POLISH ACADEMY OF SCIENCES
Institute of Fundamental Technological Research

Doctoral Thesis

PARAMETRIC AND NON-PARAMETRIC METHODS TO ADDRESS COMPLEXITY OF CELLULAR SIGNALING PATHWAYS

KAROL NIENAŁTOWSKI

A Dissertation supervised by
dr hab. **Michał Komorowski**

Warsaw, August 2021

Dedicated to
my father
R.I.P.

CONTENTS

LIST OF PUBLICATIONS INCLUDED IN THE THESIS

Article (1) **Nienałtowski K**, Jetka T, Komorowski M, Chapter 14. Sensitivity Analysis. In: Munsky, B., Tsimring, L. S. & Hlavacek, W. S. (eds.). *Quantitative Biology: Theory, Computational Methods, and Models*. The MIT Press, Cambridge, MA, 2018, pp. 293-320.

Article (2) **Nienałtowski, K.**, Włodarczyk, M., Lipniacki, T., & Komorowski, M. (2015). Clustering reveals limits of parameter identifiability in multi-parameter models of biochemical dynamics. *BMC systems biology*, 9(1), 1-9.
https://dx.doi.org/10.1186%2Fs12918-015-0205-8

Article (3) **Nienałtowski, K.**, Rigby, R. E., Walczak, J., Zakrzewska, K. E., Głów, E., Rehwinkel, J., & Komorowski, M. (2021). Fractional response analysis reveals logarithmic cytokine responses in cellular populations. *Nature Communications*, 12(1), 1-10.
https://doi.org/10.1038/s41467-021-24449-2

OTHER RESEARCH ARTICLES NOT DESCRIBED IN THE THESIS

1. Vahdat, Z., **Nienałtowski, K.**, Farooq, Z., Komorowski, M., & Singh, A. (2020, May). Information processing in unregulated and autoregulated gene expression. In 2020 European Control Conference (ECC) (pp. 258-263). IEEE.
https://ieeexplore.ieee.org/abstract/document/9143689

2. Jetka, T., Winarski, T., **Nienałtowski, K.**, Błoński, S. & Komorowski, M. Information-theoretic analysis of multivariate single-cell signalling responses using SLEMI. *PLOS Computational Biology*. 15(7), e1007132 (2019).
https://doi.org/10.1371/journal.pcbi.1007132

3. Jetka, T., **Nienałtowski, K.**, Filippi, S., Stumpf, M. & Komorowski, M. An information-theoretic framework for deciphering pleiotropic and noisy biochemical signalling. *Nature Communications*. 9(1), 4591 (2018).
https://doi.org/10.1038/s41467-018-07085-1

4. Jetka T, **Nienałtowski K**, Komorowski M, Chapter 15. Experimental Design. In: Munsky, B., Tsimring, L. S. & Hlavacek, W. S. (eds.). *Quantitative Biology: Theory, Computational Methods, and Models*. The MIT Press, Cambridge, MA, 2018, pp. 321-334.

5. Wronowska, W., Charzyńska, A., **Nienałtowski, K.**, & Gambin, A. (2015). Computational modeling of sphingolipid metabolism. *BMC systems biology*, 9(1), 1-16.
https://link.springer.com/article/10.1186/s12918-015-0176-9

LIST OF SOFTWARE DEVELOPED FOR THE THESIS

(1) FRA – R package for fractional response analysis, https://github.com/sysbiosig/FRA.

(2) ClusteringIdentifiability – R package for analysis of parameters similarity and identifiability in models of signaling pathways, https://github.com/stork119/ClusteringIdnetifiability.

SUMMARY

Doctoral thesis: Karol Nienałtowski, ***Parametric and non-parametric methods to address complexity of cellular signaling pathways***, 2021.

All processes in the organisms, including regulation of development, immunity, tissue homeostasis, and repair, require coordination between cells. This coordination occurs through cell-to-cell communication using particular signaling molecules, including cytokines, hormones, and other ligands. Briefly, cells synthesize and release signaling molecules, which then bind to specific membrane receptors. Such activated receptors trigger a series of intracellular reactions, which form a signaling pathway. The signal is eventually transferred to the nucleus, leading to the activation of cellular effectors, e.g., the creation of response genes.

However, many components of signaling pathways are functionally pleiotropic, and signaling responses are marked with substantial cell-to-cell heterogeneity. For instance, a distinct signaling molecule often activates multiple different effectors, and a particular effector can be triggered by various signaling molecules, which results in cross-wired signaling. Moreover, biochemical signaling processes are intrinsically stochastic, and the responses of apparently identical cells can vary significantly.

Cell-to-cell variability of signaling effectors, signaling dynamics, and cross-wiring structure of pathways all give rise to signaling complexity. Therefore, studies of cellular signaling require comprehensive quantitative support. There are two general approaches to these obstacles, i.e., parametric and non-parametric modeling.

The first approach intends to explore in detail reactions of signaling pathways. Here, the signaling pathway is represented as a dynamical model, which describes mathematically changes in time of the reactants state, e.g., concentration of proteins. The model's dynamics, properties, and interpretation of results strictly depend on its parameters. Methods to understand the relationship between parameters and model behavior is of particular interest in the context of biochemical dynamics and related phenomena. These methods include sensitivity and identifiability analysis. The sensitivity analysis provides valuable insights into how robust the model is to changes of biological parameters and which model inputs are the key factors that affect the model outputs. On the other hand, the identifiability analysis determines parameter estimation accuracy and the model's predictions. Both techniques have proven their importance for utilizing modeling in physics and engineering. However, models of cellular signaling differ from conventional models in several ways. They are significantly more complex, consist of a higher number of parameters, and the accuracy of their estimations is limited due to the quality or the lack of experimental data.

The second, non-parametric approach neglect any assumptions regarding the structure or character of intracellular processes. The relation between cellular response (output) to an external signal (input) is represented in a

probabilistic manner. This relationship can be explained within the Shannon information theory in terms of communication channels. In such an approach, the information flow via the input-output channel is analyzed. However, conventional communication channels differ considerably from signaling pathways. The first ones have limited noise, operate in optimal or defined conditions, and transfer long sequences of discreet symbols.
In my Ph.D., I developed parametric and non-parametric methods that can deal with the complexity and stochasticity of the cellular signaling pathways. In the parametric approach, I focused on developing techniques that explain relations between parameters and model dynamics. Firstly, in the chapter *"Sensitivity Analysis in quantitative biology models"* [1] of the book *"Quantitative Biology: Theory, Computational Methods and Examples of Models"* [9], we have identified currently used sensitivity analysis methods to describe cellular processes on underlying parameters. We have presented specific tools that analyze parameter dependence locally and globally in deterministic and stochastic regimes. Moreover, we have shown on illustrative examples that sensitivity analysis leads to a deeper understanding of studied phenomena and helps in manipulating complex models. Secondly, in the article [2], we have proposed methods for dissecting the impact of individual parameters on the behavior of complex multi-parameter models. Precisely, we have introduced an information-theory measure of similarity between groups of parameters. Based on this measure, we have established a novel definition of identifiable parameters. Moreover, we have developed a tool, which uses hierarchical clustering for finding a maximal possible number of identifiable parameters. We have shown that our methodology helps design biological experiments, which can improve models' accuracy.
Moreover, in the article [3], we have developed non-parametric methods to introduce fractional response analysis (FRA), which quantifies changes in fractions of cells with given response levels. We have demonstrated on several biological examples that outcomes of physiological processes depend on the number of cells with specific responses rather than on mean or median. Proposed fractional response analysis can be universally performed for heterogeneous, multivariate, and dynamic measurements and quantifies otherwise hidden patterns in single-cell data. Furthermore, we have shown that FRA has a rigorous mathematical definition in terms of Rényi min-information; hence, FRA estimates the overall signaling fidelity. However, compared to the conventional Shannon information, properties of Rényi min-information are more appropriate to describe information flow in biochemical systems.
Methods developed within this thesis have been implemented within freely available software packages: ClusteringIdentifiability and FRA.

SUMMARY (POLISH)

Praca doktorska: Karol Nienałtowski, ***Parametryczne i nieparametryczne metody analizy biochemicznych szlaków sygnałowych***, 2021.

Wszystkie procesy wewnątrz żywych organizmów wymagają koordynacji działań między komórkami. Koordynacja odbywa się poprzez komunikację pomiędzy komórkami za pomocą specjalnych cząsteczek sygnałowych, takich jak: cytokiny, hormony i inne ligandy. W dużym uproszczeniu cząsteczki sygnałowe są produkowane i wypuszczane przez komórkę do przestrzeni międzykomórkowej. Następnie cząsteczka sygnałowa wiąże się ze specyficznymi receptorami znajdującymi się na błonach komórkowych. W wyniku połączenia receptor zmienia swoją konformację, wywołując szereg reakcji wewnątrzkomórkowych, które tworzą tzw. ścieżkę sygnałową. Sygnał ostatecznie jest przenoszony do jądra komórkowego, gdzie aktywowana jest odpowiedź komórkowa na sygnał, np. ekspresja odpowiednich białek efektorowych.

Należy jednak zauważyć, że wiele elementów szlaków sygnałowych jest funkcjonalnie plejotropowych, a odpowiedź komórkowa na sygnał cechuje się znaczną heterogenicznością pomiędzy komórkami. Na przykład cząsteczka sygnałowa często aktywuje wiele różnych efektorów, a konkretny efektor może być aktywowany przez różne cząsteczki sygnałowe, co skutkuje sygnałowniem krzyżowym z innymi szlakami sygnałowymi. Ponadto reakcje molekularne są z natury stochastyczne co powoduje, że odpowiedź komórkowa pozornie identycznych komórek może znacznie się między sobą różnić.

Heterogeniczność odpowiedzi komórkowej na sygnał, dynamika sygnałowania oraz sygnałowanie krzyżowe z innymi szlakami prowadzą łącznie do znacznej złożoności procesów sygnałowania. Dlatego analiza sygnałowania w komórkach wymaga wsparcia przy użyciu metod ilościowych. Istnieją dwa ogólne podejścia do tego problemu, tj. modelowanie parametryczne i nieparametryczne.

Pierwsze podejście ma na celu szczegółowe zbadanie reakcji tworzących ścieżki sygnałowe. W tym podejściu ścieżkę sygnałową reprezentuje się w postaci modelu matematycznego, który opisuje zmiany stężeń cząsteczek w czasie, w wyniku reakcji biochemicznych. Dynamika modelu, jego właściwości i interpretacja wyników ściśle zależą od parametrów modelu. Metody pozwalające zrozumieć związek między parametrami a zachowaniem modelu są szczególnie interesujące w kontekście szlaków sygnałowych. Metody te obejmują między innymi analizę wrażliwości (*ang. sensitivity analysis*) i identyfikowalności (*ang. identifiability analysis*) parametrów. Analiza wrażliwości dostarcza cennych informacji na temat zmienności dynamiki modelu wynikających ze zmiany wartości parametrów oraz wskazuje, które parametry najbardziej wpływają na zachowanie modelu. Natomiast analiza identyfikowalności określa dokładność estymacji parametrów oraz wskazuje stopień poprawności predykcji modelu. Obie techniki z powodzeniem są wykorzysty-

wane w procesie modelowania w fizyce czy w inżynierii. Jednakże modele sygnałowania komórkowego znacząco różnią się od typowych modeli z innych dziedzin. Zazwyczaj modele szlaków sygnałowych są znacznie bardziej złożone, składają się z większej liczby parametrów, a dokładność ich estymacji jest ograniczona ze względu na jakość, bądź brak danych eksperymentalnych.

W przypadku podejścia nieparametrycznego pomija się wszelkie założenia dotyczące struktury lub charakteru procesów wewnątrzkomórkowych. Związek między odpowiedzią komórkową (wyjściem), a sygnałem (wejściem) jest reprezentowany w sposób probabilistyczny. Zależność tę można opisać, korzystając z teorii informacji Shannona, jako kanał komunikacyjny. W tak rozumianym szlaku sygnałowym, analizowany jest przepływ informacji pomiędzy wejściem do kanału, a jego wyjściem. Jednakże standardowe kanały komunikacyjne różnią się od ścieżek sygnałowych. W przeciwieństwie do sygnałowania komórkowego, kanały komunikacyjne mają ograniczony szum, działają w optymalnych lub określonych warunkach, a przesyłane wiadomości składają się z długich sekwencji dyskretnych symboli.

W mojej pracy doktorskiej opracowałem parametryczne i nieparametryczne metody, które są dostosowane do złożoności i stochastyczności szlaków sygnałowych.

W podejściu parametrycznym skupiłem się na opracowaniu technik wyjaśniających relacje między parametrami a zachowaniem modelu. Po pierwsze, w rozdziale *"Sensitivity Analysis in quantitative biology models"* [1] książki *"Quantitative Biology: Theory, Computational Methods and Examples of Models"* [9], zidentyfikowaliśmy obecnie stosowane metody analizy wrażliwości, które można wykorzystać do analizy sygnałowania komórkowego. Przedstawiliśmy konkretne metody, które służą do analizy wrażliwości parametrów lokalnie i globalnie w modelach deterministycznych i stochastycznych. Ponadto zaprezentowaliśmy na przykładach modeli biologicznych, w jaki sposób analiza wrażliwości prowadzi do głębszego zrozumienia badanych zjawisk oraz pomaga w tworzeniu i usprawnianiu złożonych modeli. Po drugie, w artykule [2] zaproponowaliśmy metody analizy wpływu poszczególnych parametrów na zachowanie złożonych modeli wieloparametrowych. Korzystając z teorii informacji i analizy korelacji kanonicznych, zaproponowaliśmy nową miarę podobieństwa między grupami parametrów. Na podstawie tej miary zdefiniowaliśmy metodę analizy identyfikowalności parametrów modelu. Ponadto opracowaliśmy narzędzie, które wykorzystuje grupowanie hierarchiczne w celu znalezienia maksymalnej możliwej liczby parametrów identyfikowalnych. Wykazaliśmy, że nasza metodologia pomaga w projektowaniu eksperymentów biologicznych, które mogą poprawić dokładność modeli.

