a homeomorphism, a diffeomorphism, or a topological embedding. Embeddings are allowed to increase the dimensionality of the space, so even if X and Y have probability density functions, $f(X)$ and $g(Y)$ do not need to have them. Neither of these deformations change MI. Thus, using Theorem 2.1, we can create more expressive distributions $P_{f(X)g(Y)}$ by sampling from P_{XY} and transforming the samples to obtain a data set $((f(x_1), g(y_1)), \dots, (f(x_N), g(y_N))) \in (\mathcal{X}' \times \mathcal{Y}')^N$.

While offering a tool to construct expressive distributions, another valuable perspective to consider with this problem is estimation invariance: although MI is invariant to proposed changes, the estimate may not be. More precisely, applying an estimator to the transformed data set results in the induced r.v. $E_N^{f(X)g(Y)}$ and its distribution. If the estimator were truly invariant, one would expect that $\mathbb{E}[E_N^{f(X)g(Y)}]$ should equal $\mathbb{E}[E_N^{XY}]$. This invariance is rarely questioned and often implicitly assumed as given (see e.g. Tschannen et al. [2020] or Murphy [2023, Sec. 32.2.2.3]), but as we will see in Sec. 4.3, finite-sample estimates are not generally invariant.

3 Proposed Benchmark

In this section, we outline forty tasks included in the benchmark that cover a wide family of distributions, including varying tail behaviour, sparsity of interactions, multiple modes in PDF, and transformations that break colinearity. As our base distributions, we selected multivariate normal and Student distributions, which were transformed with continuous injective mappings.

We decided to focus on the following phenomena:

1. **Dimensionality.** High-dimensional data are collected in machine learning and natural sciences [Bühlmann and van de Geer, 2011, Ch. 1]. We therefore change the dimensions of the $\mathcal{X}$ and $\mathcal{Y}$ spaces between 1 and 25.
2. **Sparsity.** While the data might be high-dimensional, the effective dimension may be much smaller, due to correlations between different dimensions or sparsity [Lucas et al., 2006]. We therefore include distributions in which some of the dimensions represent random noise,

which does not contribute to the mutual information, and perform an additional study in Section 4.1.

3. **Varying MI.** Estimating high MI is known to be difficult [McAllester and Stratos, 2020]. However, we can often bound it in advance — if there are 4000 image classes, MI between image class and representation is at most $\log 4000 \approx 8.3$ nats. In this section, we focus on problems with MI varying up to 2 nats. We additionally consider distributions with higher MI in Section 4.4.
4. **Long tails.** As Taleb [2020] and Zhang et al. [2021] argue, many real-world distributions have long tails. To model different tails we consider multivariate Student distributions as well as transformations lengthening the tails. We conduct an additional study in Section 4.2.
5. **Robustness to diffeomorphisms.** As stated in Theorem 2.1, mutual information is invariant to reparametrizations of random variables by diffeomorphisms. We however argue that when only finite samples are available, invariance in mutual information estimates may not be achieved. To test this hypothesis we include distributions obtained by using a diverse set of diffeomorphisms and continuous injective mappings. We further study the robustness to reparametrizations in Section 4.3.

While we provide a concise description here, precise experimental details can be found in Appendix D and we visualise selected distributions in Appendix F. We first describe tasks employing one-dimensional variables.

Bivariate Normal For multivariate normal variables MI depends only on the correlation matrix. We will therefore consider a centered bivariate normal distribution $P_{X,Y}$ with $\text{Cor}(X, Y) = \rho$ and $\mathbf{I}(X; Y) = -0.5 \log(1 - \rho^2)$. We chose $\rho = 0.75$.

Uniform Margins As a first transformation, we apply the Gaussian CDF F to Gaussian r.v. X and Y to obtain r.v. $X' = F(X)$ and $Y' = F(Y)$. It is a standard result in copula theory (see Lemma D.2 or e.g., Nelsen [2007] for an introduction to copulas) that $F(X) \sim \text{Uniform}(0, 1)$ resp. $F(Y) \sim \text{Uniform}(0, 1)$. The joint distribution $P_{X',Y'}$, however, is not uniform. Mutual information is preserved, $\mathbf{I}(X'; Y') = \mathbf{I}(X; Y)$. For an illustration, see Fig. 1. A distribution P transformed with Gaussian CDF F will be denoted as `Normal CDF @ P`.

Half-Cube Map To lengthen the tails we applied the half-cube homeomorphism $h(x) = |x|^{3/2} \text{sign } x$ to Gaussian variables X and Y . We visualize an example in Fig. 1 and denote a transformed distribution as `Half-cube @ P`.

Asinh Mapping To shorten the tails we applied inverse hyperbolic sine function $\text{asinh } x = \log(x + \sqrt{1 + x^2})$. A distribution transformed with this mapping will be denoted as `Asinh @ P`.

Wiggly Mapping To model non-uniform lengthscales, we applied a mapping

$$w(x) = x + \sum_i a_i \sin(\omega_i x + \varphi_i), \quad \sum_i |a_i \omega_i| < 1.$$

Due to the inequality constraint this mapping is injective and preserves MI. The parameter values can be found in Appendix D. We denote the transformed distribution as `Wiggly @ P`.

Bimodal Variables Applying the inverse CDF of a two-component Gaussian mixture model to correlated variables X and Y with uniform margins we obtained a joint probability distribution $P_{X',Y'}$ with four modes (presented in Fig. 1). We call this distribution `Bimodal` and provide technical details in Appendix D.

Additive Noise Consider independent r.v. $X \sim \text{Uniform}(0, 1)$ and $N \sim \text{Uniform}(-\varepsilon, \varepsilon)$, where ε is the noise level. For $Y = X + N$, it is possible to derive $\mathbf{I}(X; Y)$ analytically (see Appendix D for the formula and the parameter values) and we will call this distribution `Uniform (additive noise= $\varepsilon$ )`.

Swiss Roll Embedding As the last example in this section, we consider a mixed case in which a one-dimensional r.v. $X \sim \text{Uniform}(0, 1)$ is smoothly embedded into two dimensions via the Swiss roll mapping, a popular embedding used to test dimensionality reduction algorithms (see Fig. 1 for visualisation and Appendix D for the formula). Note that the obtained distribution does not have a PDF with respect to the Lebesgue measure on $\mathbb{R}^3$.

Next, we describe the tasks based on multivariate distributions.

Multivariate Normal We sampled $(X, Y) = (X_1, \dots, X_m, Y_1, \dots, Y_n)$ from the multivariate normal distribution. We model two correlation structures: “dense” interactions with all off-diagonal correlations set to 0.5 and “2-pair” interactions where $\text{Cor}(X_1, Y_1) = \text{Cor}(X_2, Y_2) = 0.8$ and there is no correlation between any other (distinct) variables (Multinormal (2-pair)). For a latent-variable interpretation of these covariance matrices we refer to Appendix D.2.

Multivariate Student To see how well the estimators can capture MI contained in the tails of the distribution, we decided to use multivariate Student distributions (see Appendix D) with dispersion matrix³ set to the identity I_{m+n} and ν degrees of freedom. Contrary to the multivariate normal case, in which the identity matrix used as the covariance would lead to zero MI, the variables X and Y can still interact through the tails.

Transformed Multivariate Distributions As the last set of benchmark tasks we decided to transform the multivariate normal and multivariate Student distributions. We apply mappings described above to each axis separately. For example, applying normal CDF to the multivariate normal distribution will map it into a distribution over the cube $(0, 1)^{m+n}$ with uniform marginals. To mix different axes we used a spiral diffeomorphism (denoted as `Spiral @ P`); we defer the exact construction to Appendix D but we visualise this transformation in Fig. 5.

Estimators In our benchmark, we include a diverse set of estimators. Following recent interest, we include four neural estimators, i.e., Donsker–Varadhan (D-V) and MINE [Belghazi et al., 2018], InfoNCE [Oord et al., 2018], and the NWJ estimator [Nguyen et al., 2007, Nowozin et al., 2016, Poole et al., 2019]. As a representative for model-based estimators, we implement an estimator based on canonical correlation analysis (CCA) [Murphy, 2023, Ch. 28], which assumes that the joint distribution is multivariate normal. For more classical estimators, we include the Kraskov, Stögbauer and Grassberger (KSG) estimator [Kraskov et al., 2004] as the most well-known k -nearest neighbour-based estimator, a recently proposed kernel-based estimator (LNN) [Gao et al., 2017b], as well as an estimator based on histograms and transfer entropy (both implemented in Julia’s `TransferEntropy` library). A detailed overview of the different classes of estimators is provided in Appendix C. Further, we describe the hyperparameter selection in Appendix E.

Preprocessing As a data preprocessing step, we standardize each dimension using the mean and standard deviation calculated from the sample. Other preprocessing strategies are also possible, but we did not observe a consistent improvement of one preprocessing strategy over the others (see Appendix E).

3.1 Benchmark Results

We show the results for $N = 10\,000$ data points in Fig. 2. Neural estimators and KSG perform better than alternative estimators on most problems, with KSG having often low sample requirements (see Appendix E). The simple model-based estimator CCA obtained excellent performance at low sample sizes (see Appendix E), even for slightly transformed multivariate normal distributions. Finally, LNN and the two Julia estimators (histogram and transfer entropy), work well in low dimension but are not viable in medium- and high-dimensional problems ($\dim X + \dim Y \geq 4$).

Overall, we observed that for most 1×1 -dimensional and multivariate normal problems MI can be reliably estimated. Difficulties arise for sparse interactions, atypical (Student) distributions, and for significantly transformed distributions. This suggests that the typical evaluations of MI estimators are overly optimistic, focusing on relatively simple problems.

KSG is able to accurately estimate MI in multivariate normal problems, however performance drops severely on tasks with sparse interactions (2-pair tasks). For 5×5 dimensions the estimate is about 70% of the true MI, and for 25×25 dimensions it drops to 10%, which is consistent with previous observations [Marx and Fischer, 2022]. Meanwhile, performance of neural estimators is stable. We study the effect of sparsity in more detail in Sec. 4.

Student distributions have proven challenging to all estimators. This is partially due to the fact that long tails can limit estimation quality (see Sec. 4), and indeed, applying a tail-shortening asinh transformation improves performance. However, performance still remains far below multivariate normal distributions with similar dimensionality and information content, showing that long tails are

³The dispersion matrix of multivariate Student distribution is different from its covariance, which does not exist for $\nu \leq 2$.

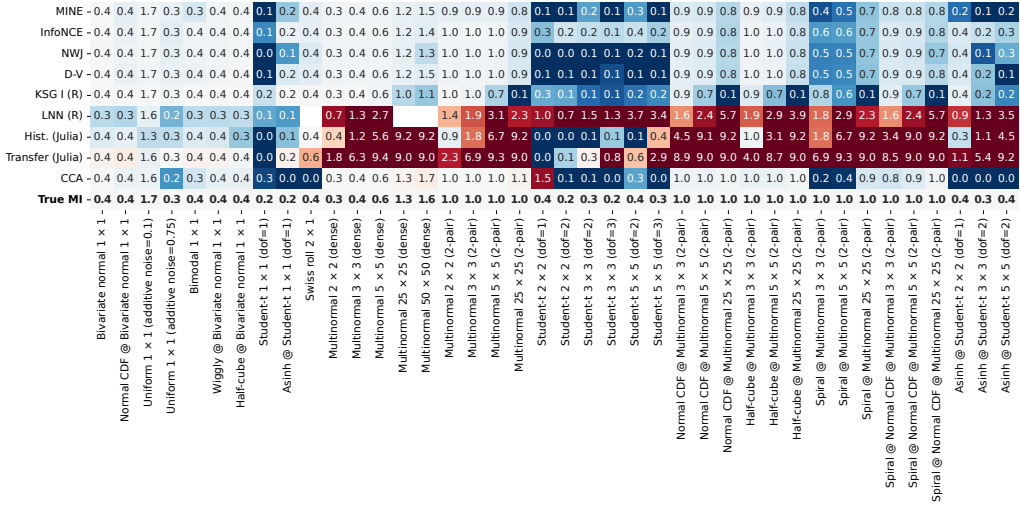


Figure 2: Mean MI estimates of nine estimators over $n = 10$ samples with $N = 10\,000$ points each against the ground-truth value on all benchmark tasks grouped by category. Color indicates relative negative bias (blue) and positive bias (red). Blank entries indicate that an estimator experienced numerical instabilities.

only part of the difficulty. For neural estimators, the issue could be explained by the fact that the critic is not able to fully learn the pointwise mutual information (see Appendix E).

Interestingly, the performance of CCA, which assumes a linear Gaussian model, is often favorable compared to other approaches and has very low sample complexity (see Appendix E). This is surprising, since in the case of transformed distributions the model assumptions are not met. The approach fails for Student distributions (for which the correlation matrix either does not exist or is the identity matrix), the Swiss roll embedding and multivariate normal distributions transformed using the spiral diffeomorphism. Since the spiral diffeomorphism proved to be the most challenging transformation, we study it more carefully in Sec. 4.

4 Selected Challenges

In this section we investigate distributions which proved to be particularly challenging for the tested estimators in our benchmark: sparse interactions, long tailed distributions and invariance to diffeomorphisms. Additionally we investigate how the estimators perform for high ground truth MI.

4.1 Sparsity of Interactions

In the main benchmark (Fig. 2) we observed that estimation of MI with “2-pair” type of interactions is considerably harder for the KSG estimator than the “dense” interactions. On the other hand, neural estimators were able to pick the relevant interacting dimensions and provide better estimates of MI in these sparse cases.

We decided to interpolate between the “dense” and “2-pair” interaction structures using a two-stage procedure with real parameters α and λ and the number of strongly interacting components K . While the precise description is in Appendix D.2, we can interpret α as controlling the baseline strength of interaction between every pair of variables in $\{X_1, \dots, X_{10}, Y_1, \dots, Y_{10}\}$ and λ as the additional strength of interaction between pairs of variables $(X_1, Y_1), \dots, (X_K, Y_K)$. Dense interaction structure has $\lambda = 0$ and 2-pair interaction structure has $\alpha = 0$ and $K = 2$. First, we set $K = 10$ and $\lambda \approx 0$ and decrease α raising λ at the same time to maintain constant MI of 1 nat. When $\alpha = 0$, whole information is contained in the pairwise interactions $(X_1, Y_1), \dots, (X_{10}, Y_{10})$. We then decrease K and increase λ , still maintaining constant MI.

In Fig. 3 we observe that the performance of all estimators considered (apart from CCA, a model-based approach suitable for multivariate normal distributions) degrades when the interactions between

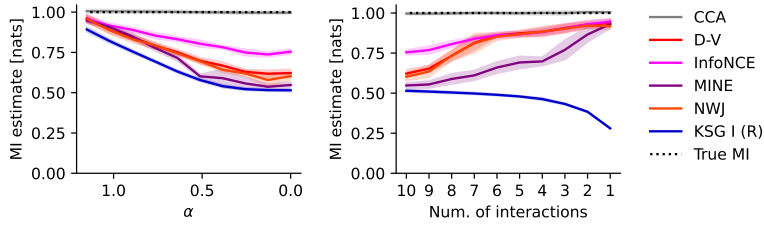


Figure 3: Two-stage interpolation between dense and sparse matrices. From left to right the sparsity of interactions increases. Shaded regions represent the sample standard deviation.

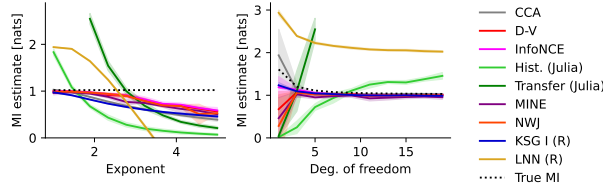


Figure 4: MI estimates as a function of the information contained in the tail of the distribution. Left: lengthening the tails via $x \mapsto |x|^k \text{sign } x$ mapping with changing k . Right: increasing the degrees of freedom in multivariate Student distribution shortens the tails. Shaded regions represent the sample standard deviation.

every pair of variables become sparser (lower α). In particular, even neural estimators are not able to model the information in ten interacting pairs of variables. However, when we decrease the number of interacting pairs of variables, the performance of the neighborhood-based KSG estimator is qualitatively different from the neural estimators: performance of KSG steadily deteriorates, while neural estimators can find (some of the) relevant pairs of variables.

This motivates us to conclude that in the settings where considered variables are high-dimensional and the interaction structure is sparse, neural estimators can offer an advantage over neighborhood-based approaches.

4.2 Long-Tailed Distributions

In Fig. 4 we investigate two different ways to lengthen the tails. First, we lengthen the tails of a multivariate normal distribution using the mapping $x \mapsto |x|^k \text{sign } x$ (with $k \geq 1$) applied to each dimension separately. In this case, we see that the performance of CCA, KSG, and neural estimators is near-perfect for k close to 1 (when the distribution P_{XY} is close to multivariate normal), but the performance degrades as the exponent increases. Second, we study multivariate Student distributions varying the degrees of freedom: the larger the degrees of freedom, the lower the information content in the tails of the distribution. Again, we see that the performance of CCA, KSG, and neural estimators is near-perfect for high degrees of freedom for which the distribution is approximately normal. For low degrees of freedom, these estimators significantly underestimate MI, with the exception of CCA which gives estimates with high variance, likely due to the correlation matrix being hard to estimate.

Overall, we see that long tails complicate the process of estimating MI. The Student distribution is particularly interesting, since even after the tails are removed (with asinh or preprocessing strategies described in Appendix E) the task remains challenging. We suspect that the MI might be contained in regions which are rarely sampled, making the MI hard to estimate. When neural estimators are used, pointwise mutual information learned by the critic does not match the ground truth (see Appendix E).

Hence, we believe that accurate estimation of MI in long-tailed distributions is an open problem. Practitioners expecting long-tailed distributions should approach this task with caution.

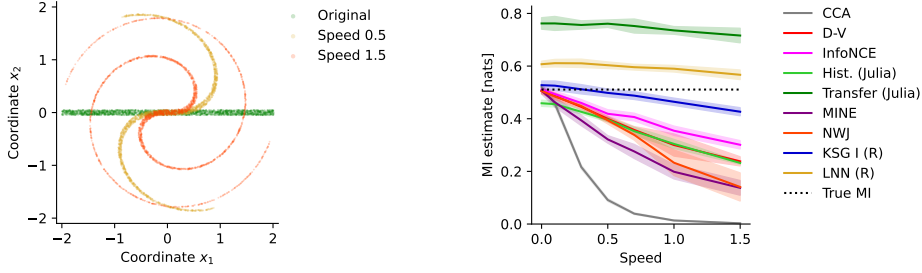


Figure 5: Left: Spiral diffeomorphism with different speed parameter v . Right: Performance of considered estimators for increasing speed.

4.3 Invariance to Diffeomorphisms

The benchmark results (Fig. 2) as well as the experiments with function $x \mapsto |x|^k \text{sign } x$ lengthening tails (Fig. 4) suggest that even if ground-truth MI is preserved by a transformation (see Theorem 2.1), the finite-sample estimates may not be invariant. This can be demonstrated using the spiral diffeomorphism.

We consider isotropic multivariate normal variables $X = (X_1, X_2) \sim \mathcal{N}(0, I_2)$ and $Y \sim \mathcal{N}(0, 1)$ with $\text{Cor}(X_1, Y) = \rho$. Then, we apply a spiral diffeomorphism $s_v: \mathbb{R}^2 \rightarrow \mathbb{R}^2$ to X , where

$$s_v(x) = \begin{pmatrix} \cos v \|x\|^2 & -\sin v \|x\|^2 \\ \sin v \|x\|^2 & \cos v \|x\|^2 \end{pmatrix}$$

is the rotation around the origin by the angle $v\|x\|^2$ (see Fig. 5; for a theoretical study of spiral diffeomorphisms see Appendix B). In Fig. 5, we present the estimates for $N = 10\,000$, $\rho = 0.8$ and varying speed v .

In general, the performance of the considered estimators deteriorates with stronger deformations. This is especially visible for the CCA estimator, as MI cannot be reliably estimated in terms of correlation.

4.4 High Mutual Information

Gao et al. [2015] proves that estimation of MI between strongly interacting variables requires large sample sizes. Poole et al. [2019], Song and Ermon [2020] investigate estimation of high MI using multivariate normal distributions in the setting when the ground-truth MI is changing over the training time. We decided to use our family of expressive distributions to see the practical limitations of estimating MI. We chose three one-parameter distribution families: the 3×3 multivariate normal distribution with correlation $\rho = \text{Cor}(X_1, Y_1) = \text{Cor}(X_2, Y_2)$, and its transformations by the half-cube mapping and the spiral diffeomorphism with $v = 1/3$. We changed the parameter to match the desired MI.

In Fig. 6 we see that for low enough MI neural estimators and KSG reliably estimate MI for all considered distributions, but the moment at which the estimation becomes inaccurate depends on the distribution. For example, for Gaussian distributions, neural estimators (and the simple CCA baseline) are accurate even up to 5 nats, while after the application of the spiral diffeomorphism, the estimates become inaccurate already at 2 nats.

We see that in general neural estimators perform well on distributions with high MI. The performance of CCA on the multivariate normal distribution suggests that well-specified model-based estimators may require lower number of samples to estimate high mutual information than model-free alternatives (cf. Gao et al. [2015]).

5 Related Work

Existing Benchmarks Khan et al. [2007] used one-dimensional variables and an additive noise assumption to evaluate the estimators based on histograms, nearest neighbors, and kernels. However,

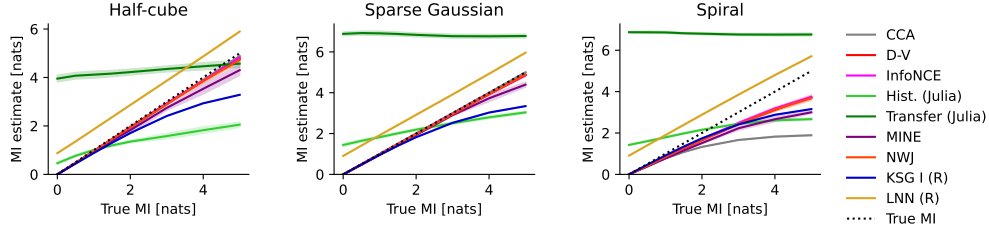


Figure 6: We select three one-parameter families of distributions varying the MI value (x -axis and dashed line). We use $N = 10\,000$ data points and $n = 5$ replicates of each experiment.

Table 1: Comparison between benchmarks.

	Khan et al. [2007]	Poole et al. [2019]	Song and Ermon [2020]	Proposed Benchmark
Gaussian Distrib.	✓	✓	✓	✓
Long-Tailed Distrib.	(✓)	(✓)	✗	✓
Sparse/Dense Inter.	✗	✗	✗	✓
High MI	✗	(✓)	(✓)	✓
Invariance to Homeom.	✗	✗	✗	✓
Self-Consistency	✗	✗	✓	✗
Language Indep.	✗	✗	✗	✓
Code Available	✗	✗	✓	✓

the original code and samples are not available. Song and Ermon [2020] proposed to evaluate neural estimators based on variational lower bounds [Oord et al., 2018, Belghazi et al., 2018] on image data. However, besides the multivariate Gaussian which they also evaluate, the ground-truth MI is not available. Therefore, they proposed self-consistency checks, which include $I(X; Y) = 0$ for independent X and Y , and preservation of the data processing inequality. Poole et al. [2019] study neural estimators in settings employing Gaussian variables, affine transformation followed by axis-wise cubic transformation and investigate the high MI cases. In our benchmark, we focus on distributions with known MI and information-preserving continuous injective mappings such as homeomorphisms, evaluate long tail distributions, and compare representative estimators from different classes and platforms. We summarize the differences between the benchmarks in Table 1.⁴

Causal Representation Learning and Nonlinear ICA Recent work on disentanglement and non-linear independent component analysis (ICA) aims at identifying causal sources up to an ambiguous transformation [Xi and Bloem-Reddy, 2022]. Examples include orthogonal transformations [Zimmermann et al., 2021], permutations and diffeomorphisms applied along specific axes [Khemakhem et al., 2020], or the whole diffeomorphism group [Von Kügelgen et al., 2021, Daunhawer et al., 2023]. To check whether the learned representations indeed capture underlying causal variables, it is important to use estimators which are invariant to the ambiguous transformations specific to the model. Our results suggest that such requirements are not met by MI estimators and practitioners should carefully choose the used metrics, so they are not only theoretically invariant, but also can be robustly estimated in practice.

6 Conclusions and Limitations

Venturing beyond the typical evaluation of mutual information estimators on multivariate normal distributions, we provided evidence that the evaluation on Gaussian distributions results in overly optimistic and biased results for existing estimators. In particular, a simple model-based approach as

⁴Khan et al. [2007] consider two one-dimensional r.v. related by additive noise, in which one of the r.v. is parametrized with a cubic function. This implicitly lengthens the tails, so we marked this with (✓); similarly, Poole et al. [2019] consider cubic transformation of a multivariate normal variable. The experimental setup of estimation of MI considered by Poole et al. [2019] and Song and Ermon [2020] requires changing ground truth MI over training time of a neural estimator. This procedure differs from the usual approaches to estimating MI, so we marked it with (✓).

CCA is already sufficient if the distribution is (close to) multivariate normal, highlighting the effect of using available prior information about the problem to design efficient estimators. To provide a basis for more thorough evaluation, we designed a benchmark, compatible with any programming language, with forty general tasks focusing on selected problems, as estimation with sparse interactions, the effects of long tails, estimation in problems with high MI, and (only theoretical) invariance to diffeomorphisms.

Our results suggest that in low- and medium-dimensional problems the KSG estimator remains a solid baseline, requiring a lower number of samples than the alternatives and no parameter tuning. In high-dimensional settings where the interactions are sparse, neural estimators offer an advantage over classical estimators being able to more consistently pick up the signal hidden in the high-dimensional space. We further provide evidence that although MI is invariant to continuous injective transformations of variables, in practice the estimates can differ substantially. Appropriate data preprocessing may mitigate some of these issues, but no strategy clearly outperforms the others (see Appendix E).

We hope that the proposed benchmark encourages the development of new estimators, which address the challenges that we highlight in this paper. It can be used to assess invariance to homeomorphisms, understand an estimator’s behavior in the situations involving distributions of different complexity, low-data behaviour, high mutual information or to diagnose problems with implementations. We believe that some parts of the benchmark could also be used to assess whether other estimators of other statistical quantities, such as kernel dependence tests, transfer entropy or representation similarity measures, can be reliably estimated on synthetic problems and if they are robust to selected classes of transformations.

Limitations and Open Problems Although the proposed benchmark covers a wide family of distributions, its main limitation is the possibility to only transform the distribution P_{XY} via transformations $f \times g$ to $P_{f(X)g(Y)}$ starting from multivariate normal and Student distributions. In particular, not every possible joint distribution is of this form and extending the family of distributions which have known MI and allow efficient sampling is the natural direction of continuation. We are, however, confident that due to the code design it will be easy to extend the benchmark with the new distributions (and their transformations) when they appear.

Estimation of information in multivariate Student distributions remains a challenge for most estimators. Although we provide the evidence that distributions with longer tails pose a harder challenge for considered estimators, shortening the tails with asinh transform did not alleviate all the issues. In neural estimators this can be partly explained by the fact that the critic may not learn to approximate pointwise mutual information (up to an additive constant) properly, as demonstrated in Appendix E, but the question of what makes the neighborhood-based estimator, KSG, fail, remains open.

As noted, unless strong inductive biases can be exploited (as illustrated for CCA), estimation of high MI is an important challenge. An interesting avenue towards developing new estimators can be by incorporating a stronger inductive bias, which could also take a form of development of better data normalization strategies and, in case of neural estimators, critic architectures. Better normalization strategies may make the estimators more invariant to considered transformations.

Acknowledgments and Disclosure of Funding

We would like to thank Craig Hamilton for the advice on scientific writing and help with the abstract. FG was supported by the Norwegian Financial Mechanism GRIEG-1 grant operated by Narodowe Centrum Nauki (National Science Centre, Poland) 2019/34/H/NZ6/00699. PC and AM were supported by a fellowship from the ETH AI Center.

References

- Ziqiao Ao and Jinglai Li. Entropy estimation via normalizing flow. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(9):9990–9998, 2022.
- R.B. Arellano-Valle, J.E. Contreras-Reyes, and M.G. Genton. Shannon entropy and mutual information for multivariate skew-elliptical distributions. *Scandinavian Journal of Statistics*, 49:42–62, 2013.
- Mohamed Ishmael Belghazi, Aristide Baratin, Sai Rajeshwar, Sherjil Ozair, Yoshua Bengio, Aaron Courville, and Devon Hjelm. Mutual information neural estimation. In *International conference on machine learning*, pages 531–540. PMLR, 2018.
- James Bradbury, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, George Necula, Adam Paszke, Jake VanderPlas, Skye Wanderman-Milne, and Qiao Zhang. JAX: composable transformations of Python+NumPy programs, 2018. URL <http://github.com/google/jax>.
- Peter Bühlmann and Sara van de Geer. *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer Publishing Company, Incorporated, 1st edition, 2011. ISBN 3642201911.
- Vincent Cabeli, Louis Verny, Nadir Sella, Guido Uguzzoni, Marc Verny, and Hervé Isambert. Learning clinical networks from medical records based on information estimates in mixed-type data. *PLOS Computational Biology*, 16(5):e1007866, 2020.
- Anthony Carapetis. Rotating spherical shells doesn’t change volume. Mathematics Stack Exchange, 2022. URL <https://math.stackexchange.com/q/4500294>. URL:<https://math.stackexchange.com/q/4500294> (version: 2022-07-26).
- Christopher J Cellucci, Alfonso M Albano, and Paul E Rapp. Statistical validation of mutual information calculations: Comparison of alternative numerical algorithms. *Physical review E*, 71(6):066208, 2005.
- Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley-Interscience New York, 2006.
- Georges A Darbellay and Igor Vajda. Estimation of the information by an adaptive partitioning of the observation space. *IEEE Transactions on Information Theory*, 45(4):1315–1321, 1999.
- Carsten O Daub, Ralf Steuer, Joachim Selbig, and Sebastian Kloska. Estimating mutual information using b-spline functions—an improved similarity measure for analysing gene expression data. *BMC bioinformatics*, 5(1):1–12, 2004.
- Imant Daunhawer, Alice Bizeul, Emanuele Palumbo, Alexander Marx, and Julia E Vogt. Identifiability results for multimodal contrastive learning. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=U_2kuqoTcB.
- Shuyang Gao, Greg Ver Steeg, and Aram Galstyan. Efficient estimation of mutual information for strongly dependent variables. In *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 277–286. PMLR, 2015.
- Weihao Gao, Sreeram Kannan, Sewoong Oh, and Pramod Viswanath. Estimating mutual information for discrete-continuous mixtures. In *Advances in Neural Information Processing Systems*, pages 5986–5997, 2017a.
- Weihao Gao, Sewoong Oh, and Pramod Viswanath. Density functional estimators with k-nearest neighbor bandwidths. In *IEEE International Symposium on Information Theory (ISIT)*, pages 1351–1355. IEEE, 2017b.
- Weihao Gao, Sewoong Oh, and Pramod Viswanath. Demystifying fixed k-nearest neighbor information estimators. *IEEE Transactions on Information Theory*, 64(8):5629–5661, 2018.

- Frederic Grabowski, Paweł Czyż, Marek Kochańczyk, and Tomasz Lipniacki. Limits to the rate of information transmission through the MAPK pathway. *Journal of The Royal Society Interface*, 16 (152):20180792, 2019. doi: 10.1098/rsif.2018.0792. URL <https://royalsocietypublishing.org/doi/abs/10.1098/rsif.2018.0792>.
- Søren Hauberg. Only Bayes should learn a manifold (on the estimation of differential geometric structure from data), 2019.
- Caroline M Holmes and Ilya Nemenman. Estimation of mutual information for real-valued data with error bars and controlled bias. *Physical Review E*, 100(2):022404, 2019.
- J. Kay. Feature discovery under contextual supervision using mutual information. In *IJCNN International Joint Conference on Neural Networks*, volume 4, pages 79–84, 1992. doi: 10.1109/IJCNN.1992.227286.
- A. Kechris. *Classical Descriptive Set Theory*. Graduate Texts in Mathematics. Springer New York, 2012. ISBN 9781461241904. URL <https://books.google.ch/books?id=WR3SBwAAQBAJ>.
- Dustin Keys, Shukur Kholikov, and Alexei A. Pevtsov. Application of mutual information methods in time–distance helioseismology. *Solar Physics*, 290(3):659–671, Mar 2015. ISSN 1573-093X. doi: 10.1007/s11207-015-0650-y. URL <https://doi.org/10.1007/s11207-015-0650-y>.
- Shiraj Khan, Sharba Bandyopadhyay, Auroop R Ganguly, Sunil Saigal, David J Erickson III, Vladimir Protopopescu, and George Ostrouchov. Relative performance of mutual information estimation methods for quantifying the dependence among short and noisy data. *Physical Review E*, 76(2): 026209, 2007.
- Ilyes Khemakhem, Diederik Kingma, Ricardo Monti, and Aapo Hyvarinen. Variational autoencoders and nonlinear ICA: a unifying framework. In *International Conference on Artificial Intelligence and Statistics*, pages 2207–2217. PMLR, 2020.
- Jussi Klemelä. Multivariate histograms with data-dependent partitions. *Statistica sinica*, pages 159–176, 2009.
- Mike Koeman and Tom Heskes. Mutual information estimation with random forests. In *International Conference on Neural Information Processing*, pages 524–531. Springer, 2014.
- L. F. Kozachenko and N. N. Leonenko. Sample estimate of the entropy of a random vector. *Problemy Peredachi Informatsii*, 23:9–16, 1987.
- Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. Estimating mutual information. *Physical Review E*, 69(6):066138, 2004.
- Thanard Kurutach, Aviv Tamar, Ge Yang, Stuart J Russell, and Pieter Abbeel. Learning plannable representations with causal infogan. *Advances in Neural Information Processing Systems*, 31, 2018.
- Bo Li, Yifei Shen, Yezhen Wang, Wenzhen Zhu, Dongsheng Li, Kurt Keutzer, and Han Zhao. Invariant information bottleneck for domain generalization. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 7399–7407, 2022.
- Warren M Lord, Jie Sun, and Erik M Bollt. Geometric k-nearest neighbor estimation of entropy and mutual information. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 28(3):033114, 2018.
- Chien Lu and Jaakko Peltonen. Enhancing nearest neighbor based entropy estimator for high dimensional distributions via bootstrapping local ellipsoid. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 5013–5020, 2020.
- Joseph Lucas, Carlos Carvalho, Quanli Wang, Andrea Bild, Joseph R. Nevins, and Mike West. *Sparse Statistical Modelling in Gene Expression Genomics*, page 155–176. Cambridge University Press, 2006. doi: 10.1017/CBO9780511584589.009.
- Alexander Marx and Jonas Fischer. Estimating mutual information via geodesic knn. In *Proceedings of the SIAM International Conference on Data Mining (SDM)*, pages 415–423. SIAM, 2022.