Ponadto w artykule [3] opracowaliśmy nieparametryczne metody frakcyjnej analizy odpowiedzi (FRA), która ilościowo określa zmiany we frakcjach komórek o danych poziomach odpowiedzi. Wykazaliśmy na kilku biologicznych przykładach, że odpowiedź systemu biologicznego zależy od liczby komórek z określonymi odpowiedziami (ich frakcji), a nie od średniej lub mediany odpowiedzi całej populacji. Zaproponowana metoda może być zastosowana do analizy danych eksperymentalnych ze ścieżek sygnałowych i innych systemów biologicznych bez względu na ich heterogeniczność, mnogość efektorów, czy dynamikę w czasie. Pokazaliśmy też, że FRA pomaga wykryć zjawiska na poziomie po-

jedynczych komórek, których nie dało się zauważyć korzystając ze standardowych technik. Dodatkowo wykazaliśmy, że FRA posiada ścisłą matematyczną interpretację w postaci min-informacji Rényi'ego (*ang. Rényi min-information*). Zatem FRA może być wykorzystywana, jako miara dokładności przepływu informacji w szlakach sygnałowych. Pokazaliśmy, że w porównaniu ze standardową informacją Shannona, właściwości min-informacji Rényi'ego są odpowiedniejsze do opisu przepływu informacji w systemach biochemicznych. Metody opracowane w ramach tej pracy zostały zaimplementowane w ramach ogólnodostępnych pakietów oprogramowania: ClusteringIdentifiability i FRA.

1 INTRODUCTION

According to the most recent estimates, a human consists of $3.7 \cdot 10^{13}(\pm 0.8)$ cells [10]. Among them, there can be distinguished over 200 different cells types [11] that perform a staggering variety of functions. The functions they play in the organism pretend inter alia their morphology, including shape and size. Cells that serve for storage, like fat cells and oocytes, have huge volumes, respectively, 6000 and 40000 times bigger than the most minor, red cells [12, 13]. The role of red cells is the delivery of oxygen to the body tissues from the pulmonary alveolus. Hence, their small size enables them to squeeze through narrow capillaries and biconcave disk shape, which maximizes the surface area to volume ratio, making oxygen transport efficient [14]. All of the cell types are among the most complex and fascinating biological systems in nature [15]. Cells need to control their growth, proliferation, metabolism, calcium balance, and many other cell-type-specific processes throughout their whole life. What is worse, all these happen in messy biological conditions with genetic mutations and variations, phenotypic noise and variability, as well as environmental changes [16]. Hence, cells seem to be a buzzing crucible of ongoing processes and complex biochemical reactions that somehow do not prevent cells from performing specific functions in organisms in cooperation with other cells.

The complexity of these mechanisms leads to questions related to intra- and inter-cellular processes within a cell and in cell populations. Firstly, how do cells interact to develop and maintain higher levels of organization and function [17]? Secondly, how do the components within a cell interact to bring about its structure and realize its functioning? Although experimental molecular biology has uncovered a massive amount of biological facts, it is not sufficient for interpreting biological systems and interactions between them [15]. This insufficiency is because of biological systems' inherent complexity and the intrinsic and extrinsic stochasticity of molecular events.

Typically, the questions above are recognized by integrating experimental insights into isolated parts of a cell or organism. Such a system-wide perspective and thus system-level understanding of biochemical activity is achieved using mathematical and computational methods and is called Systems Biology [18, 19].

1.1 BIOCHEMICAL SIGNALING PATHWAYS

Cells communicate with each other to regulate development, tissue repair, and immunity, as well as normal tissue homeostasis. Generally, cellular interaction consists of six steps: (i) synthesis and (ii) release of the signaling molecules including cytokines, hormones, morphogens, and other ligands; (iii) transport of the signal to neighboring cell; (iv) detection of the signaling molecule by various cell-surface receptors; (v) inter-cellular reactions triggered by the receptor-signal complex, e.g., phosphorylation of transduction factors; and (vi) removal of the signal, which often terminates the cellular response [20, 21, 22].

In systems biology, biochemical processes responsible for decoding the extracellular signal and leading to cellular response, i.e., steps iv-vi, are conceptualized as signaling pathways. Simplistically, the signaling pathway can be view as a pipeline that links cellular behavior with the outside environment. The initial stimuli are processed in a cascade of biochemical reactions, e.g., protein phosphorylation catalyzed by protein kinases or enzymatic reactions. This transmission leads the signal to cell decision centers and culminates in cellular response. At the molecular level, these responses include modification in genes' transcription or translation and changes in proteins, both post-translational and conformational, as well as their location.

One example of signaling pathways that play a crucial role in the immunity system is interferons (IFNs) signaling. The viral infection initiates a complex regulatory system of innate and adaptive immune responses, which goal is to deal with pathogens. One of the most critical reactions to the viral disease is the production and secretion of IFNs into extra-cellular space [23]. In a typical scenario, an infected cell releases interferons causing nearby cells to heighten their immune defenses. IFNs play a key role in antitumor immunity, the regulation of immune responsiveness, and tissue integrity under homeostatic conditions [24]. We identify more than twenty distinct IFN genes divide into specific classes, i.e., Type I IFN, Type II IFN and Type III.

The group of type I IFNs consists of many members of the IFN-α family (among others IFN-αI and IFN-αII, or IFN-ω and IFN-τ) and IFN-β. The IFN type I activates a pathway via a homologous receptor complex termed IFNAR, which consists of two subunits IFNAR-1 and IFNAR-2 [23]. Binding of IFN type I to the IFNAR results in activation of the Janus protein tyrosine kinases (Jaks), Jak1, and Tyk2 [23]. Further, phosphorylated Jaks leads to the tyrosine phosphorylation of the proteins from the signal transducers' family and activators of transcription, STAT, specifically STAT1 and STAT2. Then, phosphorylated STATs proteins form two transcriptional-activator complexes: (i) IFN-α-activated factor (AAF), and (ii) and IFN-stimulated gene factor 3 (ISGF3). The AAF is a homodimer of activated STAT1 proteins, whereas ISGF3 is a heterodimer of STAT1 and STAT2 associated with the interferon regulatory factor (IRF) [23]. ISGF3 binds DNA elements called interferon-sensitive response elements (ISREs) and subsequently activates interferon-stimulated genes, including genes encoding antiviral proteins such as Mx1 and OAS and various transcription factors including interferon-regulatory factors (IRFs) [23]. The type I IFNs are main cytokines for **innate immune response** against viral infections [25].

The IFN type II is structurally unrelated to IFN type I, and the mechanisms that regulate its production are also different. In contrary to the IFNs type I, there is only one member of this group, IFN-γ. IFN-γ activates the signaling pathway by binding to its cell-surface receptor, which consists of two subunits IFNGR1 and IFNGR2 [26]. Ligand binding induces the assembly and activation of the IFNGR complex, leading to the cross-phosphorylation and activation of Jaks, JAK1, and JAK2, and phosphorylation of the cytoplasmic domain, providing docking sites for the SH2 domains of STATs [27]. The latter sequence is as follows; phosphorylated STAT1 proteins form a homodimer through reciprocal phosphotyrosine-SH2 interactions, then STAT1 complex translocates to the nucleus and binds to IFN-γ activated sequence (GAS) elements in the promoters of most IFN-γ responsive genes (ISGs) [27]. IFN-γ is involved in the regulation of both **innate and adaptive immune responses**. The vital biological roles are antiviral and antiproliferative properties, macrophage activation, control of apoptosis, and promotion of antigen processing, presentation, and T-helper type 1 differentiation [28]. The IFN-γ signaling regulates over 200 hundred genes, which are involved in induction, repression, stabilization, or destabilization of cell functioning [28]. Therefore, the proper regulation of the IFN-γ signaling is essential for maintaining homeostasis and eliciting thoughtful immune responses [29].

The group of IFN type III include interleukin (IL)-28A (IFN-λ2), IL-28B (IFN-λ3), and IL-29 (IFN-λ1). Type I and III IFNs share a signaling cascade that is activated via different cell-surface receptors [30]. Type III IFNs bind to a heterodimeric receptor which consists of the IL-10 receptor 2 (IL10R2) and the interferon λ receptor 1 (IFNLR1), which is structurally different from the IFNAR [30]. Even though both IFNs use the same signaling cascades, type I and III IFNs evoke distinct biological activities [30]. However, it is still unclear why stimulation with type I and III IFNs result in divergent antiviral and immune-modulatory states [30, 31, 32, 33, 34]. IFNs induced signaling pathways are schematically presented in Figure 1.

The description above illustrates the complexity of a particular signaling pathway. However, the cellular system comprises many signaling pathways that create robust channels that effectively process external signals. All signaling pathways together create the signaling network, which tightly entwines all cellular components. Moreover, many signaling pathways are functionally pleiotropic [36, 37, 38]. Pleiotropy comes from signaling pathways interconnect via proteins and cascade, forming a tangled signaling network [39, 40]. It might seem that signaling networks contain prohibitively wide connected and interconnected components, which should foreclose the cell from operating accurately. Indeed, whether genetic, biochemical, or environmentally induced, perturbations of these signaling pathways can manifest in various diseases.

Dysfunctions in signalling may have tangible consequences for an organism [41, 42, 43, 44]. For instance, cancer cells exhibit altered signaling what contributes to their uncontrolled growth [45, 46, 47].

The IFNs signaling has a crucial function in orchestrating the immune system; hence, any dysregulation of the pathway may result in various immune disorders, especially hematological malignancies [48]. Different IFNs pathway dysfunctions lead to other diseases: (i) mutation (V617F) in the JH2

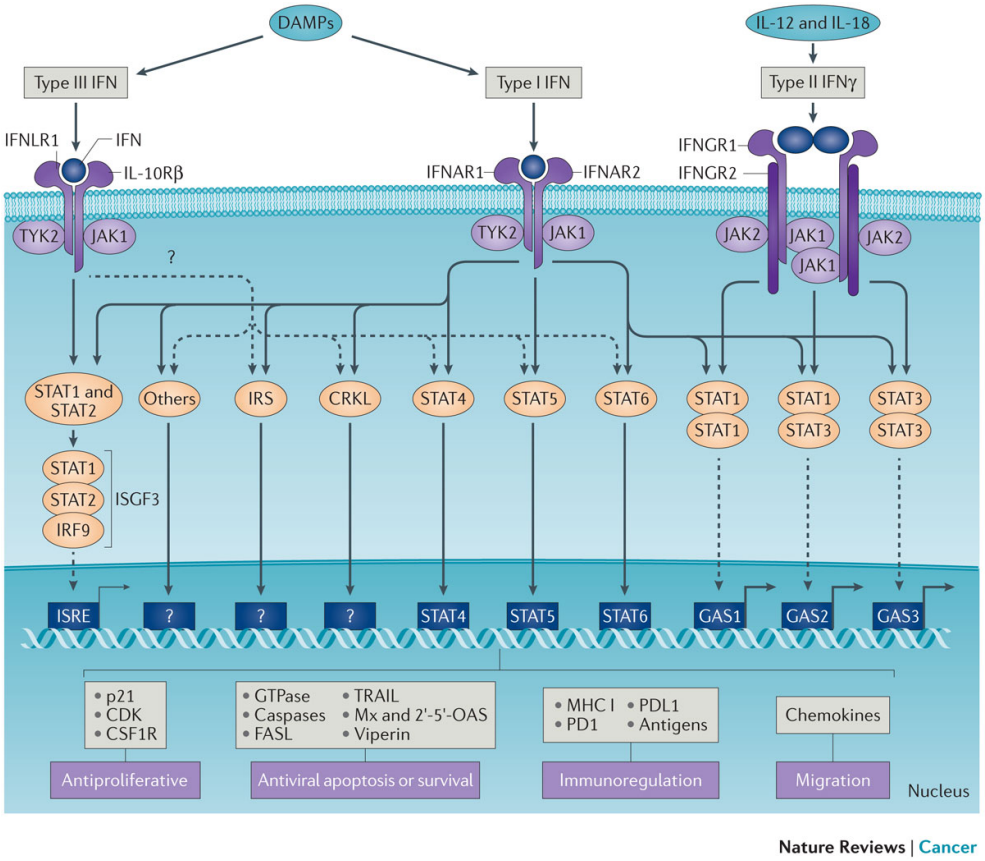


Figure 1: Figure from [35]. Precise model of signaling pathways of the interferons that mediate antitumor responses. Mechanisms presented in the figure are partially described in the main text. The scheme shows an extremely complex, cross-wired network.

pseudo-kinase domain of JAK2 leads to the constitutive activation of STAT5, and a malignant phenotype was shown in 80% of the patients with myeloproliferative neoplasms [49, 50]; (ii) rare JAK2 gene rearrangement can result in an atypical form of chronic myeloid leukemia [51, 52, 53]; (iii) several acquired JAK1 missense mutations can lead to adult lymphoblastic leukemia [54, 55, 56]; (iv) dysregulation of STAT proteins also contributes to the pathogenesis of several types of lymphoid malignancies, e.g., T-cell large granular lymphocytic leukemia, anaplastic large cell lymphoma, T-cell angioimmunoblastic lymphoma, and Sezary syndrome [57, 58, 59].

The biochemical system tries to limit the effects of signaling disorders using various defense mechanisms. The IFNs signaling is regulated by several complexes, such as protein tyrosine phosphatases (PTPs), suppressor of cytokine signaling (SOCS), and the protein inhibitor of activated STAT (PIAS) families. Another defense mechanism is the multifunctionality of biological systems that increases the capacity to cope with diverse challenges [60], including ones that have never been encountered before [61, 62]. For instance, cross-talks between signaling pathways enable cells to adjust cellular processes to intra-, and inter- environmental changes, like stress, crucial for evolutionary adaptation [16].

Studying signaling processes remains crucial for improving treatments, for instance, against cancer [63, 64]. First of all, signaling pathways are natural

ways of targeting molecular components; hence, transduction molecules represent attractive targets for various medical therapies [35, 65]. Secondly, a better understanding of signaling processes may help recognize the origins of their aberrations that alter cellular functioning [66]. Decades of experimental research have resulted in a detailed qualitative description of the most important signaling pathways, especially the oncogenic ones [67, 68]. However, quantitatively understanding how information about a complex mixture of extra-cellular stimuli is processed and translated into distinct cellular responses remains elusive [69, 16].

One of the difficulties with understanding biochemical signaling arises from the natural stochasticity of intracellular processes. The randomness of response to the signal is formed by (i) intrinsic variability, which arises from the probabilistic nature of the timing of collision events between reacting biological molecules and (ii) extrinsic variability, which results from varying components upstream to the system of interest, e.g., imbalance of homeostasis state [35, 65]. It has been shown that the latter dominates, especially in eukaryotic systems[70, 71].

Another challenge comes from the immense complexity of signaling processes [69, 16]. Typically, the signaling pathway consists of several autoregulatory mechanisms, e.g., negative and positive feedback loops and delays. Additionally, many components are functionally pleiotropic, e.g., various receptors often activate the same pathways [72]. For instance, as mentioned, IFNs type I and type III activate a shared signaling pathway via distinct cell-surface receptors.