- Alexander Marx and Jilles Vreeken. Testing conditional independence on discrete data using stochastic complexity. In *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 496–505, 2019.
- Alexander Marx, Lincen Yang, and Matthijs van Leeuwen. Estimating conditional mutual information for discrete-continuous mixtures using multi-dimensional adaptive histograms. In *Proceedings of the SIAM International Conference on Data Mining (SDM)*, pages 387–395, 2021.
- David McAllester and Karl Stratos. Formal limitations on the measurement of mutual information. In Silvia Chiappa and Roberto Calandra, editors, *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pages 875–884. PMLR, 26–28 Aug 2020. URL <https://proceedings.mlr.press/v108/mcallester20a.html>.
- Octavio César Mesner and Cosma Rohilla Shalizi. Conditional Mutual Information Estimation for Mixed, Discrete and Continuous, Data. *IEEE Transactions on Information Theory*, 2020.
- Young-II Moon, Balaji Rajagopalan, and Upmanu Lall. Estimation of mutual information using kernel density estimators. *Phys. Rev. E*, 52:2318–2321, Sep 1995. doi: 10.1103/PhysRevE.52.2318. URL <https://link.aps.org/doi/10.1103/PhysRevE.52.2318>.
- Kevin P. Murphy. *Probabilistic Machine Learning: Advanced Topics*. MIT Press, 2023. URL <http://probml.github.io/book2>.
- F Mölder, KP Jablonski, B Letcher, MB Hall, CH Tomkins-Tinch, V Sochat, J Forster, S Lee, SO Twardziok, A Kanitz, A Wilm, M Holtgrewe, S Rahmann, S Nahnsen, and J Köster. Sustainable data analysis with snakemake [version 1; peer review: 1 approved, 1 approved with reservations]. *F1000Research*, 10(33), 2021. doi: 10.12688/f1000research.29032.1.
- R.B. Nelsen. *An Introduction to Copulas*. Springer Series in Statistics. Springer New York, 2007. ISBN 9780387286785. URL <https://books.google.ch/books?id=B30NT5rBv0wC>.
- XuanLong Nguyen, Martin J Wainwright, and Michael Jordan. Estimating divergence functionals and the likelihood ratio by penalized convex risk minimization. In J. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems*, volume 20. Curran Associates, Inc., 2007. URL https://proceedings.neurips.cc/paper_files/paper/2007/file/72da7fd6d1302c0a159f6436d01e9eb0-Paper.pdf.
- Sebastian Nowozin, Botond Cseke, and Ryota Tomioka. F-gan: Training generative neural samplers using variational divergence minimization. In *Proceedings of the 30th International Conference on Neural Information Processing Systems, NIPS’16*, page 271–279, Red Hook, NY, USA, 2016. Curran Associates Inc. ISBN 9781510838819.
- Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.
- Liam Paninski. Estimation of entropy and mutual information. *Neural Computation*, 15(6):1191–1253, 2003.
- Liam Paninski and Masanao Yajima. Undersmoothed kernel entropy estimators. *IEEE Transactions on Information Theory*, 54(9):4384–4388, 2008.
- F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- M.S. Pinsker and A. Feinstein. *Information and Information Stability of Random Variables and Processes*. Holden-Day series in time series analysis. Holden-Day, 1964. ISBN 9780816268047.
- Stephen M Pizer, E Philip Amburn, John D Austin, Robert Cromartie, Ari Geselowitz, Trey Greer, Bart ter Haar Romeny, John B Zimmerman, and Karel Zuiderveld. Adaptive histogram equalization and its variations. *Computer Vision, Graphics, and Image Processing*, 39(3):355–368, 1987.

- Y. Polyanskiy and Y. Wu. *Information Theory: From Coding to Learning*. Cambridge University Press, 2022. Book draft.
- Ben Poole, Sherjil Ozair, Aaron Van Den Oord, Alex Alemi, and George Tucker. On variational bounds of mutual information. In *International Conference on Machine Learning*, pages 5171–5180. PMLR, 2019.
- Ruggero Ragonesi, Riccardo Volpi, Jacopo Cavazza, and Vittorio Murino. Learning unbiased representations via mutual information backpropagation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2729–2738, 2021.
- Jangir Selimkhanov, Brooks Taylor, Jason Yao, Anna Pilko, John Albeck, Alexander Hoffmann, Lev Tsimring, and Roy Wollman. Accurate information transmission through dynamic biochemical signaling networks. *Science*, 346(6215):1370–1373, 2014.
- Bernard W Silverman. *Density estimation for statistics and data analysis*. Routledge, 2018.
- Victor Solo. On causality and mutual information. In *2008 47th IEEE Conference on Decision and Control*, pages 4939–4944. IEEE, 2008.
- Jiaming Song and Stefano Ermon. Understanding the limitations of variational mutual information estimators. In *International Conference on Learning Representations*, 2020.
- Joe Suzuki. An estimator of mutual information and its application to independence testing. *Entropy*, 18(4):109, 2016.
- Nassim Nicholas Taleb. Statistical consequences of fat tails: Real world preasymptotics, epistemology, and applications, 2020. URL <https://arxiv.org/abs/2001.10488>.
- Michael Tschannen, Josip Djolonga, Paul K. Rubenstein, Sylvain Gelly, and Mario Lucic. On mutual information maximization for representation learning. In *International Conference on Learning Representations*, 2020.
- Shinsuke Uda. Application of information theory in systems biology. *Biophysical Reviews*, 12(2):377–384, Apr 2020. ISSN 1867-2469. doi: 10.1007/s12551-020-00665-w. URL <https://doi.org/10.1007/s12551-020-00665-w>.
- Gregory Valiant and Paul Valiant. Estimating the unseen: an $n/\log(n)$ -sample estimator for entropy and support size, shown optimal via new clts. In *Proceedings of the annual ACM symposium on Theory of computing (STOC)*, pages 685–694, 2011.
- Martin Vejmelka and Kateřina Hlaváčková-Schindler. Mutual information estimation in higher dimensions: A speed-up of a k-nearest neighbor based estimator. In *International Conference on Adaptive and Natural Computing Algorithms*, pages 790–797. Springer, 2007.
- Nguyen Xuan Vinh, Jeffrey Chan, and James Bailey. Reconsidering mutual information based feature selection: A statistical significance view. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2014.
- Julius Von Kügelgen, Yash Sharma, Luigi Gresele, Wieland Brendel, Bernhard Schölkopf, Michel Besserve, and Francesco Locatello. Self-supervised learning with data augmentations provably isolates content from style. *Advances in neural information processing systems*, 34:16451–16467, 2021.
- Quanhan Xi and Benjamin Bloem-Reddy. Indeterminacy in latent variable models: Characterization and strong identifiability. *arXiv preprint arXiv:2206.00801*, 2022.
- Alexander L. Young, Willem van den Boom, Rebecca A. Schroeder, Vijay Krishnamoorthy, Karthik Raghunathan, Hau-Tieng Wu, and David B. Dunson. Mutual information: Measuring nonlinear dependence in longitudinal epidemiological data. *PLOS ONE*, 18(4):1–16, 04 2023. doi: 10.1371/journal.pone.0284904. URL <https://doi.org/10.1371/journal.pone.0284904>.
- Yifan Zhang, Bingyi Kang, Bryan Hooi, Shuicheng Yan, and Jiashi Feng. Deep long-tailed learning: A survey, 2021. URL <https://arxiv.org/abs/2110.04596>.

Roland S Zimmermann, Yash Sharma, Steffen Schneider, Matthias Bethge, and Wieland Brendel.
Contrastive learning inverts the data generating process. In *International Conference on Machine Learning*, pages 12979–12990. PMLR, 2021.

Appendix

Table of Contents

A	Invariance to Continuous Injective Mappings	17
B	Spiral Diffeomorphisms	18
C	Overview of Estimators	20
D	Additional Information on Benchmark Tasks	22
D.1	Task Descriptions	22
D.2	Covariance Matrix Family	23
D.3	Technical Results	24
E	Additional Experiments	25
E.1	Pointwise Mutual Information	25
E.2	Sample Size Requirements	26
E.3	Preprocessing	27
E.4	Hyperparameter Selection	27
F	Visualisation of Selected Tasks	29
G	Author Contributions	30

A Invariance to Continuous Injective Mappings

In this section, we provide a proof for Theorem 2.1, formally proving the invariance of mutual information with respect to continuous injective mappings. First, we recall the standard definitions from Chapters 1 and 2 of Pinsker and Feinstein [1964].

Definition A.1. Let $(\Omega, \mathcal{A})$ be a measurable space. A family of sets $\mathcal{E} \subseteq \mathcal{A}$ is a partition of Ω if for every two distinct E_1 and E_2 in $\mathcal{E}$ it holds that $E_1 \cap E_2 = \emptyset$ and that $\bigcup \mathcal{E} = \Omega$. If $\mathcal{E}$ is finite, we call it a finite partition.

Definition A.2 (Mutual Information). Mutual information between r.v. X and Y is defined as

$$\mathbf{I}(X; Y) = \sup_{\mathcal{E}, \mathcal{F}} \sum_{i,j} P_{XY}(E_i \times F_j) \log \frac{P_{XY}(E_i \times F_j)}{P_X(E_i)P_Y(F_j)},$$

where the supremum is taken over all finite partitions $\mathcal{E} = \{E_1, E_2, \dots\}$ of $\mathcal{X}$ and $\mathcal{F} = \{F_1, F_2, \dots\}$ of $\mathcal{Y}$ with the usual conventions $0 \log 0 = 0$ and $0 \log 0/0 = 0$.

Remark A.3. If P_{XY} is absolutely continuous with respect to $P_X \otimes P_Y$ (e.g., X and Y are finite r.v.), mutual information is equal to the following Kullback–Leibler divergence:

$$\mathbf{I}(X; Y) = \mathbf{D}_{\text{KL}}(P_{XY} \parallel P_X \otimes P_Y) = \int \log f \, dP_{XY},$$

where f is the Radon–Nikodym derivative $f = dP_{XY}/d(P_X \otimes P_Y)$. If the absolute continuity does not hold, then $\mathbf{I}(X; Y) = +\infty$ [Pinsker and Feinstein, 1964, Theorem 2.1.2].

Theorem 2.1. Let $\mathcal{X}$, $\mathcal{X}'$, $\mathcal{Y}$ and $\mathcal{Y}'$ be standard Borel spaces (e.g., smooth manifolds with their Borel σ -algebras) and $f: \mathcal{X} \rightarrow \mathcal{X}'$ and $g: \mathcal{Y} \rightarrow \mathcal{Y}'$ be continuous injective mappings. Then, for every $\mathcal{X}$ -valued r.v. X and $\mathcal{Y}$ -valued r.v. Y it holds that $\mathbf{I}(X; Y) = \mathbf{I}(f(X); g(Y))$.

Proof. First, note that a continuous mapping between Borel spaces is measurable, so random variables $f(X)$ and $g(Y)$ are well-defined.

Then, observe that it suffices to prove that

$$\mathbf{I}(f(X); Y) = \mathbf{I}(X; Y).$$

It is well-known that we have $\mathbf{I}(f(X); Y) \leq \mathbf{I}(X; Y)$ for every measurable mapping f , so we will need to prove the reverse inequality.

Recall from Definition A.2 that

$$\mathbf{I}(X; Y) = \sup_{\mathcal{E}, \mathcal{F}} \sum_{i,j} P_{XY}(E_i \times F_j) \log \frac{P_{XY}(E_i \times F_j)}{P_X(E_i)P_Y(F_j)},$$

where the supremum is taken over all finite partitions $\mathcal{E}$ of $\mathcal{X}$ and $\mathcal{F}$ of $\mathcal{Y}$.

The proof will go as follows: for every finite partition $\mathcal{E}$ of $\mathcal{X}$ we will construct a finite partition of $\mathcal{X}'$, which (together with the same arbitrary partition $\mathcal{F}$ of $\mathcal{Y}$) will give the same value of the sum. This will prove that $\mathbf{I}(f(X); Y) \geq \mathbf{I}(X; Y)$.

Take any finite partitions $\{E_1, \dots, E_n\}$ of $\mathcal{X}$ and $\{F_1, \dots, F_m\}$ of $\mathcal{Y}$. First, we will show that sets

$$\begin{aligned} G_0 &= \mathcal{X}' \setminus f(\mathcal{X}) \\ G_1 &= f(E_1) \\ G_2 &= f(E_2) \\ &\vdots \end{aligned}$$

form a partition of $\mathcal{X}'$.

It is easy to see that they are pairwise disjoint as f is injective and it is obvious that their union is the whole $\mathcal{X}'$. Hence, we only need to show that they are all Borel.

For $i \geq 1$ the Lusin–Suslin theorem [Kechris, 2012, Theorem 15.1] guarantees that G_i is Borel (continuous injective mappings between Polish spaces map Borel sets to Borel sets). Similarly, $f(\mathcal{X})$ is Borel and hence $G_0 = \mathcal{X}' \setminus f(\mathcal{X})$ is Borel as well.

Next, we will prove that

$$\sum_{i=1}^n \sum_{j=1}^m P_{XY}(E_i \times F_j) \log \frac{P_{XY}(E_i \times F_j)}{P_X(E_i)P_Y(F_j)} = \sum_{i=0}^n \sum_{j=1}^m P_{f(X)Y}(G_i \times F_j) \log \frac{P_{f(X)Y}(G_i \times F_j)}{P_{f(X)}(G_i)P_Y(F_j)}.$$

Note that $P_{f(X)}(G_0) = P_X(f^{-1}(G_0)) = P_X(\emptyset) = 0$, so that we can ignore the terms with $i = 0$.

For $i \geq 1$ we have

$$P_{f(X)}(G_i) = P_X(f^{-1}(G_i)) = P_X(f^{-1}(f(E_i))) = P_X(E_i)$$

as f is injective.

Similarly, $f \times 1_Y$ is injective and maps $E_i \times F_j$ onto $G_i \times F_j$, so

$$P_{f(X)Y}(G_i \times F_j) = P_{XY}((f \times 1_Y)^{-1}(G_i \times F_j)) = P_{XY}(E_i \times F_j).$$

This finishes the proof. $\square$

B Spiral Diffeomorphisms

In this section we will formally define the spiral diffeomorphisms and prove that they preserve the centred Gaussian distribution $\mathcal{N}(0, \sigma^2 I_n)$. They were inspired by the “swirling construction”, discussed by [Hauberg \[2019\]](#).

Definition B.1. Let $O(n)$ be the orthogonal group and $\gamma: \mathbb{R} \rightarrow O(n)$ be any smooth function. The mapping

$$f: \mathbb{R}^n \rightarrow \mathbb{R}^n$$

given by $f(x) = \gamma(\|x\|^2)x$ will be called the spiral.

Example B.2. Let $\mathfrak{so}(n)$ be the vector space of all skew-symmetric matrices. For every two matrices $R \in O(n)$ and $A \in \mathfrak{so}(n)$ and any smooth mapping $g: \mathbb{R} \rightarrow \mathbb{R}$ we can form a path

$$\gamma(t) = R \exp(g(t)A).$$

In particular, $R = I_n$ and $g(t) = vt$ yields the spirals defined earlier.

From the above definition it is easy to prove that spirals are diffeomorphisms. We formalize it by the following lemma.

Lemma B.3. Every spiral is a diffeomorphism.

Proof. Let γ be the path leading to the spiral f . As matrix inverse is a smooth function, we can define a smooth function $h(x) = \gamma(\|x\|^2)^{-1}x$.

Note that $\|f(x)\|^2 = \|x\|^2$, as orthogonal transformation $\gamma(\|x\|^2)$ preserves distances.

Then

$$\begin{aligned} h(f(x)) &= \gamma(\|f(x)\|^2)^{-1} f(x) \\ &= \gamma(\|x\|^2)^{-1} f(x) \\ &= \gamma(\|x\|^2)^{-1} \gamma(\|x\|^2) x \\ &= x. \end{aligned}$$

Analogously, we prove that $f \circ h$ also is the identity mapping. $\square$

Finally, we will give the proof that spirals leave the centred normal distribution $\mathcal{N}(0, \sigma^2 I_n)$ invariant. This will follow from a slightly stronger theorem:

Theorem B.4. Let P be a probability measure on $\mathbb{R}^n$ with a smooth probability density function p which factorizes as:

$$p(x) = u(\|x\|^2)$$

for some smooth function $u: \mathbb{R} \rightarrow \mathbb{R}$.

Then, every spiral $f: \mathbb{R}^n \rightarrow \mathbb{R}^n$ preserves the measure P :

$$f_{\#}P = P,$$

where $f_{\#}P = P \circ f^{-1}$ is the pushforward measure.

In particular, for $P = \mathcal{N}(0, \sigma^2 I_n)$ (a centered isotropic Gaussian distribution), if $X \sim \mathcal{N}(0, \sigma^2 I_n)$, then $f(X) \sim \mathcal{N}(0, \sigma^2 I_n)$.

Proof. As we have already observed, for every $x \in \mathbb{R}^n$ it holds that $\|f(x)\|^2 = \|x\|^2$, so the PDF evaluated at both points is equal

$$p(x) = p(f(x))$$

Hence, we need to prove that for every point $x \in \mathbb{R}^n$, $|\det f'(x)| = 1$, where $f'(x)$ is the Jacobian matrix of f evaluated at the point x .

The proof of this identity was given by Carapetis [2022], but for the sake of completeness we supply a variant of it below.

First, we calculate the Jacobian matrix:

$$\begin{aligned} \frac{\partial f_i(x)}{\partial x_j} &= \frac{\partial}{\partial x_j} \sum_k \gamma(\|x\|^2)_{ik} x_k \\ &= \sum_k \gamma(\|x\|^2)_{ik} \frac{\partial x_k}{\partial x_j} + \sum_k x_k \frac{\partial \gamma(\|x\|^2)_{ik}}{\partial x_j} \\ &= \gamma(\|x\|^2)_{ij} + \sum_k x_k \cdot 2x_j \cdot \gamma'(\|x\|^2)_{ik} \\ &= \gamma(\|x\|^2)_{ij} + 2 \left(\sum_k \gamma'(\|x\|^2)_{ik} x_k \right) x_j \end{aligned}$$

Hence,

$$f'(x) = \gamma(\|x\|^2) + 2\gamma'(\|x\|^2)xx^T.$$

Now fix point x and consider the matrix $\gamma(\|x\|^2) \in O(n)$. Define a mapping $y \mapsto \gamma(\|x\|^2)^{-1} f(y)$. As we have already observed, multiplication by $\gamma(\|x\|^2)^{-1}$ does not change the value of the PDF. It also cannot change the $|\det f'(x)|$ value as because $\gamma(\|x\|^2)^{-1}$ is a linear mapping, its derivative is just the mapping itself and because $\gamma(\|x\|^2)^{-1}$ is an orthogonal matrix, we have $|\det \gamma(\|x\|^2)^{-1}| = 1$. Hence, without loss of generality we can assume that $\gamma(\|x\|^2) = I_n$ and

$$f'(x) = I_n + 2Axx^T,$$

where $A := \gamma'(\|x\|^2) \in \mathfrak{so}(n)$ is a skew-symmetric matrix.

If $x = 0$ we have $|\det f'(x)| = |\det I_n| = 1$. In the other case, we will work in coordinates induced by an orthonormal basis $e_1, \dots, e_n$ in which the first basis vector is $e_1 = x/\|x\|$. Hence,

$$f'(x) = I_n + 2\|x\|^2 A e_1 e_1^T.$$

For $i \geq 2$ we have $e_1^T e_i = 0$ and consequently

$$f'(x)e_i = e_i + 2\|x\|^2 A e_1 e_1^T e_i = e_i.$$

This shows that in this basis we have

$$f'(x) = \begin{pmatrix} a_1 & 0 & 0 & \dots & 0 \\ a_2 & 1 & 0 & \dots & 0 \\ a_3 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_n & 0 & 0 & \dots & 1 \end{pmatrix},$$

where the block on the bottom right is the identity matrix I_{n-1} and $\det f'(x) = a_1$.

The first column is formed by the coordinates of

$$f'(x)e_1 = e_1 + 2\|x\|^2 A e_1$$

and we have

$$a_1 = e_1^T f'(x)e_1 = 1 + 2\|x\|^2 e_1^T A e_1 = 1 + 0 = 1$$

as $A \in \mathfrak{so}(n)$ is skew-symmetric. $\square$

C Overview of Estimators

There exists a broad literature on (conditional) mutual information estimators ranging from estimators on discrete data [Paninski, 2003, Valiant and Valiant, 2011, Vinh et al., 2014, Marx and Vreeken, 2019], over estimators for continuous data [Kraskov et al., 2004, Darbellay and Vajda, 1999, Belghazi et al., 2018] to estimators that can consider discrete-continuous mixture data [Gao et al., 2017a, Marx et al., 2021, Mesner and Shalizi, 2020]. Here, we focus on the most commonly used class, which also finds applications in representation learning: estimators for continuous data. For continuous data, the most widely propagated classes of estimators are histogram or partition-based estimators, nearest neighbor-based density estimators, kernel methods, and neural estimators. In the following, we briefly discuss a representative selection of estimators from these classes, as well as model-based estimation.

Histogram-Based Estimators A natural way to approximate mutual information, as stated in Definition A.2, is by partitioning the domain space of X and Y into segments and estimate the densities locally. This approach is also known as histogram density estimation, where the probability mass function is approximated locally for each hyper-rectangle E_i by $\hat{P}_X(E_i) = \frac{n_i/N}{\text{Vol}_i}$, where Vol_i is the volume of E_i , and n_i refers to the number of data points residing in E_i .

To estimate mutual information, we explore the space of allowed partitionings and select the partitioning that maximizes mutual information [Darbellay and Vajda, 1999, Cellucci et al., 2005]

$$E_{\text{Disc}, N}^{XY} = \sup_{\mathcal{E}, \mathcal{F}} \sum_{i,j} \hat{P}_{XY}(E_i \times F_j) \log \frac{\hat{P}_{XY}(E_i \times F_j)}{\hat{P}_X(E_i) \hat{P}_Y(F_j)}.$$

Many estimators follow this idea. Darbellay and Vajda [1999] use adaptive partitioning trees and propose to split bins until the density within a bin is uniform, which they check with the χ^2 test. Daub et al. [2004] use B-Splines to assign data points to bins, and other authors use regularization based on the degrees of freedom of partitioning to avoid overfitting [Suzuki, 2016, Cabeli et al., 2020, Vinh et al., 2014, Marx et al., 2021].

A well-known property of histogram-based estimators is that they are consistent for $N \rightarrow \infty$ if the volume of the largest hypercube goes to zero, while the number of hypercubes grows sub-linearly with the number of samples, N (see e.g. Cover and Thomas [2006]). For low-dimensional settings, histogram density estimation can provide accurate results, but the strategy suffers from the curse of dimensionality [Pizer et al., 1987, Klemelä, 2009].

In our evaluation, we resort to partitioning estimators based on a fixed number of grids (10 equal-width bins per dimension), since more recent partitioning-based estimators focus on estimating conditional MI and only allow X and Y to be one-dimensional [Suzuki, 2016, Marx et al., 2021, Cabeli et al., 2020], or do not have a publicly available implementation [Darbellay and Vajda, 1999].

kNN Estimators An alternative to histogram-based methods is to use k -nearest neighbor (k NN) distances to approximate densities locally. Most generally, k NN estimators express mutual information as the sum $H(X) + H(Y) - H(X, Y)$, which can be done under the assumptions discussed in Remark A.3, and estimate the individual entropy terms as

$$\hat{H}((x_1, \dots, x_N)) = -\frac{1}{N} \sum_{i=1}^N \log \frac{k(x_i)/N}{\text{Vol}_i},$$

where $k(x_i)$ is the number of data points other than x_i in a neighborhood of x_i , and Vol_i is the volume of the neighborhood [Kozachenko and Leonenko, 1987]. The quality of the estimator depends on two

factors: i) the distance metric used to estimate the nearest neighbor distances, and ii) the estimation of the volume terms.

The most well-known estimator is the Kraskov, Stögbauer and Grassberger (KSG) estimator [Kraskov et al., 2004]. In contrast to the first k NN-based estimators [Kozachenko and Leonenko, 1987], Kraskov et al. [2004] propose to estimate the distance to the k -th neighbor $\rho_{i,k}/2$ for each data point i only on the joint space of $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$, with

$$D_Z(z_i, z_j) = \max\{D_X(x_i, x_j), D_Y(y_i, y_j)\},$$

where D_X and D_Y can refer to any norm-induced metric (e.g., Euclidean ℓ_2 , Manhattan ℓ_1 or Chebyshev ℓ_∞).⁵ To compute the entropy terms of the sub-spaces $\mathcal{X}$ and $\mathcal{Y}$, we count the number of data points with distance smaller than $\rho_{i,k}/2$, i.e. we define $n_{x,i}$ as

$$n_{x,i} = \left| \left\{ x_j : D_X(x_i, x_j) < \frac{\rho_{i,k}}{2}, i \neq j \right\} \right|, \quad (1)$$

and specify $n_{y,i}$ accordingly. As a consequence, all volume-related terms cancel and the KSG estimator reduces to

$$E_{\text{KSG},N}^{XY} = \psi(k) + \psi(n) - \frac{1}{n} \sum_{i=1}^n \psi(n_{x,i} + 1) + \psi(n_{y,i} + 1). \quad (2)$$

The digamma function⁶ ψ was already introduced by Kozachenko and Leonenko [1987] to correct for biases induced by the volume estimation. For a detailed exposition of the issue, we refer to Kraskov et al. [2004, Sec. II.A].

There exist various extensions of the classical KSG estimator. When selecting the L_∞ norm as a distance measure for D_X and D_Y , $E_{\text{KSG},N}^{XY}$ can be efficiently computed with the K-d trie method [Vejmelka and Hlaváčková-Schindler, 2007]. Other approaches aim to improve the estimation quality by explicitly computing the volume terms via SVD [Lord et al., 2018] or PCA [Lu and Peltonen, 2020], or use manifold learning [Marx and Fischer, 2022] or normalizing flows [Ao and Li, 2022] to estimate the nearest neighbors more efficiently in high-dimensional settings.

Kernel Estimators A smaller class of estimators builds upon kernel density estimation (KDE) [Moon et al., 1995, Paninski and Yajima, 2008]. Classical approaches compute the density functions for the individual entropy terms via kernel density estimation, which is well-justified under the assumption of P_{XY} being absolutely continuous with respect to $P_X \otimes P_Y$ (cf. Remark A.3). i.e., the density for a d -dimensional random vector X is approximated as

$$\hat{p}(x) = \frac{1}{Nh^d} \sum_{i=1}^N K\left(\frac{x - x_i}{h}\right),$$

where h is the smoothing parameter or bandwidth of a Gaussian kernel [Silverman, 2018].

Moon et al. [1995] propose to estimate mutual information using a Gaussian kernel, for which the bandwidth can be estimated optimally [Silverman, 2018]. Similar to histogram-based approaches, KDE-based estimators with fixed bandwidth h are consistent if $Nh^d \rightarrow \infty$ as $h \rightarrow 0$.

A more recent proposal, LNN [Gao et al., 2017b], uses a data-dependent bandwidth based on nearest-neighbor distances, to avoid fixing h globally.

Neural Estimators Neural network approaches such as MINE [Belghazi et al., 2018] use a variational lower bound on mutual information using the Donsker–Varadhan theorem: let $f: \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$ be a bounded measurable function and

$$V_f(X, Y) = \sup_f \mathbb{E}_{P_{XY}}[f(X, Y)] - \log \mathbb{E}_{P_X \otimes P_Y}[\exp f(X, Y)].$$

Under sufficient regularity conditions on P_{XY} , $\mathbf{I}(X; Y)$ is equal to $V_f(X, Y)$, which is evaluated over all bounded measurable mappings f . In practice, the right hand side is optimized with respect

⁵In the original approach, the L_∞ is used, and Gao et al. [2018] derives theoretical properties of the estimator for the L_2 norm.

⁶The digamma function is defined as $\psi(x) = d \log \Gamma(x) / dx$. It satisfies the recursion $\psi(x+1) = \psi(x) + \frac{1}{x}$ with $\psi(1) = -\gamma$, where $\gamma \approx 0.577$ is the Euler–Mascheroni constant.

to the parameters of a given parametric predictive model f , as a neural network. Approaches following this idea also count as discriminative approaches since the neural network is trained to distinguish points from the joint and marginal distribution. The literature on variational lower bounds of mutual information is growing [Oord et al., 2018, Song and Ermon, 2020] and the idea of using a discriminative approach to estimate mutual information has also been explored by Koeman and Heskes [2014], who used random forests. For an excellent theoretical overview of different neural estimators we refer to Poole et al. [2019].

Model-based Estimators Finally, if one assumes that P_{XY} is a member of a family of distributions $Q_{XY}(\theta)$ with known mutual information, one can estimate θ basing on the data $((x_1, y_1), \dots, (x_N, y_N))$ and calculate the corresponding mutual information. For example, if one assumes that there exist jointly multivariate normal variables $\tilde{X} = (\tilde{X}_1, \dots, \tilde{X}_r, E_1, \dots, E_{m-r})$ and $\tilde{Y} = (\tilde{Y}_1, \dots, \tilde{Y}_r, \tilde{E}_1, \dots, \tilde{E}_{n-r})$ such that the cross-covariance terms are zero apart from $\text{Cov}(\tilde{X}_i, \tilde{Y}_i)$, and that $X = A\tilde{X} + a$ and $Y = B\tilde{Y} + b$ for invertible matrices A and B and vectors $a \in \mathbb{R}^m$ and $b \in \mathbb{R}^n$, then the mutual information can be found as

$$\mathbf{I}(X; Y) = - \sum_{i=1}^r \log \sqrt{1 - \text{Cor}(\tilde{X}_i, \tilde{Y}_i)^2}.$$

The model parameters A , B , a , and b , as well as the pairs of corresponding r.v. $\tilde{X}_i$ and $\tilde{Y}_i$, can be found by maximising likelihood using canonical correlation analysis (CCA) [Murphy, 2023, Ch. 28], which has been previously explored by Kay [1992].

D Additional Information on Benchmark Tasks

D.1 Task Descriptions

Bimodal Variables Consider a r.v. X distributed according to a Gaussian mixture model with CDF $F = w_1 F_1 + w_2 F_2$, where F_1 and F_2 are CDFs of two Gaussian distributions and w_1 and w_2 are positive weights constrained as $w_1 + w_2 = 1$. The CDF is everywhere positive, continuous, and strictly increasing. Hence, it has an inverse, the quantile function, $F^{-1}: (0, 1) \rightarrow \mathbb{R}$, which is continuous and injective as well (see Lemma D.1). Thus, we can define r.v. $X' = F^{-1}(X)$, where $X \sim \text{Uniform}(0, 1)$ and $\mathbf{I}(X; Y) = \mathbf{I}(X'; Y)$ for every r.v. Y . We generated two CDFs of Gaussian mixtures and implemented a numerical inversion in JAX [Bradbury et al., 2018] defining both X' and Y' to follow a bimodal distribution.

For the experimental values, we used

$$\begin{aligned} F_X(x) &= 0.3F(x) + 0.7F(x - 5) \\ F_Y(y) &= 0.5F(x + 1) + 0.5F(x - 3), \end{aligned}$$

where F is the CDF of the standard normal distribution $\mathcal{N}(0, 1)$.

Wiggly Mapping As the “wiggly mapping” we used

$$\begin{aligned} w_X(x) &= x + 0.4 \sin x + 0.2 \sin(1.7x + 1) + 0.03 \sin(3.3x - 2.5) \\ w_Y(y) &= y - 0.4 \sin(0.4y) + 0.17 \sin(1.3y + 3.5) + 0.02 \sin(4.3y - 2.5) \end{aligned}$$

and we used visualisations to make sure that they do not heavily distort the distribution.

Additive Noise For the additive noise model of $Y = X + N$ where $X \sim \text{Uniform}(0, 1)$ and $N \sim \text{Uniform}(-\varepsilon, \varepsilon)$ it holds that

$$\mathbf{I}(X; Y) = \begin{cases} \varepsilon - \log(2\varepsilon) & \text{if } \varepsilon \leq 1/2 \\ (4\varepsilon)^{-1} & \text{otherwise} \end{cases}$$

implying that $\mathbf{I}(X; Y) = 1.7$ for $\varepsilon = 0.1$ and $\mathbf{I}(X; Y) = 0.3$ for $\varepsilon = 0.75$.

Swiss Roll Embedding Consider a bivariate distribution P_{XY} with uniform margins $X \sim \text{Uniform}(0, 1)$ and $Y \sim \text{Uniform}(0, 1)$ and $\mathbf{I}(X; Y) = 0.8$. The Swiss roll mapping is an embedding $i: (0, 1)^2 \rightarrow \mathbb{R}^3$ given by $i(x, y) = (e(x), y)$, where $e: (0, 1) \rightarrow \mathbb{R}^2$ is given by

$$e(x) = \frac{1}{21}(t(x) \cos t(x), t(x) \sin t(x)), \quad t(x) = \frac{3\pi}{2}(1 + 2x).$$

Note that $P_{X'Y'}$ does not have a PDF with respect to the Lebesgue measure on $\mathbb{R}^3$. We visualise the Swiss roll distribution in Fig. 1.

Spiral Diffeomorphism For the precise definition and properties of the general spiral diffeomorphism consult Appendix B.

In our benchmark we applied the spiral $x \mapsto \exp(vA\|x\|^2)x$ to both X and Y variables. We used an $m \times m$ skew-symmetric matrix A_X with $(A_X)_{12} = -(A_X)_{21} = 1$ and an $n \times n$ skew-symmetric matrix A_Y with $(A_Y)_{23} = -(A_Y)_{32} = 1$. Each of these effectively “mixes” only two dimensions.

The speed parameters v_X and v_Y were chosen as $v_X = 1/m$ and $v_Y = 1/n$ to partially compensate for the fact that if $X \sim \mathcal{N}(0, I_m)$, then $\|X\|^2 \sim \chi_m^2$ and has mean m .

Multivariate Student Let Ω be an $(m+n) \times (m+n)$ positive definite dispersion matrix and ν be a positive integer (the degrees of freedom). By sampling an $(m+n)$ -dimensional random vector $(\tilde{X}, \tilde{Y}) \sim \mathcal{N}(0, \Omega)$ and a random scalar $U \sim \chi_\nu^2$ one can define rescaled variables $X = \tilde{X} \sqrt{\nu/U}$ and $Y = \tilde{Y} \sqrt{\nu/U}$, which are distributed according to the multivariate Student distribution. For $\nu = 1$ this reduces to the multivariate Cauchy distribution (and the first two moments are not defined), for $\nu = 2$ the mean exists, but covariance does not, and for $\nu > 2$ both the mean and covariance exist, with $\text{Cov}(X, Y) = \frac{\nu}{\nu-2}\Omega$, so that the tail behaviour can be controlled by changing ν . In particular, for $\nu \gg 1$ because of the concentration of measure phenomenon U has most of its probability mass around ν and the variables (X, Y) can be well approximated by $(\tilde{X}, \tilde{Y})$.