Moreover, signaling pathways interact with each other. For instance, although different types of IFN transmit signals through distinct receptor complexes, the IFN type I receptor component, IFNAR1, facilitates the efficient assembly of IFN type II activated transcription factors [40]. The cross-talk between IFNs type I and type II signaling pathways provides a molecular basis for their overlapping roles [40]. Moreover, IFN type I signaling has an important role in strengthening IFN type II signaling by increasing the association of IFNAR1 with IFNGR2 and maintaining the docking sites in IFNAR1 for STAT1 [40]. IFN type II could achieve stronger activation effects based on this dependence by enhancing responsiveness to other IFNs type I stimuli. This mechanism is called *priming* phenomenon [73].

Systems biology ensures computational and mathematical support for analyzing complex, stochastic biochemical systems. Quantitative methods integrate theoretical knowledge with experimental data to create predictive models of signaling systems. However, computational models are not the one-to-one equivalent of biochemical systems because computers' performance is not enough for the computational complexity of the 'precise' models. Moreover, there are still many cellular processes that have not been sufficiently studied experimentally. The lack of experimental data decreases the predictivity of 'precise' models. Therefore, the computational models present a simplification of the cellular processes.

The typical study of signaling systems involves examining how the intensity of a stimulus, e.g., cytokine dose, translates into the activity of signaling effectors, e.g., transcription factors[74, 75, 76, 77, 78, 79, 80]. This is usually done by exposing cells to a range of doses and measuring responses either

in bulk or at the single-cell level. Results of such experiments are then represented and interpreted in terms of dose-response curves. The standard dose-response curve depicts how the mean or median of response changes with the increasing dose and provides a basic, first-order model of how a signaling system operates. Commonly, the dose-response curves are enriched by the error bars with standard deviations of signaling effectors to depict the heterogeneity of biochemical responses. However, there are many drawbacks of such representation of signaling pathways. Firstly, cells exposed to the same stimulus show great cell-to-cell heterogeneity [81, 82, 83, 84]. Hence, the mean response can result from a small fraction of strongly responding cells or a significant fraction of weakly responding cells [74, 75]. Moreover, the standard deviation is not able to represent different distributions of cell-to-cell heterogeneity. Secondly, the highly interconnected architecture typical for mammalian signaling usually results in a single stimulus activating several primary signaling effectors or downstream genes [85, 86, 87, 88, 89]. Finally, it was shown that the stimulation alters temporal profiles of signaling responses [90, 91]. These problems can be addressed in two ways:

1. replacing the first-order model with a more complex parametric model,
2. interpreting the simple input-output relation as a communication channel.

The parametric models assume the structure of the biological system and describe its dynamics using various mathematical techniques. The most common modeling formalism assumes the deterministic behavior of molecules and represents their dynamics as a set of ordinary differential equations (ODEs). ODEs systems represent the production and consumption rates of individual species in the model; hence, the change in concentration for each species is represented by a single ODE. The ODEs models are obtained using mass-action kinetics that assumes that the rate of a chemical reaction is proportional to the concentration of the reactants [92]. The difficulty with mass-action kinetics is that it is only valid for elementary monomolecular or pseudo-monomolecular reactions. Elementary reactions constitute fundamental processes; hence, they should not be decomposed into smaller steps. Additional assumptions on the law of mass action provide novel kinetics. For instance, Michaelis-Menten kinetics is used to describe enzyme-catalyzed reactions [93]. In contrast, Hill function is used to describe the membrane diffusion process [94] and simple gene expression regulation [95]. Deterministic modeling is an excellent tool for organizing and formalize knowledge about the cellular processes [8]. It also provides a mechanism for integrating different experimental sources in a predictive model [96].
However, novel, high-throughput experimental techniques allow studying the stochasticity of intra-cellular processes on the single-cell level. If we do not want to lose the additional information hidden in variability, we need another stochastic approach. A wide range of methods has been developed to model intrinsic and extrinsic sources of variability in biological systems. Intrinsic variability is a relatively well-understood aspect of biological models. Traditionally, it is modeled using the chemical master equations (CMEs), which is the foundation for modeling stochastic dynamics in most physical, chemical, and biological phenomena [97]. However, analytic solutions

of CMEs can be found only for a few trivial models. A good alternative for studying stochastic models is the exact simulation of processes using Gillespie's framework [98, 99]. However, to obtain a distribution resulting from the intrinsic variability, many trajectories of the Gillespie algorithm need to be simulated, which can be computationally expensive. Another option is to solve CME using approximation methods, e.g., the finite state projection [100], or linear noise approximation [101].

Significant effort has been invested in modeling intrinsic variability in systems biology, although it has been shown that extrinsic variability generally dominates, especially in eukaryotic systems[70, 102]. In the literature, there is no commonly accepted method for modeling extrinsic variability. Despite several intense mathematical and theoretical studies of intrinsic and extrinsic variability [70, 103, 104, 105, 102], computational modeling efforts that combine intrinsic and extrinsic variability are still rare. In the literature, we could find several solutions [106, 107, 108, 102]. Unfortunately, application of most of them is limited, for example, due to the high computational cost [106], the requirement of analytical solutions of extrinsic factor [107] or restriction to only one source of extrinsic variability [108].

The inherent limitation of the parametric modeling approach is associated with their fixed model structure. The prior knowledge and modeling assumptions can lead to potential bias. Moreover, the dynamics of biological models are susceptible to the kinetics parameter. Their values are estimated from experimental data so that the model's dynamic replicates the behavior of the biological system with minimum error. However, the accuracy of the estimation process is commonly limited due to the lack of experimental data. Numerous tools have been proposed to address problems with models' parameters, like identifiability and sensitivity analysis. However, many of them are inherited from engineering and theoretical statistics; hence they are not solely suited to the more complex models in quantitative biology [1, 7]. Therefore, the necessity of developing new methods that specifically address problems with parametrical modeling of biological systems.

Information theory provides a novel paradigm for analyzing biochemical networks. Here, the complex signaling network is reduced to the input-output information channel [109, 110]. Therefore, we do not adjust a detailed description of each reaction but treat the system as a black box and focus on the resulting response. This simplification protects from bias that comes from the model's structure and parameters estimation. Information theory provides sophisticated tools to describe the statistical relation between input and output of the system. This methodology was developed to analyze electronic communication, e.g., the accuracy of encoding and decoding schemes, precision, and size of data compression [111, 112]. In the biological systems, this formalism achieves interesting biologically relevant interpretations [113, 78, 5, 6]. However, it can be debated that metrics used to quantify the information flow in the communication channels are the most suitable for cellular signaling relays. For instance, communication channels are developed to reduce signal noise and enable the perfect decoding of discrete signals. On the other hand, the biological systems are characterized by considerable variability of all processes; hence, they do not require exact decoding of the input signal.

In my research, I focused on developing analysis techniques in both paradigms, i.e., parametrical modeling and information theory. In parametrical modeling, I have proposed a novel metric for analyzing the relations between groups of parameters. Based on this metric, I have created a tool for identifiability analysis, which can improve parameters estimation and help meaningful design experiments. Moreover, I have depicted the usefulness of local and global sensitivity analysis in the context of quantitative biology. My work in the field of information theory resulted in novel measures of information flow that consider the nature of biological systems. Finally, I have proposed a simple extension of the conventional dose-response curve that enables describing biological systems with multiple effectors and takes into account the variability of response.

1.2 MATHEMATICAL MODELS OF BIOCHEMICAL SIGNALING PATHWAYS

Biochemical signaling networks are immensely complex and stochastic systems; therefore, understanding their processes requires various quantitative strategies to enable biological insights. One of the approaches that deal with this problem is computational modeling. Briefly, mathematical models of signaling networks quantitatively describe changes of molecular state, e.g., copy numbers of molecules, in time after some biochemical reaction, e.g., activation of the signaling receptor through binding specific cytokine. Mathematical models present a simplified abstraction of signaling networks under different levels of detail. Most signaling network reconstructions focus on particular nodes, modules, or pathways in a given network (see Figure 2) [114]. Models can be abstract constructs used to emphasize some critical features of signaling pathways [115, 116] as well can describe in detail the whole dynamics of specific pathways in specific organisms [117, 118].

We can distinguish two roles of the modeling approach: descriptive and predictive. In the first role, models help to integrate experimental knowledge to decipher phenomena observed in experiments. Here, the model reconstructs the theoretical explanation of phenomena. Such models enable us to confront our prior knowledge of the biochemical system with the actual behavior of cells. It is done by comparing the model's dynamics with experimental data. In the latter role, modeling can be used to specify different, often mutually exclusive hypotheses and then explore the possible emerging behaviors [9]. For instance, such techniques enable us to determine the exact conditions under which different complex behaviors, such as oscillations or bistability, can occur; it is also used to simulate experiments that would be difficult or impossible to perform in the laboratory. The usage of modeling has the most sense if it meets both issues.

The process of designing mathematical models is a complex procedure that involves a series of interconnected steps of model improvement: (i) construction, (ii) verification, (iii) calibration, and (iv) validation. In the construction step, we need to define the scope and level of modeling abstraction and mathematical methods to represent the molecular processes. In the verification step, we check if the model's behavior is in line with theoretical knowledge. The calibration and validation of the model require the usage of experimental data. These steps are cyclically iterated until the improved model reaches expected conditions. In my Ph.D., I was working on improving methods used in the process of designing mathematical models. These methods are applicable at various stages of model development. Below, I will introduce the reader to the fundamental mathematical techniques related to the modeling of signaling pathways. I hope it will help to understand the contribution of my work to mathematical modeling and systems biology.

In general, a mathematical model of a biological system can be seen as a mapping between a vector of model parameters, $\theta = (\theta_1, \ldots, \theta_l) \in \mathbb{R}^l$, to model responses, $\mathbf{x} = (x_1, \ldots, x_N) \in \mathbb{R}^N$. In other words, θ and $\mathbf{x}$ can be thought of as the model input (e.g., kinetic rates and initial conditions) and the model output (e.g., mRNA, protein levels, and receptor states). The form of a re-

lation between the output, $\mathbf{x}$, and the parameter vector, θ, depends on the assumptions about the specified biological system as well as our goal according to the model. The main division is regarding attitude to the sources of response diversity. In the deterministic regime, i.e., assuming homogeneity within the cell and within a cell population, parameters-output relation is described as a function

$$\mathbf{x} = \mathbf{F}(\theta). \tag{1}$$

In the stochastic setting model can be represented in two forms. Assuming homogeneity within a cell population, the modeled relation takes the form of probabilistic dependence

$$\mathbf{x} \sim \mathcal{P}(\mathbf{x}|\theta). \tag{2}$$

Otherwise, some a prior distribution $\theta \sim P(\theta)$ is introduced to describe diversity in cell population. In such setting the output is considered over all possible environmental conditions, what mathematically is expressed by marginal likelihood

$$\mathbf{x} \sim \mathcal{P}(\mathbf{x}) = \int \mathcal{P}(\mathbf{x}|\theta)P(\theta)d\theta. \tag{3}$$

1.2.1 Deterministic modeling

The assumption of homogeneity, both within the cell and within a cell population, significantly simplifies the mathematical description of a biological system. Such models describe dynamics of *dependent variables* (model's output), i.e., changes of variables state, as a function of *independent variables* (e.g., time and/or space). Here, the dynamics of the variables are strictly determined by a set of primary conditions, i.e., kinetics parameters and initial values of variables. Commonly, signaling pathways are represented as spatially homogeneous time-variant systems described by ordinary differential equations (ODEs). Here, the only independent variable is time. This simplification facilitates the analysis of the effects of multiple inputs, feedback loops, and pathway crosstalk on the dynamics of complex signaling networks [119, 120, 121].

Generally, ODEs describe how the a set of N variables, $x = (x_1, \ldots, x_N)$, change with time t given an initial condition $x(t_0) = x_0$,

$$\frac{dx(t)}{dt} = f(x(t), \theta). \tag{4}$$

Here, $x(t)$ is the state of the system, for example, the concentration of some entities, t is an independent variable, and $f(\cdot)$ is a possibly nonlinear function that describes how the parameters θ affect the rate of the state's change. The output of the model can be then defined as a concatenated vector of values of $\mathbf{x}$ at specified times, $(t_1, \ldots, t_n)$,

$$\mathbf{x} = (x(t_1), \ldots, x(t_n))^T. \tag{5}$$

Generally, the solution of the Eq. 5, $\mathbf{x}(t) = F(\theta, t)$ can be understood as a time-dependent function. This solution settles down to a constant value in biological models following an initial transient phase in specific conditions.

This particular state of the system is called a **steady state**. It holds for the state $\mathbf{x} = x^*$ when the right side of the Eq. 5, $f(\cdot, \cdot)$, satisfy the condition

$$0 = f(x^*, \theta). \tag{6}$$

The stability of the ODEs solution plays an essential role in analyzing the behavior of the model.

ODE models represent the rates of production and consumption of individual biomolecular species, $\frac{dx(t)}{dt}$, in terms of mass action kinetics, an empirical law stating that rates of a reaction are proportional to the concentrations of the reacting species [9]. Each biochemical transformation is therefore represented by an elementary reaction with forwarding and reverse rate constants. Changes in localization, a central feature of biological pathways, are represented by compartmentalization. For example, cytoplasm and nucleus can be considered as two separate, well-mixed compartments [9]. A set of ODEs gives the dynamics of the system

$$\frac{dx(t)}{dt} = S \cdot w(x), \tag{7}$$

where S is the *stoichiometric matrix*, whereas $w(x)$ is a vector of reactions kinetic functions. The stoichiometric matrix is obtained based on the principle of mass balance. To derive a stoichiometric matrix for each species, we have to track what produces it (collect the reactions where it acts as a product) and consume it (collect the reactions where it acts as substrate). Summing up these processes while respecting their stoichiometry gives the corresponding rate of change of this species. A vector of reactions kinetic functions, $w(x) = [w_1, \ldots, w_R]$, describes how fast each j^{th} reaction occurs. The particular functional forms of $w(x)$ depend on the reaction kinetic laws. In the literature, we can find a large variety of reaction kinetic laws [122], however the most crucial in modeling signaling pathways are law of mass action and Michaelis-Menten kinteics.

The solution of ODEs representing biochemical model solutions is generally obtained by numerical integration. Although for linear or saturated kinetics, stationary concentration profiles can be obtained analytically [123, 124].

1.2.2 Uncertatintity and identifiability of parameters in the biological models

Characteristically, quantitative models are intrinsically dependent on parameters. Compared to engineering or physics problems, biological models depend on a relatively large number of parameters [125]. Unfortunately, parameters of biochemical models rarely can be measured directly. Therefore, the choice of parameter values requires methodological support.