Arellano-Valle et al. [2013] proved that $\mathbf{I}(X; Y)$ can be computed via the sum of the mutual information of the Gaussian distributed basis variables and a correction term, i.e., $\mathbf{I}(X; Y) = \mathbf{I}(\tilde{X}; \tilde{Y}) + c(\nu, m, n)$, where

$$c(\nu, m, n) = f(\nu) + f(\nu+m+n) - f(\nu+m) - f(\nu+n), \quad f(x) = \log \Gamma\left(\frac{x}{2}\right) - \frac{x}{2} \psi\left(\frac{x}{2}\right),$$

and ψ is the digamma function.

Contrary to the Gaussian case, even for $\Omega = I_{m+n}$, the mutual information $\mathbf{I}(X; Y) = c(\nu, m, n)$ is non-zero, as U provides some information about the magnitude. In the benchmark we use this dispersion matrix to quantify how well the estimators can use the information contained in the tails, rather than focusing on estimating the Gaussian term.

D.2 Covariance Matrix Family

We generated jointly multivariate normal variables $X = (X_1, \dots, X_m)$ and $Y = (Y_1, \dots, Y_n)$ with the following procedure.

Consider i.i.d. Gaussian variables

$$U_{\text{all}}, U_X, U_Y, Z_1, \dots, Z_K, E_1, \dots, E_m, F_1, \dots, F_n, E'_1, \dots, E'_m, F'_{n-K}, \dots, F'_n \sim \mathcal{N}(0, 1).$$

Now let $\varepsilon_X, \varepsilon_Y, \eta_X, \eta_Y, \alpha, \beta_X, \beta_Y, \lambda \in \mathbb{R}$ be hyperparameters and for $l = 1, 2, \dots, K$ define

$$X_l = \varepsilon_X E_l + \alpha U_{\text{all}} + \beta_X U_X + \lambda Z_l, \quad Y_l = \varepsilon_Y F_l + \alpha U_{\text{all}} + \beta_Y U_Y + \lambda Z_l,$$

for $l = K+1, \dots, m$ define

$$X_l = \varepsilon_X E_l + \alpha U_{\text{all}} + \beta_X U_X + \eta_X E'_l$$

and for $l = K+1, \dots, n$ define

$$Y_l = \varepsilon_Y F_l + \alpha U_{\text{all}} + \beta_Y U_Y + \eta_Y F'_l.$$

The interpretation of the hyperparameters is the following:

1. α contributes to the correlations between all the possible pairs of variables.
2. β_X contributes to the correlations between the X variables.
3. β_Y contributes to the correlations between the Y variables.
4. λ gives additional “interaction strength” between pairs of variables X_l and Y_l for $l = 1, \dots, K$.
5. ε_X and ε_Y control part of the variance non-shared with any other variable.
6. η_X and η_Y can be used to increase the variance of X_l and Y_l for $l > K$ to match the variance of variables $l \leq K$ due to the λ .

Every term of the covariance matrix is easy to calculate analytically:

$$\text{Cov}(X_i, X_j) = \alpha^2 + \beta_X^2 + \mathbf{1}[i = j](\varepsilon_X^2 + \lambda^2 \mathbf{1}[i \leq K] + \eta_X^2 \mathbf{1}[i > K])$$

$$\text{Cov}(Y_i, Y_j) = \alpha^2 + \beta_Y^2 + \mathbf{1}[i = j](\varepsilon_Y^2 + \lambda^2 \mathbf{1}[i \leq K] + \eta_Y^2 \mathbf{1}[i > K])$$

$$\text{Cov}(X_i, Y_j) = \alpha^2 + \lambda^2 \mathbf{1}[i = j] \mathbf{1}[i \leq K]$$

In the following, we analyse some special cases.

Dense Interactions Consider $\lambda = \eta_X = \eta_Y = \beta_X = \beta_Y = 0$ and write $\varepsilon := \varepsilon_X = \varepsilon_Y$. We have

$$\text{Cov}(X_i, X_j) = \text{Cov}(Y_i, Y_j) = \alpha^2 + \varepsilon^2 \mathbf{1}[i = j]$$

$$\text{Cov}(Y_i, Y_j) = \alpha^2$$

Hence, between every pair of distinct variables we have the same correlation $\frac{\alpha^2}{\alpha^2 + \varepsilon^2}$.

Sparse Interactions Consider $\alpha = 0$, $\varepsilon := \varepsilon_X = \varepsilon_Y$, $\eta_X = \eta_Y = \lambda$ and $\beta_X = \beta_Y = 0$

$$\text{Cov}(X_i, X_j) = \text{Cov}(Y_i, Y_j) = \mathbf{1}[i = j](\varepsilon^2 + \lambda^2)$$

$$\text{Cov}(X_i, Y_j) = \lambda^2 \mathbf{1}[i = j] \mathbf{1}[i \leq K]$$

All the variables have the same variance, but the covariance structure is now different. Between (distinct) X_i and X_j the correlation is zero and similarly for (distinct) Y_i and Y_j .

Correlations between X_i and Y_j are zero unless $i = j \leq K$, when $\text{Cor}(X_i, Y_i) = \lambda^2 / (\varepsilon^2 + \lambda^2)$.

Interpolation In Section 4 we discussed a two-stage interpolation from dense matrices to sparse matrices: in the first stage α is decreased increasing λ at the same time to keep the mutual information constant. In the second stage we decrease the number of interacting variables K increasing λ . The selected covariance matrices from this two-stage process are visualised in Fig. 7.

D.3 Technical Results

In this subsection we add the technical results, related to the copula theory [Nelsen, 2007] and uniformization of random variables.

Lemma D.1. *Let X be a real-valued r.v. with a strictly increasing, positive everywhere, and continuous CDF F . Then*

$$F: \mathbb{R} \rightarrow (0, 1)$$

is a homeomorphism.

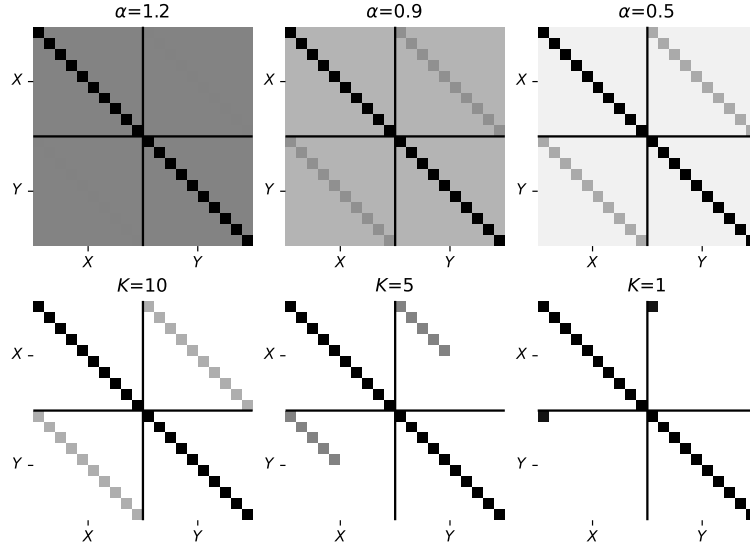


Figure 7: Interpolation between dense and sparse covariance matrices.

Proof. As F is a CDF we have $F(\mathbb{R}) \subseteq [0, 1]$. Note that 0 is excluded from the image as F is positive everywhere and 1 is excluded as F is strictly increasing. Hence, $F(\mathbb{R}) \subseteq (0, 1)$. Using continuity, strict monotonicity, and

$$\lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow +\infty} F(x) = 1$$

we note that F is a bijection onto $(0, 1)$. Analogously, we prove that

$$F((a, b)) = (F(a), F(b))$$

for all $a < b$, so F is an open map. As a continuous bijective open map F is a homeomorphism. $\square$

Lemma D.2. *Let X be a real-valued r.v. with a CDF F which is continuous, strictly increasing and positive everywhere. Then, $F(X)$ is uniformly distributed on $(0, 1)$.*

Proof. Our assumptions on F imply that it is a continuous homeomorphism $\mathbb{R} \rightarrow (0, 1)$.

Let now $y \in (0, 1)$. We have

$$P(F(X) \leq y) = P(X \leq F^{-1}(y)) = F(F^{-1}(y)) = y.$$

$\square$

E Additional Experiments

E.1 Pointwise Mutual Information

To understand why neural estimators are not able to estimate the mutual information contained in the Student distribution, we decided to qualitatively study the learned critic.

In the NWJ approach [Nguyen et al., 2007], for $N \rightarrow \infty$ data points the optimal critic $f(x, y)$ is given by

$$f(x, y) = 1 + \text{PMI}(x, y),$$

where

$$\text{PMI}(x, y) = \log \frac{p_{XY}(x, y)}{p_X(x)p_Y(y)}$$

is the pointwise mutual information [Pinsker and Feinstein, 1964, Ch. 2] and p_{XY} , p_X and p_Y are the PDFs of measures P_{XY} , P_X and P_Y , respectively.

to solve the task if for every $\ell' \geq \ell$ the estimator provides the estimate between $2/3$ and $3/2$ of the ground-truth mutual information.

We present the minimal number of samples ℓ as a function of benchmark task and considered estimator in Fig. 9. Note that CCA and KSG typically need very small amount of samples to solve the task. As neural estimators use batch size of 256 and split the sample into training and test data sets, it is not possible for them to have $\ell < 1000$, which is the smallest number of samples considered greater than $2 \cdot 256 = 512$.

E.3 Preprocessing

In Fig. 2 and Fig. 9 we transformed all samples by removing the mean and scaling by the standard deviation in each dimension calculated empirically from the sample. This strategy ensures that the scales on different axes are similar. Note that this strategy is not always suitable (e.g., if the distance metric used in the k NN approach should weight select dimensions differently) and makes the observed data points not i.i.d. (as the “shared” information, sample mean and variances, has been used to rescale the variables), although the effect of it seems to be negligible on most tasks in the benchmark.

Additionally, we considered two other preprocessing strategies, implemented using the `scikit-learn` package [Pedregosa et al., 2011]: making each marginal distribution uniform $\text{Uniform}(0, 1)$ by calculating empirical CDF along each axis (see Section D for the technical lemmata discussing this operation) and making each marginal distribution a Gaussian variable $\mathcal{N}(0, 1)$. These approaches may affect the effective sample size more strongly than the previous rescaling strategy, as estimating empirical CDF requires more data points than estimating mean and standard deviation.

We used $N = 10\,000$ data points and rerun the benchmark (see Fig. 10). Overall, the impact of the “uniformization” strategy appeared to be insubstantial for most tasks, and in certain cases negative. The “Gaussianisation” strategy improved estimation for Student distributions and helped histogram based approaches in low-dimensional tasks. Interestingly, both approaches worsen the performance of CCA on Student distributions.

While the improvements gained from “Gaussianisation” were minor, the approach is simplistic, treating each axis separately. Ideally, appropriate preprocessing would be able to transform P_{XY} into a better-behaved distribution $P_{f(X)g(Y)}$, and thus simplify the estimation. In certain cases such preprocessing could (partially) undo transformations considered in our benchmark, and in effect make the estimators invariant (or more robust) to these transformations (cf. Section 4). We consider the design of informed preprocessing strategies an important open problem.

E.4 Hyperparameter Selection

CCA We used the implementation provided in the `scikit-learn` [Pedregosa et al., 2011] Python package (version 1.2.2). We set the latent space dimension to the smaller of the dimensions of the considered random vectors X and Y .

Histograms Adaptive estimators such as JIC [Suzuki, 2016] or MIIC [Cabeli et al., 2020] were not applicable to the high-dimensional data we considered, hence we resort to using an equal-width estimator implemented in `TransferEntropy` library (version 1.10.0). We fixed the number of bins to 10, although the optimal number of bins is likely to depend on the dimensionality of the problem.

KSG As a default, we used the KSG-1 variant from the `rmi` package with 10 neighbors. However, as shown in our ablation in Fig. 11, changing the number of neighbors or using the KSG-2 variant did not considerably change the estimates. We compared the used implementation with a custom Python implementation and one from `TransferEntropy` library. The version implemented in the `TransferEntropy` suffers from strong positive bias.

LNN We used the implementation from the `rmi` package (version 0.1.1) with 5 neighbors and truncation parameter set to 30. Ablations can be seen in Fig. 11.

Neural Estimators We split the data set into two equal-sized parts (training and test) and optimized the statistics network on the training data set using the Adam optimiser with initial learning rate set to 0.1. We used batch size of 256 and ran each algorithm for up to 10 000 training steps with early stopping (checked every 250 training steps). We returned the highest estimate on the test data set.

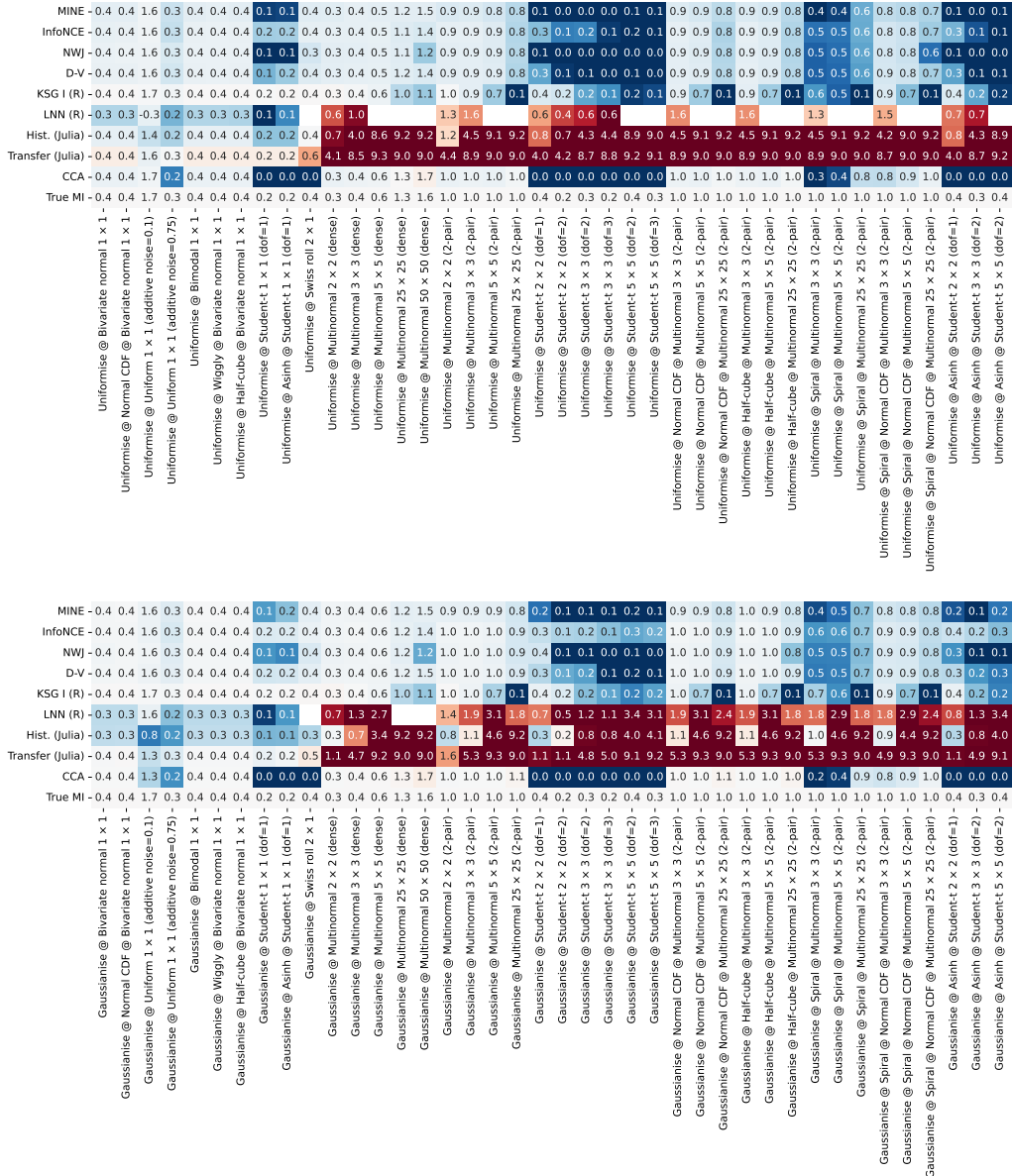


Figure 10: Estimates for preprocessed data with $N = 10,000$ samples. Top: margins transformed into uniform distributions. Bottom: margins transformed into normal distributions.

We chose the architecture using a mini-benchmark with four ReLU neural networks considered on two neural estimators (see Fig. 11) and chose a neural network with two hidden layers (network M in Table 2).

As neural estimators rely on stochastic gradient descent training, we implemented heuristic diagnostics used to diagnose potential problems with overfitting and non-convergence. We present the number of runs susceptible to these issues in Fig. 12. We did not include these runs when computing statistics in the main benchmark figures (Fig. 2 and Fig. 9).

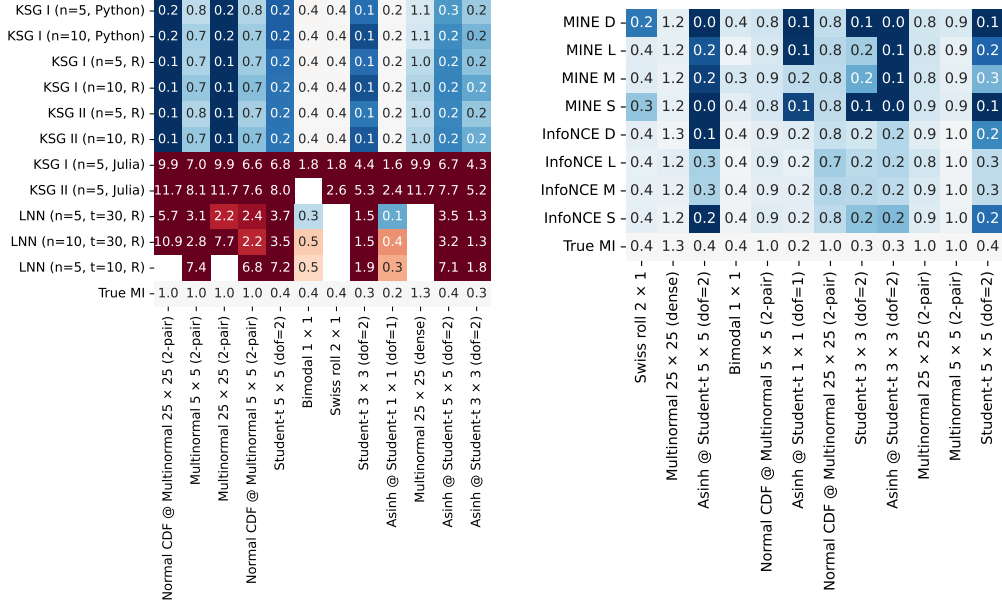


Figure 11: Left: ablation of hyperparameters of the KSG and LNN estimators (with n neighbors and threshold t). Right: ablation of neural network architectures.

Architecture	Hidden layers
D	8, 8, 8
L	24, 12
M	16, 8
S	10, 5

Table 2: Network architectures and their corresponding hidden layer specifications.

F Visualisation of Selected Tasks

In the following, we visualise a selection of benchmark tasks. Each figure visualises a sample of $N = 10\,000$ points from P_{XY} . The upper diagonal represents the scatter plots $P_{X_i X_j}$, $P_{X_i Y_j}$ and $P_{Y_i Y_j}$ for all possible combinations of $i \neq j$. The lower diagonal represents the empirical density fitted using a kernel density estimator with 5 levels. The diagonal represents the histograms of the marginal distributions P_{X_i} and P_{Y_i} for all possible i .

We present selected 1×1 -dimensional tasks in Fig. 13 and Fig. 14; the 2×1 -dimensional Swiss-roll distribution in Fig. 15 and selected 2×2 -dimensional tasks in Fig. 16. Finally, in Fig. 17 we present two 3×3 -dimensional problems, visualising the effect of the normal CDF and the spiral diffeomorphism.

MINE	-	0	0	0	0	0	0	0	7	2
InforNCE	-	0	0	0	0	0	0	4	0	0
NWJ	-	0	0	0	0	0	0	8	3	0
D-V	-	0	0	0	0	0	0	5	6	1
Bivariate normal 1 x 1 -	-	0	0	0	0	0	0	6	2	0
Bivariate normal 1 x 1 -	-	0	0	0	0	0	0	4	0	0
Uniform 1 x 1 (additive noise=0.1) -	-	0	0	0	0	0	0	3	3	0
Uniform 1 x 1 (additive noise=0.75) -	-	0	0	0	0	0	0	6	2	0
Bimodal 1 x 1 -	-	0	0	0	0	0	0	2	3	0
Wiggly @ Bivariate normal 1 x 1 -	-	0	0	0	0	0	0	0	0	0
Half-cube @ Bivariate normal 1 x 1 -	-	0	0	0	0	0	0	1	0	0
Student-t 1 x 1 (dof=1) -	-	0	0	0	0	0	0	0	0	0
Asinh @ Student-t 1 x 1 (dof=1) -	-	0	0	0	0	0	0	0	0	0
Swiss roll 2 x 1 -	-	0	0	0	0	0	0	0	0	0
Multinormal 2 x 2 (dense) -	-	0	0	0	0	0	0	0	0	0
Multinormal 3 x 3 (dense) -	-	0	0	0	0	0	0	0	0	0
Multinormal 5 x 5 (dense) -	-	0	0	0	0	0	0	0	0	0
Multinormal 25 x 25 (dense) -	-	0	0	0	0	0	0	0	0	0
Multinormal 50 x 50 (dense) -	-	0	0	0	0	0	0	0	0	0
Multinormal 2 x 2 (2-pair) -	-	0	0	0	0	0	0	0	0	0
Multinormal 3 x 3 (2-pair) -	-	0	0	0	0	0	0	0	0	0
Multinormal 5 x 5 (2-pair) -	-	0	0	0	0	0	0	0	0	0
Multinormal 25 x 25 (2-pair) -	-	0	0	0	0	0	0	0	0	0
Student-t 2 x 2 (dof=1) -	-	0	0	0	0	0	0	0	0	0
Student-t 2 x 2 (dof=2) -	-	0	0	0	0	0	0	0	0	0
Student-t 3 x 3 (dof=2) -	-	0	0	0	0	0	0	0	0	0
Student-t 3 x 3 (dof=3) -	-	0	0	0	0	0	0	0	0	0
Student-t 5 x 5 (dof=2) -	-	0	0	0	0	0	0	0	0	0
Student-t 5 x 5 (dof=3) -	-	0	0	0	0	0	0	0	0	0
Multinormal 3 x 3 (2-pair) -	-	0	0	0	0	0	0	0	0	0
Multinormal 5 x 5 (2-pair) -	-	0	0	0	0	0	0	0	0	0
Multinormal 25 x 25 (2-pair) -	-	0	0	0	0	0	0	0	0	0
Half-cube @ Multinormal 3 x 3 (2-pair) -	-	0	0	0	0	0	0	0	0	0
Half-cube @ Multinormal 5 x 5 (2-pair) -	-	0	0	0	0	0	0	0	0	0
Half-cube @ Multinormal 25 x 25 (2-pair) -	-	0	0	0	0	0	0	0	0	0
Spiral @ Multinormal 3 x 3 (2-pair) -	-	0	0	0	0	0	0	0	0	0
Spiral @ Multinormal 5 x 5 (2-pair) -	-	0	0	0	0	0	0	0	0	0
Spiral @ Multinormal 25 x 25 (2-pair) -	-	0	0	0	0	0	0	0	0	0
Spiral @ Normal CDF @ Multinormal 3 x 3 (2-pair) -	-	0	0	0	0	0	0	0	0	0
Spiral @ Normal CDF @ Multinormal 5 x 5 (2-pair) -	-	0	0	0	0	0	0	0	0	0
Spiral @ Normal CDF @ Multinormal 25 x 25 (2-pair) -	-	0	0	0	0	0	0	0	0	0
Asinh @ Student-t 2 x 2 (dof=2) -	-	0	0	0	0	0	0	0	0	0
Asinh @ Student-t 3 x 3 (dof=2) -	-	0	0	0	0	0	0	0	0	0
Asinh @ Student-t 5 x 5 (dof=2) -	-	0	0	0	0	0	0	0	0	0

Figure 12: Number of runs (out of 10) when our convergence and overfitting diagnostics diagnosed potential problems in neural estimator training.

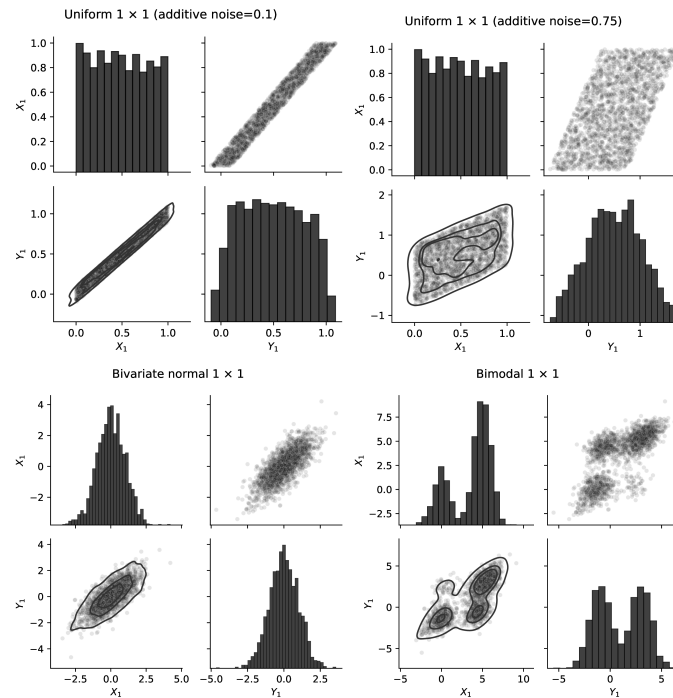


Figure 13: Top row: additive noise model. Bottom row: bivariate normal distribution and transformation making each variable a bimodal mixture of Gaussians.

G Author Contributions

Contributions according to the Contributor Roles Taxonomy:

1. Conceptualization: PC, FG, AM, NB.
2. Methodology: FG, PC.
3. Software: PC, FG.
4. Validation: FG, PC.
5. Formal analysis: PC, FG.
6. Investigation: FG, PC, AM.

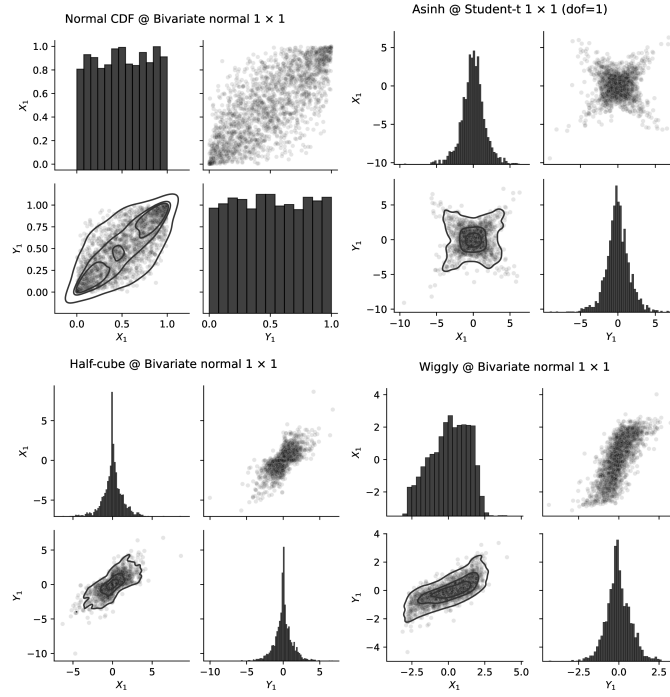


Figure 14: Top row: bivariate normal distribution transformed so that the margins are uniform and asinh-transformed bivariate Student distribution with 1 degree of freedom. Bottom row: bivariate normal distribution transformed with the half-cube and the wiggly mappings.

7. Resources: FG, NB.
8. Data Curation: FG, PC.
9. Visualization: FG, PC.
10. Supervision: AM, NB.
11. Writing – original draft: PC, AM, FG.
12. Writing – review and editing: PC, AM, FG, NB, JEV.

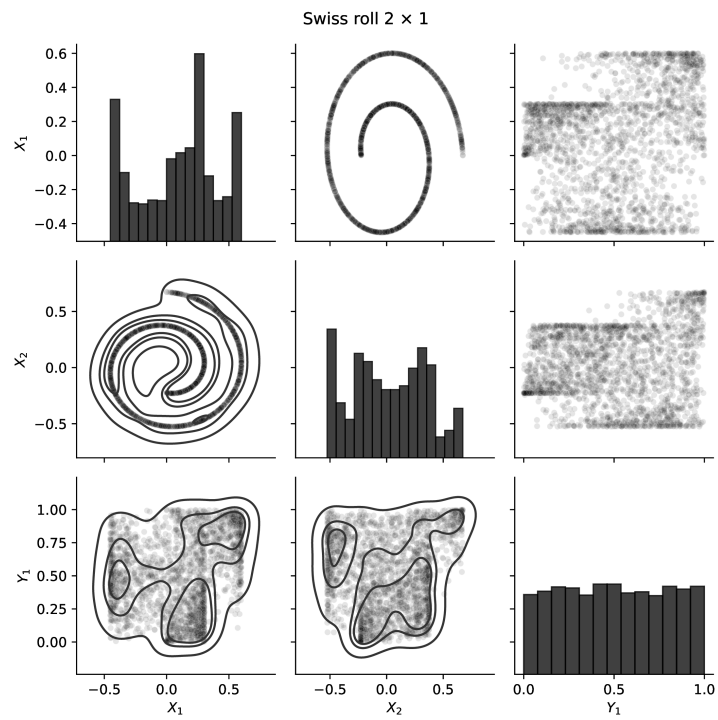


Figure 15: Visualisation of the Swiss-roll distribution.

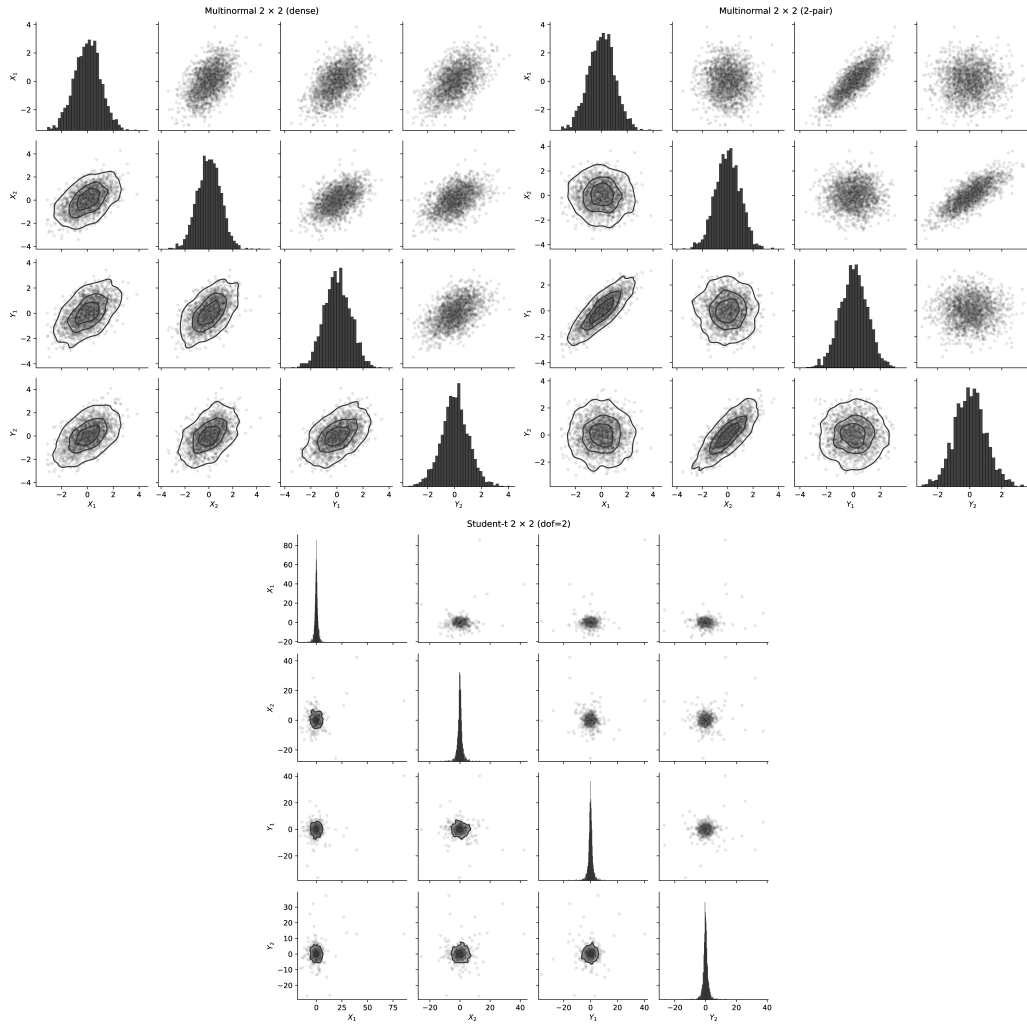


Figure 16: Visualisations of selected 2×2 -dimensional problems. Top row: multivariate normal distributions with dense and sparse interactions. Bottom row: multivariate Student distribution with 2 degrees of freedom.

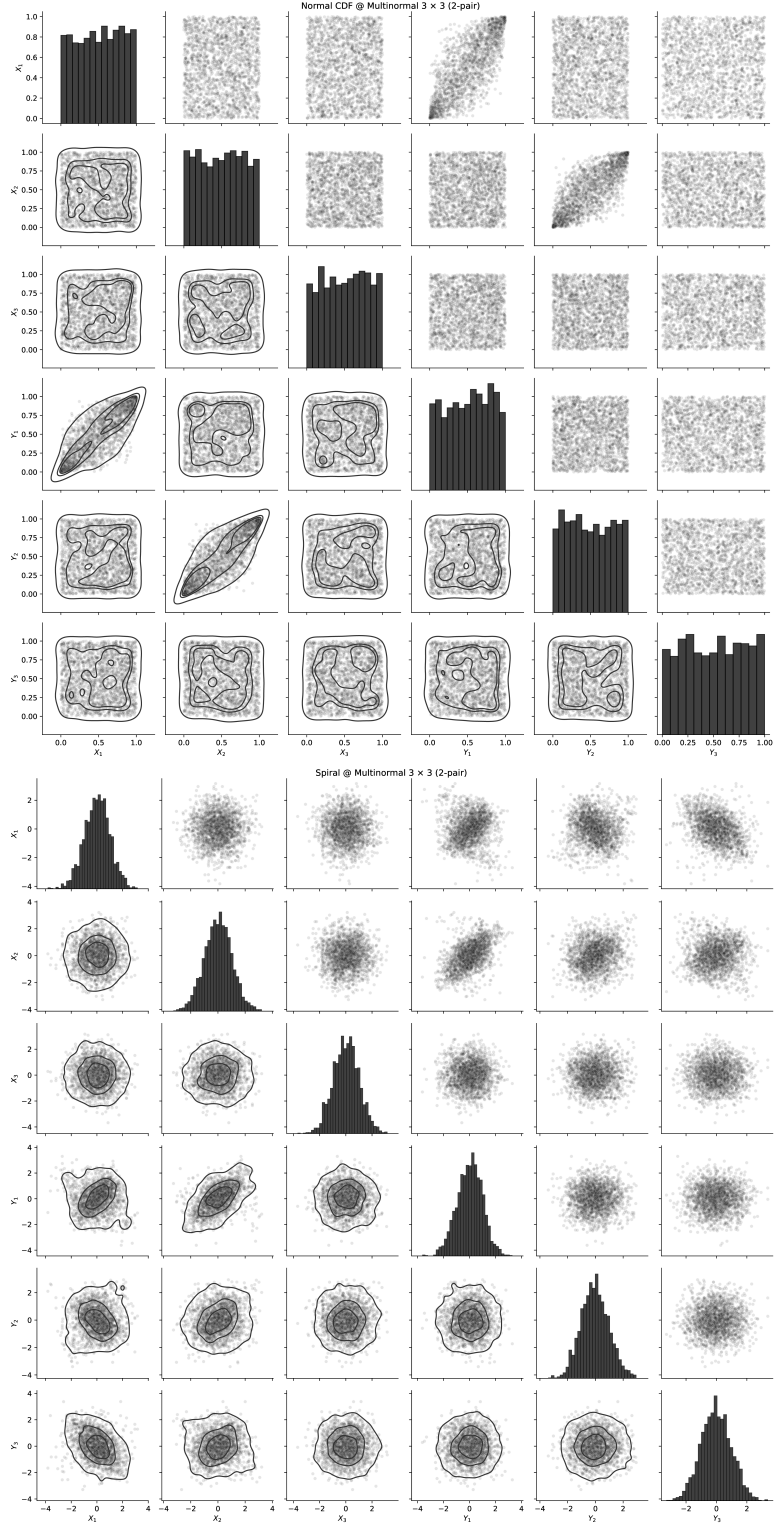


Figure 17: Visualisations of selected 3×3 -dimensional problems. Top: multivariate normal distribution transformed axis-wise with normal CDF, to uniformise margins. Bottom: multivariate normal distribution transformed with the spiral diffeomorphism.

On the Properties and Estimation of Pointwise Mutual Information Profiles

Paweł Czyż*

*ETH AI Center and Department of Biosystems Science and Engineering
ETH Zurich
Zurich, Switzerland*

pawelpiotr.czyz@ai.ethz.ch

Frederic Grabowski*

*Institute of Fundamental Technological Research
Warsaw, Poland*

frederic.grabowski@ippt.pan.pl

Julia E. Vogt

*Department of Computer Science
ETH Zurich
SIB Swiss Institute for Bioinformatics
Zurich, Switzerland*

julia.vogt@inf.ethz.ch

Niko Beerenwinkel†

*Department of Biosystems Science and Engineering
ETH Zurich
SIB Swiss Institute for Bioinformatics
Basel, Switzerland*

niko.beerenwinkel@bsse.ethz.ch

Alexander Marx†

*Research Center Trustworthy Data Science and Security of the University Alliance Ruhr
Department of Statistics
TU Dortmund University
Dortmund, Germany*

alexander.marx@tu-dortmund.de

Reviewed on OpenReview: <https://openreview.net/forum?id=LdflD41Gn8>

Abstract

The pointwise mutual information profile, or simply profile, is the distribution of pointwise mutual information for a given pair of random variables. One of its important properties is that its expected value is precisely the mutual information between these random variables. In this paper, we analytically describe the profiles of multivariate normal distributions and show that for an expressive family of distributions, termed *Bend and Mix Models*, the profile can be accurately estimated using Monte Carlo methods. We then show how Bend and Mix Models can be used to study the limitations of existing mutual information estimators, investigate the behavior of neural critics used in variational estimators, and understand the effect of experimental outliers on mutual information estimation. Finally, we show how Bend and Mix Models can be used to obtain model-based Bayesian estimates of mutual information, suitable for problems with available domain expertise in which uncertainty quantification is necessary. The accompanying code is available at <https://github.com/cbg-ethz/bmi>.

*Equal contribution. †Joint supervision.

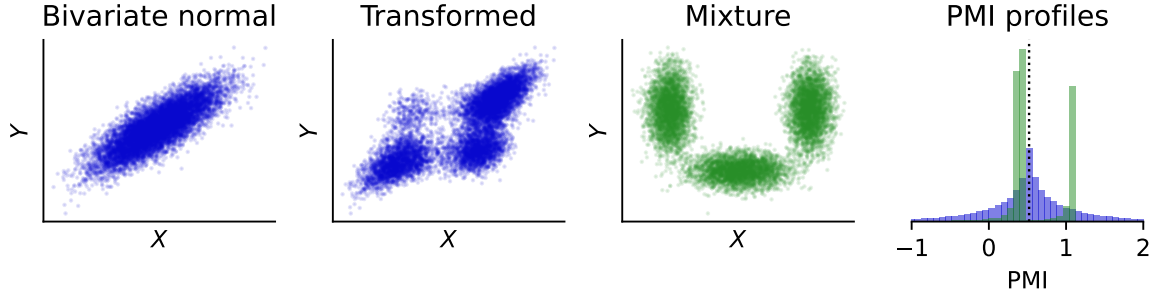


Figure 1: First two panels: samples from a bivariate normal distribution and the same distribution with marginals transformed. Both distributions have the same PMI profile (blue histogram in the fourth panel). Third panel: mixture distribution, which cannot be obtained as a transformation of the normal distribution due to a distinct PMI profile (green histogram in the fourth panel). All three distributions have the same mutual information, marked with the black line in panel four.

1 Introduction

Mutual information (MI) is a key measure of dependence between two random variables (r.v.s). Due to its theoretical properties, including invariance to reparametrization, it has found many applications (Polyanskiy & Wu, 2022). However, estimating mutual information from a finite sample remains challenging (McAllester & Stratos, 2020) and the development of mutual information estimators remains an active area of research (Kay, 1992; Kraskov et al., 2004; Oord et al., 2018; Belghazi et al., 2018; Carrara & Ernst, 2023). In response to the introduction of new estimators, various benchmarks have been proposed to assess trade-offs (Khan et al., 2007; Poole et al., 2019; Czyż et al., 2023).

In this manuscript we study the pointwise mutual information profile (PMI profile), a closely related statistical measure which has received much less attention. The PMI profile is a distribution whose expected value is the mutual information between two considered r.v.s. We determine the PMI profile for multivariate normal distributions analytically and show that the PMI profile, similarly to mutual information, is invariant to reparametrization. This last property has an important implication for MI benchmarks (Czyż et al., 2023) which transform simple distributions with known MI to create more complex benchmarking tasks: using the invariance of the profile, we demonstrate that the family of distributions obtained in this manner is inherently limited. For example, distributions such as those studied by Grabowski et al. (2019) and Carrara & Ernst (2023), cannot be obtained by transforming a normal distribution.

To overcome this limitation, we study a family of distributions, here termed Bend and Mix Models (BMMs), for which the PMI profile can be efficiently approximated. By extension, since MI is the expected value of the PMI profile, it can also be readily estimated using Monte Carlo methods. BMMs include highly-expressive Gaussian mixture models, and thus allow us to model complex distributions. Through a series of numerical experiments we demonstrate the usefulness of BMMs for creating non-trivial benchmark tasks; in particular, we discuss robustness of mutual information estimators to inlier and outlier noise. Additionally, BMMs allow us to investigate the properties of PMI profiles directly. Based on empirical evidence, we argue that existing estimators based on neural variational lower bounds (Nguyen et al., 2007; Belghazi et al., 2018) implicitly estimate the PMI profile, even though they do not reliably estimate the pointwise mutual information *function*.

Finally, we show that BMMs can provide model-based Bayesian estimates of both the MI and the PMI profile, generalizing the earlier approaches of Kay (1992) and Brillinger (2004) to a large class of models. Although this approach is not universal, we demonstrate that in low-dimensional problems with available domain expertise, BMMs provide state-of-the-art point estimates. Additionally, they offer a principled approach to uncertainty quantification in MI estimation, which has been lacking in other approaches.

To summarize, the main contributions of this manuscript are as follows:

- We describe the PMI profile, an invariant generalizing mutual information, determine the PMI profile for normal distributions and prove further characterising properties (Sec. 2.1).

- We show that for an expressive family of distributions, called Bend and Mix Models, the PMI profile, and hence mutual information, is tractable using Monte Carlo methods (Sec. 2.2).
- We use Bend and Mix Models to construct novel benchmarking tasks (Sec. 3.1–3.2 and Appendix C.2), study variational MI estimators (Sec. 3.3), and propose model-based Bayesian MI estimators (Sec. 3.4).

2 Theoretical framework

In this section, we first introduce PMI profiles and then define Bend and Mix Models, an expressive family of distributions for which the PMI profile can be estimated with Monte Carlo methods.

2.1 Pointwise mutual information profiles

Consider random variables X and Y valued in spaces $\mathcal{X}$ and $\mathcal{Y}$. Provided that the joint distribution P_{XY} is sufficiently regular (in the sense of belonging to the family $\mathcal{P}(\mathcal{X}, \mathcal{Y})$ defined in Appendix A.1), the mutual information is given as follows (Pinsker & Feinstein, 1964):

Definition 1. Let X and Y be random variables valued in spaces $\mathcal{X}$ and $\mathcal{Y}$, respectively, such that $P_{XY} \in \mathcal{P}(\mathcal{X}, \mathcal{Y})$. The mutual information (MI) is defined as $\mathbf{I}(X; Y) = \mathbb{E}_{(x,y) \sim P_{XY}} [\text{PMI}_{XY}(x, y)]$, where the pointwise mutual information (PMI) function is given as¹

$$\text{PMI}_{XY}(x, y) = \log \frac{p_{XY}(x, y)}{p_X(x) p_Y(y)},$$

where p_{XY} , p_X and p_Y are PDFs (or PMFs) of the joint and marginal distributions, respectively.

While mutual information measures whether two variables are independent (which is equivalent to $\mathbf{I}(X; Y) = 0$), the pointwise mutual information function can be understood as a measure of similarity of two events: it is positive if they jointly occur at a higher frequency than if they were sampled independently from the marginal distributions. It is also often associated with variational MI estimators (Nguyen et al., 2007; Belghazi et al., 2018; Oord et al., 2018), where one aims to approximate it using a neural network (cf. Sec. 3.3). However, in this work we focus on the distribution of its values, i.e., its profile:

Definition 2. The pointwise mutual information profile² Prof_{XY} is defined as the distribution of the random variable $T = \text{PMI}_{XY}(X, Y)$.

We discuss generalizations of the PMI profile in Appendix A.5. A key property of the PMI profile is that its mean is equal to mutual information, $\mathbf{I}(X; Y) = \mathbb{E}_{T \sim \text{Prof}_{XY}} [T]$. In the following, we characterize the profile for multivariate normal and discrete distributions, and study its invariance properties. It is known that MI is invariant under diffeomorphisms (Kraskov et al., 2004). More generally, in Appendix A.1, we show that the entire profile is invariant:

Theorem 3. Let $P_{XY} \in \mathcal{P}(\mathcal{X}, \mathcal{Y})$ and $f: \mathcal{X} \rightarrow \mathcal{X}$ and $g: \mathcal{Y} \rightarrow \mathcal{Y}$ be diffeomorphisms. Then for $X' = f(X)$ and $Y' = g(Y)$ it holds that $P_{X'Y'} \in \mathcal{P}(\mathcal{X}, \mathcal{Y})$ and $\text{Prof}_{X'Y'} = \text{Prof}_{XY}$.

We demonstrate this result in Fig. 1. The profile of the mixture distribution cannot be obtained as a transformation of the bivariate normal distribution. This, in particular, implies that benchmarks relying on transforming normal distributions for generating more complex problems (Czyż et al., 2023), cannot create distributions with PMI profiles that differ from the normal distributions being transformed.

In fact, for multivariate normal distributions we can express the PMI profile analytically based on the notion of canonical correlations (Hotelling, 1936):

Theorem 4. Let X and Y be r.v.s such that the joint distribution $P_{XY} \in \mathcal{P}(\mathbb{R}^m, \mathbb{R}^n)$ is multivariate normal. If $k = \min(m, n)$ and $\rho_1, \rho_2, \dots, \rho_k$ are canonical correlations between X and Y , then the profile Prof_{XY} is a

¹We use the natural logarithm, meaning that all quantities are measured in nats.

²Although we are not aware of a prior formal definition and studies of the PMI profile, histograms of approximate PMI between words have been studied before in the computational linguistics community (Allen & Hospedales, 2019).

generalized χ^2 distribution, namely the distribution of the variable

$$T = \mathbf{I}(X; Y) + \sum_{i=1}^k \frac{\rho_i}{2} (Q_i - Q'_i),$$

where $\{Q_i, Q'_i\}_{i=1, \dots, k}$ are independent and identically distributed (i.i.d.) variables drawn from the χ^2_1 distribution. In particular, Prof_{XY} is symmetric around its median, which coincides with the mean

$$\mathbf{I}(X; Y) = -\frac{1}{2} \sum_{i=1}^k \log(1 - \rho_i^2),$$

and all moments of this distribution exist. Its variance is equal to $\text{Var}[T] = \rho_1^2 + \dots + \rho_k^2$.

Additionally, in Appendix A.2, we derive bounds on the variance of the PMI profile of multivariate normal distributions given the MI. In the same place, we prove the following results, characterising PMI profiles when MI is zero and for discrete random variables:

Proposition 5. *Let X and Y be r.v.s with joint distribution $P_{XY} \in \mathcal{P}(\mathcal{X}, \mathcal{Y})$. Then, $\mathbf{I}(X; Y) = 0$ if and only if $\text{Prof}_{XY} = \delta_0$ is the Dirac measure with a single atom at 0.*

Proposition 6. *If X and Y are discrete r.v.s with $P_{XY} \in \mathcal{P}(\mathcal{X}, \mathcal{Y})$, then the PMI profile is discrete:*

$$\text{Prof}_{XY} = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{XY}(x, y) \delta_{\text{PMI}_{XY}(x, y)}.$$

For more complex distributions, analytic formulae governing the PMI profiles are not known. Below we address this issue by offering a Monte Carlo approximation for a wide family of distributions.

2.2 Bend and Mix Models

As discussed above, PMI profiles are analytically known only for a small number of basic distributions and their reparametrizations, which do not change the PMI profile (Fig. 1). To generate a wider family of distributions with distinct PMI profiles, we introduce *Bend and Mix Models* which combine two strategies: *bending* a distribution (transforming with diffeomorphisms; see also the literature on normalizing flows, e.g., Kobyzev et al. (2021) and Papamakarios et al. (2021)) and *mixing* (combining multiple models into a mixture model). Although the mixing operation generally leads to distributions whose PMI profile is not available analytically, we can efficiently construct numerical approximations via Monte Carlo approaches. To ensure this, we require the following property:

Definition 7 (informal). *Every distribution $P_{XY} \in \mathcal{P}(\mathcal{X}, \mathcal{Y})$ for which we can efficiently sample $(X, Y) \sim P_{XY}$ and numerically evaluate the densities $p_{XY}(x, y)$, $p_X(x)$ and $p_Y(y)$ at every point $(x, y) \in \mathcal{X} \times \mathcal{Y}$ is considered a Bend and Mix Model (BMM).*

Any distribution that satisfies this definition can be used as a basic building block for a BMM. Examples include discrete distributions as well as multivariate normal and Student distributions. More complex distributions, such as a mixture of discrete and continuous random variables (see discussion and example in Appendix B), can then be constructed using bending and mixing operations:

Proposition 8. *If P_{XY} is a BMM and f and g are diffeomorphisms with numerically tractable Jacobians (e.g., normalizing flows), then $P_{f(X)g(Y)}$ is a BMM.*

Proposition 9. *Consider BMMs $P_{X_1Y_1}, \dots, P_{X_KY_K}$. If $w_1, \dots, w_K$ are positive weights, such that $w_1 + \dots + w_K = 1$, then the mixture distribution $P_{X'Y'} = w_1 P_{X_1Y_1} + \dots + w_K P_{X_KY_K}$ is a BMM.*

We prove both propositions in Appendix A.3. Note that since Gaussian mixture models are universal approximators of smooth densities, and they belong to the BMM family, in principle one could use these constructions to approximate any distribution. In practice, the evaluation time of PMI grows linearly with

the number of components and can become costly for a high number of mixture components (or when large normalizing flow architectures are used).

Unfortunately, the MI of a mixture cannot be computed naively by a weighted sum of the information shared within the individual components. To illustrate this, we study two special cases in which constructing a mixture *decreases*, as well as *increases* the overall MI compared to the weighted sum of the components:

Proposition 10. *Consider r.v.s (X_k, Y_k) such that $\mathbf{I}(X_k; Y_k) < \infty$ for $k = 1, \dots, K$. Let (X', Y') be their mixture with weights $w_1, \dots, w_K$. Then,*

$$0 \leq \mathbf{I}(X'; Y') \leq \sum_{k=1}^K w_k \mathbf{I}(X_k; Y_k) + \log K.$$

Moreover, these inequalities are tight:

1. There exists a mixture such that $\mathbf{I}(X'; Y') = \log K$ even though $\mathbf{I}(X_k; Y_k) = 0$ for all k .
2. There exists a mixture such that $\mathbf{I}(X'; Y') = 0$ even though $\mathbf{I}(X_k; Y_k) > 0$ for all k .

More general bounds on MI in mixture models have also been studied by [Haussler & Oppor \(1997\)](#) and [Kolchinsky & Tracey \(2017\)](#), but for completeness we include a proof in [Appendix A.4](#).

The properties of BMMs are chosen so that we can estimate the MI using Monte Carlo approaches up to an arbitrary accuracy, rather than relying on upper and lower bounds. We can sample $T \sim \text{Prof}_{XY}$ by sampling a data point (x, y) from the joint distribution P_{XY} and evaluating $t = \text{PMI}_{XY}(x, y)$. Then, MI can be approximated with a Monte Carlo estimate of the integral $\mathbf{I}(X; Y) = \mathbb{E}[T]$. Assuming $\mathbf{I}(X; Y) < \infty$, the Monte Carlo estimator of MI is guaranteed to be unbiased. However, the quality of the estimation depends on the variance of the PMI profile and the number of Monte Carlo samples used. For example, if the variance is infinite, the Monte Carlo estimates (even though unbiased) will not be adequate due to enormous variance and several independent runs may be necessary to diagnose such problems. Additionally, the relative error of the Monte Carlo approximation can be large when the number of samples is not large enough, but the ground-truth MI is close to zero. For a more detailed discussion of Monte Carlo standard error (MCSE) under different regularity conditions we refer to [Flegal et al. \(2008\)](#) and [Koehler et al. \(2009\)](#).

Analogously, we can estimate the PMI profile with a histogram: for a bin $B \subset \mathbb{R}$ one can introduce its indicator function $\mathbf{1}_B$ and integrate $\mathbb{E}[\mathbf{1}_B(T)]$. Its cumulative density function can be approximated with an empirical sample using the expectations $\mathbb{E}[\mathbf{1}_{(-\infty, a_k]}(T)]$ for a given increasing sequence (a_k) . Since the indicator functions are bounded between 0 and 1, the Monte Carlo estimator for both quantities is unbiased. For N samples, the standard error is bounded above by $1/\sqrt{4N}$, according to the inequality of [Popoviciu \(1935\)](#).

3 Case studies

In this section, we apply Bend and Mix Models to four distinct problems. In [Sec. 3.1](#) we demonstrate how they can be used to extend existing benchmarks of MI estimators (with an explicit benchmark constructed in [Appendix C.2](#)). In [Sec. 3.2](#) we show how BMMs can be used to investigate the robustness of MI estimation to outliers and inliers. In [Sec. 3.3](#) we apply BMMs to investigate the biases and sample efficiency of variational estimators employing neural critics. Finally, in [Sec. 3.4](#) we show that if the distribution P_{XY} is known to be well-approximated by a family of BMMs, we can provide Bayesian estimates of both MI and the PMI profile.

3.1 Novel distributions for estimator evaluation

To illustrate how BMMs can be used to create expressive benchmark tasks, we implemented a benchmark of 26 continuous distributions in TensorFlow Probability on JAX ([Dillon et al., 2017](#); [Bradbury et al., 2018](#)) ([Appendix C.2](#)). Additionally, in [Appendix B](#), we consider distributions involving discrete variables. In [Fig. 2](#) we visualise samples from four example distributions: the X distribution is a mixture of two bivariate normal distributions. The marginal distributions P_X and P_Y are normal, although the joint distribution P_{XY} is not. The AI distribution is a mixture of six bivariate normal distributions, illustrating how expressive

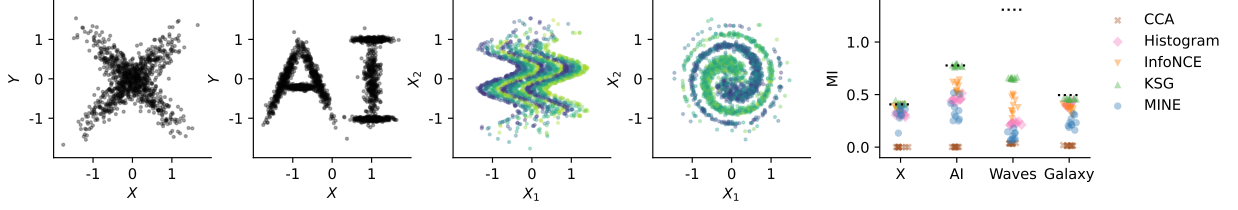


Figure 2: Samples from the example distributions. Distributions X and AI represent one-dimensional variables X and Y . Distributions Waves and Galaxy plot two-dimensional X variable using spatial coordinates, while one-dimensional Y variable is represented by color. The rightmost plot presents estimates according to different mutual information algorithms using independently generated data sets with $N = 5000$ points each, compared to the ground-truth MI of the distribution (dotted line).

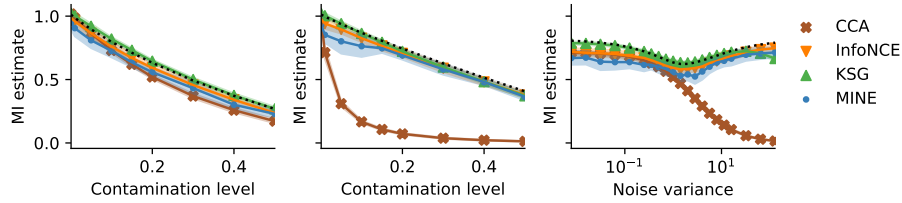


Figure 3: Left: increasing the contamination level α with inlier noise distribution. Middle: increasing the contamination level α with outlier noise distribution. Right: increasing the variance of the noisy normal distribution for constant contamination of 20%. Outliers have less impact than inliers.

BMMs can be. The two final distributions consist of a two-dimensional X variable and a one-dimensional Y variable, allowing us to apply more complex transformations to the X variable. The Waves distribution is a mixture of twelve multivariate normal distributions with the X variable transformed, while the Galaxy distribution is a mixture of two multivariate normal distributions with the X variable transformed by the spiral diffeomorphism (Czyż et al., 2023). A detailed description of these distributions can be found in Appendix C.1.

For each BMM we sampled ten data sets with $N = 5000$ points and applied five estimators: the histogram-based estimator (Cellucci et al., 2005; Darbellay & Vajda, 1999), the popular KSG estimator (Kraskov et al., 2004), canonical correlation analysis (CCA; Kay (1992); Brillinger (2004)), and two neural estimators: InfoNCE (Oord et al., 2018) and MINE (Belghazi et al., 2018) (see Appendix C.3 for hyperparameters used). The estimates are shown in Fig. 2.

Even though problems in Fig. 2 are low-dimensional and do not encode more information than 1.5 nats, they pose a considerable challenge for the estimators. The KSG estimator, which performs well in low-dimensional tasks, gave the best estimate in all tasks. However, the Waves task was not solved by any estimator. The CCA estimator, excelling at distributions that are close to multivariate normal (Czyż et al., 2023), is not able to capture any information at all. This suggests that BMMs can provide a rich set of distributions that can be used to test MI estimators. In Appendix C.2 we evaluate KSG, CCA and four neural estimators on the compiled list of 26 proposed benchmark tasks. Particularly interesting are the tasks including inliers, which we study in the next section.

3.2 Modeling inliers and outliers

In this section, we use BMMs to study the effect of inliers and outliers on MI estimation. Consider an electric circuit or a biological system modeled as a communication channel $p_{Y|X}(y|x)$. The researcher controls the input variable X , which results in a distribution P_X , and subsequently measures the outcome variable Y . The mutual information $\mathbf{I}(X;Y)$ is then estimated from the experimental samples (Nałecz-Jawecki et al., 2023).

However, every experimental system can suffer from occasional failures. We model the output of a failing system with a noise distribution with a PDF $n(y)$. If the probability of system failure, denoted by α , and the distribution of noise $n(y)$ do not depend on the input value x , the communication channel becomes a mixture $p_{Y'|X}(y|x) = (1 - \alpha)p_{Y|X}(y|x) + \alpha n(y)$. By multiplying both sides by the density $p_X(x)$, the distribution of the channel inputs provided by the scientist, we arrive at the mixture distribution $P_{XY'}$ with PDF $p_{XY'}(x, y) = (1 - \alpha)p_{XY}(x, y) + \alpha n(y)p_X(x)$. If the system failure is unnoticed, one can only measure Y' , rather than Y . It is therefore of interest to understand how much $\mathbf{I}(X; Y)$ and $\mathbf{I}(X; Y')$ can differ under realistic assumptions on the noise $n(y)$ and whether the standard MI estimation techniques are robust to it. In Appendix A.4 we prove the following upper bound:

Proposition 11. *Let $\alpha \in [0, 1]$ be a parameter. Consider variables X , Y and Y' , s.t. $p_{XY'}(x, y) = (1 - \alpha)p_{XY}(x, y) + \alpha n(y)p_X(x)$. Then, $\mathbf{I}(X; Y') \leq (1 - \alpha)\mathbf{I}(X; Y)$.*

Alternatively, we can model the system as a BMM and evaluate this quantity using the Monte Carlo approximation. For example, consider a setting with a two-dimensional input variable X and two-dimensional output variables Y (perfect output) and Y' (contaminated output). As the joint density p_{XY} we use a multivariate normal with unit scale and correlations $\text{corr}(X_1, Y_1) = \text{corr}(X_2, Y_2) = 0.8$ and for the noise $n(y)$ we use a multivariate normal distribution with covariance $\sigma^2 I_2$. If $\sigma^2 \approx 1$ this results in inliers, where the noise distribution is hard to distinguish from the signal. For $\sigma^2 \ll 1$ the system failures are all close to 0, while outliers are present for $\sigma^2 \gg 1$.

In Fig. 3 we present the results of three experiments: in the first two, we changed the contamination level $\alpha \in [0, 0.5]$ for $\sigma^2 = 1$ (inlier noise) and $\sigma^2 = 5^2$ (outlier noise) respectively. In the third experiment, we fixed $\alpha = 0.2$ and varied $\sigma^2 \in [2^{-7}, 2^8]$. We see that the inlier noise results in a slightly faster decrease of mutual information, while the outlier noise decreases almost linearly following the upper bound $(1 - \alpha)\mathbf{I}(X; Y)$. Interestingly, in this low-dimensional setting, the KSG, MINE, and InfoNCE estimators reliably estimate the mutual information $\mathbf{I}(X; Y')$, which can significantly differ from $\mathbf{I}(X; Y)$. Although CCA would be the preferred method to estimate $\mathbf{I}(X; Y)$ without any noise in this linear setting (Czyż et al., 2023), even a small number of outliers ($\alpha = 5\%$) can result in unreliable estimates.

3.3 Variational estimators and the PMI profile

Variational estimators of mutual information are frequently used in self-supervised learning (Oord et al., 2018) and optimize a critic function $f: \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$ to obtain an approximate lower bound on mutual information. For example, Belghazi et al. (2018) use the Donsker–Varadhan loss, $I_{\text{DV}}(f) = \mathbb{E}_{P_{XY}}[f] - \log \mathbb{E}_{P_X \otimes P_Y}[\exp f]$, which is a lower bound on $\mathbf{I}(X; Y)$ for any bounded function f . This lower bound becomes tight when $f = \text{PMI}_{XY} + c$ where c is any real number, i.e., $I_{\text{DV}}(f) = \mathbf{I}(X; Y)$. Hence, one can approach MI estimation by optimizing I_{DV} over a flexible family of functions f , usually parameterized by a neural network. Other examples include the NWJ estimator (Nguyen et al., 2007), which becomes tight for $f = \text{PMI}_{XY} + 1$, and InfoNCE (Oord et al., 2018), which has a functional degree of freedom, i.e., $f(x, y) = \text{PMI}_{XY}(x, y) + c(x)$ for a function c . We provide more details on these methods in Appendix C.4. Typically, these losses are interpreted as approximate lower bounds on MI and connect f to the PMI function (Poole et al., 2019; Song & Ermon, 2020) in the infinite sample limit. In practice, however, only a finite sample is available and the critic f is modeled via some parametric family (e.g., a neural network), which has to be learned from the data available. Since the ideal critic function of a neural estimator is an approximation of PMI, we investigate how well the PMI function can be learned by the previously described estimators, as well as investigate how they behave under misspecification.

Critics and the PMI function We simulated $N = 5000$ data points from a mixture of four bivariate normal distributions (Fig. 4) with $\mathbf{I}(X; Y) = 0.36$ and fitted the neural critics (see Appendix C) to half of the data, retaining the latter half as the test set, on which the final estimates were obtained, yielding $I_{\text{NWJ}} = 0.33$, $I_{\text{DV}} = 0.32$, $I_{\text{NCE}} = 0.35$. This (minor) difference can be attributed to the learned critic not matching the true PMI (up to the required constants), estimation error due to evaluation on a finite batch, or both. In Fig. 4 we plot the PMI function and the optimized critics. As the Donsker–Varadhan estimator estimates PMI only up to an additive constant, we normalized the critic and the PMI plots to have zero mean. Analogously, we removed the functional degree of freedom $c(x)$ from the InfoNCE estimator by removing

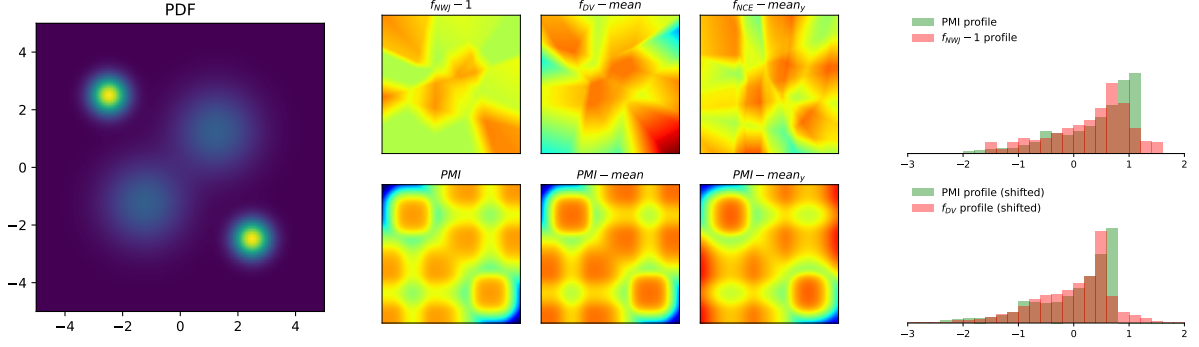


Figure 4: Left: PDF of the considered distribution. Middle: neural critic and PMI values. Right: normalized neural critic and PMI profiles.

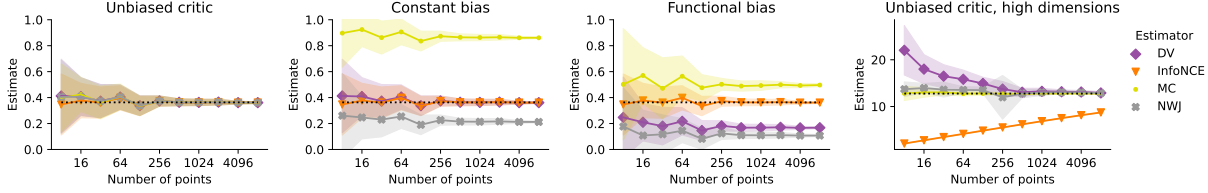


Figure 5: Estimation of mutual information using a function approximating PMI as a function of sample size for Monte Carlo (MC), InfoNCE, Donsker-Varadhan and NWJ losses. From left to right: true PMI function from Fig. 4 is used, a constant bias is added, a functional bias is added. The rightmost plot: true PMI function for a different, high-dimensional problem is used.

the mean calculated along the y dimension. Overall we saw a mismatch, suggesting that neural critics do not capture the PMI function of the distribution well. However, we can compare the PMI profile with the histogram of the values predicted by the critic in Fig. 4 (shifting the PMI profile and the histogram to have mean 0 for the Donsker–Varadhan estimator; for InfoNCE it is not possible to compare the PMI profile with the critic values due to the functional degree of freedom $c(x)$). For both NWJ and DV, we see little discrepancy between the (shifted) PMI profile and learned values. *This suggests that although the neural critics may not learn the PMI function properly in regions with low density, the PMI profile (and, hence, the MI) can still be approximated well.*

Robustness to misspecification of the critic In the next experiment (Fig. 5), we evaluated I_{DV} , I_{NWJ} , I_{InfoNCE} , and the simple Monte Carlo estimator (MC) for increasing sample size N where we provided them with an “oracle” critic f , which we varied as follows: $f(x, y) - \text{PMI}_{XY}(x, y) \in \{0, c, \sin(x^2)\}$. Using a perfect critic, $f = \text{PMI}_{XY}$, we saw that all estimators performed well in this problem. When a constant bias was added, $f = \text{PMI}_{XY} + c$, the Monte Carlo and NWJ estimators became biased. Finally, InfoNCE was the only unbiased estimator when a functional degree of freedom was added, $f(x, y) = \text{PMI}_{XY}(x, y) + \sin(x^2)$, confirming the known theoretical limitations of the estimators.