Generally, the choice of the parameters values, i.e., their estimation, is performed by comparing model's output with experimental data. However, biological systems are observed partially, hence dimensions of observations are smaller than the dimensions of internal model states. The comparison be-

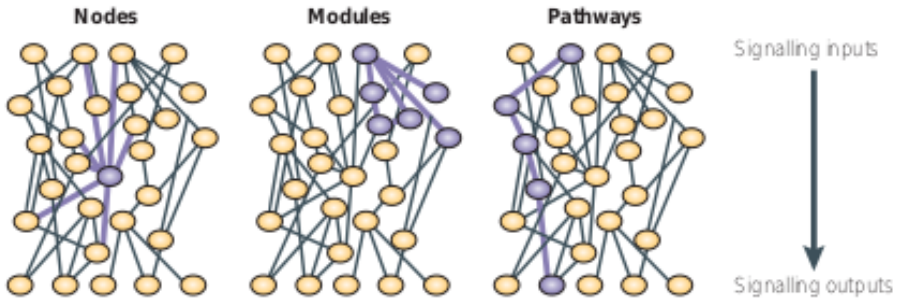


Figure 2: Figure from [114]. Scope of the model construction. Owing to a lack of comprehensive data regarding the interactions in a network, most signalling network reconstructions focus on particular nodes, modules or pathways in a given network. These reconstructions usually consist of a list of associations between network components. Network nodes describe the many interactions that a given molecules participates in. Network modules consist of a group of related reactions that often incorporate feedback mechanisms. Network pathways connect a signalling input to a signalling output. Each of these types of reconstruction has distinct advantages for analytical purposes. However, some properties emerge from the interconnectivity of the nodes, modules and pathways with other network components.

tween model and experimental data is done using function $g(\cdot,\cdot)$, that map the internal model states, $x(t)$, to the observables $y(t)$

$$y(t) = g(x(t), \vartheta) + \varepsilon(t), \tag{8}$$

where ϑ denotes scaling and offset parameters, whereas $\varepsilon(t)$ represents the measurement noise. The measure, that rate how well a model with particular parameters values describes the experimental data y^D is called a goodness-of-fit (GOF). There are many possible choices for goodness-of-fit function [9]. In the case of the biochemical models, researches commonly use the chi-square function, $\chi^2(\theta)$. Characteristically, for normally distributed noise $\varepsilon(t) \sim \mathcal{N}\left(0, \sigma^2\right)$, $\chi^2(\theta)$ is a function of the model likelihood

$$\chi^2(\theta) \propto -2 \cdot L(y^D|\theta), \tag{9}$$

where $L(y^D|\theta) = \log P(y^D|\theta))$ denotes the log-likelihood.

The larger the value of goodness-of-fit, the greater difference between the model's output and the experimental data. Thus, minimizing of goodness-of-fit leads to the optimal (estimated) parameter values, $\hat{\theta} = \arg\min\left[\chi^2(\theta)\right]$. Finding the optimal parameter, $\hat{\theta}$ is called an optimization problem and can be solved using various numerical algorithms.

The most straightforward optimization algorithms start at some initial point (e.g., prior or sampled parameters values) and move around on the landscape of parameters space to obtain a better fit. A straightforward strategy is to move downhill as quickly as possible: known gradient of GOF descent, this algorithm repeatedly estimates the local gradient and moves a short distance in the direction that most decreases the GOF. The method above is called local because the algorithm penetrates the neighborhood of the cur-

rent parameters estimates. There are some extensions of this method that decrease the number of steps of the algorithm or can be more efficient by taking advantage of the least-squares structure of our minimization problem, e.g., Levenberg-Marquardt algorithm. However, the landscape can be arbitrarily complicated, with sudden sharp cliffs or even multiple local minima, making such algorithms ineffective or leading to incorrect results. Another global approach is to choose various initial points and apply local methods randomly. Popular global strategies include simulated annealing, e.g., Markov Chain Monte Carlo (MCMC), and genetic algorithms inspired by genetic recombination and natural selection.

When the model sufficiently describes measured data, it can be applied to the purposes for which it was derived. Usually, the goal is to predict some experimentally non-measurable values, e.g., rate constants or initial concentrations, time-courses of experimentally unobserved species concentrations, or the system behavior under changed environmental conditions such as altered network structure or different external stimulation. Therefore, researchers need to define the level of confidence of the model's prediction. As the model intrinsically depends on the parameters' values, the level of prediction's confidence is determined by the level of their uncertainty [126]. The uncertainty of the parameters' values comes from the model's structure and 'so-called' practical problems with their estimation. Structural uncertainty comes from (1) compensative effects of parameter changes, i.e., there is a function, which describes the relationship between the model's parameters, also known as parameters collinearity; (2) insensitivity of individual parameters, i.e., a change of the parameter value does not affect the model's output. The practical uncertainty happens when the amount and quality of experimental data are insufficient to estimate parameters with high precision. The 'weakness' of experimental data comes from various limitations, e.g., availability of specific anti-bodies, technical constraints to the number of species observed at once, or huge experimental costs (counted in time and money). Consequently, biological systems are often only partially observed. The parameter is non-identifiable if it cannot be precisely estimated due to model's or experimental limitations [118]. Contrarily, the parameter is identifiable if its value can be determined explicitly.

Identifiability of parameters are determined within some range, called confidence intervals (CI), which contain the "true" value of the parameter with the desired probability [127]. Thus, the relevant issue is also the width of the CIs, which indicates a model prediction reliability.

Mathematically, a confidence interval $[\hat{\theta}_i - \sigma_i, \hat{\theta}_i + \sigma_i]$ of a parameter estimate $\hat{\theta}_i$ denotes that with a confidence level α signifies that the true value θ_i^* is located within this interval with probability α.

Asymptotically, CIs are derived from the curvature of the likelihood, as the Hessian matrix of the likelihood function, i.e., $H = \bigtriangledown^T \bigtriangledown \chi^2(\hat{\theta}_i)$. Then asymptotic confidence intervals are equal

$$\sigma_i = \sqrt{\chi^2(\alpha, df) \cdot C_{ii}}, \tag{10}$$

where C is a covariance matrix of the parameter estimates, i.e., $C = 2 \cdot H^{-1}$, $\chi^2(\alpha, df)$ is the α quantile of the χ^2-distribution with df degrees of free-

dom [128]. The approximation of the actual uncertainty of parameters using asymptotic CI is possible if the amount of experimental data is large compared to the number of the parameters and the measurement noise is relatively small. Such conditions are challenging to be satisfied in systems biology. Therefore, the solution is to approximate CIs using a threshold in the likelihood. In this approach, CIs are defined by a confidence region [129]

$$\left\{\theta|\chi^2(\theta) - \chi^2(\hat{\theta}) < \chi^2(\alpha, df)\right\}, \tag{11}$$

where confidence region borders represent CIs.

The analysis of CIs helps to recognize the identifiability of the parameters. First of all, identifiable parameters characterize with finite confidence intervals. On the other hand, the structural and practical non-identifiability indicates non-finite confidence intervals.

Structural non-identifiability is related to the model structure independent from experimental data. In such conditions, the parameter estimates $\hat{\theta}$ are not uniquely identified, and consequently, the likelihood function, $\chi^2(\theta)$, obtains its minimum for different parameters values. As a remedy for structural non-identifiability, most approaches aim to select an optimal subset of parameters that is both sufficiently sensitive and has the lowest collinearity. The identifiable subset can then be estimated jointly with the remaining parameters assumed fixed.

In practical non-identifiability, insufficient amount and/or quality of experimental data manifests in infinite CIs. Precisely, according to a definition proposed in [126], a parameter estimate $\hat{\theta}_i$ is practically non-identifiable if the likelihood-based confidence region is infinitely extended in increasing and/or decreasing direction of θ_i. However, the likelihood has a unique minimum for this parameter.

Following [126], we illustrate parameters identifiability using contour plots of the likelihood function, $\chi(\theta)$, for a two-dimensional parameter space, as in Figure 3.A. Fistly in te left panel, structural non-identifiability results in a perfect flat valley infinitely extended along with the corresponding functional relation between parameters. Furter, the middle panel shows a practical non-identifiability that leads to the relatively flat valley, which is infinitely extended in one direction. In contrast to the structural identifiability, a single lowest point represents parameters to estimate, $\hat{\theta}$. Finally, in the right panel, the confidence interval, expressed by a thick white line, is limited. Thus, this plot represents a case where both parameters are identifiable.

However, it becomes challenging to discriminate between structural and practical identifiability using this approach. Difficulties appear in multiparametric models with sparse or low-quality experimental data. Typically, in the case of practical non-identifiability, there are observed non-single minima of the likelihood function. Therefore, in [126] authors proposed a convenient and systematic approach for analysis parameters identifiability by using profile likelihoods, χ^2_{PL}. Therein, the parameter space is explored for each parameter in the direction of the slightest increase in χ^2.

Precisely, the profile likelihood method works by systematically tracing an optimal path over the likelihood or in the case of MAP estimation, the probability density function. Initially, parameters are estimated together, $\theta^* = \hat{\theta}$,

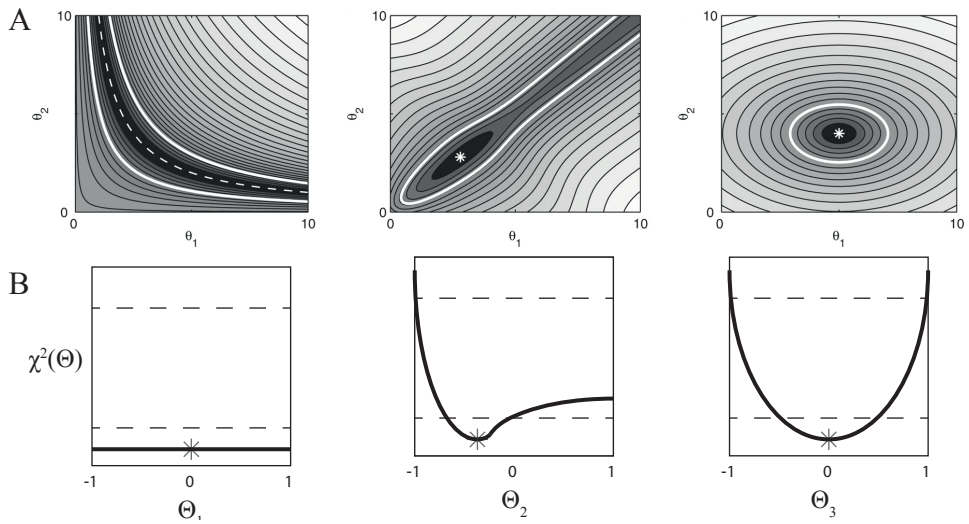


Figure 3: **A** Contour plots of $\chi^2(\theta)$ for a two-dimensional parameter space, shown in non-logarithmic scale for illustrative reasons. Shades from black to whitecorrespond to low and high values of χ^2, respectively. Thick white lines display likelihood-based confidence regions and white stars the optimal parameter estimates $\hat{\theta}$. Left panel: a structural non-identifiability along the functional relation $h(\theta) = \theta_1 \cdot \theta_2 = 10$ (dashed line). The likelihood-based confidence regionis infinitely extended. Middle panel: a practical non-identifiability. The likelihood-based confidence region is infinitely extended for $\theta_1 \to \infty$ and $\theta_2 \to -\infty$, lower confidence bounds can be derived. Right panel: both parameters are identifiable. Figure from [126]
B Black lines display profile likelihood versus parameter in the hypothetic example. Gray stars display the estimated values of the parameters θ. Dashed lines represent thresholds for confidence intervals. Left panel: a structural non-identifiability modification of parameter value can be compensated by changing other parameters, which keep the likelihood value constant. Middle panel: practical non-identifiability. Right panel: identifiability - changing the value of parameter leads to decreasing the likelihood.

after which each parameter θ_i is profiled separately. The profiled parameter is subsequently changed within some range, i.e., $\theta_i \in [\theta_i^* - \sigma_{PL}, \theta_i^* + \sigma_{PL}]$. Subsequently, for each parameter's value, all the unchanged parameters are re-estimated. In such case, the profile likelihood of parameter θ_i at point $\theta_i = \tilde{\theta}_i$ is equal

$$\chi^2_{PL,i}(\tilde{\theta}_i) = \min_{\theta_{-i}} \left[\chi(\theta|\theta_i = \tilde{\theta}_i)\right], \tag{12}$$

where θ_{-i} denotes, that re-optimisation of likelihood function is with respect all parameters except θ_i. Analysis of the whole profile likelihood, i.e., for each value within range $[\theta_i^* - \sigma_{PL}, \theta_i^* + \sigma_{PL}]$, enables to determine the identifiability of the patameter. If parameter θ_i is structurally non-identifiable, then its profile likelihood is flat, because small changes in θ_i can be easily compensated by changes of other parameters (see illustrative Figure 3B). The profile likelihood of a practically non-identifiable parameter has a minimum but is not excessing a threshold $\chi^2(\alpha, df)$ within a range of values of parameter θ_i. Finally, the profile likelihood of an identifiable parameter exceeds $\chi^2(\alpha, df)$ for both increasing and decreasing values of θ_i.

Analysis of the PL enables to distinguish non-identifiability among calibrated parameters. As it was mentioned, non-identifiability parameters alter the level of confidence in the model's prediction. Hence, it should be taken into account, especially during experimental design. Poorly planned experiments might lead to practical non-identifiability and might not increase the model's predictivity. Calculation of PL requires experimental data; therefore, this method can only be used to verify the identifiability of the parameters. In the article, [2] we introduced the asymptotic version of profile likelihoods (APL) that enables to predict non-identifiability without experimental data. This technique can be used for designing the most efficient experiments from the modeling perspective.