Increasing the Dimensionality In low dimensions variational approximations perform comparably to the Monte Carlo estimator when the correct critic function is used, and are more robust to misspecification due to additional degrees of freedom. However, this additional robustness results in a bias in higher dimensions (Fig. 5, right panel). We simulated a data set using a multivariate normal distribution with 25 strongly interacting components, $\text{corr}(X_i, Y_i) = 0.8$, which results in $\mathbf{I}(X; Y) \approx 12.8$. Although the Monte Carlo estimator and NWJ estimator (which are not robust to additional degrees of freedom) provided estimates close to the ground-truth (with Monte Carlo slightly outperforming NWJ in both bias and variance), Donsker–Varadhan estimator resulted in a strong positive bias for small batch sizes and InfoNCE had a negative bias with very low variance which has been observed in previous studies and has been related to the $O(\log N)$ behavior of this approximate lower bound (Oord et al., 2018; Song & Ermon, 2020).

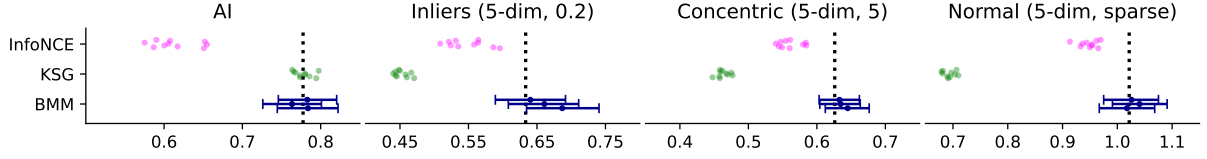


Figure 6: BMMs provide uncertainty estimates in the form of credible intervals and, when domain expertise is available, can be more accurate than generic black-box alternatives. We plot the posterior mean and an interval created using 10th and 90th percentile of the posterior distribution. Ground truth is represented with a dashed line.

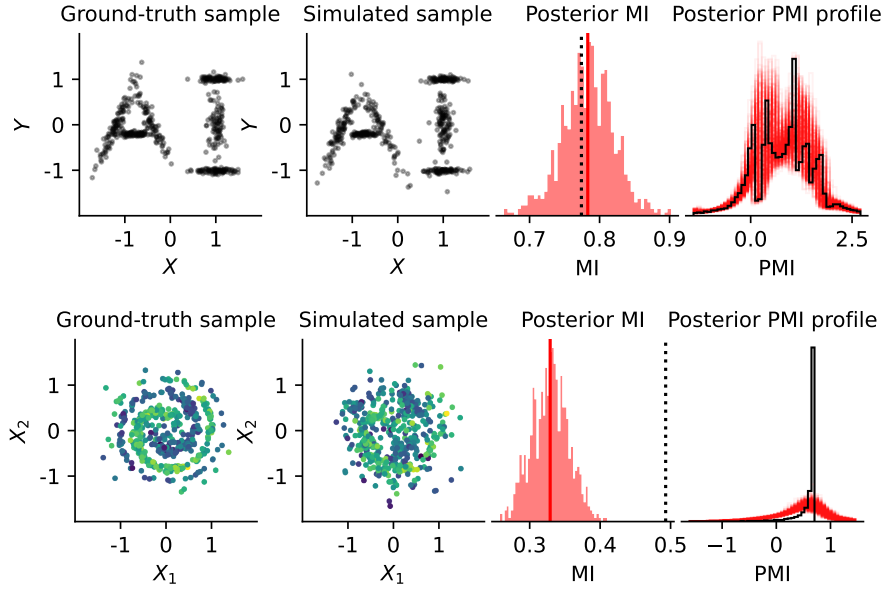


Figure 7: Top row: data generated according to the AI distribution, data generated according to a single MCMC sample, posterior distribution of the mutual information (red line denotes posterior mean and black line denotes the ground-truth value), and the posterior distribution of the PMI profile (black curve denotes the ground-truth value). For a well-specified model the posterior reflects epistemic uncertainty well. Bottom row: experiment with a misspecified model fitted to the Galaxy distribution. Misspecification results in biased inferences (two right-most panels) with miscalibrated posterior, although it can be diagnosed by comparing true data with the data generated from the model.

3.4 Model-based mutual information estimation

As the final application of BMMs, we investigate the problem of Bayesian model-based MI estimation.

The general idea Consider a family of distributions $\{P_\theta \mid \theta \in \Theta\}$, such that for every value of the parameter vector θ , the mutual information contained in P_θ is known. That is, for $(X_\theta, Y_\theta) \sim P_\theta$ we can compute $\mathbf{I}(X_\theta; Y_\theta)$, which we denote by $\mathbf{I}(P_\theta)$. If the observed data are modeled as i.i.d. r.v.s $(X_i, Y_i) \sim P_\theta$, one can estimate θ and use $\mathbf{I}(P_\theta)$ as an estimate of the mutual information. For example, Brillinger (2004) interprets the CCA-based MI estimator of Kay (1992) as the plug-in estimator $\mathbf{I}(P_{\hat{\theta}})$, where $\hat{\theta}$ is the maximum likelihood estimate (MLE) in the model in which P_θ is a multivariate normal distribution. For discrete data it is well-known the plug-in estimator is biased (Goebel et al., 2005; Suzuki, 2016; Bu et al., 2018; Marx & Vreeken, 2019) and one particular way of regularizing the estimate is to approximate the Bayesian posterior over the MI (Hutter, 2001). We note that one can construct a Bayesian alternative for the CCA-based estimator by using a prior on the covariance matrix (Lewandowski et al., 2009) and Markov chain Monte Carlo (Gelman et al., 2013, Ch. 11) to provide samples $\theta_1, \dots, \theta_M$ from the posterior $P(\theta \mid X_1, Y_1, \dots, X_N, Y_N)$ and construct a sample-based approximation to the posterior on mutual information, $\mathbf{I}(P_{\theta_1}), \dots, \mathbf{I}(P_{\theta_M})$.

More generally, consider a statistical model $\{P_\theta \mid \theta \in \Theta\}$ such that all P_θ are BMMs, and a prior $P(\theta)$. One can then apply Bayesian inference algorithms to construct a sample $\theta_1, \dots, \theta_M$ from the posterior. Although the exact values for $\mathbf{I}(P_{\theta_1}), \dots, \mathbf{I}(P_{\theta_M})$ are not available, they can be approximated as in Sec. 2. Hence, we can construct an approximate posterior distribution and quantify epistemic uncertainty of the estimate (Fig. 7). Since BMMs include both mixture models and normalizing flows, this approach could, in principle, be used as a general technique for building model-based mutual information estimators, where the generative model can be constructed using domain knowledge. Moreover, this is the *first Bayesian estimator of the PMI profile*: as all distributions P_{θ_m} are BMMs, one can construct M histograms (or CDFs) approximating the profile.

Proof of concept To illustrate this approach we implemented a sparse Gaussian mixture model (see Appendix C.5) in NumPyro (Phan et al., 2019) and used the NUTS sampler (Hoffman & Gelman, 2014) to obtain the Bayesian posterior for selected low-dimensional problems from the proposed benchmark (see Table 2 in Appendix C.5.2). We visualise the predictions for four distributions in Fig. 6: we stress that other methods can only provide a single point estimate (for each given data set), while BMMs can provide uncertainty quantification in the form of posterior distribution, visualised by credible intervals. Interestingly, even for low-dimensional problems where KSG has been considered state-of-the-art (Czyż et al., 2023), the BMM-based estimator provided better estimates.

On the role of assumptions All parametric statistical methods rely on assumptions (Gelman et al., 2020, Sec. 2). In particular, Bayesian inference can result in unreliable inferences when the model is misspecified, i.e., the true data-generating process P_{XY} does not belong to the assumed family P_θ (Watson & Holmes, 2016). To illustrate the role of the assumptions, we obtained a Bayesian posterior conditioned on 500 data points from the AI and Galaxy distributions (see Fig. 7). We see that for the AI distribution, the posterior is concentrated around the ground-truth mutual information value and the ground-truth PMI profile is well-approximated by the posterior samples.

The misspecification in the Galaxy distribution can be immediately diagnosed via posterior predictive checking (Gelman et al., 2013, Ch. 6): in Fig. 7 we see that a data sample simulated from the model looks substantially different from the observed data, meaning that the model did not capture the distribution well (cf. Appendix C.5). This provides a clear indication that the estimates should not be trusted: most of the probability mass of the Bayesian posterior is far from the ground-truth mutual information. Similarly, the posterior on the PMI profile is biased. We therefore recommend using model-checking techniques such as posterior predictive checks and discriminator-based validation (Sankaran & Holmes, 2023, Sec. 4) to understand the deficiencies of the model. To mitigate the risk of overfitting, which can bias the model (see Appendix C.5), we recommend cross-validation (Piironen & Vehtari, 2017).

In summary, although BMMs are not generic estimators applicable to problems without available domain knowledge, in some situations, principled Bayesian modeling can be beneficial to obtain better estimates along with uncertainty quantification.

4 Discussion

In this article, we have studied pointwise mutual information profiles (PMI profiles), determining them analytically for multivariate normal distributions (Theorem 4), and proposed the family of Bend and Mix Models (BMMs), which include multivariate normal and Student distributions, mixture models and normalizing flows, for which the PMI profile can be approximated using Monte Carlo methods. We showed how BMMs can be used to provide novel benchmark tasks to test MI estimators and calculate MI transmitted through a communication channel in the presence of inliers and outliers (which can be used in the experimental design in electrical and biological sciences). BMMs allowed us to study how well neural critics approximate PMI profiles. Finally, we showed how Bayesian estimates of MI between continuous r.v.s can be performed using BMMs; additionally, the proposed method estimates the PMI profiles. Although this approach is not universal, we find it suitable for problems with precise domain knowledge available (which can be used to construct the generative distribution P_θ and provide the prior $P(\theta)$) and in which uncertainty quantification is desired.

Limitations and further research Generalized PMI profiles (see Appendix A.5) allow one to provide a wide family of invariants, which can be related to f -divergences (Nowozin et al., 2016). We leave the exploration of potential ramifications of this generalization to future work.

Although in Sec. 3.4 we propose BMMs as a basis for model-based Bayesian inference of mutual information, model misspecification (Watson & Holmes, 2016) as well as underfitting and overfitting can bias inference. We therefore recommend, as usual in Bayesian statistics, to validate the model (Gelman et al., 2020). Overfitting and model misspecification may be more difficult to detect especially in high-dimensional situations or when more expressive models are constructed, for example using normalizing flows, as Bayesian inference for models involving neural networks is known to be challenging (Izmailov et al., 2021). We hypothesise that tempering the likelihood (Grünwald & van Ommen, 2017) or employing alternative inference schemes (Lyddon et al., 2018) may improve robustness to model misspecification, although we leave this topic to future work.

In Sec. 3.2, we study a channel in which the probability of failure α is constant and independent of the input variable X . However, in both biological and electric systems this assumption does not necessarily hold: for large values of X the system may result in an outlier more easily. Unfortunately, cases in which α depends on X are not easily modeled by BMMs. Moreover, BMMs do not allow modeling additive noise. Although adding Gaussian noise can be viewed as a continuous mixture, marginalizing the latent variable to obtain an analytic form of $\log p_{Y'}(y)$ is not possible. Employing unbiased estimators, such as SUMO (Luo et al., 2020), may hence become an interesting direction for future research.

More generally, replacing analytical formulae with accurate numerical estimates has the potential to extend the family of distributions with tractable MI beyond BMMs. In Sec. 2.2 we construct BMMs in a bottom-up fashion, where simple distributions are transformed and mixed together to obtain more expressive distributions. The introduced operations allow us to evaluate the probability densities and directly sample from the joint distribution, resulting in a Monte Carlo estimator for mutual information.

Alternatively, one could imagine a top-down approach, where for a given distribution P_{XY} one derives a family of unbiased log-density estimators (Luo et al., 2020). Namely, let $L_{XY}(x, y)$ be a family of estimators indexed by $(x, y) \in \mathcal{X} \times \mathcal{Y}$, such that $\mathbb{E}[L_{XY}(x, y)] = \log p_{XY}(x, y)$ for all points (x, y) in the support of P_{XY} . Similarly, let $L_X(x)$ and $L_Y(y)$ be families of estimators such that $\mathbb{E}[L_X(x)] = \log p_X(x)$ and $\mathbb{E}[L_Y(y)] = \log p_Y(y)$. We can then construct an unbiased estimator of the PMI function, $\mathbb{E}[L_{XY}(x, y) - L_X(x) - L_Y(y)] = \text{PMI}_{XY}(x, y)$, and aim to approximate the PMI profile and the ground-truth MI via nested Monte Carlo sampling (Rainforth et al., 2018). However, the feasibility of this approach relies on construction of the required families L_{XY} , L_X and L_Y , as well as on ensuring that the variance of the resulting estimate is sufficiently small.

Similarly, Monte Carlo integration of the PMI function with respect to the P_{XY} measure can, in principle, be replaced by other integration methods, such as Markov chain Monte Carlo (Brooks et al., 2011) or sequential Monte Carlo samplers (Del Moral et al., 2006). These integration methods could allow one to integrate the PMI_{XY} function for distributions in which the log-density functions $\log p_{XY}$, $\log p_X$ and $\log p_Y$ (or their unbiased estimates) are available, but Monte Carlo sampling from P_{XY} is not possible. However, this approach relies on ensuring whether the employed integration method is accurate enough, e.g., by controlling the convergence diagnostics and the effective sample size (Vehtari et al., 2021; Vats & Knudson, 2021).

We hope that our analysis of pointwise mutual information profiles provides a fresh perspective on mutual information estimators, and that BMMs will prove to be a useful tool for benchmark design and will renew interest in Bayesian approaches to mutual information estimation.

Reproducibility To facilitate reproducibility, the experiments have been implemented as Snakemake workflows (Mölder et al., 2021). The code available at <https://github.com/cbg-ethz/bmi> can be used to reproduce all experimental results and figures.

Acknowledgments FG was supported by the OPUS 18 grant operated by Narodowe Centrum Nauki (National Science Centre, Poland) 2019/35/B/NZ2/03898. PC was supported by a doctoral fellowship from the ETH AI Center. AM was supported by a postdoctoral fellowship from the ETH AI Center and a postdoctoral position at the Computational Biology Group. We would like to thank Paweł Nałęcz-Jawecki, Julia Kostin and anonymous reviewers for helpful suggestions on the manuscript.

References

- Carl Allen and Timothy Hospedales. Analogies explained: Towards understanding word embeddings. In Kamalika Chaudhuri and Ruslan Salakhutdinov (eds.), *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 223–231. PMLR, 09–15 Jun 2019. URL <https://proceedings.mlr.press/v97/allen19a.html>. 3
- Mohamed Ishmael Belghazi, Aristide Baratin, Sai Rajeshwar, Sherjil Ozair, Yoshua Bengio, Aaron Courville, and Devon Hjelm. Mutual information neural estimation. In *International conference on machine learning*, pp. 531–540. PMLR, 2018. 2, 3, 6, 7, 34
- James Bradbury, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, George Neca, Adam Paszke, Jake VanderPlas, Skye Wanderman-Milne, and Qiao Zhang. JAX: composable transformations of Python+NumPy programs, 2018. URL <http://github.com/google/jax>. 5
- David R. Brillinger. Some data analyses using mutual information. *Brazilian Journal of Probability and Statistics*, 18(2):163–182, 2004. ISSN 01030752, 23176199. URL <http://www.jstor.org/stable/43601047>. 2, 6, 9, 28, 34
- S. Brooks, A. Gelman, G. Jones, and X.L. Meng. *Handbook of Markov Chain Monte Carlo*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. CRC Press, 2011. ISBN 9781420079425. URL <https://books.google.ch/books?id=qfRsAIKZ4rIC>. 11
- Yuheng Bu, Shaofeng Zou, Yingbin Liang, and Venugopal V Veeravalli. Estimation of KL divergence: Optimal minimax rate. *IEEE Transactions on Information Theory*, 64(4):2648–2674, 2018. 9
- Nick Carrara and Jesse Ernst. Using Monte Carlo tree search to calculate mutual information in high dimensions, 2023. 2, 28
- Christopher J Cellucci, Alfonso M Albano, and Paul E Rapp. Statistical validation of mutual information calculations: Comparison of alternative numerical algorithms. *Physical review E*, 71(6):066208, 2005. 6, 34
- Paweł Czyż, Frederic Grabowski, Julia Vogt, Niko Beerenwinkel, and Alexander Marx. Beyond normal: On the evaluation of mutual information estimators. In A. Oh, T. Neumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 16957–16990. Curran Associates, Inc., 2023. 2, 3, 6, 7, 10, 28, 31, 32, 33, 34
- Georges A Darbellay and Igor Vajda. Estimation of the information by an adaptive partitioning of the observation space. *IEEE Transactions on Information Theory*, 45(4):1315–1321, 1999. 6, 34
- Pierre Del Moral, Arnaud Doucet, and Ajay Jasra. Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):411–436, 2006. doi: <https://doi.org/10.1111/j.1467-9868.2006.00553.x>. URL <https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2006.00553.x>. 11
- Joshua V. Dillon, Ian Langmore, Dustin Tran, Eugene Brevdo, Srinivas Vasudevan, Dave Moore, Brian Patton, Alex Alemi, Matthew D. Hoffman, and Rif A. Saurous. Tensorflow distributions. *CoRR*, abs/1711.10604, 2017. URL <http://arxiv.org/abs/1711.10604>. 5
- James M. Flegal, Murali Haran, and Galin L. Jones. Markov Chain Monte Carlo: Can we trust the third significant figure? *Statistical Science*, 23(2):250 – 260, 2008. doi: 10.1214/08-STS257. URL <https://doi.org/10.1214/08-STS257>. 5
- Sylvia Frühwirth-Schnatter and Gertraud Malsiner-Walli. From here to infinity: sparse finite versus Dirichlet process mixtures in model-based clustering. *Advances in Data Analysis and Classification*, 13(1):33–64, Mar 2019. ISSN 1862-5355. doi: 10.1007/s11634-018-0329-y. URL <https://doi.org/10.1007/s11634-018-0329-y>. 35

- Kurt G. Radon-Nikodym derivative and push forward of measures. Mathematics Stack Exchange, 2022. URL <https://math.stackexchange.com/q/4413000>. URL:https://math.stackexchange.com/q/4413000 (version: 2024-05-14). 27
- Weihao Gao, Sreeram Kannan, Sewoong Oh, and Pramod Viswanath. Estimating mutual information for discrete-continuous mixtures. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. 28, 29
- A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, and D.B. Rubin. *Bayesian Data Analysis, Third Edition*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis, 2013. ISBN 9781439840955. URL <https://books.google.pl/books?id=ZXL6AQAAQBAJ>. 9, 10
- Andrew Gelman, Aki Vehtari, Daniel Simpson, Charles C. Margossian, Bob Carpenter, Yuling Yao, Lauren Kennedy, Jonah Gabry, Paul-Christian Bürkner, and Martin Modrák. Bayesian workflow, 2020. 10, 11
- B. Goebel, Z. Dawy, J. Hagenauer, and J.C. Mueller. An approximation to the distribution of finite sample size mutual information estimates. In *IEEE International Conference on Communications, 2005. ICC 2005*. 2005, volume 2, pp. 1102–1106 Vol. 2, 2005. 9
- Frederic Grabowski, Paweł Czyż, Marek Kochańczyk, and Tomasz Lipniacki. Limits to the rate of information transmission through the MAPK pathway. *Journal of The Royal Society Interface*, 16(152):20180792, 2019. doi: 10.1098/rsif.2018.0792. URL <https://royalsocietypublishing.org/doi/abs/10.1098/rsif.2018.0792>. 2, 28
- Peter Grünwald and Thijs van Ommen. Inconsistency of Bayesian Inference for Misspecified Linear Models, and a Proposal for Repairing It. *Bayesian Analysis*, 12(4):1069 – 1103, 2017. doi: 10.1214/17-BA1085. URL <https://doi.org/10.1214/17-BA1085>. 11
- David Haussler and Manfred Oppen. Mutual information, metric entropy and cumulative relative entropy risk. *The Annals of Statistics*, 25(6):2451 – 2492, 1997. doi: 10.1214/aos/1030741081. URL <https://doi.org/10.1214/aos/1030741081>. 5, 25
- Matthew D. Hoffman and Andrew Gelman. The No-U-Turn Sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(47):1593–1623, 2014. URL <http://jmlr.org/papers/v15/hoffman14a.html>. 10, 35
- Harold Hotelling. Relations between two sets of variates. *Biometrika*, 28(3-4):321–377, 12 1936. ISSN 0006-3444. doi: 10.1093/biomet/28.3-4.321. URL <https://doi.org/10.1093/biomet/28.3-4.321>. 3
- Marcus Hutter. Distribution of mutual information. In T. Dietterich, S. Becker, and Z. Ghahramani (eds.), *Advances in Neural Information Processing Systems*, volume 14. MIT Press, 2001. 9, 28
- J. P. Imhof. Computing the distribution of quadratic forms in normal variables. *Biometrika*, 48(3/4):419–426, 1961. ISSN 00063444. URL <http://www.jstor.org/stable/2332763>. 21
- Pavel Izmailov, Sharad Vikram, Matthew D Hoffman, and Andrew Gordon Wilson. What are Bayesian neural network posteriors really like? In Marina Meila and Tong Zhang (eds.), *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 4629–4640. PMLR, 18–24 Jul 2021. URL <https://proceedings.mlr.press/v139/izmailov21a.html>. 11
- Takoua Jendoubi and Korbinian Strimmer. A whitening approach to probabilistic canonical correlation analysis for omics data integration. *BMC Bioinformatics*, 20(1):15, Jan 2019. ISSN 1471-2105. doi: 10.1186/s12859-018-2572-9. URL <https://doi.org/10.1186/s12859-018-2572-9>. 20
- J. Kay. Feature discovery under contextual supervision using mutual information. In *IJCNN International Joint Conference on Neural Networks*, volume 4, pp. 79–84, 1992. doi: 10.1109/IJCNN.1992.227286. 2, 6, 9, 33, 34

- Shiraj Khan, Sharba Bandyopadhyay, Auroop R Ganguly, Sunil Saigal, David J Erickson III, Vladimir Protopopescu, and George Ostrouchov. Relative performance of mutual information estimation methods for quantifying the dependence among short and noisy data. *Physical Review E*, 76(2):026209, 2007. 2
- I. Kobyzev, S. D. Prince, and M. A. Brubaker. Normalizing flows: An introduction and review of current methods. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 43(11):3964–3979, nov 2021. ISSN 1939-3539. doi: 10.1109/TPAMI.2020.2992934. 4
- E. Koehler, E. Brown, and S.J.-P.A. Haneuse. On the assessment of Monte Carlo error in simulation-based statistical analyses. *The American Statistician*, 63(2):155–162, 2009. doi: 10.1198/tast.2009.0030. 5
- Artemy Kolchinsky and Brendan D. Tracey. Estimating mixture entropy with pairwise distances. *Entropy*, 19(7), 2017. ISSN 1099-4300. doi: 10.3390/e19070361. URL <https://www.mdpi.com/1099-4300/19/7/361>. 5, 25
- Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. Estimating mutual information. *Physical Review E*, 69(6):066138, 2004. 2, 3, 6, 18, 33, 34
- J.M. Lee. *Introduction to Smooth Manifolds*. Graduate Texts in Mathematics. Springer, 2nd edition, 2012. ISBN 9781441999825. URL <https://doi.org/10.1007/978-1-4419-9982-5>. 18, 29
- Daniel Lewandowski, Dorota Kurowicka, and Harry Joe. Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, 100(9):1989–2001, 2009. ISSN 0047-259X. doi: <https://doi.org/10.1016/j.jmva.2009.04.008>. URL <https://www.sciencedirect.com/science/article/pii/S0047259X09000876>. 9, 35
- Yucen Luo, Alex Beatson, Mohammad Norouzi, Jun Zhu, David Duvenaud, Ryan P. Adams, and Ricky T. Q. Chen. SUMO: unbiased estimation of log marginal probability for latent variable models. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net, 2020. URL <https://openreview.net/forum?id=SyIkYeHtwr>. 11
- Simon Lyddon, Stephen Walker, and Chris C Holmes. Nonparametric learning from Bayesian models with randomized objective functions. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. 11
- Alexander Marx and Jilles Vreeken. Testing conditional independence on discrete data using stochastic complexity. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pp. 496–505. PMLR, 2019. 9
- Alexander Marx, Lincen Yang, and Matthijs van Leeuwen. Estimating conditional mutual information for discrete-continuous mixtures using multi-dimensional adaptive histograms. In *Proceedings of the SIAM International Conference on Data Mining (SDM)*, pp. 387–395, 2021. 29
- David McAllester and Karl Stratos. Formal limitations on the measurement of mutual information. In Silvia Chiappa and Roberto Calandra (eds.), *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, volume 108 of *Proceedings of Machine Learning Research*, pp. 875–884. PMLR, 26–28 Aug 2020. URL <https://proceedings.mlr.press/v108/mcallester20a.html>. 2, 34
- Kevin P. Murphy. *Probabilistic Machine Learning: Advanced Topics*. MIT Press, 2023. URL <http://probml.github.io/book2>. 30
- F Mölder, KP Jablonski, B Letcher, MB Hall, CH Tomkins-Tinch, V Sochat, J Forster, S Lee, SO Twardziok, A Kanitz, A Wilm, M Holtgrewe, S Rahmann, S Nahnsen, and J Köster. Sustainable data analysis with Snakemake. *F1000Research*, 10(33), 2021. doi: 10.12688/f1000research.29032.1. 11
- Paweł Nałęcz-Jawecki, Paolo Armando Gagliardi, Marek Kochańczyk, Coralie Dessauges, Olivier Pertz, and Tomasz Lipniacki. The MAPK/ERK channel capacity exceeds 6 bit/hour. *PLOS Computational Biology*, 19(5):1–21, 05 2023. doi: 10.1371/journal.pcbi.1011155. URL <https://doi.org/10.1371/journal.pcbi.1011155>. 6

- XuanLong Nguyen, Martin J Wainwright, and Michael Jordan. Estimating divergence functionals and the likelihood ratio by penalized convex risk minimization. In J. Platt, D. Koller, Y. Singer, and S. Roweis (eds.), *Advances in Neural Information Processing Systems*, volume 20. Curran Associates, Inc., 2007. 2, 3, 7, 34
- Sebastian Nowozin, Botond Cseke, and Ryota Tomioka. f-GAN: Training generative neural samplers using variational divergence minimization. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016. URL https://proceedings.neurips.cc/paper_files/paper/2016/file/cedebb6e872f539bef8c3f919874e9d7-Paper.pdf. 11, 28
- Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018. 2, 3, 6, 7, 8, 34
- George Papamakarios, Eric Nalisnick, Danilo Jimenez Rezende, Shakir Mohamed, and Balaji Lakshminarayanan. Normalizing flows for probabilistic modeling and inference. *Journal of Machine Learning Research*, 22(57):1–64, 2021. URL <http://jmlr.org/papers/v22/19-1028.html>. 4
- K. B. Petersen and M. S. Pedersen. The matrix cookbook, November 2012. URL <http://www2.compute.dtu.dk/pubdb/pubs/3274-full.html>. Version 20121115. 20
- Du Phan, Neeraj Pradhan, and Martin Jankowiak. Composable effects for flexible and accelerated probabilistic programming in NumPyro. *arXiv preprint arXiv:1912.11554*, 2019. 10, 35
- Juho Piironen and Aki Vehtari. Comparison of Bayesian predictive methods for model selection. *Statistics and Computing*, 27(3):711–735, May 2017. ISSN 1573-1375. doi: 10.1007/s11222-016-9649-y. URL <https://doi.org/10.1007/s11222-016-9649-y>. 10
- M.S. Pinsker and A. Feinstein. *Information and Information Stability of Random Variables and Processes*. Holden-Day series in time series analysis. Holden-Day, 1964. ISBN 9780816268047. 3
- Dimitris N Politis. *On the entropy of a mixture distribution*. Purdue University. Department of Statistics, 1991. 29
- Y. Polyanskiy and Y. Wu. *Information Theory: From Coding to Learning*. Cambridge University Press, 2022. Book draft. 2, 26, 28, 34
- Ben Poole, Sherjil Ozair, Aaron Van Den Oord, Alex Alemi, and George Tucker. On variational bounds of mutual information. In *International Conference on Machine Learning*, pp. 5171–5180. PMLR, 2019. 2, 7, 34
- Tiberiu Popoviciu. Sur les équations algébriques ayant toutes leurs racines réelles. *Mathematica (Cluj)*, 9: 129–145, 1935. 5
- Tom Rainforth, Rob Cornish, Hongseok Yang, Andrew Warrington, and Frank Wood. On nesting monte carlo estimators. In *International Conference on Machine Learning*, pp. 4267–4276. PMLR, 2018. 11
- Kris Sankaran and Susan P. Holmes. Generative models: An interdisciplinary perspective. *Annual Review of Statistics and Its Application*, 10(1):325–352, 2023. doi: 10.1146/annurev-statistics-033121-110134. URL <https://doi.org/10.1146/annurev-statistics-033121-110134>. 10
- Jiaming Song and Stefano Ermon. Understanding the limitations of variational mutual information estimators. In *International Conference on Learning Representations*, 2020. 7, 8, 34
- Joe Suzuki. An estimator of mutual information and its application to independence testing. *Entropy*, 18(4): 109, 2016. 9
- Dootika Vats and Christina Knudson. Revisiting the Gelman–Rubin diagnostic. *Statistical Science*, 36(4):518 – 529, 2021. doi: 10.1214/20-STS812. URL <https://doi.org/10.1214/20-STS812>. 11

Aki Vehtari, Andrew Gelman, Daniel Simpson, Bob Carpenter, and Paul-Christian Bürkner. Rank-Normalization, Folding, and Localization: An Improved $\hat{R}$ for Assessing Convergence of MCMC (with Discussion). *Bayesian Analysis*, 16(2):667 – 718, 2021. doi: 10.1214/20-BA1221. URL <https://doi.org/10.1214/20-BA1221>. 11, 36

James Watson and Chris Holmes. Approximate Models and Robust Decisions. *Statistical Science*, 31(4):465 – 489, 2016. doi: 10.1214/16-STS592. URL <https://doi.org/10.1214/16-STS592>. 10, 11

Appendix

Table of Contents

A Technical results	18
A.1 Proof of the invariance of the pointwise mutual information profile	18
A.2 Derivations of pointwise mutual information profiles	19
A.3 Constructing new Bend and Mix Models	23
A.4 Mutual information in finite mixtures	24
A.5 Divergence profiles	26
B Distributions involving discrete variables	28
B.1 Known distributions	28
B.2 Novel distributions	30
C Experimental protocols and additional experiments	30
C.1 Specification of the proposed distributions	30
C.2 A benchmark extension proposal	32
C.3 Estimator hyperparameters	34
C.4 Variational estimators of mutual information	34
C.5 Bayesian estimation of Gaussian mixture models	35

A Technical results

In this section we formalize the results described in Sec. 2. In Sec. A.1 we precisely define the family of $\mathcal{P}(\mathcal{X}, \mathcal{Y})$ distributions, which is then used to prove that the PMI profile is invariant to diffeomorphisms. In Sec. A.2 we derive the PMI profiles in the cases where they are analytically tractable. Then, in Sec. A.3 we formalize the properties of Bend and Mix Models. In Sec. A.4 we derive the bounds on mutual information in finite mixture models. Finally, in Sec. A.5 we provide a measure-theoretic analysis of the PMI profile, which offers more general invariance results, though it requires more complex mathematical machinery.

A.1 Proof of the invariance of the pointwise mutual information profile

In this section, we consider random variables valued in smooth manifolds without boundary (Lee, 2012, Ch. 1) equipped with their Borel σ -algebras and given reference measures. In particular, the results in Section 2 apply to both discrete and continuous r.v.s, as the considered manifolds include open subsets of the Euclidean space $\mathbb{R}^n$ with the Lebesgue measure as well as zero-dimensional discrete spaces $\{1, \dots, m\}$ with the counting measure.

For a given pair of smooth manifolds $\mathcal{X}$ and $\mathcal{Y}$, we equip their product $\mathcal{X} \times \mathcal{Y}$ with the product measure and define the set $\mathcal{P}(\mathcal{X}, \mathcal{Y})$ to consist of all probability measures P_{XY} on $\mathcal{X} \times \mathcal{Y}$ such that the joint measure P_{XY} , as well as the marginal measures $P_X(A) = P_{XY}(A \times \mathcal{Y})$ and $P_Y(B) = P_{XY}(\mathcal{X} \times B)$, have smooth and positive PDFs (or PMFs) p_{XY} , p_X and p_Y with respect to the reference measures on $\mathcal{X} \times \mathcal{Y}$, $\mathcal{X}$ and $\mathcal{Y}$, respectively. In particular, the PMI function (Definition 1) exists (without the need to define expressions of the form $\log 0/0$) and is a smooth function from the manifold $\mathcal{X} \times \mathcal{Y}$ to the set of real numbers, $\mathbb{R}$. Note that as smooth functions are measurable, the PMI profile, introduced in Definition 2, indeed exists.

More generally, it is possible to work with standard Borel spaces (instead of smooth manifolds) and prove invariance of the profile with respect to arbitrary continuous injective mappings (instead of diffeomorphisms). We discuss this approach in Appendix A.5.

Recall a well-known result (Kraskov et al., 2004, Appendix):

Lemma 12 (Invariance of PMI). *Let $X' = f(X)$ and $Y' = g(Y)$, where f and g are diffeomorphisms. Then for every x' and y' we have*

$$\text{PMI}_{X'Y'}(x', y') = \text{PMI}_{XY}(x, y),$$

where $x = f^{-1}(x')$ and $y = g^{-1}(y')$.

Proof. From

$$p_{X'Y'}(x', y') = p_{XY}(x, y) |\det D(f^{-1} \times g^{-1})(x', y')|$$

and analogous quantities we conclude that $p_{X'Y'}$, as well as $p_{X'}$ and $p_{Y'}$ are smooth and everywhere positive functions, so that $\text{PMI}_{X'Y'}$ is well-defined. As $D(f^{-1} \times g^{-1})(x', y')$ is a block matrix with $Df^{-1}(x')$ and $Dg^{-1}(y')$ blocks on the diagonal and other blocks zero, we have $\det D(f^{-1} \times g^{-1})(x', y') = \det Df^{-1}(x') \cdot \det Dg^{-1}(y')$. $\square$

Now we can prove invariance of the PMI profile:

Theorem 3. *Let $P_{XY} \in \mathcal{P}(\mathcal{X}, \mathcal{Y})$ and $f: \mathcal{X} \rightarrow \mathcal{X}$ and $g: \mathcal{Y} \rightarrow \mathcal{Y}$ be diffeomorphisms. Then for $X' = f(X)$ and $Y' = g(Y)$ it holds that $P_{X'Y'} \in \mathcal{P}(\mathcal{X}, \mathcal{Y})$ and $\text{Prof}_{XY} = \text{Prof}_{X'Y'}$.*

Proof. From the proof of Lemma 12 we conclude that $P_{X'Y'} \in \mathcal{P}(\mathcal{X}, \mathcal{Y})$. Then, we note the profile is the pushforward measure

$$\text{Prof}_{XY} := (\text{PMI}_{XY})_{\#} P_{XY}.$$

Now let $B \subseteq \mathbb{R}$ be any set in the Borel σ -algebra and $\mathbf{1}_B$ be its characteristic function. Using the change of variables formula for pushforward measure and invariance of PMI:

$$\begin{aligned}
\text{Prof}_{X'Y'}(B) &= \int \mathbf{1}_B(t) d((\text{PMI}_{X'Y'})_{\#} P_{X'Y'})(t) \\
&= \int \mathbf{1}_B(\text{PMI}_{X'Y'}(x', y')) dP_{X'Y'}(x', y') \\
&= \int \mathbf{1}_B(\text{PMI}_{X'Y'}(f(x), g(y))) dP_{XY}(x, y) \\
&= \int \mathbf{1}_B(\text{PMI}_{XY}(x, y)) dP_{XY}(x, y) \\
&= \text{Prof}_{XY}(B)
\end{aligned}$$

□

Remark 13. The above proof can be understood also in the following manner. Lemma 12 allows us to express the PMI function of the transformed variables as the composition $\text{PMI}_{X'Y'} = \text{PMI}_{XY} \circ (f^{-1} \times g^{-1})$. As $\text{Prof}_{X'Y'} = (\text{PMI}_{X'Y'})_{\#} P_{X'Y'}$ and $P_{X'Y'} = (f \times g)_{\#} P_{XY}$, the $f \times g$ cancels out with its inverse:

$$\begin{aligned}
\text{Prof}_{X'Y'} &= (\text{PMI}_{XY} \circ (f^{-1} \times g^{-1}))_{\#} (f \times g)_{\#} P_{XY} \\
&= (\text{PMI}_{XY} \circ (f^{-1} \times g^{-1}) \circ (f \times g))_{\#} P_{XY} \\
&= (\text{PMI}_{XY} \circ \text{identity})_{\#} P_{XY} \\
&= (\text{PMI}_{XY})_{\#} P_{XY} = \text{Prof}_{XY}.
\end{aligned}$$

This perspective allows one to generalize PMI profile invariance to more general mappings, such as continuous injective functions. We prove a general result in Appendix A.5.

A.2 Derivations of pointwise mutual information profiles

The following result shows that the distributions from the class $\mathcal{P}(\mathcal{X}, \mathcal{Y})$ with zero mutual information have the same profile:

Proposition 5. Let X and Y be r.v.s with joint distribution $P_{XY} \in \mathcal{P}(\mathcal{X}, \mathcal{Y})$. Then, $\mathbf{I}(X; Y) = 0$ if and only if $\text{Prof}_{XY} = \delta_0$ is the Dirac measure with a single atom at 0.

Proof. If $\text{Prof}_{XY} = \delta_0$, then the expected value is $\mathbf{I}(X; Y) = 0$. To prove the converse, if $\mathbf{I}(X; Y) = 0$, then X and Y are independent. As the probability density functions of distributions in the $\mathcal{P}(\mathcal{X}, \mathcal{Y})$ class are smooth, we have $p_{XY}(x, y) = p_X(x)p_Y(y)$ at every point $(x, y) \in \mathcal{X} \times \mathcal{Y}$ and $\text{PMI}_{XY}(x, y) = 0$ everywhere. □

The following result characterizes the PMI profiles for discrete r.v.s:

Proposition 6. If X and Y are discrete r.v.s with $P_{XY} \in \mathcal{P}(\mathcal{X}, \mathcal{Y})$, then the PMI profile is discrete:

$$\text{Prof}_{XY} = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{XY}(x, y) \delta_{\text{PMI}_{XY}(x, y)}.$$

Proof. The measure P_{XY} is discrete and given by

$$P_{XY} = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{XY}(x, y) \delta_{(x, y)},$$

so its pushforward by the PMI_{XY} function has the form

$$\text{Prof}_{XY} = (\text{PMI}_{XY})_{\#} P_{XY} = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{XY}(x, y) \delta_{\text{PMI}_{XY}(x, y)}.$$

□

The next results describe the properties of PMI profiles associated with multivariate normal variables:

Theorem 4. *Let X and Y be r.v.s such that the joint distribution $P_{XY} \in \mathcal{P}(\mathbb{R}^m, \mathbb{R}^n)$ is multivariate normal. If $k = \min(m, n)$ and $\rho_1, \rho_2, \dots, \rho_k$ are canonical correlations between X and Y , then the profile Prof_{XY} is a generalized χ^2 distribution, namely the distribution of the variable*

$$T = \mathbf{I}(X; Y) + \sum_{i=1}^k \frac{\rho_i}{2} (Q_i - Q'_i),$$

where $\{Q_i, Q'_i\}_{i=1, \dots, k}$ are independent and identically distributed (i.i.d.) variables drawn from the χ_1^2 distribution. In particular, Prof_{XY} is symmetric around its median, which coincides with the mean

$$\mathbf{I}(X; Y) = -\frac{1}{2} \sum_{i=1}^k \log(1 - \rho_i^2),$$

and all moments of this distribution exist. Its variance is equal to $\text{Var}[T] = \rho_1^2 + \dots + \rho_k^2$.

Proof. Without loss of generality assume that $m \leq n$. As the PMI profile is invariant to diffeomorphisms (Theorem 3), we can also assume that variables X and Y have been whitened by applying canonical correlation analysis (Jendoubi & Strimmer, 2019), that is $\mathbb{E}[X] = 0$, $\mathbb{E}[Y] = 0$ and the covariance matrix is given by

$$\Sigma = \begin{pmatrix} I_m & \Sigma_{XY} \\ \Sigma_{XY}^T & I_n \end{pmatrix} = \begin{pmatrix} I_m & R & 0 \\ R & I_m & 0 \\ 0 & 0 & I_{n-m} \end{pmatrix}$$

where

$$\Sigma_{XY} = \begin{pmatrix} R & 0_{m \times (n-m)} \end{pmatrix}$$

is an $m \times n$ matrix with the last $n - m$ columns being zero vectors and $R = \text{diag}(\rho_1, \dots, \rho_m)$ being the $m \times m$ diagonal matrix representing canonical correlations.

We can write the inverse in the block form

$$\Sigma^{-1} = \begin{pmatrix} \Lambda_X & \Lambda_{XY} \\ \Lambda_{XY}^T & \Lambda_Y \end{pmatrix} = \begin{pmatrix} \Lambda_X & \tilde{R} & 0 \\ \tilde{R} & \Lambda_X & 0 \\ 0 & 0 & I_{n-m} \end{pmatrix},$$

where the blocks have been calculated using the formula from Petersen & Pedersen (2012, Sec. 9.1):

$$\begin{aligned} \Lambda_X &= (I_m - \Sigma_{XY} \Sigma_{XY}^T)^{-1} = \text{diag}(u_1, \dots, u_m) \\ \Lambda_Y &= (I_n - \Sigma_{XY}^T \Sigma_{XY})^{-1} = \text{diag}(u_1, \dots, u_m, 1, \dots, 1) \\ \Lambda_{XY} &= -\Sigma_{XY} \Lambda_Y = \begin{pmatrix} \tilde{R} & 0_{m \times (n-m)} \end{pmatrix}, \end{aligned}$$

where $\tilde{R} = -\text{diag}(u_1 \rho_1, \dots, u_m \rho_m)$ and $u_i = 1 / (1 - \rho_i^2)$.

We define a quadratic form

$$\begin{aligned} s(x, y) &= x^T \Lambda_X x + y^T \Lambda_Y y + 2x^T \Lambda_{XY} y \\ &= \sum_{i=1}^m u_i (x_i^2 + y_i^2 - 2\rho_i x_i y_i) + \sum_{j=m+1}^n y_j^2 \end{aligned}$$

which can be used to calculate log-PDFs:

$$\begin{aligned} \log p_{XY}(x, y) &= -\frac{1}{2} s(x, y) - \frac{1}{2} \log \det \Sigma - \frac{m+n}{2} \log 2\pi, \\ \log p_X(x) &= -\frac{1}{2} x^T \Lambda_X x - \frac{m}{2} \log 2\pi, \\ \log p_Y(y) &= -\frac{1}{2} y^T \Lambda_Y y - \frac{n}{2} \log 2\pi. \end{aligned}$$

Hence,

$$\text{PMI}_{XY}(x, y) = \frac{x^T x + y^T y - s(x, y)}{2} - \frac{1}{2} \log \det \Sigma.$$

We recognize the last summand as

$$\mathbf{I}(X; Y) = \frac{1}{2} \log \left(\frac{\det I_m \cdot \det I_n}{\det \Sigma} \right) = -\frac{1}{2} \log \det \Sigma.$$

Define quadratic form

$$\begin{aligned} q(x, y) &= 2(\text{PMI}_{XY}(x, y) - \mathbf{I}(X; Y)) = x^T x + y^T y - s(x, y) \\ &= \sum_{i=1}^m \left((1 - u_i) (x_i^2 + y_i^2) + 2\rho_i u_i x_i y_i \right), \end{aligned}$$

which has a corresponding matrix

$$Q = \begin{pmatrix} K & F & 0 \\ F & K & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

where

$$K = \text{diag}(1 - u_1, \dots, 1 - u_m)$$

and

$$F = \text{diag}(\rho_1 u_1, \dots, \rho_m u_m).$$

We are interested in the distribution of

$$q(X, Y) = \begin{pmatrix} X^T & Y^T \end{pmatrix} Q \begin{pmatrix} X \\ Y \end{pmatrix},$$

where $(X, Y) \sim \mathcal{N}(0, \Sigma)$.

[Imhof \(1961\)](#) presents a general approach to evaluating the distributions of such quadratic forms. Consider a r.v.

$$Z = \begin{pmatrix} \eta \\ \epsilon \\ \xi \end{pmatrix} \sim \mathcal{N}(0, I_{m+n})$$

which is split into blocks of sizes m , m and $n - m$. If we construct a linear transformation A such that

$$\begin{pmatrix} X \\ Y \end{pmatrix} = A \begin{pmatrix} \eta \\ \epsilon \\ \xi \end{pmatrix},$$

then the distribution of $q(X, Y)$ is the distribution of

$$Z^T (A^T Q A) Z, \quad Z \sim \mathcal{N}(0, I_{m+n}).$$

We can construct A as

$$A = \begin{pmatrix} P_- & P_+ & 0 \\ -P_- & P_+ & 0 \\ 0 & 0 & I_{n-m} \end{pmatrix}$$

where

$$P_- = \text{diag} \left(\sqrt{\frac{1 - \rho_1}{2}}, \dots, \sqrt{\frac{1 - \rho_m}{2}} \right), \quad P_+ = \text{diag} \left(\sqrt{\frac{1 + \rho_1}{2}}, \dots, \sqrt{\frac{1 + \rho_m}{2}} \right).$$

We calculate

$$AA^T = \begin{pmatrix} P_-^2 + P_+^2 & P_-^2 - P_+^2 & 0 \\ P_+^2 - P_-^2 & P_-^2 + P_+^2 & 0 \\ 0 & 0 & I_{n-m} \end{pmatrix} = \begin{pmatrix} I_m & R & 0 \\ R & I_m & 0 \\ 0 & 0 & I_{n-m} \end{pmatrix} = \Sigma$$

and

$$A^TQA = \begin{pmatrix} 2P_-^2(K-F) & 0 & 0 \\ 0 & 2P_+^2(K+F) & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

where

$$2P_-^2(K-F) = \text{diag}(-\rho_1, \dots, -\rho_m), \quad 2P_+^2(K+F) = \text{diag}(\rho_1, \dots, \rho_m).$$

Hence, the distribution of $q(X, Y)$ is the same as the distribution of

$$\sum_{i=1}^m \rho_i(-\eta_i^2 + \epsilon_i^2) + \sum_{j=1}^{n-m} 0 \cdot \xi_j^2,$$

where $(\eta, \epsilon, \xi) \sim \mathcal{N}(0, I_{m+n})$. To summarize, let $Q_1, \dots, Q_m, Q'_1, \dots, Q'_m$ be i.i.d. random variables distributed according to the χ_1^2 distribution. The quadratic form $q(X, Y)$ has the same distribution as

$$\sum_{i=1}^m \rho_i(Q_i - Q'_i),$$

which can also be written as

$$q(X, Y) \sim \sum_{i=1}^m (\rho_i \chi_1^2 - \rho_i \chi_1^2).$$

Note that this distribution is symmetric around 0. We can now reconstruct the profile from $q(X, Y)$:

$$\text{Prof}_{XY} = \mathbf{I}(X; Y) + \sum_{i=1}^m \left(\frac{\rho_i}{2} \chi_1^2 - \frac{\rho_i}{2} \chi_1^2 \right),$$

which is symmetric around $\mathbf{I}(X; Y)$ and, in agreement with Proposition 5, degenerates to the atomic distribution δ_0 if and only if $\mathbf{I}(X; Y) = 0$, which is equivalent to $\rho_i = 0$ for all i .

As a linear combination of independent χ_1^2 variables, the profile has all finite moments. Using the fact that the variance of χ_1^2 distribution is 2, and quadratic scaling of variance, each term has variance $2 \cdot (\rho_i/2)^2 = \rho_i^2/2$. As variances of independent variables are additive, we can sum up all the $2m$ terms to obtain $\rho_1^2 + \dots + \rho_m^2$. $\square$

Proposition 14. *For a multivariate normal distribution with fixed mutual information, the variance of the PMI profile is maximized when $\rho_1^2 = \rho_2^2 = \dots = \rho_m^2$.*

Proof. Let $a_i = 1 - \rho_i^2$. To maximize the variance we equivalently have to minimize $a_1 + \dots + a_m$ preserving given constraint on mutual information and $a_i \in (0, 1]$.

The constraint on mutual information takes the form

$$\mathbf{I}(X; Y) = -\frac{1}{2} \sum_{i=1}^m \log(1 - \rho_i^2) = -\frac{1}{2} \log(a_1 \cdots a_m).$$

Hence, the product $a_1 \cdots a_m$ has to be constant. Denote this constant by A^m for $A \in (0, 1]$ as well. Let $a_1, \dots, a_m$ be such that $a_1 \cdots a_m = A^m$ and $a_i \in (0, 1]$. From the inequality between arithmetic and geometric means we note that

$$\frac{a_1 + \dots + a_m}{m} \geq \sqrt[m]{a_1 \cdots a_m} = A,$$

where the equality holds only if $a_1 = \dots = a_m = A$. Hence, this is the unique minimum under the constraints provided. It follows that $\rho_1^2 = \dots = \rho_m^2$. $\square$

Corollary 15. *For a multivariate normal distribution with fixed mutual information and m canonical correlations, the variance of the PMI profile does not exceed*

$$V = m(1 - \exp(-2\mathbf{I}(X; Y)/m)).$$

Proof. The variance is maximized when $\rho_1^2 = \dots = \rho_m^2$. Writing ρ^2 for the common value, we have

$$\rho^2 = 1 - \exp(-2\mathbf{I}(X; Y)/m)$$

and

$$V = m\rho^2 = m(1 - \exp(-2\mathbf{I}(X; Y)/m)).$$

The mutual information can also be written as function of variance

$$\mathbf{I}(X; Y) = -\frac{1}{2}m \log(1 - \rho^2) = -\frac{1}{2}m \log(1 - V/m).$$

□

Proposition 16. *For a multivariate normal distribution with fixed non-zero mutual information the variance of the PMI profile is minimized when $\rho_i \neq 0$ for exactly one i .*

Proof. Let $a_i \in (0, 1]$ be any numbers. We have

$$(1 - a_1)(1 - a_2) \geq 0$$

which is equivalent to

$$1 + a_1a_2 \geq a_1 + a_2$$

where the equality holds if and only if $a_1 = 1$ or $a_2 = 1$.

Using the principle of mathematical induction one can prove a more general inequality:

$$\begin{aligned} a_1a_2 \cdots a_m + (m-1) &= 1 + a_1(a_2 \cdots a_m) + (m-2) \\ &\geq a_1 + (a_2 \cdots a_m + (m-2)) \\ &\geq a_1 + a_2 + (a_3 \cdots a_m + (m-3)) \\ &\vdots \\ &\geq a_1 + a_2 + \cdots + a_m. \end{aligned}$$

Let us analyze when the equality can hold. To obtain equality in the first step, we need $a_1 = 1$ or $a_2 \cdots a_m = 1$, which, given the constraints $a_i \in (0, 1]$ would mean that $a_2 = \dots = a_m = 1$. Reasoning inductively, one proves that equality holds only when at least $m-1$ among these numbers are 1.

Now define $a_i = 1 - \rho_i^2$ and note that we are solving a maximization problem $a_i \in (0, 1]$ under a constraint

$$a_1 \cdots a_m = P.$$

Using the argument above we note that all the maxima for the above problem are permutations of the sequence $P, 1, 1, \dots, 1$. This proves that at most one $\rho_i^2 \neq 0$. Hence, we have at most one $\rho_i \neq 0$. □

A.3 Constructing new Bend and Mix Models

In this section we prove Proposition 8 and Proposition 9.

Proposition 8. *If P_{XY} is a BMM and f and g are diffeomorphisms with numerically tractable Jacobians (e.g., normalizing flows), then $P_{f(X)g(Y)}$ is a BMM.*

Proof. From the proof of Lemma 12 and the assumption of tractability of the Jacobians of f and g we obtain the tractability of the formulae for the densities $p_{f(X)g(Y)}$, $p_{f(X)}$ and $p_{g(Y)}$. Sampling $(f(X), g(Y))$ amounts to sampling (X, Y) and then transforming the sample using f and g . $\square$

Proposition 9. Consider BMMs $P_{X_1Y_1}, \dots, P_{X_KY_K}$. If $w_1, \dots, w_K$ are positive weights, such that $w_1 + \dots + w_K = 1$, then the mixture distribution $P_{X'Y'} = w_1P_{X_1Y_1} + \dots + w_KP_{X_KY_K}$ is a BMM.

Proof. We can evaluate the densities $p_{X'Y'}(x, y)$, $p_{X'}(x)$ and $p_{Y'}(y)$ of the mixture distribution using the weighted sums of $p_{X_kY_k}(x, y)$, $p_{X_k}(x)$ and $p_{Y_k}(y)$, respectively. To sample $(X', Y') \sim P_{X'Y'}$ from the mixture distribution we can sample an auxiliary variable $Z \sim \text{Categorical}(K; w_1, \dots, w_K)$. Then, we have $((X', Y') \mid Z = k) = (X_k, Y_k)$, so that we can sample from $P_{X_kY_k}$. $\square$

A.4 Mutual information in finite mixtures

In this section we discuss the counterintuitive properties of mutual information in mixture models. We start with the proof of the failing channel inequality:

Proposition 11. Let $\alpha \in [0, 1]$ be a parameter. Consider variables X , Y and Y' , s.t. $p_{XY'}(x, y) = (1 - \alpha)p_{XY}(x, y) + \alpha n(y)p_X(x)$. Then, $\mathbf{I}(X; Y') \leq (1 - \alpha)\mathbf{I}(X; Y)$.

Proof. Let $Z \sim \text{Bernoulli}(1 - \alpha)$ be an auxiliary variable. We have $(X, Y') \mid (Z = 1) \sim P_{XY}$ and $(X, Y') \mid (Z = 0) \sim P_X \otimes N_Y$.

From the data processing inequality and the chain rule we conclude that

$$\mathbf{I}(X; Y') \leq \mathbf{I}(X; Y', Z) = \mathbf{I}(X; Z) + \mathbf{I}(X; Y' \mid Z).$$

Now note that X and Z are independent, so $\mathbf{I}(X; Z) = 0$.

Hence,

$$\begin{aligned} \mathbf{I}(X; Y') &\leq \mathbf{I}(X; Y' \mid Z) \\ &= \alpha \mathbf{D}_{\text{KL}}(P_{XY' \mid Z=0} \parallel P_{X \mid Z=0} \otimes P_{Y' \mid Z=0}) + (1 - \alpha) \mathbf{D}_{\text{KL}}(P_{XY' \mid Z=1} \parallel P_{X \mid Z=1} \otimes P_{Y' \mid Z=1}) \\ &= \alpha \mathbf{D}_{\text{KL}}(P_{XY' \mid Z=0} \parallel P_X \otimes N_Y) + (1 - \alpha) \mathbf{D}_{\text{KL}}(P_{XY} \parallel P_X \otimes P_Y) \\ &= \alpha \mathbf{D}_{\text{KL}}(P_X \otimes N_Y \parallel P_X \otimes N_Y) + (1 - \alpha) \mathbf{I}(X; Y) \\ &= (1 - \alpha) \mathbf{I}(X; Y). \end{aligned}$$

$\square$

In this example, mixing the signal with a noise component, not encoding any information, decreased the information. However, mixing with independent noise components can also *increase* mutual information.

Example 17. Let $A = (0, 1)$ and $B = (1, 2)$ be two disjoint intervals of unit length. We define two pairs of random variables:

$$(X_1, Y_1) \sim \text{Uniform}(A \times A), \quad (X_2, Y_2) \sim \text{Uniform}(B \times B).$$

Note that

$$\mathbf{I}(X_1; Y_1) = \mathbf{I}(X_2; Y_2) = 0.$$

If $(X, Y) \sim 0.5P_{X_1Y_1} + 0.5P_{X_2Y_2}$ is distributed according to a mixture, we have

$$p_{XY}(x, y) = \frac{1}{2} \mathbf{1}[(x, y) \in A \times A \cup B \times B]$$

and

$$p_X(x) = \frac{1}{2} \mathbf{1}[x \in A \cup B], \quad p_Y(y) = \frac{1}{2} \mathbf{1}[y \in A \cup B].$$

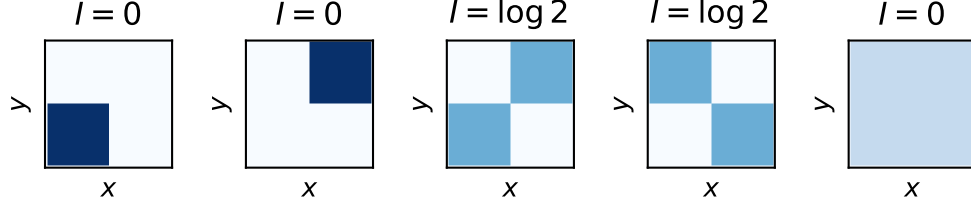


Figure 8: Mixing two distributions with zero MI can result in a mixture distribution with MI equal to one bit. Mixing two one-bit distributions can result in a mixture distribution with no mutual information.

Hence,

$$\text{PMI}_{XY}(x, y) = \log 2 \cdot \mathbf{1}[(x, y) \in A \times A \cup B \times B]$$

and

$$\mathbf{I}(X; Y) = \frac{1}{2} \log 2 + \frac{1}{2} \log 2 = \log 2.$$

Mixing two distributions with one bit of information each, can however result in independent random variables:

Example 18. Recall the distribution constructed as above:

$$p_{XY}(x, y) = \frac{1}{2} \mathbf{1}[(x, y) \in A \times A \cup B \times B]$$

and a symmetric one

$$p_{UV}(x, y) = \frac{1}{2} \mathbf{1}[(x, y) \in A \times B \cup B \times A].$$

We have

$$\mathbf{I}(X; Y) = \mathbf{I}(U; V) = \log 2.$$

On the other hand, the mixture distribution

$$(Z, T) \sim 0.5P_{XY} + 0.5P_{UV}$$

has zero mutual information, $\mathbf{I}(Z; T) = 0$, as

$$\begin{aligned} p_{ZT}(x, y) &= \frac{1}{4} \mathbf{1}[(x, y) \in (A \cup B) \times (A \cup B)] \\ &= \frac{1}{2} \mathbf{1}[x \in A \cup B] \cdot \frac{1}{2} \mathbf{1}[y \in A \cup B] \\ &= p_Z(x) \cdot p_T(y). \end{aligned}$$

Both examples have been visualised in Fig. 8. Note that similarly counterintuitive examples can be constructed using BMMs by using mixtures of multivariate normal distributions.

This demonstrates that mixtures can create and destroy information. There is however an upper bound on the amount of information mixtures can create (see also [Haussler & Opper \(1997\)](#) and [Kolchinsky & Tracey \(2017\)](#)):

Proposition 10. Consider r.v.s (X_k, Y_k) such that $\mathbf{I}(X_k; Y_k) < \infty$ for $k = 1, \dots, K$. Let (X', Y') be their mixture with weights $w_1, \dots, w_K$. Then,

$$0 \leq \mathbf{I}(X'; Y') \leq \sum_{k=1}^K w_k \mathbf{I}(X_k; Y_k) + \log K.$$

Moreover, these inequalities are tight:

1. There exists a mixture such that $\mathbf{I}(X'; Y') = \log K$ even though $\mathbf{I}(X_k; Y_k) = 0$ for all k .
2. There exists a mixture such that $\mathbf{I}(X'; Y') = 0$ even though $\mathbf{I}(X_k; Y_k) > 0$ for all k .

Proof. The examples provide explicit constructions of distributions with the specified properties. To prove the upper bound, consider a variable $Z \sim \text{Categorical}(K; w_1, \dots, w_K)$. The random variables corresponding to the mixture distribution, (X', Y') , have conditional distributions

$$(X', Y') \mid Z = k \sim P_{X_k Y_k}.$$

From the data processing inequality and chain rule we have

$$\mathbf{I}(X'; Y') \leq \mathbf{I}(X'; Y', Z) = \mathbf{I}(X'; Z) + \mathbf{I}(X'; Y' \mid Z).$$

As Z is discrete, the first summand, $\mathbf{I}(X'; Z)$, is bounded from above by the entropy $H(Z)$ (Polyanskiy & Wu, 2022, Th. 3.4e), which cannot exceed $\log K$ (Polyanskiy & Wu, 2022, Th. 1.4b). The second summand can be written as

$$\begin{aligned} \mathbf{I}(X'; Y' \mid Z) &= \sum_{k=1}^K P(Z = k) \mathbf{D}_{\text{KL}}(P_{X'Y'|Z=k} \parallel P_{X'|Z=k} \otimes P_{Y'|Z=k}) \\ &= \sum_{k=1}^K w_k \mathbf{D}_{\text{KL}}(P_{X_k Y_k} \parallel P_{X_k} \otimes P_{Y_k}) = \sum_{k=1}^K w_k \mathbf{I}(X_k; Y_k). \end{aligned}$$

□

A.5 Divergence profiles

In this section we show how to define the PMI profile in a more general setting of a *divergence profile* and prove a general invariance result (Theorem 19). The main ideas stay similar to the proof of Theorem 3.

We consider a standard Borel space $\mathcal{M}$ equipped with two probability measures, P and Q , such that $P \ll Q$. For example, if one analyses the mutual information between random variables X and Y , they can define $\mathcal{M} = \mathcal{X} \times \mathcal{Y}$ with $P = P_{XY}$ representing the joint distribution and $Q = P_X \otimes P_Y$ representing the product of marginals. Then, a sufficient condition for $P \ll Q$ is to assume that $\mathbf{I}(X; Y) < \infty$.

As P is absolutely continuous with respect to Q , the Radon–Nikodym derivative $f = dP/dQ$ can be defined. It is a measurable function $f: \mathcal{M} \rightarrow [0, \infty)$ defined up to a Q -null set.

This allows one to define the Kullback–Leibler divergence:

$$\mathbf{D}_{\text{KL}}(P \parallel Q) = \int f \log f \, dQ = \int \log f \, dP,$$

where $0 \log 0 = 0$ in the first expression. In the second expression, one can notice that as $P \ll Q$, $\log 0$ can occur only on a P -null set. Additionally, using another Radon–Nikodym derivative (differing from f on a Q -null set), does not change the value of the integral, so that Kullback–Leibler divergence is indeed well-defined.

These properties correspond to the following properties of the *divergence profile*, defined as the pushforward measure

$$\text{Prof}_{P \parallel Q} = (\log f)_\# P.$$

Namely, this distribution does not depend on the choice of f (i.e., we can replace f by a function differing on a Q -null set) and it is easy to prove that $\text{Prof}_{P \parallel Q}(\{-\infty, \infty\}) = 0$. As this distribution does not have atoms at the infinite values, we can treat it as a distribution over real numbers, $\mathbb{R}$.

Now consider a measurable mapping $i: \mathcal{M} \rightarrow \mathcal{M}'$ between standard Borel spaces, allowing us to define the pushforward measures $i_\# P$ and $i_\# Q$ on $\mathcal{M}'$. We are interested in investigating the profile defined using these pushforward distributions. For example, if $\mathcal{M} = \mathcal{X} \times \mathcal{Y}$, $\mathcal{M}' = \mathcal{X}' \times \mathcal{Y}'$, $P = P_{XY}$, $Q = P_X \otimes P_Y$ and one

considers random variables $X' = i_1(X)$ and $Y' = i_2(Y)$, then the mapping $i = i_1 \times i_2: \mathcal{M} \rightarrow \mathcal{M}'$ defines the pushforward measures $P_{X'Y'} = i_{\#}P_{XY}$ and $P_{X'} \otimes P_{Y'} = i_{\#}(P_X \otimes P_Y)$.

Note that the profile $\text{Prof}_{i_{\#}P \parallel i_{\#}Q}$ can be defined in this case: as $P \ll Q$, then also $i_{\#}P \ll i_{\#}Q$, because $i_{\#}Q(B) = 0$ is equivalent to $Q(i^{-1}(B)) = 0$, which then implies $P(i^{-1}(B)) = 0$.

In general, $\text{Prof}_{i_{\#}P \parallel i_{\#}Q}$ can differ from $\text{Prof}_{P \parallel Q}$: for example, using a constant mapping i results in $\text{Prof}_{i_{\#}P \parallel i_{\#}Q} = \delta_0$, independently on the original profile $\text{Prof}_{P \parallel Q}$. However, we can generalize Theorem 3 as follows:

Theorem 19. *Let $i: \mathcal{M} \rightarrow \mathcal{M}'$ be a measurable mapping between standard Borel spaces with a measurable left inverse. If P and Q are two probability distributions on $\mathcal{M}$ such that $P \ll Q$, then $\text{Prof}_{i_{\#}P \parallel i_{\#}Q} = \text{Prof}_{P \parallel Q}$.*

To prove it, we use the following version of Lemma 12, which generalizes a result of G. (2022):

Lemma 20. *Let $i: \mathcal{M} \rightarrow \mathcal{M}'$ be a measurable mapping between standard Borel spaces such that there exists a measurable left inverse $a: \mathcal{M}' \rightarrow \mathcal{M}$. If P and Q are two probability distributions on $\mathcal{X}$ such that $P \ll Q$ and $f = dP/dQ$ is the Radon–Nikodym derivative, then $i_{\#}P \ll i_{\#}Q$ and the Radon–Nikodym derivative is given by $di_{\#}P/di_{\#}Q = f \circ a$.*

Proof of Lemma 20. Let $a: \mathcal{M}' \rightarrow \mathcal{M}$ be any measurable left inverse of i . To prove that $f \circ a: \mathcal{M}' \rightarrow [0, \infty)$ is the Radon–Nikodym derivative we need to show that

$$i_{\#}P(B) = \int_B f \circ a \, di_{\#}Q$$

for every Borel subset $B \subseteq \mathcal{M}'$. Using the change of variables formula:

$$\begin{aligned} \int_B f \circ a \, di_{\#}Q &= \int_{i^{-1}(B)} f \circ a \circ i \, dQ \\ &= \int_{i^{-1}(B)} f \, dQ \\ &= \int_{i^{-1}(B)} dP \\ &= P(i^{-1}(B)) = i_{\#}P(B), \end{aligned}$$

where the second equality follows from the definition of the left inverse, $a \circ i = \text{id}_{\mathcal{M}}$.

Note that even if a is substituted for another left inverse, the Radon–Nikodym derivative $f \circ a$ is still determined uniquely up to an $i_{\#}Q$ -null set. $\square$

Proof of Theorem 19. The profile is defined as $\text{Prof}_{i_{\#}P \parallel i_{\#}Q} = (\log di_{\#}P/di_{\#}Q)_{\#}(i_{\#}P)$. Take any measurable left inverse a of i and write

$$\log \frac{di_{\#}P}{di_{\#}Q} = \log f \circ a,$$

where $f = dP/dQ$. We have

$$\text{Prof}_{i_{\#}P \parallel i_{\#}Q} = (\log f \circ a)_{\#} i_{\#}P = (\log f \circ a \circ i)_{\#}P = (\log f)_{\#}P = \text{Prof}_{P \parallel Q}.$$

$\square$

Remark 21. *The above proof suggests that for $P \ll Q$ on a standard Borel space $\mathcal{M}$ and sufficiently well-behaved functions $u: [0, \infty) \rightarrow \mathbb{R}$, one could define a generalized profile*

$$\left(u \circ \frac{dP}{dQ} \right)_{\#} P,$$

which stays invariant under measurable mappings $i: \mathcal{M} \rightarrow \mathcal{M}'$ with measurable left inverses.

These quantities can be used to study whether a distribution could not be obtained as a simple reparametrization of another one, similarly as the PMI profiles were used in Fig. 1.

We suspect that it may also be possible to generalize the profile to use a general f -divergence (Polyanskiy & Wu, 2022, Ch. 7), rather than the Kullback–Leibler divergence studied in this manuscript. Nowozin et al. (2016) related f -divergences and their variational lower bounds (cf. Sec. 3.3) to generative adversarial networks (GANs), although a relation between a “ f -divergence profile” and the GAN training remains unclear to us.

Theorem 19 proves invariance under the existence of a measurable left inverse. Czyż et al. (2023, Theorem 2.1) prove invariance of the mutual information under reparametrizations by continuous injective mappings. As the lemma below shows, Theorem 19 holds also for continuous injective mappings:

Lemma 22. *Let $i: \mathcal{M} \rightarrow \mathcal{M}'$ be a continuous injective mapping between two standard Borel spaces. Then, i admits a measurable left inverse.*

Proof. Choose an arbitrary point $x_0 \in \mathcal{M}$ and define a function $a: \mathcal{M}' \rightarrow \mathcal{M}$ in the following manner:

$$a(y) = \begin{cases} x_0 & \text{if } y \notin i(\mathcal{M}) \\ x & \text{if } x \text{ is the (unique) point such that } i(x) = y \end{cases}$$

This function is well-defined due to the fact that i is injective and it is a left inverse: $a(i(x)) = x$ for all $x \in \mathcal{M}$. Now we need to prove that it is measurable.

Take any Borel set $B \subseteq \mathcal{M}$ and consider its preimage $a^{-1}(B) = \{y \in \mathcal{M}' \mid a(y) \in B\}$. If $x_0 \notin B$, we have $a^{-1}(B) = i(B)$, which is Borel by Lusin–Suslin theorem. If $x_0 \in B$, we can write

$$\begin{aligned} a^{-1}(B) &= a^{-1}(B \setminus \{x_0\}) \cup a^{-1}(\{x_0\}) \\ &= i(B \setminus \{x_0\}) \cup \{i(x_0)\} \cup (\mathcal{M}' \setminus i(\mathcal{M})), \end{aligned}$$

which is Borel as a finite union of Borel sets. □

B Distributions involving discrete variables

The formalism in Section 2 is applicable to both continuous and discrete random variables, although in Section 3.1 we focus on the distributions in which both X and Y are continuous. If X and Y are discrete, mutual information $\mathbf{I}(X; Y)$ can be calculated analytically from the joint probability matrix and there exist numerous approaches to estimate it from collected samples (Hutter, 2001; Brillinger, 2004).