1.2.3 Sensitivity analysis of parameters

Analysis of the identifiability of the parameters improves the model's accuracy, i.e., the precision of description of the system's behavior and the ability to predict novel phenomenons. When the model's parameters are determined, we can investigate how parameter values impact quantitative characteristics of model behavior. The sensitivity analysis provides rigorous tools for this purpose. A key concept in sensitivity analysis is the sensitivity coefficient, which quantifies how sensitive the model output is to changes in a given parameter. The sensitivity coefficient can be calculated for small changes around nominal values. Such a technique is called local sensitivity analysis. For deterministic models, a natural way to evaluate the local sensitivity is to estimate the derivative using a Taylor expansion for small perturbations around some nominal value, θ^*

$$x(\theta^* + \Delta) \approx x(\theta^*) + \frac{\partial x}{\partial \theta} \cdot \Delta. \tag{13}$$

The higher raise of the output, $x(\theta^* + \Delta)$, to the small perturbation of parameter θ, the more significant its impact on the model. This relation is described by the sensitivy coefficient $\frac{\partial x}{\partial \theta}$. Parameters that have a more substantial effect on the model are called sensitive. For multi-parameter model we analyse the relation between output and all parameters, $\partial\theta = (\partial\theta_1, ..., \partial\theta_l)$. If changes in parameters are small, the problem is solved by finding the derivative of a solution of the equation (4), $x(t)$, with respect to the parameter θ_i, $z_i(t) = \frac{\partial x(t)}{\partial \theta_i}$. Evaluation of $z_i(t)$ at the times and components of interests defines the sensitivity vector $S_i = \left(z_i^{(q)}(t_1), \ldots, z_i^{(q)}(t_n)\right)$ of the parameter θ_i. The sensitivity vector describes the shift in Y in response to perturbation in the parameter θ_i, $\partial Y = S_i \partial\theta_i$. A collection of the sensitivity vectors for all $i = 1, \ldots, l$ constitutes the sensitivity matrix $S = (S_1, ..., S_l)$, which summarises the change in Y in response to perturbation of all of the model parameters $\partial X = S\partial\theta$. The sensitivity matrix, S, is directly linked with the concept of Fisher information. Given that Y is observed with the Gaussian unit variance error the FI can be written as

$$FI(\theta) = S^T S. \tag{14}$$

Therefore the FI contains information regarding the size of a perturbation, $\|\partial X\| = \sqrt{\partial\theta^T FI(\theta)\partial\theta}$. The pairwise similarity between parameters, quantified as the cosine between the S_i and S_j vectors, is also given by elements of the FI, $\cos(S_i, S_j) = S_i^T S_j / \|S_i\|\|S_j\|$.

In the global sensitivity analysis, the sensitivity coefficient is calculated arbitrary changes within some range [130]. In contrary to the local version, global analysis enables the comparison of parameter impact across parameter space.

Learning which parameters have more robust and weaker relationships to model behavior is valuable in itself. However, it provides other insights about model properties. For instance, sensitivity analysis shows how strong the system's behavior is to external perturbations of parameters [131, 132]. Moreover, if experiments exhibit unexpected variability, sensitivity analysis helps verify if parameter value fluctuations can explain such observations. Knowing sensitivities helps understand the effects of uncertainties in parameter estimates. Specifically, sensitive parameters can be inferred from data with higher accuracy than insensitive parameters. Therefore, calculating sensitivities is helpful to select experiments that yield more informative data to constrain unknown parameter values. Finally, knowledge of parameter sensitivities can also be used to identify and eliminate insensitive parameters, leading to simplified models [133].

There are a large number of methods of sensitivity analysis [134]. Most of them are inherited from engineering and theoretical statistics; they are not entirely suited to the specificity of the biochemical models. Notably, many techniques are limited only to the deterministic regime. In the article, [1] I present local and global sensitivity methods analysis with their application to biochemical models.

1.3 INFORMATION THEORY IN ANLYSING SINGLAING PATHWAYS

Claude Shannon developed information theory to find fundamental limits on signal processing and communication operations. However, it might seem confusing that information can be quantified, not a physical entity but an abstract concept.

The definition of the amount of information is provided by Shannon's information theory [112, 111]. Firstly, he proposed to quantify the information that is kept in some random data, by the average description length that is needed to specify the microstate of the system. Precisely, for a discrete random variable X, with possible outcomes $x_1, \ldots, x_n$, which description length is equal $l_1, \ldots, l_n$ and the probability of occurence $p(x_1), \ldots, p(x_n)$ the average description length of the data has the lower band

$$\sum_i p(x_i) l_i \geqslant -\sum_i p(x_i) \log p(x_i) = H(X), \tag{15}$$

where $H(X)$ is called entropy of random variable X. This inequality indicates that the average description length for any representation of x_i cannot be reduced to less than the value of entropy. Thus, entropy can be used as a measure of the amount of information contained in a distribution (or data which is obtained from distribution). The entropy, $H(X)$, is expressed in bits, and $2^{n \cdot H(X)}$ equals an average number of states of data set of length n with entropy $H(X)$. Entropy is interpreted as the uncertainty regarding data values when there is no other source of knowledge. Therefore, entropy is the highest when the input distribution is uniform because the probability that we guess the input value without any additional knowledge is the lowest. On the other hand, entropy takes a maximum value when x can bring only one value.

The notion of the information proposed by Shannon can be interpreted in a broad sense to include all systems in which the state of one part affects the state of another element [112, 111]. Such a system can be seen as a model that consists of the input (sender) X, the output (receiver) Y, and a channel $X \to Y$, that describes the transformation of input into output. Shannon provided a formal definition of a communication system. The communication channel is described as a stochastic model, where X and Y are random variables, whereas channel $X \to Y$ by a conditional random variable $Y|X$. Random variables X, Y follow distributions $P(x)$ and $P(y)$, whereas the channel $Y|X$ is described by conditional distribution $P(y|x)$.

When the output is observed, the uncertainty regarding the input values is described by the conditional Shannon entropy

$$H(X|Y = y) = -\sum_{i=1} P(x_i|Y = y) \log_2 P(x_i|Y = y), \tag{16}$$

where $P(x_i|Y = y)$ decscribe the probability that the input value x_i is generated by the output y

$$P(x_i|Y = y) = \frac{P(y|X = x_i)P(x_i)}{\sum_{j=1}^{m} P(y|X = x_j)P(x_j)}. \tag{17}$$

Entropy above describes uncertainty regarding input value after observing a specific output value, y, and is not representative of all possible output values. The average entropy of the input after observing the output is given by

$$H(X|Y) = \int_{y \in \mathcal{Y}} H(X|Y = y)P(y)dy, \tag{18}$$

over some output space $\mathcal{Y}$.

The difference between initial, a priori, entropy of the input, $H(X)$, and average entropy of the input given the output, a posteriori, $H(X|Y)$, quantifies the information gain that defines the Shannon information, or mutual information

$$I(X, Y) = H(X) - H(X|Y) = \sum_i \sum_j P(x_i, y_j) \log \frac{P(y_j|x_i)}{P(y_j)}. \tag{19}$$

The mutual information can be interpeted as the amount of uncertatinty regarding random variable X, which can be expained having knowledge about random variable Y.

The transmission via the channel $Y \rightarrow X$ depends on the communication system and the input distribution $P(X)$. The upper bound for the mutual information, given an input distribution $P(X)$, is called the channel capacity

$$C^* = \max_{P(X)} I(X, Y) \tag{20}$$

Shannon defined information capacity to quantify $\log_2$ of the maximal number of discrete symbols that can be transferred in a single transmission with a negligible error when messages are encoded in terms of long sequences of discrete symbols through a communication channel described by the probability distribution $P(Y \mid X)$. Such interpretation of Shannon capacity is known as Shannon coding theorem [111]. The distribution for which the maximum of mutual information is achieved is called the optimal input distribution and denoted as $P^*(X)$. The information capacity, C^*, is expressed in bits, and 2^{C^*} can be interpreted, within Shannon's coding theorem, as the number of input values that the system can effectively resolve based on the information contained in the output. For instance, $C^* = 2$ denotes the existence of four states that can be distinguished, on average, with negligible error.

1.3.1 Measure of information flow in the signaling pathways

Systems biology applies information theory for two primary purposes. Firstly, it is used to quantify information transmission in cells [113, 135, 135, 78, 80]. Secondly, mutual information can help to identify complex and nonlinear relationships within signaling network [136, 137, 138]. It comes from the fact

that mutual information is a measure of statistical dependence between two variables that, in contrast to classical correlation measures, is robust to dimensionality, nonlinearity, and the exact analytical form of the underlying distribution [139, 140, 141, 142]. In my thesis, I was focused on the first approach.

Following Berg and Purcell [143], probabilistic modeling has been applied to examine the fidelity of receptors, as well as more complex biochemical signaling systems [113]. In this way, the signaling pathway can be interpreted as an information transmission problem suitable to be described from the perspective of mathematical information theory.

Regardless of specific details of a signaling pathway, a signaling system can be considered as an input-output device described using a probability distribution $P(Y|X = x)$ that for a given level of input, x, elicits stochastic output, Y (see Figure 4). In a typical example, the input, x, is the concentration of a ligand, e.g., cytokine, that activates a receptor. The output, Y, is an activity of one or more signaling effectors, e.g., of transcription factors quantified over time. As cellular signaling systems are inherently stochastic, the information about the input contained in the output is imprecise, and only a limited number of input values can be resolved [143, 79, 144, 145]. In a typical experiment aimed to quantify fidelity, input values, $x_1 \leqslant x_2, \ldots \leqslant x_m$, ranging from 0 to saturation are considered, that follow a certain distribution $P(X) = (P(x_1), \ldots, P(x_m))$. Formally, responses corresponding to each of the inputs, x_i, are assumed to follow a probability distribution

$$y \sim P(Y|X = x_i). \tag{21}$$

Schematically, the representation of the signaling pathway as a channel is presented in the Figure 4.

It might seem that modeling of signaling pathways as the input-output channel reduces a complex mechanism to a simple dose-response relationship. However, these are the most relevant factors that describe biological context mechanism [109]. In contrast to parametric models, such representation rejects any prior assumptions about the system. Moreover, information theory takes into account the probabilistic character of biological phenomena. Hence, we do not need to build models as the structure of the complex mechanism is somehow hidden in the heterogeneity of responses. Information theory elucidates this complexity in terms of the discrimination of different inputs based on the responses.

Representation of the signaling pathway as a communication channel leads to a novel interpretation of information measures. In the context of biochemical signaling entropy of the input, $H(X)$ quantifies the prior uncertainty regarding the stimulants when we do not observe the response. On the other hand, when the response to the stimulation is observed, then remaining uncertainty regarding the stimulants is quantified by the conditional entropy $H(X|Y)$. Thus, the information transfer of a given signaling system is equal to uncertainty reduction when we observe the signaling response. The definition above strictly corresponds to the mutual information, $I(X, Y)$. In other words, the mutual information tells us what amount of knowledge about the input the signaling system can express (write) in the signaling response in

certain environmental conditions, described by the distribution $P(X)$. When applying these concepts to biological situations, the in vivo input distribution is often difficult to measure, so the channel capacity is usually calculated instead of evaluating the mutual information from the input distribution. In the biological systems, we could interpret the mutual information as a measure of the adaptation to the environmental conditions, represented as the input distribution [145]. Therefore, it is said that the optimal distribution of stimulants, $P^*(x)$, describes the most suitable environmental conditions for which the biological system has adapted during the evolutionary process [145]. Therefore, the channel capacity quantifies how many distinguishable states the signaling systems sense, which provides the overall signaling fidelity in the best environmental conditions, represented as an optimal occurrence of the set of stimulant concentrations. According to the capacity values, we can distinguish three types of cellular systems:

- cells can distinct only presence and absence of the stimulus, $2^{C^*} = 2$,
- cells distinguish the stimulant with mistakes, $2^{C^*} < 2$,
- cells distinct additionally the level of stimulant doses, e.g., small or high level, $2^{C^*} > 2$.

Estimation of the mutual information requires a large sample size for the data set; therefore, many single cells need to be measured. In such a case, we interpret information as "information at the population level". It assumes that cellular systems do not differ significantly between cells. On the other hand, if a single cell can be stimulated and measured multiple times, we obtain a single cell system description. Hence, analyzed information refers to "information at the single-cell level". Here, we assume that the cellular system can differ between different cells. The interpretation of the information depends on the problem settings involved in encoding and decode systems.

1.3.2 Applications of information theory to systems biology

Information theory has been extensively used to examine information processing in signaling transduction [113, 135, 135, 78, 80]. Here, we present some of the essential works in this area.
In [113] authors studied the channel capacity and mutual information between an upstream transcription factor, Bicoid, and a downstream target gene product, Hunchback, during early embryogenesis in Drosophila flies. The authors showed that in vivo Bicoid/Hunchback system uses a distribution of the input, Bicoid, that leads to the mutual information that differs only by 10% from the channel capacity. If the close agreement in mutual information values and the channel capacity could be generalized, it would imply a design principle in cellular information processing.
In [135] authors examined the channel capacity between tumor necrosis factor (input) and the transcription factor NFκB and activating transcription factor–2 (ATF-2) (outputs). They found that a combination of the effects of NF-κB and ATF-2 increases the information capacity. Moreover, the authors predicted potential structures of the network of information transmission. They defined their properties and the consequences for the cellular system.

In [78] authors calculated the channel capacity of the continuous-time course of molecular concentrations in response to stimulation measured using live imaging. The novel method of calculation of the channel capacity uses the k-nearest neighbors method. The combined effect can increase the channel capacity compared with using only a snapshot at a single time point.

In [5] authors proposed a novel algorithm for quantification of mutual information and channel capacity in signaling systems with high-dimensional outputs and a large number of input values. The provided solution is efficient in terms of computational time as well as the sample size needed for accurate estimation.

In [80] compared the mutual information in the single-cell with population level in G protein-coupled receptor signaling. Authors stimulated human kidney embryo cells, HEK293, with different doses of acetylcholine (Ach), which activated the muscarinic acetylcholine receptor M3R - the only member of the acetylcholine GPCR group expressed in HEK293 cells - resulting in Ca 2+ intracellular influx, which was measured as the signaling output in individual cells using a live imaging technique. Individual cells were exposed 5 times to each of 7 increasing doses of Ach so that 35 measurements were taken for each cell so that individual quantitative characteristics of individual cells were obtained, examining 433 cells in total. For each dose, the data enabled the construction of response distributions both for the population of cells as well as for individual cells. The authors showed huge differences between single cells, suggesting that the information transmission system differs between individual cells. However, the law number of samples used to estimate the channel capacity leads to questions regarding the overestimation of this measure. Moreover, the authors showed that the average channel capacity at the single-cell level (almost 2 bits) was greater than the channel capacity at the population level, which is expected.

1.3.3 Bottleneck of Shannon information for describing information flow in signaling

Shannon information gained widespread applicability in engineering and has also been adapted to quantify information transfer in cellular signaling pathways. However, it can be debated whether it is the most suitable measure for quantification of information flow along with cellular signaling relays.

Cellular signaling systems do not transfer information with messages encoded in terms of long sequences of discreet symbols, which makes the interpretation of Shannon information difficult to formulate under the circumstances in which cellular signaling operates.