In this section we consider the mixed case, in which one variable is continuous and the other one is discrete. For example, Carrara & Ernst (2023) describe a particle physics experiment in which X is an 18-dimensional random variable, but Y is binary. Grabowski et al. (2019) consider a cell transmitting information through the MAPK signalling pathway, assuming the input signal X to be discrete and the measured response Y to be continuous.

B.1 Known distributions

There are only a few distributions P_{XY} with known ground-truth mutual information assuming this discrete-continuous case. Gao et al. (2017, Sec. 5) describe a discrete random variable X which is uniformly sampled from the set $\mathcal{X} = \{0, \dots, m-1\}$ and the continuous Y variable is sampled as

$$(Y \mid X = x) \sim \text{Uniform}(x, x + 2),$$

which is therefore distributed on $\mathcal{Y} = (0, m + 1)$. Gao et al. (2017) prove that mutual information in this case is

$$\mathbf{I}(X; Y) = \log m - \frac{m}{m-1} \log 2$$

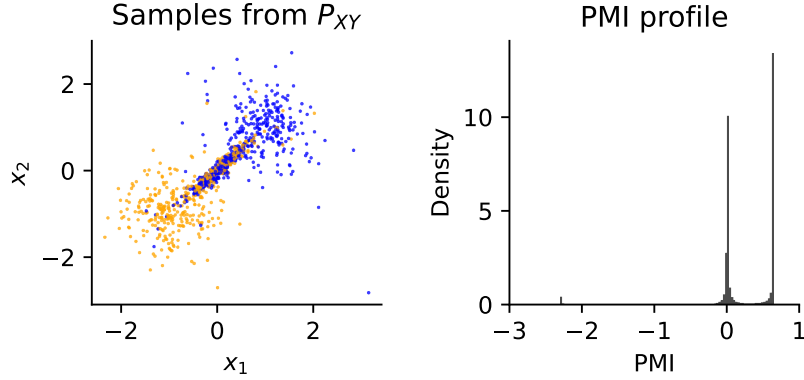


Figure 9: Left: samples from the P_{XY} distribution, colored by the value of the binary variable Y . Right: the PMI profile of P_{XY} distribution.

for $m \geq 2$ and $\mathbf{I}(X; Y) = 0$ for $m = 1$ and consider a multivariate analogue of this distribution, in which pairs of variables (X_k, Y_k) for $k = 1, \dots, K$ are sampled independently using the above procedure. Then, they are concatenated into multivariate vectors $(X_1, \dots, X_K)$ and $(Y_1, \dots, Y_K)$ with K times larger mutual information.

We will show how to relate the bivariate example to the framework of Bend and Mix Models (multivariate case can be constructed analogously). Note that the joint distribution P_{XY} is not strictly in $\mathcal{P}(\mathcal{X}, \mathcal{Y})$, as it is supported on the one-dimensional manifold $\mathcal{M} \subset \mathcal{X} \times \mathcal{Y}$ with m connected components (cf. Politis (1991) and Marx et al. (2021)), on which

$$p_{XY}(x, y) = \frac{1}{2m} \mathbf{1}[y \in (x, x + 2)],$$

but it is possible to extend the definitions of Section 2, so that this technical difficulty is resolved. The marginal distributions are tractable and admit PDFs:

$$\begin{aligned} p_X(x) &= 1/m, \\ p_Y(y) &= \sum_{x=0}^{m-1} p_{XY}(x, y). \end{aligned}$$

Although p_Y is not smooth on $\mathcal{Y}$ (at integer points), this technical difficulty can also be resolved as this set is of measure zero. Hence, although this distribution is not strictly a BMM, it can still be modeled using the introduced framework.

Next, Gao et al. (2017, Sec. 5) consider the zero-inflated Poissonization of the exponential random variable valued in $\mathcal{X} = \{x \in \mathbb{R} \mid x \geq 0\}$, with: $X \sim \text{Exp}(1)$ and Y being a discrete random variable valued in the set of non-negative integers $\mathcal{Y} = \{0, 1, 2, \dots\}$:

$$(Y \mid X = x) \sim p \delta_0 + (1 - p) \text{Poisson}(x).$$

They show that

$$\mathbf{I}(X; Y) = (1 - p) \left(2 \log 2 - \gamma - \sum_{k=0}^{\infty} \log k \cdot 2^{-k} \right),$$

where γ is the Euler–Mascheroni constant.

To use BMMs in this case, one has to formally extend the definition of $\mathcal{P}(\mathcal{X}, \mathcal{Y})$ as $\mathcal{X}$ is a manifold with boundary (Lee, 2012, Ch. 1). The joint probability distribution is then given by

$$p_{XY}(x, y) = p_X(x) \cdot p_{Y|X}(y \mid x) = e^{-x} \cdot \left(p \cdot \mathbf{1}[y = 0] + (1 - p) \frac{e^{-x} x^y}{y!} \right)$$

and the PMF function of the Y variable is also analytically known:

$$p_Y(y) = p \cdot \mathbf{1}[y = 0] + (1 - p) \cdot 2^{-(1+y)}$$

Hence, the framework of BMMs, with minor technical adjustments, can accommodate the above distributions.

B.2 Novel distributions

However, Bend and Mix Models also allow one to create more expressive distributions, for which analytical formulae for ground-truth mutual information are not available, but can be approximated with the Monte Carlo methods as explained in Section 2: consider a continuous random variable X and a discrete random variable Y . To introduce a dependency between X and Y variable, we can use a mixture of distributions in which the component variables are independent, i.e., $P_{X_k Y_k} = P_{X_k} \otimes P_{Y_k}$. Therefore, we consider a graphical model $X \leftarrow Z \rightarrow Y$, in which the distributions $P_{X_k} = P_{X|Z=k}$ are known and have tractable PDFs. The distributions of $P_{Y_k} = P_{Y|Z=k}$ are given by probability tables. Monte Carlo estimators of Section 2 can then be used to estimate $\mathbf{I}(X; Y)$ with high accuracy. Note also that this general procedure includes the case $X \leftarrow Y$ by setting $Z = Y$.

We illustrate it in a simple example with $\mathcal{X} = \mathbb{R}^2$, $\mathcal{Y} = \{0, 1\}$ and $K = 3$ components. The first component models a cluster in the $\mathcal{X}$ space, strongly associated with $Y = 1$ value. For P_{X_1} we use a bivariate Student distribution centered at $(1, 1)$ with isotropic dispersion $\Omega = 0.2 \cdot I_2$ and 8 degrees of freedom. We take P_{Y_1} to be the Bernoulli variable with probability $P(Y_1 = 1) = 0.95$.

Analogously, we define a second cluster, strongly associated with $Y = 0$ value: P_{X_2} is a bivariate Student distribution with the same dispersion matrix, but centered at $(-1, -1)$ and with 8 degrees of freedom. Then, Y_2 is a Bernoulli variable with $P(Y_2 = 1) = 0.05$.

We then define a third component using a bivariate normal distribution centered at $(0, 0)$ and with covariance matrix

$$\Sigma = 0.1 \begin{pmatrix} 1 & 0.95 \\ 0.95 & 1 \end{pmatrix}.$$

This component is not informative of Y , that is $P(Y_3 = 1) = 0.5$.

We used weights $w_1 = w_2 = 1/4$ and $w_3 = 1/2$, what resulted in the distribution visualised in Fig. 9. We estimated both the profile and the mutual information using $N = 10^6$ samples and obtained $\mathbf{I}(X; Y) = 0.224$ with MCSE of $5.1 \cdot 10^{-4}$.

In principle, the above construction can be used to generate realistic high-dimensional data sets (e.g., audio or image) with known ground-truth mutual information, by assuming the generative model $Y \rightarrow X$ (i.e., $Z = Y$) and modeling each P_{X_k} using a normalizing flow or an autoregressive model (Murphy, 2023, Ch. 22) trained on an auxiliary data set with fixed label $Y_k = y_k$. Hence, at least in principle, one could obtain highly-expressive generative process $p_{X|Y}(x | y)$ with tractable probability and sampling. Pairing this with an arbitrary probability vector $p_Y(y)$ one can obtain $p_{XY}(x, y)$ and $p_X(x)$ even for high-dimensional data sets, so that Monte Carlo estimator can be used to determine the ground-truth mutual information. However, we anticipate possible practical difficulties with scaling up the proposed approach to high-dimensional data and we leave empirical investigation of this topic to future work.

C Experimental protocols and additional experiments

In this section we provide detailed information on experimental protocols used (such as hyperparameter choices) and supplement the findings of Sec. 3 with additional experiments.

C.1 Specification of the proposed distributions

In this section we provide the details of the distributions introduced in Sec. 3.1. To estimate the ground-truth mutual information we used the Monte Carlo approach described in Sec. 2 with $N = 200\,000$ samples.

C.1.1 The X distribution

We constructed the X distribution as a mixture of bivariate normal distributions with equal weights, zero mean and covariance matrices specified by

$$\Sigma_{\pm} = 0.3 \begin{pmatrix} 1 & \pm 0.9 \\ \pm 0.9 & 1 \end{pmatrix}.$$

Note that the marginal distributions of each of component distributions is $\mathcal{N}(0, 0.3^2)$ and subsequently their mixture has exactly the same marginal distributions. This is therefore an interesting example of a distribution in which the joint probability distribution is not multivariate normal, although the marginal distributions of X and Y variables are normal individually. The mutual information is in this case $\mathbf{I}(X; Y) = 0.41$ nats.

C.1.2 The AI distribution

The AI distribution was constructed as an equally-weighted mixture of six bivariate normal distributions with equal weights and the following parameters:

$$\begin{aligned} \mu_1 &= (1, 0) \\ \Sigma_1 &= \text{diag}(0.01, 0.2) \\ \mu_2 &= (1, 1) \\ \Sigma_2 &= \text{diag}(0.05, 0.001) \\ \mu_3 &= (1, -1) \\ \Sigma_3 &= \text{diag}(0.05, 0.001) \\ \mu_4 &= (-0.8, -0.2) \\ \Sigma_4 &= \text{diag}(0.03, 0.001) \\ \mu_5 &= (-1.2, 0) \\ \Sigma_5 &= \begin{pmatrix} 0.04 & 0.085 \\ 0.085 & 0.2 \end{pmatrix} \\ \mu_6 &= (-0.4, 0) \\ \Sigma_6 &= \begin{pmatrix} 0.04 & -0.085 \\ -0.085 & 0.2 \end{pmatrix} \end{aligned}$$

The mutual information of this distribution is $\mathbf{I}(X; Y) = 0.78$ nats.

C.1.3 The Galaxy distribution

The Galaxy distribution was constructed as an equally-weighted mixture of isotropic multivariate normal distributions with $\mu_{\pm} = \pm(1, 1, 1)$ and unit covariance matrix and the X variable was transformed using the spiral diffeomorphism with $v = 0.5$, which is described by [Czyż et al. \(2023\)](#). The Galaxy distribution contains $\mathbf{I}(X; Y) = 0.49$ nats.

C.1.4 The Waves distribution

The Waves distribution was created as an equally-weighted mixture of 12 multivariate normal distributions with equal covariance matrices $\Sigma = \text{diag}(0.1, 1, 0.1)$ and mean vectors

$$\mu_i = (x, 0, x \bmod 4), \quad i \in \{0, 1, \dots, 11\}.$$

This construction results in a distribution where different vertical components of the X variable are assigned Y values calculated modulo 4. Then, we transformed the X variable with a continuous injection

$$f(x_1, x_2) = (x_1 + 5 \sin(3x_2), x_2),$$

which does not change the mutual information. Finally, we applied the affine mappings

$$a_1(x) = 0.1x - 0.8, \quad a_2(y) = 0.5y,$$

to make the range of the typical values comparable with other distributions. This distribution encodes $\mathbf{I}(X; Y) = 1.31$ nats.

C.2 A benchmark extension proposal

In Sec. 3.1 and Appendix C.1 we provide examples of novel low-dimensional distributions, which pose a significant challenge to the mutual information estimators. Therefore, we would like to propose a new benchmark of mutual information estimators, extending the benchmark of Czyż et al. (2023).

C.2.1 Included distributions

The original benchmark of Czyż et al. (2023) consists of 40 distributions. On several problems the estimator performance was similar, so that choosing a representative from a group could allow us to reduce the computation amount. Additionally, we decided to remove the tasks which we considered too difficult: for example, the Student distributions are mapped through the asinh transform to remove the tails. Together with the new tasks, based on BMMs the new version of the benchmark consists of 26 distributions grouped as follows:

One-dimensional variables We include four distributions such that $\mathcal{X} = \mathcal{Y} = \mathbb{R}$. From the original benchmark we decided to retain the additive noise task with $\varepsilon = 0.75$ and the asinh-transformed version of the centered bivariate Student distribution with one degree of freedom and the identity matrix for the dispersion (see Czyż et al. (2023, Appendix D) for precise descriptions). Additionally, we include X and AI distributions described in Appendix C.1.

Embeddings We retained the Swiss roll embedding distribution (Czyż et al., 2023, Appendix D). In this problem the distribution P_{XY} is supported on a two-dimensional surface embedded in $\mathbb{R}^3$. This distribution does not have a density with respect to the Lebesgue measure on $\mathbb{R}^3$.

Many-versus-one distributions We consider multiple distributions with $\mathcal{X} = \mathbb{R}^m$ and $\mathcal{Y} = \mathbb{R}$. For $m = 2$ we consider the Waves and Galaxy distributions described in Appendix C.1. For other m we propose the concentric isotropic multivariate normal distributions. Namely, let K be a parameter, specifying the number of components of a BMM. Each component has independent (multivariate) normal variables $X_k \sim \mathcal{N}(0, k^2 I_m)$ and $Y_k \sim \mathcal{N}(k, 10^{-4})$. To obtain non-zero mutual information we mix these distributions with equal proportions, i.e., the weights vector is given by $w_k = 1/K$ for all k . We consider four tasks obtained by varying $m \in \{3, 5\}$ and $K \in \{5, 10\}$. Additionally, we consider a high-dimensional distribution with $m = 25$ and $K = 5$ components.

Multivariate normal and Student distributions We selected five problems from multivariate normal distributions (Czyż et al., 2023). We use $\mathcal{X} = \mathbb{R}^m$ and $\mathcal{Y} = \mathbb{R}^n$ and the dense interaction model jointly changing $m = n \in \{5, 25, 50\}$ dimensions and the sparse interactions model (with two pairs of interacting dimensions) in $m = n \in \{5, 25\}$ dimensions.

Additionally, we selected three multivariate Student distributions with the dispersion being the identity matrix. For $m = n = 2$ we consider a distribution with one degree of freedom (i.e., the multivariate Cauchy distributions, which does not have first two moments) and for $m = n \in \{3, 5\}$ we consider a distribution with two degrees of freedom (for which the first moment is defined, but not the second). As described above, the Student distributions had been transformed with the asinh mapping to reduce the tails.

Spiral diffeomorphism As described in Czyż et al. (2023), one can transform the multivariate normal distribution with the spiral diffeomorphism, which results in empirically challenging problems. We selected the distributions corresponding to the sparse interactions in $m = n \in \{3, 5\}$ dimensions.

Table 1: Benchmark results. New benchmark problems are marked in green. Best-performing estimator in each row has been marked with bold font.

	CCA	DV	InfoNCE	KSG	MINE	NWJ	True MI
AI	0.00 ± 0.01	0.57 ± 0.06	0.61 ± 0.03	0.78 ± 0.02	0.49 ± 0.07	0.53 ± 0.10	0.78
X	0.00 ± 0.01	0.37 ± 0.02	0.38 ± 0.02	0.42 ± 0.02	0.33 ± 0.04	0.36 ± 0.02	0.41
Additive	0.18 ± 0.01	0.31 ± 0.02	0.30 ± 0.02	0.32 ± 0.01	0.32 ± 0.02	0.32 ± 0.01	0.33
Swiss roll	0.02 ± 0.01	0.37 ± 0.02	0.37 ± 0.03	0.42 ± 0.02	0.37 ± 0.03	0.37 ± 0.02	0.41
Galaxy	0.01 ± 0.01	0.31 ± 0.04	0.36 ± 0.03	0.46 ± 0.01	0.26 ± 0.06	0.23 ± 0.04	0.49
Waves	0.04 ± 0.01	0.14 ± 0.07	0.34 ± 0.13	0.66 ± 0.01	0.13 ± 0.10	0.11 ± 0.07	1.31
Concentric (3-dim, 10)	0.00 ± 0.01	0.50 ± 0.02	0.50 ± 0.02	0.51 ± 0.01	0.45 ± 0.02	0.44 ± 0.03	0.56
Concentric (3-dim, 5)	0.00 ± 0.01	0.41 ± 0.03	0.41 ± 0.02	0.42 ± 0.02	0.38 ± 0.04	0.36 ± 0.03	0.46
Concentric (5-dim, 10)	0.00 ± 0.01	0.65 ± 0.03	0.66 ± 0.03	0.55 ± 0.01	0.62 ± 0.04	0.56 ± 0.02	0.75
Concentric (5-dim, 5)	0.00 ± 0.01	0.55 ± 0.03	0.56 ± 0.02	0.46 ± 0.01	0.53 ± 0.03	0.42 ± 0.05	0.63
Concentric (25-dim, 5)	0.00 ± 0.01	0.67 ± 0.06	0.68 ± 0.04	0.12 ± 0.01	0.65 ± 0.04	0.28 ± 0.06	1.19
Student (1-dim)	0.00 ± 0.01	0.17 ± 0.07	0.20 ± 0.02	0.22 ± 0.01	0.18 ± 0.01	0.18 ± 0.06	0.22
Student (2-dim)	0.00 ± 0.01	0.36 ± 0.03	0.37 ± 0.03	0.36 ± 0.02	0.27 ± 0.10	0.24 ± 0.08	0.43
Student (3-dim)	0.00 ± 0.01	0.16 ± 0.08	0.20 ± 0.03	0.17 ± 0.01	0.09 ± 0.08	0.03 ± 0.01	0.29
Student (5-dim)	0.01 ± 0.01	0.21 ± 0.07	0.28 ± 0.03	0.23 ± 0.01	0.13 ± 0.12	0.03 ± 0.03	0.45
Inliers (25-dim, 0.2)	0.58 ± 0.03	0.34 ± 0.04	0.27 ± 0.09	0.12 ± 0.03	0.39 ± 0.05	0.41 ± 0.04	0.63
Inliers (25-dim, 0.5)	0.24 ± 0.02	-0.11 ± 0.09	-0.08 ± 0.04	0.05 ± 0.02	0.08 ± 0.02	0.06 ± 0.04	0.27
Inliers (5-dim, 0.2)	0.52 ± 0.02	0.55 ± 0.03	0.55 ± 0.03	0.45 ± 0.02	0.54 ± 0.05	0.56 ± 0.08	0.63
Inliers (5-dim, 0.5)	0.18 ± 0.01	0.19 ± 0.02	0.18 ± 0.03	0.19 ± 0.01	0.20 ± 0.04	0.21 ± 0.04	0.27
Normal (25-dim, dense)	1.35 ± 0.02	1.09 ± 0.07	1.06 ± 0.06	1.05 ± 0.02	1.16 ± 0.05	0.18 ± 0.38	1.29
Normal (5-dim, dense)	0.60 ± 0.02	0.55 ± 0.03	0.54 ± 0.03	0.56 ± 0.01	0.56 ± 0.03	0.56 ± 0.02	0.59
Normal (50-dim, dense)	1.87 ± 0.02	1.26 ± 0.07	1.20 ± 0.09	1.25 ± 0.02	1.45 ± 0.05	0.62 ± 0.78	1.62
Normal (25-dim, sparse)	1.08 ± 0.02	0.69 ± 0.05	0.67 ± 0.07	0.18 ± 0.02	0.78 ± 0.05	0.79 ± 0.03	1.02
Normal (5-dim, sparse)	1.03 ± 0.02	0.95 ± 0.02	0.95 ± 0.02	0.69 ± 0.02	0.92 ± 0.05	0.95 ± 0.02	1.02
Spiral (3-dim)	0.24 ± 0.02	0.50 ± 0.03	0.53 ± 0.04	0.65 ± 0.02	0.43 ± 0.05	0.45 ± 0.03	1.02
Spiral (5-dim)	0.39 ± 0.02	0.48 ± 0.03	0.52 ± 0.03	0.50 ± 0.01	0.45 ± 0.03	0.48 ± 0.03	1.02

Inliers Finally, we implemented four distributions based along the principles described in Sec. 3.2. Namely, we took the multivariate normal distributions with sparse interactions (described above) with $m = n \in \{5, 25\}$ dimensions P_{XY} and then constructed a mixture $(1 - \alpha)P_{XY} + \alpha P_X \otimes P_Y$ for the inlier fraction $\alpha \in \{0.2, 0.5\}$.

C.2.2 Experimental protocol

Determining the ground-truth value The ground-truth mutual information for distributions proposed by Czyż et al. (2023) is analytically available. For the tasks based on BMMs we used $N = 200,000$ Monte Carlo samples to provide an estimate, as described in Section 2.2. The additional computational cost of estimating the ground-truth mutual information for these distributions is minor: on a standard laptop, estimating the mutual information of the Galaxy task to be 0.4953 ± 0.0012 (which results in the relative MCSE smaller than 0.5%, using 200,000 samples) takes less than 2.5 seconds. Similarly, estimating the mutual information of the Concentric task (25-dimensional variant with 5 components) to be 1.1911 ± 0.0015 (using 200,000 samples again) takes less than 3.5 seconds.

Included estimators We included a total of six estimators in this benchmark: four variational estimators, the neighborhood-based KSG estimator (Kraskov et al., 2004), and the simple CCA-based estimator of Kay (1992). We chose them as the most promised approaches: Czyż et al. (2023) argue that KSG is the preferred estimator for low-dimensional problems, neural estimators are preferred choices for high-dimensional problems if only a few dimensions are interacting, and the CCA-based estimator is the optimal choice for multivariate normal distributions.

Reported uncertainty For each distribution we sampled $N' = 5,000$ data points $S = 10$ times to obtain different data sets on which the estimators were run. For each estimator and a task we then calculated the mean and the standard deviation (basing on the $S - 1$ degrees of freedom), which we reported in the table up to two decimal digits. We rounded the standard deviation upwards to two decimal digits to not underestimate it.

C.2.3 Benchmark results

We present the obtained results in Table 1. Results for the BMM estimator on low-dimensional tasks are shown in Table 2.

Problems not solved Interestingly, the Waves task is not solved by any estimator, with the best-performing one being KSG and underestimating the MI by 50%. Similarly, a high-dimensional Concentric distribution

(25-dimensional variables with 5 components) is not solved by any estimator. This contrasts with using a single component (Normal (25-dim, dense) and Normal (25-dim, sparse)), for which the CCA estimator can be used. Problems involving the spiral diffeomorphism remain unsolved.

KSG For low-dimensional problems, the KSG estimator seems preferable, which agrees with the conclusions from Czyż et al. (2023).

Neural estimators In high-dimensional problems (25-dimensional or 50-dimensional variables), neural estimators generally outperform KSG. However, they typically still underestimate the ground-truth mutual information and the model-based CCA estimate may be preferable in the problems involving multivariate normal distribution.

CCA In the proposed benchmark it is visible that a CCA-based estimator is not competitive on majority of the problems. The concentric multivariate normal distributions, designed as adversarial examples, indeed result in CCA not being able to find any mutual information. Moreover, we see that larger numbers of inliers result in worse predictions. Interestingly, for $N = 5,000$ data points and high-dimensional problems the CCA-based estimator can overestimate the result, by overfitting to the noise. This supports the view that regularizing may improve the estimates (see Sec. 3.4).

C.3 Estimator hyperparameters

Czyż et al. (2023, Appendix E.4) study the effects of hyperparameters on mutual information estimators. We decided to use the histogram-based estimator (Cellucci et al., 2005; Darbellay & Vajda, 1999) with a fixed number of 10 bins per dimension and the popular KSG estimator (Kraskov et al., 2004) with $k = 10$ neighbors. Canonical correlation analysis (Kay, 1992; Brillinger, 2004) does not have any hyperparameters. Finally, we variational estimators with the neural critic being a ReLU network of variant M (with 16 and 8 hidden neurons), as it obtained competitive performance in the benchmark of Czyż et al. (2023, Appendix E.4). As a preprocessing strategy, we followed Czyż et al. (2023, Appendix E.3) and transformed all samples to have zero empirical mean and unit variance along each dimension. Our code is based on the MIT-licensed code associated with the Czyż et al. (2023) publication.

C.4 Variational estimators of mutual information

In Sec. 3.3 we study the loss of Belghazi et al. (2018) as well as two other loss functions. This section provides a more detailed overview of these approaches. At the same time, our review is not exhaustive: many more variational lower bounds exist and have been described in the literature. For a detailed overview we refer to the articles by Poole et al. (2019) and Song & Ermon (2020), as well as to Chapters 4 and 7 of the textbook by Polyanskiy & Wu (2022).

Belghazi et al. (2018) use the Donsker–Varadhan loss,

$$I_{DV}(f) = \mathbb{E}_{P_{XY}}[f] - \log \mathbb{E}_{P_X \otimes P_Y}[\exp f],$$

which is a lower bound on $\mathbf{I}(X; Y)$ for any bounded function f . Hence, they decide to use a neural critic, i.e., a neural network $f: \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$ maximizing the functional form to obtain an (approximate) lower bound on mutual information.

As Poole et al. (2019) describe, this bound becomes tight if c is any real number and $f = \text{PMI}_{XY} + c$, i.e., $I_{DV}(f) = \mathbf{I}(X; Y)$, so one can approach MI estimation to optimisation I_{DV} over a flexible family of functions f parameterized by neural networks. However, we note that as only a finite sample is available, the expectation value cannot be calculated exactly, so that any provided estimate does not need to be a lower bound on MI. Moreover, if no split into training and test set is used, then f may overfit and provide biased estimates (McAllester & Stratos, 2020).

Another lower bound, introduced in Nguyen et al. (2007),

$$I_{NWJ}(f) = \mathbb{E}_{P_{XY}}[f] - \mathbb{E}_{P_X \otimes P_Y}[\exp(f - 1)],$$

becomes tight for $f = \text{PMI}_{XY} + 1$.

Oord et al. (2018) propose a variational approximation loss which uses a batch of $(x_i, y_i)_{i=1, \dots, n}$ samples from P_{XY} to estimate

$$I_{\text{NCE}}(f) = \mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n \log \frac{\exp f(x_i, y_i)}{\frac{1}{n} \sum_{j=1}^n \exp f(x_i, y_j)} \right]$$

Here, if $f(x, y) = \text{PMI}_{XY}(x, y) + c(x)$, where c is any function, then $I_{\text{NCE}}(f) \rightarrow \mathbf{I}(X; Y)$ as $n \rightarrow \infty$.

C.5 Bayesian estimation of Gaussian mixture models

In this section we provide additional details on using BMMs to estimate mutual information and the pointwise mutual information profile.

C.5.1 Model description

Recall from Sec. 3.4 that Bayesian estimation of mutual information consists of the following steps:

1. Propose a parametric generative model of the data, $P_\theta := P(X, Y \mid \theta)$, and assume a prior $P(\theta)$ on the parameter space.
2. Use a Markov chain Monte Carlo method to obtain a sample $\theta^{(1)}, \dots, \theta^{(m)}$ from the posterior $P(\theta \mid X_1, Y_1, \dots, X_N, Y_N)$.
3. Estimate mutual information (and the PMI profile) for each $\theta^{(m)}$ using the Monte Carlo method described in Sec. 2.
4. Validate the findings using e.g., posterior predictive checks and cross-validation.

We consider the following sparse Gaussian mixture model with $K = 10$ components:

$$\begin{aligned} \pi &\sim \text{Dirichlet}(K; 1/K, 1/K, \dots, 1/K), \\ Z_n \mid \pi &\sim \text{Categorical}(\pi), & n = 1, \dots, N, \\ \mu_k &\sim \mathcal{N}(0, 3^2 I_D), & k = 1, \dots, K, \\ \Sigma_k &\sim \text{ScaledLKJ}(1, 1), & k = 1, \dots, K, \\ (X_n, Y_n) \mid Z_n, \{\mu_k, \Sigma_k\} &\sim \mathcal{N}(\mu_{Z_n}, \Sigma_{Z_n}), & n = 1, \dots, N. \end{aligned}$$

Sampling a single covariance matrix Σ from $\text{ScaledLKJ}(\sigma, \eta)$ distribution corresponds to sampling the correlation matrix R from the Lewandowski-Kurowicka-Joe (LKJ) distribution (Lewandowski et al., 2009):

$$p(R) \propto (\det R)^{\eta-1},$$

sampling the scale parameters

$$\lambda_1, \lambda_2, \dots, \lambda_D \sim \text{HalfCauchy}(\text{scale}=\sigma),$$

and then constructing the covariance matrix as $\Sigma_{ij} = R_{ij} \lambda_i \lambda_j$.

The sparse Dirichlet prior is a finite-dimensional alternative to the Dirichlet process, which truncates the number of occupied clusters depending on the data (Frühwirth-Schnatter & Malsiner-Walli, 2019). In particular, the *a priori* expected number of clusters depends on the number of data points to be observed. We generally use NumPyro (Phan et al., 2019) with local latent variables Z_n marginalized out, what allowed us to run Markov chain Monte Carlo inference using the NUTS sampler (Hoffman & Gelman, 2014).

The m th sample is therefore given by

$$\theta^{(m)} = \left(\pi^{(m)}, \left(\mu_k^{(m)}, \Sigma_k^{(m)} \right)_{k=1, \dots, K} \right)$$

which is then used to parametrize a Gaussian mixture distribution $P_{\theta^{(m)}}$. Finally, using the Monte Carlo method described in Sec. 2, we then estimated mutual information $\mathbf{I}(P_{\theta^{(m)}})$ and the PMI profile.

Table 2: BMM-provided estimates for selected low-dimensional problems. Each run corresponds to estimation on a different sample and reports the mean together with a credibility interval obtained from the 10th and 90th percentile. Bold font represents problems solved by BMMs.

	True MI	Run 1	Run 2	Run 3
Additive	0.33	0.31 (0.29–0.34)	0.29 (0.26–0.31)	0.31 (0.29–0.34)
X	0.41	0.42 (0.38–0.45)	0.41 (0.38–0.45)	0.41 (0.38–0.44)
AI	0.78	0.78 (0.74–0.82)	0.76 (0.73–0.80)	0.78 (0.75–0.82)
Galaxy	0.49	0.27 (0.24–0.29)	0.29 (0.26–0.31)	0.30 (0.27–0.32)
Concentric (3-dim, 5)	0.46	0.46 (0.43–0.48)	0.46 (0.43–0.49)	0.45 (0.42–0.48)
Concentric (5-dim, 5)	0.63	0.64 (0.61–0.68)	0.63 (0.60–0.66)	0.63 (0.60–0.66)
Concentric (3-dim, 10)	0.56	0.56 (0.53–0.59)	0.57 (0.53–0.60)	0.46 (0.43–0.49)
Concentric (5-dim, 10)	0.75	0.63 (0.60–0.66)	0.62 (0.58–0.65)	0.56 (0.53–0.59)
Inliers (5-dim, 0.2)	0.63	0.69 (0.63–0.74)	0.66 (0.61–0.71)	0.64 (0.59–0.69)
Inliers (5-dim, 0.5)	0.27	0.18 (0.15–0.21)	0.29 (0.26–0.33)	0.19 (0.16–0.21)
Normal (5-dim, dense)	0.59	0.58 (0.54–0.62)	0.61 (0.57–0.65)	0.59 (0.56–0.63)
Normal (5-dim, sparse)	1.02	1.02 (0.97–1.07)	1.04 (0.99–1.09)	1.03 (0.98–1.08)
Student (1-dim)	0.22	0.24 (0.21–0.27)	0.23 (0.20–0.26)	0.22 (0.19–0.25)
Student (2-dim)	0.43	0.43 (0.39–0.47)	0.41 (0.37–0.44)	0.40 (0.37–0.44)
Student (3-dim)	0.29	0.19 (0.16–0.22)	0.20 (0.17–0.23)	0.22 (0.19–0.25)
Student (5-dim)	0.45	0.26 (0.22–0.31)	0.22 (0.19–0.25)	0.26 (0.22–0.30)

C.5.2 BMMs performance on the proposed benchmark

We evaluated the performance of the BMM described above on a subset of distributions proposed in Appendix C.2. We excluded the high-dimensional distributions due to the high computational cost of fitting Markov chains: in our preliminary experiments used to select sampling hyperparameters, the chains often failed to sample properly the posterior distribution. For lower-dimensional problems we did not notice convergence problems in our preliminary runs (apart from the label-switching issues, which do not affect the predictive distribution). To reduce the computational cost, we decided to use three (rather than ten) samples from each distribution. Then, for each sample we ran a single Markov chain with 1,000 warm-up steps and 1,000 collected samples. As we ran a single chain and the model is not identifiable due to the label switching of Z_n variables, we could not employ metrics diagnosing convergence failures, such as $\hat{R}$ (Vehtari et al., 2021). We then used 1,000 Monte Carlo samples to estimate the MI in each distribution.

In Table 2 we report the MI estimates across three runs. Each estimate is a summary of the posterior distribution, specifying the mean and the 10th and 90th percentile of the distribution.

Overall, we see that BMMs offer good performance and appropriate uncertainty quantification in low-dimensional problems, such as X, AI, normal and low-dimensional concentric distributions. Interestingly, even when the model is slightly misspecified (additive and low-dimensional Student distributions with tail-shortening transformation applied), we obtain good performance.

It is however important to note that the proposed approach is not universal: for a 5-dimensional concentric distribution and 10 components and 5-dimensional problem with 50% of inliers, the model, even though it is well-specified, does not always estimate the mutual information appropriately, underestimating the mutual information. We suppose that posterior predictive checking and sensitivity analysis can be used to diagnose problems with this model. We investigate this issue in more detail in Appendix C.5.3.

C.5.3 Posterior predictive checking and estimation of pointwise mutual information profiles

As discussed in Section 3.4, BMMs allow one to do model-based estimation of both MI and the PMI profile. At the same time, they require domain expertise to propose a generative model as well as careful model criticism. In this section we further investigate these aspects, by applying the model to the X, AI, Waves and Galaxy distributions, changing the number of data points $N \in \{125, 250, 500, 1000\}$.

For each distribution we sampled N data points once, ran a single Markov chain with 2000 warm-up steps and collected 800 samples. Then, for each sample we estimated MI and the profile via the Monte Carlo approximation using 100,000 samples. We visualise the observed sample, a single posterior predictive sample and posterior on mutual information and the PMI profile in Fig. 10, Fig. 11, Fig. 12 and Fig. 13.

Although the model performance is good for the X and AI distributions, which additionally supports the results obtained in Appendix C.5.3, we see that model misspecification results in unreliable estimates for the Waves and Galaxy distributions.

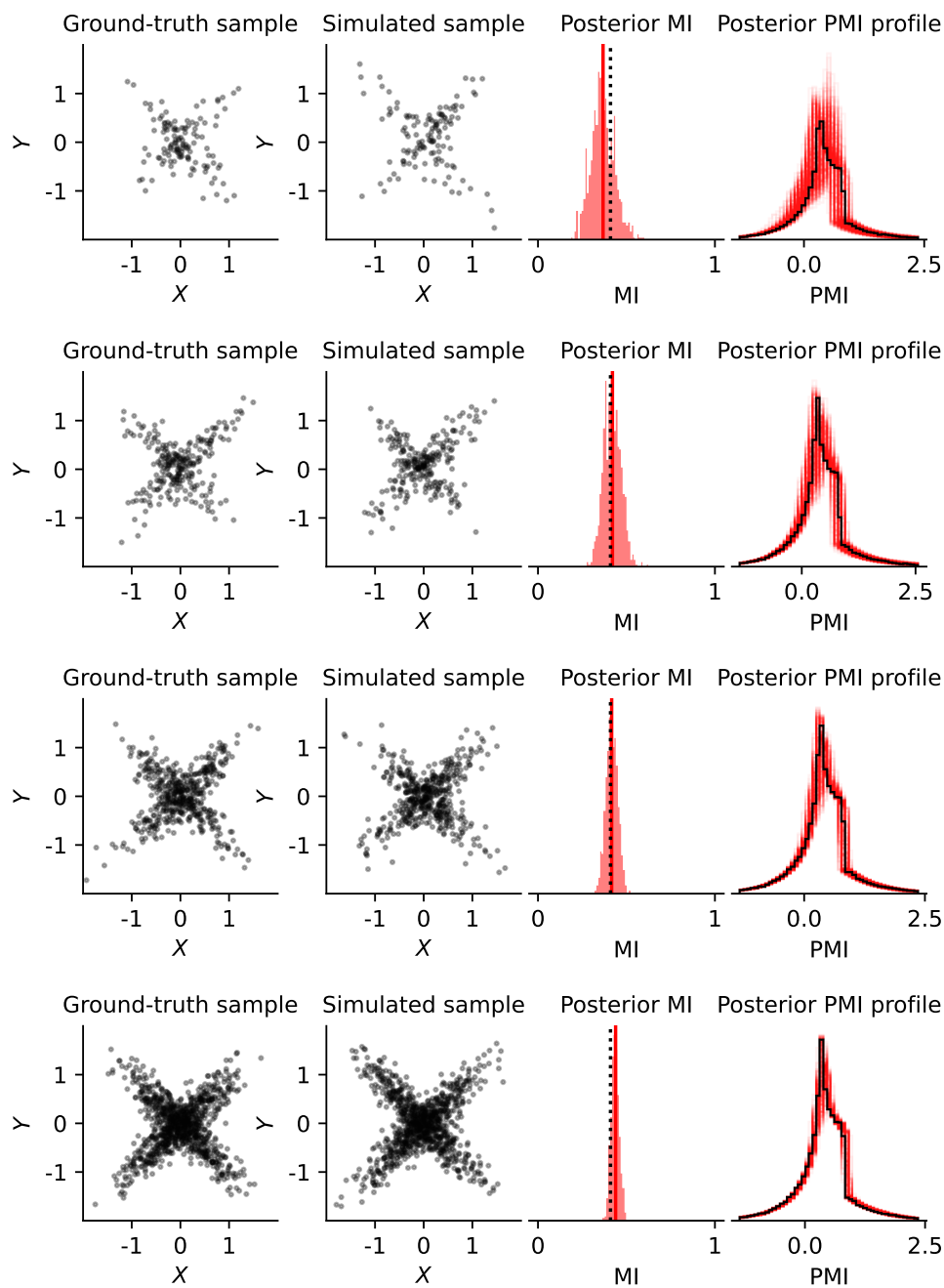


Figure 10: Gaussian mixture model fitted to the X distribution with 125, 250, 500 and 1000 samples.

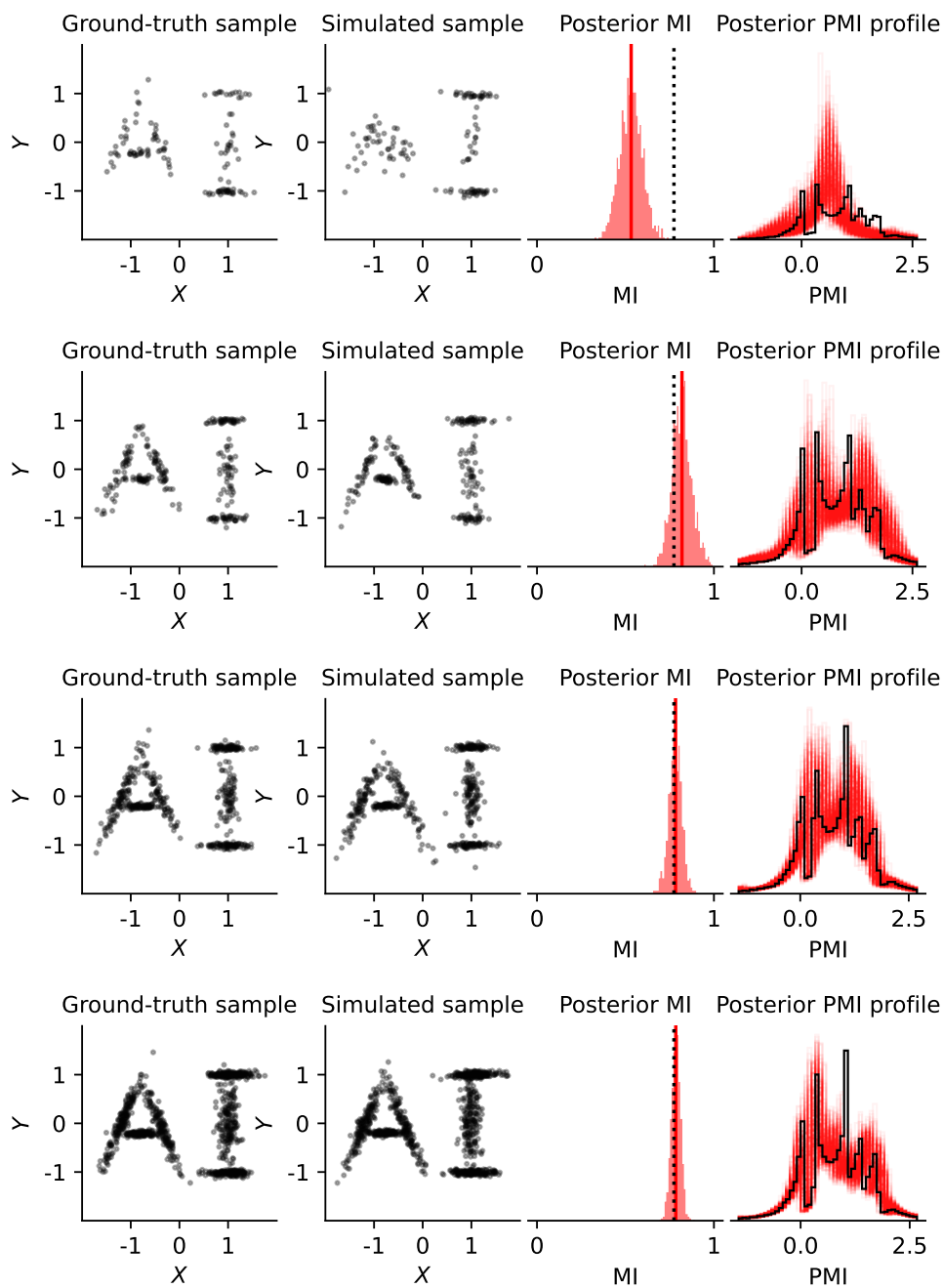


Figure 11: Gaussian mixture model fitted to the AI distribution with 125, 250, 500 and 1000 samples.

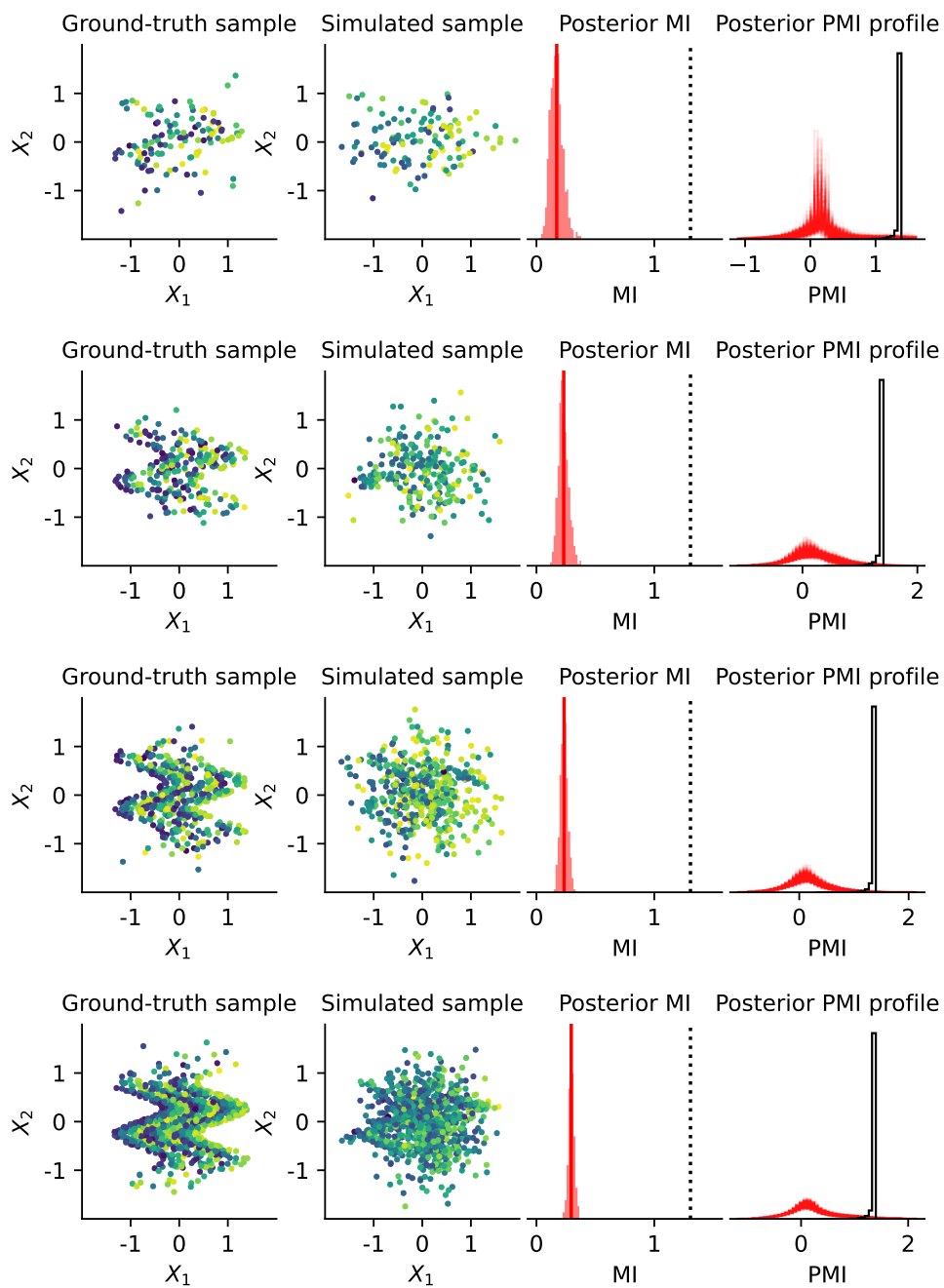


Figure 12: Gaussian mixture model fitted to the Waves distribution with 125, 250, 500 and 1000 samples.

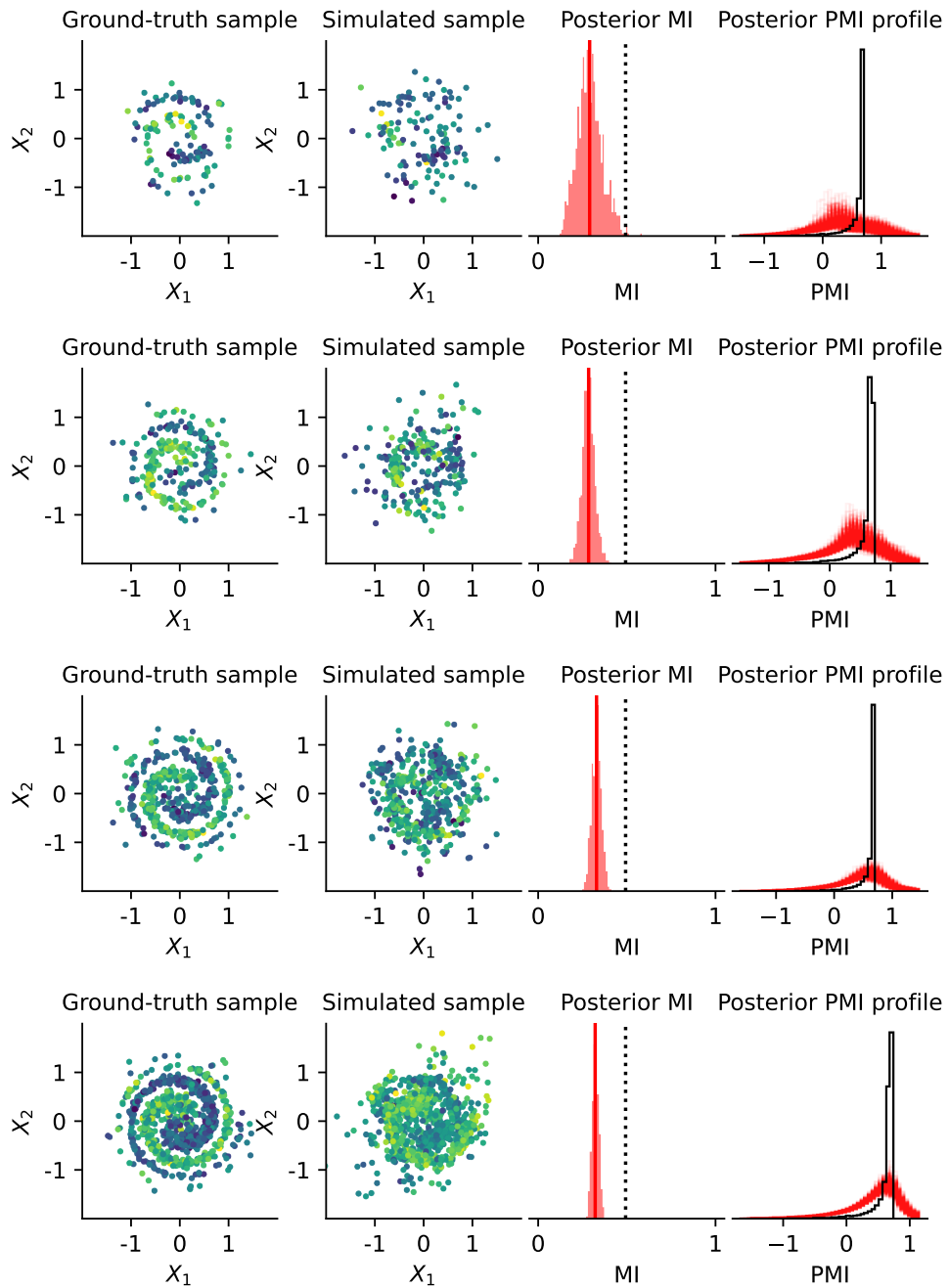


Figure 13: Gaussian mixture model fitted to the Galaxy distribution with 125, 250, 500 and 1000 samples.

Author contributions

In article **Antagonism** ★, Antagonism between viral infection and innate immunity at the single-cell level, published in *PLOS Pathogens*, 19(9), e1011597,

Frederic Grabowski (co-first author):

- designed the mathematical model,
- curated the data,
- wrote the simulator,
- managed massively parallel simulations,
- analysed data from experiments and simulations,
- wrote the manuscript, prepared main text Figures 2–7 and all supplementary materials;

Marek Kochańczyk (co-first author):

- curated the data,
- wrote the simulator,
- wrote the manuscript and prepared main text Figure 1;

Zbigniew Korwek:

- designed and performed biological experiments,
- revised the manuscript;

Maciej Czerkies:

- designed and performed biological experiments,
- revised the manuscript;

Wiktor Prus:

- designed and performed biological experiments,
- revised the manuscript;

Tomasz Lipniacki (corresponding author):

- conceived and supervised the study,
- analyzed and interpreted the results,
- wrote the manuscript.


Frederic Grabowski


Marek Kochańczyk


Zbigniew Korwek


Maciej Czerkies


Wiktor Prus


Tomasz Lipniacki

In article **Lambda** ★, Type III interferons suppress influenza A virus infection independently of STAT activation by triggering cell death, preprint available on *bioRxiv*, 2024–09,

Wiktor Prus:

- designed and performed biological experiments,
- wrote the Manuscript;

Frederic Grabowski:

- curated the data,
- performed the SIR model analysis,
- performed the statistical analysis,
- wrote the manuscript, prepared all main text figures and all supplementary materials;

Paulina Koza:

- designed and performed biological experiments,
- revised the manuscript;

Zbigniew Korwek:

- designed and performed biological experiments,
- wrote the manuscript;

Maciej Czerkies:

- designed and performed biological experiments,
- wrote the manuscript;

Marek Kochończyk:

- co-supervised the study,
- curated the data,
- wrote the manuscript;

Tomasz Lipniacki (corresponding author):

- conceived and co-supervised the study,
- analyzed and interpreted the results,
- wrote the manuscript.


Wiktor Prus


Frederic Grabowski

Signature provided
on another page.
Paulina Koza


Zbigniew Korwek


Maciej Czerkies


Marek Kochończyk


Tomasz Lipniacki

In article **Lambda** ★, Type III interferons suppress influenza A virus infection independently of STAT activation by triggering cell death, preprint available on *bioRxiv*, 2024–09,

Wiktor Prus:

- designed and performed biological experiments,
- wrote the Manuscript;

Frederic Grabowski:

- curated the data,
- performed the SIR model analysis,
- performed the statistical analysis,
- wrote the manuscript, prepared all main text figures and all supplementary materials;

Paulina Koza:

- designed and performed biological experiments,
- revised the manuscript;

Zbigniew Korwek:

- designed and performed biological experiments,
- wrote the manuscript;

Maciej Czerkies:

- designed and performed biological experiments,
- wrote the manuscript;

Marek Kocharczyk:

- co-supervised the study,
- curated the data,
- wrote the manuscript;

Tomasz Lipniacki (corresponding author):

- conceived and co-supervised the study,
- analyzed and interpreted the results,
- wrote the manuscript.

Signature provided
on another page.

Wiktor Prus

Signature provided
on another page.

Frederic Grabowski

Pauline Koza

Paulina Koza

Signature provided
on another page.

Zbigniew Korwek

Signature provided
on another page.

Maciej Czerkies

Signature provided
on another page.

Marek Kocharczyk

Signature provided
on another page.

Tomasz Lipniacki

In article **Dynamics** ★, Dynamics of COVID-19 pandemic at constant and time-dependent contact rates, published in *Mathematical Modelling of Natural Phenomena*, 15, 28,

Marek Kochańczyk:

- conceived the study,
- designed and implemented the mathematical model,
- curated the data,
- performed and analysed model simulations,
- wrote the manuscript and prepared Figures 1, 3;

Frederic Grabowski:

- conceived the study,
- designed and implemented the mathematical model,
- performed and analysed model simulations,
- wrote the manuscript and prepared Figures 2, 4, 5, 6;

Tomasz Lipniacki (corresponding author):

- conceived and supervised the study,
- analyzed and interpreted the results,
- wrote the manuscript.



Marek Kochańczyk



Frederic Grabowski



Tomasz Lipniacki

In article **Super-spreaders** ★, Super-spreading events initiated the exponential growth phase of COVID-19 with R_0 higher than initially estimated, published in *Royal Society Open Science*, 7(9), 200786,

Marek Kochańczyk:

- conceived the study,
- designed the mathematical model,
- curated the data,
- performed and analysed model simulations,
- wrote the Manuscript and prepared Figures 1, 3;

Frederic Grabowski:

- conceived the study,
- designed and implemented the mathematical model,
- performed and analysed model simulations,
- wrote the Manuscript and prepared Figure 2;

Tomasz Lipniacki (corresponding author):

- conceived and supervised the study,
- analyzed and interpreted the results,
- wrote the Manuscript.



Marek Kochańczyk



Frederic Grabowski



Tomasz Lipniacki

In article **Alpha** ★, SARS-CoV-2 variant of concern 202012/01 has about twofold replicative advantage and acquires concerning mutations, published in *Viruses*, 13(3), 392,

Frederic Grabowski (co-first author):

- conceived the study
- curated the data,
- analysed the case data and emerging substrains,
- prepared Figures 1–4, 6, 7, Table 1 and Supplementary Materials,
- wrote the Manuscript;

Grzegorz Preibisch:

- curated the data,
- performed sequence analysis,
- prepared Figure 5;

Stanisław Giziński:

- curated the data,
- performed sequence analysis,
- prepared Figure 5;

Marek Kochańczyk (co-first author, corresponding author):

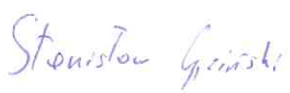
- conceived and co-supervised the study,
- curated the data,
- analyzed the sequence data,
- wrote the Manuscript

Tomasz Lipniacki (corresponding author):

- conceived and co-supervised the study,
- analyzed and interpreted the results,
- wrote the Manuscript.


Frederic Grabowski


Grzegorz Preibisch


Stanisław Giziński


Marek Kochańczyk


Tomasz Lipniacki

In article **Omicron** ★, The spread of SARS-CoV-2 variant Omicron with a doubling time of 2.0–3.3 days can be explained by immune evasion, published in *Viruses*, 14(2), 294,

Frederic Grabowski:

- conceived the study,
- curated the data,
- analysed the case and mobility data,
- performed the Bayesian analysis of Omicron emergence,
- wrote the Manuscript, prepared Figures 1–3 and Supplementary Materials;

Marek Kochańczyk:

- conceived and co-supervised the study,
- performed the two-strain model analysis,
- revised the Manuscript and prepared Figure 4

Tomasz Lipniacki (corresponding author):

- conceived and co-supervised the study,
- analyzed and interpreted the results,
- wrote the Manuscript.


Frederic Grabowski


Marek Kochańczyk


Tomasz Lipniacki

In article **Limits** ★, Limits to the rate of information transmission through the MAPK pathway, published in *Journal of the Royal Society Interface*, 16(152), 20180792,

Frederic Grabowski (co-first author):

- prepared the computational model,
- performed numerical simulations,
- analysed model simulations,
- designed the capacity optimization algorithm
- derived the theoretical approximation,
- wrote the Manuscript and prepared Figures 1, 2, 4;

Paweł Czyż (co-first author):

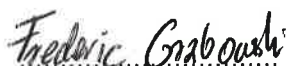
- prepared the computational model,
- prepared the massively parallel simulations,
- performed numerical simulations,
- analysed model simulations,
- designed the capacity optimization algorithm
- revised the Manuscript and prepared the Figures 1, 3, 4;

Marek Kočańczyk:

- supervised the preparation of the computational model,
- co-supervised the study,
- wrote the Manuscript;

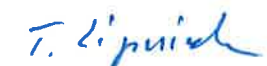
Tomasz Lipniacki (corresponding author):

- conceived and co-supervised the study,
- analyzed and interpreted the results,
- wrote the Manuscript.


Frederic Grabowski


Paweł Czyż


Marek Kočańczyk


Tomasz Lipniacki

In article **Predictive ★**, Predictive power of non-identifiable models, published in *Scientific Reports*, 13(1), 11143,

Frederic Grabowski:

- conceived the study,
- prepared the computational model,
- performed the numerical simulations,
- analysed model simulations (including sloppiness analysis),
- analyzed and interpreted the results,
- wrote the Manuscript, prepared the Figures;

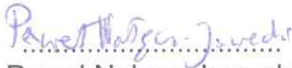
Paweł Nałęcz-Jawecki:

- conceived the study,
- analyzed the results,
- revised the Manuscript;

Tomasz Lipniacki (corresponding author):

- conceived and supervised the study,
- analyzed and interpreted the results,
- wrote the Manuscript.


Frederic Grabowski


Paweł Nałęcz-Jawecki


Tomasz Lipniacki

In article **Beyond ★**, Beyond normal: On the evaluation of mutual information estimators, published in *Advances in Neural Information Processing Systems*, 36,

Paweł Czyż (co-first author):

- conceived the study,
- wrote the software,
- designed the benchmark,
- performed the numerical computations,
- analysed benchmarking results,
- wrote the Manuscript and Appendix, prepared the Figures;

Frederic Grabowski (co-first author):

- conceived the study,
- wrote the software,
- designed the benchmark,
- performed the numerical computations,
- analysed benchmarking results,
- wrote the Manuscript, prepared the Figures;

Julia E. Vogt:

- revised the Manuscript;

Niko Beerenwinkel (co-supervision):

- conceived the study,
- revised the Manuscript;

Alexander Marx (co-supervision):

- conceived the study,
- designed the benchmark,
- analysed benchmarking results,
- wrote the Manuscript.



Paweł Czyż

Signature provided
on another page.

Niko Beerenwinkel



Frederic Grabowski

Signature provided
on another page.

Alexander Marx

Signature provided
on another page.

Julia E. Vogt

In article **Beyond ★**, Beyond normal: On the evaluation of mutual information estimators, published in *Advances in Neural Information Processing Systems*, 36,

Paweł Czyż (co-first author):

- conceived the study,
- wrote the software,
- designed the benchmark,
- performed the numerical computations,
- analysed benchmarking results,
- wrote the Manuscript and Appendix, prepared the Figures;

Frederic Grabowski (co-first author):

- conceived the study,
- wrote the software,
- designed the benchmark,
- performed the numerical computations,
- analysed benchmarking results,
- wrote the Manuscript, prepared the Figures;

Julia E. Vogt:

- revised the Manuscript;

Niko Beerenwinkel (co-supervision):

- conceived the study,
- revised the Manuscript;

Alexander Marx (co-supervision):

- conceived the study,
- designed the benchmark,
- analysed benchmarking results,
- wrote the Manuscript.

Signature provided
on another page.

Paweł Czyż

Signature provided
on another page.

Frederic Grabowski

Signature provided
on another page.

Niko Beerenwinkel

Signature provided
on another page.

Alexander Marx



Julia E. Vogt

In article **Beyond ★**, Beyond normal: On the evaluation of mutual information estimators, published in *Advances in Neural Information Processing Systems*, 36,

Paweł Czyż (co-first author):

- conceived the study,
- wrote the software,
- designed the benchmark,
- performed the numerical computations,
- analysed benchmarking results,
- wrote the Manuscript and Appendix, prepared the Figures;

Frederic Grabowski (co-first author):

- conceived the study,
- wrote the software,
- designed the benchmark,
- performed the numerical computations,
- analysed benchmarking results,
- wrote the Manuscript, prepared the Figures;

Julia E. Vogt:

- revised the Manuscript;

Niko Beerenwinkel (co-supervision):

- conceived the study,
- revised the Manuscript;

Alexander Marx (co-supervision):

- conceived the study,
- designed the benchmark,
- analysed benchmarking results,
- wrote the Manuscript.

Signature provided
on another page.

Paweł Czyż



Niko Beerenwinkel

Signature provided
on another page.

Frederic Grabowski

Signature provided
on another page.

Alexander Marx

Signature provided
on another page.

Julia E. Vogt

In article **Beyond ★**, Beyond normal: On the evaluation of mutual information estimators, published in *Advances in Neural Information Processing Systems*, 36,

Paweł Czyż (co-first author):

- conceived the study,
- wrote the software,
- designed the benchmark,
- performed the numerical computations,
- analysed benchmarking results,
- wrote the Manuscript and Appendix, prepared the Figures;

Frederic Grabowski (co-first author):

- conceived the study,
- wrote the software,
- designed the benchmark,
- performed the numerical computations,
- analysed benchmarking results,
- wrote the Manuscript, prepared the Figures;

Julia E. Vogt:

- revised the Manuscript;

Niko Beerenwinkel (co-supervision):

- conceived the study,
- revised the Manuscript;

Alexander Marx (co-supervision):

- conceived the study,
- designed the benchmark,
- analysed benchmarking results,
- wrote the Manuscript.

Signature provided
on another page.

Paweł Czyż

Signature provided
on another page.

Frederic Grabowski

Signature provided
on another page.

Julia E. Vogt

Signature provided
on another page.

Niko Beerenwinkel


Alexander Marx

In article **Mixtures** ★, On the Properties and Estimation of Pointwise Mutual Information Profiles, published in *Transactions on Machine Learning Research* (2025),

Paweł Czyż (co-first author):

- conceived the study,
- wrote the software,
- designed the new benchmarking tasks,
- performed the numerical computations,
- analysed benchmarking results,
- proposed the use of mixtures to extend the benchmark,
- proposed the use of BMMs as an Bayesian estimator,
- derived the measure-theoretical proofs and generalized profiles,
- wrote the Manuscript, prepared the Figures;

Frederic Grabowski (co-first author):

- conceived the study,
- wrote the software,
- designed the new benchmarking tasks,
- performed the numerical computations,
- analysed benchmarking results,
- proposed the PMI profile as an invariant and tool for studying MI estimators,
- proposed the use of BMMs for inlier and outlier noise and continuous-discrete tasks,
- derived the analytical formula of PMI profiles for multivariate normal distributions,
- wrote the Manuscript, prepared the Figures;

Julia E. Vogt:

- revised the Manuscript;

Niko Beerenwinkel (co-supervision):

- conceived the study,
- revised the Manuscript;

Alexander Marx (co-supervision):

- conceived the study,
- analysed benchmarking results,
- wrote the Manuscript.



Paweł Czyż

Signature provided
on another page.

Niko Beerenwinkel



Frederic Grabowski

Signature provided
on another page.

Alexander Marx

Signature provided
on another page.

Julia E. Vogt

In article **Mixtures** ★, On the Properties and Estimation of Pointwise Mutual Information Profiles, published in *Transactions on Machine Learning Research* (2025),

Paweł Czyż (co-first author):

- conceived the study,
- wrote the software,
- designed the new benchmarking tasks,
- performed the numerical computations,
- analysed benchmarking results,
- proposed the use of mixtures to extend the benchmark,
- proposed the use of BMMs as an Bayesian estimator,
- derived the measure-theoretical proofs and generalized profiles,
- wrote the Manuscript, prepared the Figures;

Frederic Grabowski (co-first author):

- conceived the study,
- wrote the software,
- designed the new benchmarking tasks,
- performed the numerical computations,
- analysed benchmarking results,
- proposed the PMI profile as an invariant and tool for studying MI estimators,
- proposed the use of BMMs for inlier and outlier noise and continuous-discrete tasks,
- derived the analytical formula of PMI profiles for multivariate normal distributions,
- wrote the Manuscript, prepared the Figures;

Julia E. Vogt:

- revised the Manuscript;

Niko Beerenwinkel (co-supervision):

- conceived the study,
- revised the Manuscript;

Alexander Marx (co-supervision):

- conceived the study,
- analysed benchmarking results,
- wrote the Manuscript.

Signature provided
on another page.

Paweł Czyż

Signature provided
on another page.

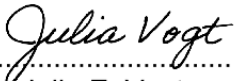
Frederic Grabowski

Signature provided
on another page.

Niko Beerenwinkel

Signature provided
on another page.

Alexander Marx


Julia E. Vogt

In article **Mixtures** ★, On the Properties and Estimation of Pointwise Mutual Information Profiles, published in *Transactions on Machine Learning Research* (2025),

Paweł Czyż (co-first author):

- conceived the study,
- wrote the software,
- designed the new benchmarking tasks,
- performed the numerical computations,
- analysed benchmarking results,
- proposed the use of mixtures to extend the benchmark,
- proposed the use of BMMs as an Bayesian estimator,
- derived the measure-theoretical proofs and generalized profiles,
- wrote the Manuscript, prepared the Figures;

Frederic Grabowski (co-first author):

- conceived the study,
- wrote the software,
- designed the new benchmarking tasks,
- performed the numerical computations,
- analysed benchmarking results,
- proposed the PMI profile as an invariant and tool for studying MI estimators,
- proposed the use of BMMs for inlier and outlier noise and continuous-discrete tasks,
- derived the analytical formula of PMI profiles for multivariate normal distributions,
- wrote the Manuscript, prepared the Figures;

Julia E. Vogt:

- revised the Manuscript;

Niko Beerenwinkel (co-supervision):

- conceived the study,
- revised the Manuscript;

Alexander Marx (co-supervision):

- conceived the study,
- analysed benchmarking results,
- wrote the Manuscript.

Signature provided
on another page.

Paweł Czyż



Niko Beerenwinkel

Signature provided
on another page.

Frederic Grabowski

Signature provided
on another page.

Alexander Marx

Signature provided
on another page.

Julia E. Vogt

In article **Mixtures** ★, On the Properties and Estimation of Pointwise Mutual Information Profiles, published in *Transactions on Machine Learning Research* (2025),

Paweł Czyż (co-first author):

- conceived the study,
- wrote the software,
- designed the new benchmarking tasks,
- performed the numerical computations,
- analysed benchmarking results,
- proposed the use of mixtures to extend the benchmark,
- proposed the use of BMMs as an Bayesian estimator,
- derived the measure-theoretical proofs and generalized profiles,
- wrote the Manuscript, prepared the Figures;

Frederic Grabowski (co-first author):

- conceived the study,
- wrote the software,
- designed the new benchmarking tasks,
- performed the numerical computations,
- analysed benchmarking results,
- proposed the PMI profile as an invariant and tool for studying MI estimators,
- proposed the use of BMMs for inlier and outlier noise and continuous-discrete tasks,
- derived the analytical formula of PMI profiles for multivariate normal distributions,
- wrote the Manuscript, prepared the Figures;

Julia E. Vogt:

- revised the Manuscript;

Niko Beerenwinkel (co-supervision):

- conceived the study,
- revised the Manuscript;

Alexander Marx (co-supervision):

- conceived the study,
- analysed benchmarking results,
- wrote the Manuscript.

Signature provided
on another page.

Paweł Czyż

Signature provided
on another page.

Frederic Grabowski

Signature provided
on another page.

Julia E. Vogt

Signature provided
on another page.

Niko Beerenwinkel


Alexander Marx