The Shannon information can be interpreted as the number of inputs that can be discriminated on average based on cellular response. However, it does not tell us directly which inputs and to what extent can be distinguished. For instance, what is the difference between two signaling systems with the number of distinguishable states equal 1.7, and 1.4, respectively? Typically, it is said that the first system deciphers the proper input with fewer errors than

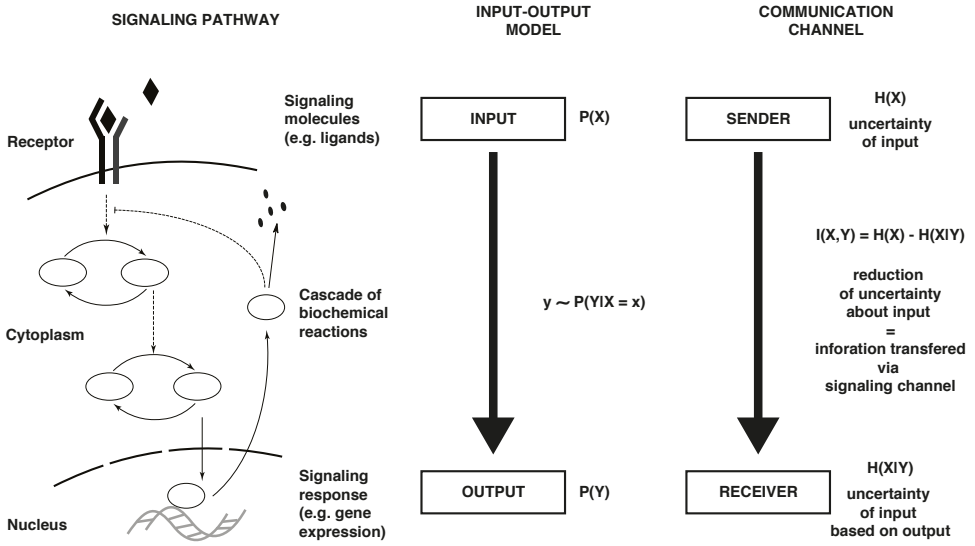


Figure 4: Signaling pathway as a communication channel.
On the left schematic representation of the signaling pathway. Extracellular signaling molecules (e.g., ligands) bind to specific membrane receptors, activating a series of intracellular reactions (signaling pathway). This process culminates in the translocation of a transcription factor into the nucleus, inducing the expression of target genes.
In the middle panel representation of signaling pathway as a simple input-output model. Input and output are represented as random variables X, Y. The signaling system can be considered an input-output device described using a probability distribution $P(Y|X = x)$.
On the right side, schematic presentation of the sender-receiver communication system. Here, the signal, X, is transferred to the receiver via a communication system. The signal is decoded to the form of cellular response, Y. The accuracy of the communication channel is described as an amount of knowledge about the signal, X, that we can obtain based on the decoded outcome, Y. The information is quantified using the notion of entropy as an uncertainty. Therefore, the accuracy of the communication channel is defined as the mutual information between sent and received signals, i.e., I(X, Y). It is equal to the difference between the entropy of the signal, H(X), and the entropy of the signal given the received signal, $H(Y|X)$, i.e., $I(X, Y) = H(X) - H(Y|X)$. The maximal mutual information depending on the signal distribution is called channel capacity, C^*.

the second one. However, there is no general formula that maps Shannon's information to errors' numbers.
Shannon information depends on the input distribution, $P(X)$, which in most biochemical applications are unknown, and a rationale for choosing a particular distribution is missing. Besides, maximization concerning the input distribution for calculating information capacity raises the question of whether cellular signaling operates under optimal conditions. Moreover, the Shannon measure's strong deprecation of the overlapping output distributions leads to the optimal distribution, consisting of the most extreme signals. However, such a repertoire of signals is extremely rare.
Taking all into account, there is a necessity to consider how to measure an information flow in signaling pathways. Current methods provide a formal description and important insights; however, their properties do not fully match the properties of biological systems.

2 MAIN GOAL AND OBJECTIVES

This thesis aims to develop quantitative methods that can deal with the complexity of the cellular signaling pathways. My work focused on two general approaches to analyzing signaling pathways, i.e., parametric and non-parametric modeling. The first approach aims to explore in detail processes of signaling pathways described as multi-parameter models. In the second approach, the complexity of the signaling mechanisms is hidden in the stochasticity of observed cellular responses.
I defined three objectives to accomplish these goals:

- **Objective A**. Identification of the potential of sensitivity analysis parametric methods for studies of cellular signaling.
- **Objective B**. Utilization of identifiability analysis approaches to examine multiparameter models of signaling pathways.
- **Objective C**. Development of non-parametric methods suited for the complexity of cellular signaling processes.

3 METHODS

3.1 CANONICAL CORRELATION ANALYSIS

Canonical correlation analysis (CCA) is a statistical method used to measure the linear relationship between two multidimensional variables [146]. It was developed by Hotelling in mid 1930s [147] and is commonly used in various areas, e.g., economics, medical studies, engineering [148, 149, 150]. Briefly, CCA retrieves bases for each variable in which the correlation matrix between the variables is diagonal, and the correlations on the diagonal are maximized. The dimensionality is reduced to the smallest dimensionality of the two variables. Characteristically, CCs are invariant with respect to affine transformations of the variables.

Let us consider two sets of vectors $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_m)$. The goal is to find bases, in which the correlations of the variables onto these basis vectors are mutually maximized. Let us look at the case where only one pair of basis vectors are sought, namely the ones corresponding to the largest canonical correlation. Those basis vectors are obtained by finding the linear combinations $\tilde{x} = x^T w_x$ and $\tilde{y} = y^T w_y$, such that their correlation is maximized, i.e.,

$$\begin{aligned} \rho &= \frac{\tilde{x}\tilde{y}}{\sqrt{\tilde{x}^2\tilde{y}^2}} \\ &= \frac{w_x^T x y^T w_y}{\sqrt{(w_x^T x x^T w_x)(w_y^T y y^T w_y)}} \\ &= \frac{w_x^T \Sigma_{xy} w_y}{\sqrt{(w_x^T \Sigma_{xx} w_x)(w_y^T \Sigma_{yy} w_y)}} \end{aligned} \tag{22}$$

Vectors $\tilde{x}$ and $\tilde{y}$ are first pair of canonical variables. The maximal correlation ρ equals first eigenvalue, whereas w_x and w_y are eigenvectors of the covaraince matrix between x and y, i.e.,

$$C = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}. \tag{23}$$

Thus, the canonincal correlation is obtained by solving the eigenvalue equations

$$\begin{cases} \Sigma_{x,x}^{-1} \Sigma_{x,y} \Sigma_{y,y}^{-1} \Sigma_{y,x} w_x &= \rho^2 w_x \\ \Sigma_{y,y}^{-1} \Sigma_{y,x} \Sigma_{x,x}^{-1} \Sigma_{x,y} w_y &= \rho^2 w_y \end{cases} \tag{24}$$

The number of non-zero solutions of eigenvalue equations equals the smallest dimensions of spaces defined by a set of vectors x and y; hence it is less than $\min m, n$. The following pairs of canonical variables are obtained by finding eigenvalues of decreasing magnitudes. The subsequent canonical correlations are uncorrelated (orthogonal) to previous solutions due to the symmentry of the covariance matrix, C. Let $\tilde{x}^{(i)}$ and $\tilde{y}^{(i)}$ be an i^{th} pair of

canonical variables, and ρ_i is corresponding canonincal correlation. Then, for any $i \neq j \neq \min(m, n)$ canonical variables satosfy the condition

$$\begin{aligned}
\tilde{x}^{(i)}\tilde{x}^{(j)} &= w_{\tilde{x}^{(i)}}^T x x^T w_{\tilde{x}^{(j)}} = w_{\tilde{x}^{(i)}}^T \Sigma_{xx} w_{\tilde{x}^{(j)}}, \\
\tilde{y}^{(i)}\tilde{y}^{(j)} &= w_{\tilde{y}^{(i)}}^T y y^T w_{\tilde{y}^{(j)}} = w_{\tilde{y}^{(i)}}^T \Sigma_{yy} w_{\tilde{y}^{(j)}}, \\
\tilde{x}^{(i)}\tilde{y}^{(j)} &= w_{\tilde{x}^{(i)}}^T x y^T w_{\tilde{y}^{(j)}} = w_{\tilde{x}^{(i)}}^T \Sigma_{xy} w_{\tilde{y}^{(j)}}.
\end{aligned} \tag{25}$$

There is a strict realiton betwwen canoncial correlations and mutual information. Let us consider two random multivariate normal variables X, Y. Suppose $X = (X_1, \ldots, X_m)$ and $Y = (Y_1, \ldots, Y_l)$. We can divide the Fisher information matrix and the covariance matrix $\Sigma = FI^{-1}$ into components corresponding to X and Y

$$\Sigma = \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{pmatrix} \tag{26}$$

and

$$FI = \begin{pmatrix} FI_{XX} & FI_{XY} \\ FI_{YX} & FI_{YY} \end{pmatrix}. \tag{27}$$

The mutual information formula for the multivariate normal can be easily computed using one of the following well-known formulae [151]

$$I(X, Y) = -\frac{1}{2} \log \left(\frac{|\Sigma|}{|\Sigma_{XX}||\Sigma_{YY}|} \right) \tag{28}$$

or

$$I(\theta_X, \theta_Y) = -\frac{1}{2} \sum_{j=1}^{k} \log(1 - (\rho_j)^2) \tag{29}$$

where $|\cdot|$ denotes the determinant of a matrix and ρ_i are canonical correlations calculated from Σ. The formulae (28) and (29) have however two strong disadvantages: involve inversion of the FI, and requires division by the matrix determinant, which can be close to zero.

In the article [2] we have shown that Eq.28 can be reformulated to avoid inversion of the FI. Precisely, we prooved that

$$\frac{|\Sigma|}{|\Sigma_X||\Sigma_Y|} = \frac{|FI|}{|FI_X||FI_Y|}. \tag{30}$$

Therefore, the mutual information (Eq. 29) can be reformulated as

$$I(X, Y) = -\frac{1}{2} \sum_{j=1}^{k} \log(1 - (\rho_j^{FI})^2), \tag{31}$$

where ρ_i are canonical correlations calculated directly from FI. These are defined as

$$\rho_i = \max_{w_X^i, w_Y^i} \left(\frac{{w_X^i}^T FI_{AB} w_Y^i}{\sqrt{{w_X^i}^T FI_{AA} w_X^i {w_Y^i}^T FI_{BB} w_Y^i}} \right) \tag{32}$$

subject to $(i \neq j)$

$$w_X^{i\,T} FI_{AA} w_X^j = 0,$$

$$w_Y^{i\,T} FI_{BB} w_Y^j = 0,$$

and

$$w_X^{i\,T} FI_{AB} w_Y^j = 0,$$

for $i = \min(m, l)$ and $j < i$.

3.2 HIERARCHICAL CLUSTERING

Clustering aims to find *similar* objects and assign them into groups, called clusters. Clustering is an unsupervised classification method, which goal is to find natural centers for a set of unlabeled objects. Clustering is fundamental statistical and data exploratory analysis used in multiple areas, including pattern recognition, image analysis, information retrieval, bioinformatics, data compression, computer graphics, and machine learning.
Hierarchical clustering is one of the types of clustering. Contrary to many other methods, hierarchical clustering does not have a pre-defined number of clusters. In general, we distinguish two clustering strategies agglomerative and divisive (see Figure 5).
Agglomerative clustering is also called bottom-up. Therein, each object starts as a one-element cluster. Iteratively pairs of the most similar clusters are merged as one and move up the hierarchy.
Divisive clustering is commonly known as a top-down approach. In this case, all objects are grouped in one cluster, recursively divided into the most distinctive clusters, which moves down the hierarchy.
In both approaches, the result is a hierarchy of nested clusters in both cases, usually presented in a dendrogram. A dendrogram is a multi-level hierarchy diagram representing a tree in which clusters from one level are linked and form larger clusters of the higher level.
The hierarchical clustering algorithm requires defining a measure of similarity between sets of objects. Usually, this is done by defining a proper metric function (e.g., Euclidean and Manhattan distance), a linkage criterion that specifies the similarity of sets as a function of the pairwise distances of objects in those sets (e.g., simple/complete-linkage clustering).

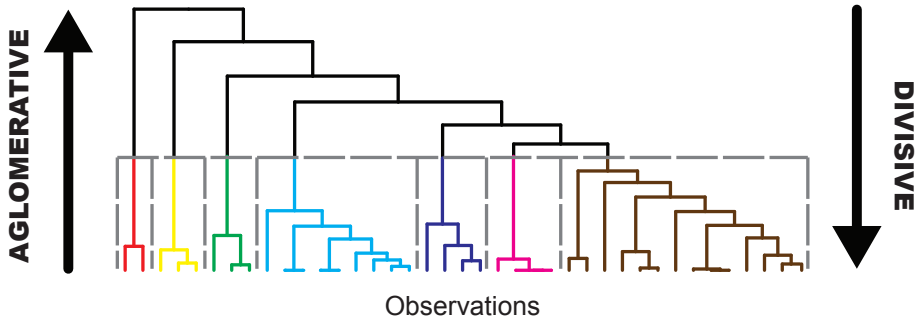


Figure 5: Representation of the hierarchical clustering with a diagram called a dendrogram. The bottom points (nodes) represent observations. The connections with black lines determine the clustering results. The height of the linkage is proportional to the dissimilarity between connected groups of parameters. The horizontal dashed line represents the cut-off that forms clusters of the most similar observations. Clusters are represented with colors. The arrows on the left and right sides of the dendrogram represent the clustering strategies, agglomerative and divisive, respectively.

3.3 GENERALISATION OF SHANNON INFORMATION IN THE RÉNYI SENSE

Eight years after publication of Shanon's concept, in 1956 Alfred Rényi generalized Shannon's entropy by questioning one of the Shannon's axioms [152]. Precisely, Shannon assumed ammong others that for any distribution $P = (p_1, p_2, \ldots, p_n)$ and for $t \in [0, 1]$, the entropy measure should follow a formula

$$H[((t \cdot p_1, (1-t) \cdot p_1, p_2, \ldots, p_n)] = H[(p_1, p_2, \ldots, p_n)] + H[(t, 1-t)]. \quad (33)$$

Rényi proposed to replace this axiom with a weaker additivity condition. Additivity of the entropy indicates that a combined experiment consisting of the performance of two independent experiments is equal to the sum of the entropies of these two experiments. Formally, for distributions P and Q, the additivity is expressed as

$$H[P * Q] = H[P] + H[Q]. \quad (34)$$

Following this concept, Rényi proposed a novel parametrized entropy measure, H_α, where the $\alpha \in [0, \infty]$. Generally speaking, α quantifies the penalty for the deviation from the scenario of equiprobable values, which maximize the entropy regardless of α. Particularly, for $\alpha = 1$ Rényi entropy takes the Shannon form.
The definition of the differential Rényi entropy, also called α entropy, is based on the generalization of the relative entropy (Kullback Leibler divergence) [153].

Definition 3.3.1 (Rényi Divergence)**.** *Let X be random variable, whereas P and Q be a probability distibutions with density functions $p(x)$ and $q(x)$, respectively. Then the Rényi divergence, D_α, between the probability distributions P and Q of a random variable X is defined as*

$$D_\alpha(P||Q) = \frac{1}{1-\alpha} \log \int \frac{p(x)^{1-\alpha}}{q(x)^{1-\alpha}} p(x) dx, \quad (35)$$

where $\alpha \in (0, 1) \cup (1, \infty)$ denotes order of Rényi divergence. Moreover, the Rényi divergences of orders $\alpha \in \{0, 1, \infty\}$ are defined as [154]

$$\begin{aligned} D_0(P||Q) &= \lim_{\alpha \downarrow 0} D_\alpha(P||Q) = -\log Q(p > 0), \\ D_1(P||Q) &= \lim_{\alpha \uparrow 1} D_\alpha(P||Q) = D(P||Q), \\ D_\infty(P||Q) &= \lim_{\alpha \uparrow \infty} D_\alpha(P||Q) = \log \left(ess\ sup_P \frac{p}{q} \right), \end{aligned} \quad (36)$$

where ess sup denotes the essential supremum[1].

Rényi divergence follows many properties of the Kullback-Leibler divergence. For instance, $D_\alpha(P||Q)$ is nonnegative for all $\alpha \geqslant 0$ [153]. Moreover,

1 The *essesntial supremum* of any random variable X with respect to the probability distribution P is defined as $\text{ess sup}_P(X) = \sup\{c | P(X > c) > 0\}$

for $\alpha \in (0, \infty)$ the Rényi divergence is equal to 0 if and only if $P_X(\cdot) = Q_X(\cdot)$. However, for $\alpha \to 0$ and $\alpha \to \infty$ $D_\alpha(P\|Q)$ is not sufficient to claim the equality of $P_X(\cdot)$ and $Q_X(\cdot)$. Morevover, the Rényi divergence is nondecreasing in α [153], i.e.,

$$D_0(P\|Q) \leqslant D_\alpha(P\|Q) \leqslant D_\infty(P\|Q) \text{ for any } \alpha \in (0, \infty). \tag{37}$$

Given the definition of Rényi divergence, we can introduce the notion of Rényi entropy.

Definition 3.3.2 (Rényi Entropy)**.** *Let* X *be a random variable on sample space* $\mathcal{X}$*, that is an interval* I *of length* n*. Let* P *be a probability distribtuion with density functions* $p(x)$*, that has a support in an interval* I*. Then the Rényi entropy of order* $\alpha \in [0, \infty]$ *of random variable* X *described by probability distribution* P *is defined as*

$$H_\alpha(X) = \log n - D_\alpha(P\|U_I), \tag{38}$$

where U_I *denotes the uniform distribution on* I*.*

The Rényi's entropy is monocite in α, which comes directly from the Eq. 38 and fact that $D_\alpha(\cdot\|\cdot)$ is nondecresaing in α Eq.37. Precisely, $H_\alpha(\cdot)$ is nonincreasing in α, i.e.,

$$H_\infty \leqslant H_\alpha \leqslant H_0. \tag{39}$$

Rényi entropy properties lead to more convenient formulae for different Rényi orders.

$$H_0(X) = \log|\{x \in \mathcal{X} | p(x) > 0\}| \qquad \text{Hartley or max-entropy} \tag{40a}$$

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \int (p(x))^\alpha dx \qquad \alpha \in (0,1) \cup (1, \infty) \tag{40b}$$

$$H_1(X) = H(X) \qquad \text{Shannon entropy} \tag{40c}$$

$$H_\infty(X) = \min_{x \in \mathcal{X}} -\log p(x) \qquad \text{min-entropy} \tag{40d}$$

Many properties of Shanon's entropy are not generalised. For instance, unlike Shanon's entropy, $H_\alpha(X)$ (for $\alpha \neq 1$) suffers from the disadvantage that the ineqaultity $H_\alpha(X_1, X_2) \leqslant H_\alpha(X_1) + H_\alpha(X_2)$ does not hold in general [155]. Moreover, if we mimic the definition of the conditional entropy, $H(X|Y)$, and define conditional Rényi entropy as $\int_y p(y) H_\alpha(X|Y = y) dy$, then the conditional version may be larger than $H_\alpha(X)$ [156]. Hence, the definitions of the mutual information and the channel capacity also cannot be translated one-to-one to Rényi version.

Due to the importance of the Rényi generalizations, [157] there have been proposed multiple remedies for this situation. We are particularly interested in the definition of generalized mutual information. Below, we review two major approaches to accomplish such generalization [157]. All of them have been designed for the application of the coding theorem. Hence, some of them are defined only to the discrete input alphabet.

To be consistent, here we consider the memoryless channel $W_d : X \to Y$ described by the by a conditional probability distribution $P(Y|X)$, where X is a random variable on finite discrete alphabet $\mathcal{X}$ ($M = |\mathcal{X}|$) and Y is a random

variable on alphabet $\mathcal{Y}$. Limitation to the discrete memoryless channel is not problematic, as the signaling pathways are usually analyzed for a limited range of the input (signal) values.
The different definitions of generalized Rényi's mutual information have been derived from different representations of their Shannon analog, i.e.,

$$\begin{aligned} I(X,Y) &= H(X) + H(Y) - H(X,Y) && \text{(41a)}\\ &= H(X) - H(X|Y) && \text{(41b)}\\ &= D(P_{Y|X}P_X \| P_Y \times P_X) && \text{(41c)}\\ &= D(P_{Y|X} \| P_Y | P_X) && \text{(41d)}\\ &= \min_{Q_Y} D(P_{Y|X}P_X \| Q_Y \times P_X) && \text{(41e)}\\ &= \min_{Q_Y} D(P_{Y|X} \| Q_Y | P_X), && \text{(41f)} \end{aligned}$$

where minimizations are taken over unconditional distributions on $\mathcal{Y}$.

3.3.1 Arimoto-Rényi mutual information

Suguru Arimoto drew his inspiration from the Eq. 41b. However, as it was mentioned, there is no simple Rényi equivalent of conditional entropy. Therefore, Arimoto proposed a novel definition of the conditional entropy [158].

Definition 3.3.3 (Conditional Arimoto-Rényi entropy)**.** *Let us consider the channel W_d. Then the conditional Arimoto-Rényi entropy, $H_\alpha^a(X|Y)$, for $\alpha \in (0,1) \cup (1,\infty)$ is defined as*

$$\begin{aligned} H_\alpha^a(X|Y) &= \frac{\alpha}{1-\alpha} \log \left(\mathbb{E}\left[\| P_{X|Y}(\cdot|Y) \|_\alpha \right] \right) && \text{(42a)}\\ &= \frac{\alpha}{1-\alpha} \log \left(\mathbb{E}\left[\left(\sum_{x \in \mathcal{X}} \left(P_{X|Y}(x|Y) \right)^\alpha \right)^{-\alpha} \right] \right), && \text{(42b)} \end{aligned}$$

where $\| \cdot \|_\alpha$ denotes the α-norm. By its continuous extension, the conditional Arimoto-Rényi entropy of orders 0,1 and ∞ are defined as

$$\begin{aligned} H_0(X|Y) &= ess\,sup H_0(P_{X|Y}(\cdot|Y)), && \text{(43a)}\\ H_1(X|Y) &= H(X|Y), && \text{(43b)}\\ H_\infty(X|Y) &= -\log \mathbb{E}\left[\max_{x \in \mathcal{X}} P_{X|Y}(x|Y) \right], && \text{(43c)} \end{aligned}$$

where ess sup denotes the essential suprememum.

Having the definition of the conditional Rényi entropy, we can define the Arimoto-Rényi mutual information [158].

Definition 3.3.4 (Arimoto-Rényi mutual information)**.** *Under the assumptions as in 3.3.3, the Arimoto-Rényi mutual information of the channel W_d is defined as*

$$I_\alpha^a(X,Y) = H_\alpha(X) - H_\alpha^a(X|Y). \tag{44}$$

The conditional entropy in Arimoto sense has an interesting interpretation because of its relation to the Bayesian hypothesis testing [155]. Following the article *Sason et. al.* [155], let us consider the aforementioned maximum-a-posteriori decision rule (MAP). Then, let denote ε_X as the minimum probability of error in guesing the value of random variable X. According to the MAP approach and the Eq.40d ε_X is equal to [155]

$$\varepsilon_X = 1 - \exp\left(-H_\infty^a(X)\right) \tag{45}$$

Then, the the minimum probability of error of X given Y, $\varepsilon_{X|Y}$, can be achieved by a deterministic function $\mathcal{L}^* : \mathcal{Y} \to \mathcal{X}$:

$$\begin{aligned}\varepsilon_{X|Y} &= \min_{\mathcal{L}:\mathcal{Y}\to\mathcal{X}} \mathbb{P}\left[X \neq \mathcal{L}(Y)\right] \\ &= \mathbb{P}\left[X \neq \mathcal{L}^*(Y)\right] \\ &= 1 - \mathbb{E}\left[\max_{x\in\mathcal{X}} P_{X|I}(x|Y)\right] \\ &= 1 - \exp\left(-H_\infty^a(X|Y)\right).\end{aligned} \tag{46}$$

3.3.2 Csiszár-Rényi mutual information

Instead of defyining the conditional Rényi entropy, Csiszár proposed to derive the generelisation from the Eq. 41f.

Definition 3.3.5 (Csiszár-Rényi Mutual Information)**.** *Let consider the memoryless channel $W : X \to Y$ described by a conditional probability distribution $P(Y|X)$, where X and Y are random variables taking values in $\mathcal{X}$ and $\mathcal{Y}$, respectively. Then, the Csiszár-Rényi mutual information of order α of channel W is defined by*

$$I_\alpha^c(X,Y) = \inf_{Q_Y} \mathbb{E}_X\left[D_\alpha\left(P_{Y|X}(\cdot|X)||Q_Y\right)\right], \tag{47}$$

where the expectency is taken over the $P(X)$ and the infinium is taken over such probability distributions Q_Y, which $Supp Q_Y = \mathcal{Y}$. If the probability distribution $P(X)$ has a density function $p(x)$, then Eq.47 takes form

$$I_\alpha^c(X,Y) = \inf_{Q_Y} \int_{\mathcal{X}} D_\alpha\left(P_{Y|X}(\cdot|X)||Q_Y\right) p(x)\,dx \tag{48}$$

The channel capacity of the Csiszár-Rényi mutual information takes a form

$$C_\alpha^c = \sup_{P(X)} \inf_{Q_Y} \mathbb{E}_X\left[D_\alpha\left(P_{Y|X}(\cdot|X)||Q_Y\right)\right], \tag{49}$$

whereas its differential version is equal

$$C_\alpha^c = \sup_{p(X)} \inf_{Q_Y} \int_{\mathcal{X}} D_\alpha\left(P_{Y|X}(\cdot|X)||Q_Y\right) p(x)\,dx. \tag{50}$$

A major shortcoming of Csiszár's definition of the mutual information is that the minimization therein is hard to solve analytically even for simple toy examples. Nevertheless, the Csiszár channel capacity plays an important role in the coding theory. It has been shown (sequently for different orders

$\alpha = 1$ [159, 160], $\alpha \in (0, \infty)$ [154], and $\alpha \in \{0, \infty\}$ [153]), that for finitie $\mathcal{Y}$ the Csiszár's channel capacity equals the *minimax Rényi divergence*

$$R_\alpha = \inf_{Q_Y} \sup_{x \in \mathcal{X}} D_\alpha(P_{Y|X} \| Q_Y). \tag{51}$$

A minimax Rényi divergence characterizes the redundancy for universal lossless compression in Campbell's setting [161].
Similarly to the Shannon approach, the distibution $P(X)$ for which C_α^c gain its maximum is called optimal or *capacity achievieng* and it is denoted $P^*(X)$ or $P_{opt}(X)$. Analogously, a distribution for which R_α obtain its minmum is calle *redundancy optimal* or *redundancy achievieng*, and is denoted $Q_{opt}(Y)$. Moreover, for finite output space $\mathcal{Y}$ a redundancy achieving distribution always exists [153].

3.4 RÉNYI–MIN INFORMATION

In the section above, we have introduced the generalization of the Shannon information proposed by Alfred Rényi. Here, we describe some properties of the Rényi-∞ information that were used in this Ph.D. thesis. The Rényi-∞ information derives from the Rényi-∞ entropy, H_∞, which characterize with the lowest penalty for the deviation from the scenario of equiprobable values among all Rényi's entropies. Therefore, it is commonly named Rényi-min information. For convenience, we will use both nomenclatures.

3.4.1 Csiszár's-Rényi min-capacity as the maximum of the minimax Rényi divergence

The minimax redundancy (Eq. 51) for $\alpha = \infty$ with assumption of countable $\mathcal{Y}$ adopts a simpler formula [161, 153]. Precisely, if we reformulate the minimax redundancy

$$\begin{aligned} R_\infty &= \inf_{Q_Y} \sup_x D_\infty(P_{Y|X} \| Q_Y) \\ &= \inf_{Q_Y} \sup_x \log \sup_y \frac{P_{Y|X}(y)}{Q_Y(y)} \\ &= \inf_{Q_Y} \sup_y \log \frac{\sup_x P_{Y|X}(y)}{Q_Y(y)} \\ &= \inf_{Q_Y} \left\{ \sup_y \left\{ -\log Q_Y(y) - \inf_x -\log P_{Y|X}(y) \right\} \right\}, \end{aligned} \tag{52}$$

we obtain the definition of the worst-regret in terms of the Minimum Description Length (MDL) [162].
Briefly, MDL is a formalization of the so-called "Occam's razor". In 13^{th} century, Franciscan friar William of Ockham, a scholastic philosopher, postulated that *"Entities should not be multiplied without necessity."*. In general, Occam's razor states that when presented with competing hypotheses that make the exact predictions, one should select the solution with the fewest

assumptions [162, 163]. However, it is not a way of choosing between hypotheses that make different predictions. The formalization of Occam's razor principle in the information-theoretic language is called MDL [162, 164]. According to this principle, the "best" hypothesis, i.e., a model together with its parameters, for a given set of data, is the one that leads to the best compression of the data. The "best" model minimizes the total number of bits needed to encode the model and the data given the model.

The problem of measuring maximal information transfer in the memoryless channel $X \to Y$ can be defined in terms of the MDL [162]. In this case model, $\mathcal{M}$, consists of the probability distributions of the output depending on the signal, i.e., $\mathcal{M} = \{p_1 = P(Y|X = x_1), \ldots, P(Y|X = x_k)\}$. Then, the regret of some code length function $L(\cdot)$ with respect to the model $\mathcal{M}$ on n-element vector of data y^n, $y^n \in \mathcal{Y}^n$, is defined as

$$R(\mathcal{M}, L, y^n) = L(y^n) - \inf_{i\in\{1,\ldots,k\}} -\log P(Y = y^n|X = x_i). \tag{53}$$

Then, the worst-regret equals $\sup_{y^n\in\mathcal{Y}^n} L(y^n) - \inf_{i\in\{1,\ldots,k\}} -\log P(Y = y^n|X = x_i)$. Therefore, with $L(y^n) = -\log Q_Y(y^n)$ the minimax Rényi divergence for $\alpha = \infty$ equals the minimal worst-regret

$$R_\infty = \inf_{Q_Y}\{R(\mathcal{M}, L, y^n)\}. \tag{54}$$

However, we know that the worst-regret is minimized with Shratkov distribution [162]. Thus, for $Q_Y = p_{NML}$ the minimax Rényi divergence is maximized and takes the form

$$R_\infty = \sum_{y^n\in\mathcal{Y}^n} \sup_{i\in\{1,\ldots,k\}} P(Y = y^n|X = x_i). \tag{55}$$

Therefore, we obtain a simple formula for calculating the channel capacity in the Rényi $\alpha = \infty$ sense.

3.4.2 Optimal input distribution of Rényi min-information. Equivalence of the Rényi-min capacity in the Arimoto's and Csiszár's sense

As mentioned above, one of the factors that affect mutual information is the input distribution, $P(X)$. The selection of the input distribution suitable for quantifying information transfer in a specific application can be problematic and provides a degree of arbitrariness. Therefore, typically, we consider the input distribution, which maximizes the mutual information. Hence, it is called the optimal distribution. Nevertheless, in the case of the Shannon information, the optimal distribution usually consists of only the most extreme signals and omits intermediate ones. However, such a repertoire of signals is relatively rare in biochemical systems. It seems more natural to consider a situation where the signal comes from some range. Here, we show that the Rényi-min information in the Arimoto sense is maximized for the equiprobable input distribution. Hence, this property both eliminates the arbitrariness of the input distribution choice and takes into account more realistic prior

conditions. In addition, we prove the equivalence of the Arimoto and Csiszár formulations on the level of channel capacity.

Theorem 3.4.1. *Let's* C_∞^a *denote maximal Arimoto-Rényi mutual information, i.e.* $C_\infty^a = \max_{P_X} I_\infty^a(X,Y)$. *Then the following statements are true:*

1. $I_\infty^a(X,Y)$ *is maximised for equiprobable distribution* $P(X)$.

2. $C_\infty^a = C_\infty^c$,

Proof.

$$\begin{aligned} I_\infty^a(X,Y) = H_\infty(X) - H_\infty(X|Y) = \log \frac{\mathbb{E}_Y\left[\max_x P(x|Y)\right]}{\max_x P(x)} = \\ = \log \frac{\sum_y \max_x P(x|y)P(y)}{\max_x P(x)} = \log \frac{\sum_y \max_x P(x,y)}{\max_x P(x)} = \\ = \log \frac{\sum_y \max_x P(y|x)P(x)}{\max_x P(x)} \leqslant \log \frac{\sum_y \max_x P(y|x)\max_x P(x)}{\max_x P(x)} = \\ = \log \sum_y \max_x P(y|x) \frac{\max_x P(x)}{\max_x P(x)} = \log \sum_y \max_x P(y|x) \end{aligned} \tag{56}$$

The inequality above define the upper bound of the mutual information defined by Arimoto. Let's notice that for equiprobable distribution $P(X) = \frac{1}{|\mathcal{X}|} = \frac{1}{M}$ the upper bound is reached, i.e.

$$\begin{aligned} I_\infty^a(X,Y) = \dots \\ = \log \frac{\sum_y \max_x P(y|x)P(x)}{\max_x P(x)} \\ = \log \frac{\sum_y \max_x P(y|x)\frac{1}{M}}{\frac{1}{M}} \\ = \log \sum_y \max_x P(y|x). \end{aligned} \tag{57}$$

Hence, the probability distribution $P(X) = \frac{1}{M}$ is an optimal distribution. Therefore, the Arimoto-Rényi-min channel capacity is equal to

$$C_\infty^a(X,Y) = \log \sum_y \max_x P(y|x). \tag{58}$$

Moreover, let's notice, that right side of the Eq.59 is equal to the worst-case redundancy (Eq. 52),

$$C_\infty^a(X,Y) = \log \sum_y \max_x P(y|x) = R_\infty^c(X,Y) = C_\infty^c(X,Y), \tag{59}$$

hence, the equivalence between Arimoto and Csiszár Rényi-min channel capacities. □

According to the equivalence of Arimoto and Csiszár-Rényi-min channel capacities, for the simplicity of the notation, below we denote the Rényi-min

channel capacity as $C_\infty(X,Y)$ or $C_{min}(X,Y)$. The proof above provides the unified formula of the value of Rényi-min channel capacity, i.e.,

$$C_{min}(X,Y) = \log \sum_y \max_x P(y|x). \tag{60}$$

Operating between two definitions of mutual information might be confusing. Their equivalence applies only to the formula and the value of the channel capacity. Nevertheless, the optimality of the input distributions as well as other properties of the mutual information remains disjunctive. For instance, the Theory 3.4.1 shows that the Rényi-min information in Arimoto sense is maximized for the equiprobable input distribution. However, it does not tell that this is the only possible optimal input distribution. On the other hand, the Rényi-min capacity in Csiszár sense does not depend on the input distribution. Therefore, we should be careful while we operate with the notion of the Rényi-min capacity.

We consider only the equiprobable input distribution in the further analysis as this distribution maximizes both Rényi-min information measures. The optimality of equiprobable distribution does not depend on the structure of the output distributions. Notably, it does not deprecate signals that give overlapped output distributions. It is opposite to the Shannon information, which strongly favors signals, for which output is the most distinct. According to Shannon approach, each signal, which gives output, that overlap with other should be strongly penalized. This difference may lead to significantly different conclusions in the analysis of the information transfer in signaling pathways.

3.4.3 Arimoto-Rényi min-information as an uncertainty measure.

This section introduces a functional relation between Rényi-min capacity and the uncertainty in the input discrimination based on the output data. Here, we consider the mutual information in the sense of Aimoto, $I^a_{min}(X,Y)$. Firstly, Rényi-min entropy defined as

$$H_{min}(X) = \min_{x\in\mathcal{X}} -\log p(x), \tag{61}$$

can be interpreted in terms of the task of guessing a signal. Indeed, according to Eq.45, H_{min} determines the probability of error in guessing the value of random variable X, i.e., $\varepsilon_X = 1 - \exp\left(-H_{min}(X)\right)$. According to Eqs. 46 and 44 analogous relation can be obtained for the Arimoto-Rényi-min mutual information. Precisely, let us denote $\varepsilon_{X|Y}$ as the minimum probability of error of guessing X given Y, then the following formula occurs

$$\varepsilon_{X|Y} = 1 - \exp\left(I^a_{min}(X|Y) - H_{min}(X)\right). \tag{62}$$

If we assume the equprobability of the input distribution, then

$$\begin{aligned} C_{min}(X,Y) &= \log(M) + \log(1-\varepsilon_{X|Y}) \\ &= \log\left(M\left(1-\varepsilon_{X|Y}\right)\right), \end{aligned} \tag{63}$$

as the Arimoto-Rényi-min information takes the maximal value in such condition. From Eq. 63 we get the convenient formulat of $\varepsilon_{X|Y}$, as

$$\varepsilon_{X|Y} = 1 - \frac{1}{M} 2^{C_{min}(X,Y)}. \tag{64}$$

The $\varepsilon_{X|Y}$ denotes the minimum probability of error in MAP decision rule, then $1 - \varepsilon_{X|Y}$ can be interpreted as maximal success, i.e., maximal probability of the correct discrimination of X based on Y. Therefore, for equiprobable input distribution, the exponent of the Rényi-min capacity, $2^{C^*_{min}}$, can be interpreted as the number of inputs that the system can resolve on average. Therefore, in the Rényi-min approach, the number of resolved input states is proportional to the maximal probability of the correct discrimination. The property above clearly indicates the Rényi-min channel capacity to measure the uncertainty of the input decoded by the observer from the output.

4 OBJECTIVES

4.1 IDENTIFICATION OF THE POTENTIAL OF SENSITIVITY ANALYSIS PARAMETRIC METHODS FOR STUDIES OF CELLULAR SIGNALING

Sensitivity analysis provides rigorous tools to investigate how parameter values impact quantitative characteristics of model behaviour [165]. It allows to obtain a deeper understanding of studied phenomena as well as helps in manipulating complex models. Basic sensitivity analysis methods have been developed to analyze engineering or physics problems that differ significantly from dynamical models in quantitative biology. Typically quantitative biology models depend on a relatively larger number of parameters; hence sensitivity analysis methods need to be adjusted to solve biochemical problems.

In the chapter *"Sensitivity Analysis in quantitative biology models"* [1] of the book *"Quantitative Biology: Theory, Computational Methods and Examples of Models"* [9], **we have introduced methods of sensitivity analysis in the context of quantitative biology**. We presented a small number of the primary techniques without superfluous technical details. The focus on major methodological concepts aims to help understand how sensitivity analysis can address specific questions arising in modeling practice. Typically, methods of sensitivity analysis are presented only in the deterministic scenario. Here, all of the methods were also described in the stochastic regime.

4.1.1 Local sensitivity analysis methods

Firstly, we have introduced the local methods of sensitivity analysis. Briefly, local methods indicate how small changes of the parameters value Δ, with respect to some nominal value, θ^*, influence the model's response (see Eq. 13). Therein, we introduced local sensitivity analysis in both deterministic and stochastic regimes. Assuming the the deterministic approach (as in Eq. 1), the sensitivity to parameter θ_i is defined by a vector of derivatives $SV_j = \frac{\partial X}{\partial \theta} = \left(\frac{\partial \mathbf{x}(t_1)}{\partial \theta_i}, \ldots, \frac{\partial \mathbf{x}(t_n)}{\partial \theta_i}\right)^T$, where each $\frac{\mathbf{x}(t_j)}{\partial \theta_i}$ is a N-dimmnesional row vector. A collection of sensitivity vectors for $i = 1, \ldots, l$ constitutes the sensitivity matrix

$$\mathbf{SM} = (\mathbf{SV}_1, \ldots, \mathbf{SV}_l)\,. \tag{65}$$

In the chapter, we have presented an interpretation of those factors as well as methods of their computation.

In the stochastic regime (as in Eq. 2), coventional derivative of the output cannot be defined. In the chapter we presented different approaches to evaluate local sensitivites, i.e.,(i) derivate of the mean, $\frac{\mathbb{E}X}{\partial \theta}$, (ii) derivateive of the vari-

ance, $\frac{\mathbb{E}(X-\mathbb{E}X)^2}{\partial\theta}$, and (iii) the Fisher Inormation Matrix, FI (see Eq. 14). The FI quantifies changes in the probability distribution due to joint changes of all model parameters. The diagonal elements of the FI are used to calculate local sensitivity coefficients,

$$\mathbf{SL}_i = \sqrt{FI_{i,i}}. \tag{66}$$

However, in the information theory, FI also measures the amount of information that random variable X carries about an unknown parameter θ. Therefore, the sensitivity coefficient tells us how much information we gain from experimental measures about the model's parameters.

Although the local sensitivity coefficients in the deterministic and stochastic scenarios exhibit many similarities, significant differences exist. The matrix $\mathbf{SM}^2$ gives sensitivities in units of the output, whereas the FI measures the relative change of unit-less probability density values. As a result, the deterministic and stochastic scenarios cannot be compared directly with each other [166].

In general, the calculation of the Fisher information matrix of multi-parameter biochemical models is challenging. However, it is tractable if we assume that the model's output is expressed as the multivariate normal distribution. Following [101] we have precisely described the method of calculation the FI in chemical master equation (CME) models approximated using linear noise approximation (LNA).

4.1.2 Global sensitivity analysis methods

Typically, in engineering, models' parameters come from a specific small range of values. In the context of the multi-parameter biochemical models, sometimes, it is not easy to define the scope of possible parameter values. Moreover, biochemical systems characterize by extrinsic variability, which results from varying components upstream to the system of interest. Varying model parameters can determine extrinsic variability. Here, rather than considering model parameters as point values, they are represented as distributions, i.e., $P(\theta)$.

On the other hand, the local sensitivities are valid only around nominal parameter values, θ^* and may be substantially different from sensitivities calculated for other points in the parameter space, $\theta \in \mathbb{R}^l$. Therefore, it is necessary to explore sensitivities in some range of parameters' values in the biochemical systems. The global sensitivity analysis (GSA) operate with a probability distribution, $\mathcal{P}(\theta)$, of plausible parameter values [130, 167]. With this distribution, one can sample parameter space and assemble a representative collection of model outputs in order to analyze the impact of parameters on the output. How the impact of parameters is assessed varies among methods and depends on the specific aim of the analysis. In the chapter, we have described three commonly used GSA techniques.

The screening technique is a simple generalization of the local method. Here, sensitivity coefficients are calculated for values on a grid, $\mathbf{\Omega}$, that represent

the considered parameter space, $\mathbb{R}^l$. In the screening method, derivative $\frac{\delta Y}{\delta \theta_{i]}}$ is replaced by so-called elementary effects, EE_i

$$EE_i = \frac{dist\left(\mathbf{Y}(\boldsymbol{\theta}), \mathbf{Y}(\boldsymbol{\theta} + \boldsymbol{\Delta}_i)\right)}{\Delta}, \tag{67}$$

where $dist(\cdot)$ is a distance function between two corresponding outputs, e.g., Euclidian distance in deterministic case or Kolmogorov-Smirnov distance in the stochastic scenario. Biochemical models are characterized by a relatively huge number of parameters and a wide range of parameter values; therefore, it is necessary to look precisely through the parameter space to obtain informative sensitivity factors. One of the efficient solutions is a penetration of parameter space using the *Morris Trajectory algorithm*. The Morris algorithm produces a representative sample of elementary effects $\{\mathbf{EE}^{(1)}, \mathbf{EE}^{(2)}, \ldots, \mathbf{EE}^{(r)}\}$, where $\mathbf{EE}^{(j)}$ is a vector of EEs for each parameter that is generated in the j^{th} step of the algorithm. Based on the series of elementary effects, we could estimate the overall influence of the parameter on the output and quantify the extent to which the influence of the parameter varies across the considered grid. Moreover, the screening tells us about the interactions of each parameter with other parameters.

Next, variance-based (or Sobol) methods quantify sensitivity of parameters by computing variance of the output, $\mathcal{P}(\mathbf{Y})$, resulting from sampling the model parameters from the distribution $\mathcal{P}(\boldsymbol{\theta})$. The variance of the distribution $\mathcal{P}(\mathbf{Y})$ can be decomposed into contributions resulting from individual parameters. Using the variance decomposition theorem [168] the overall variance can be written in two ways as the sum of the components corresponding to the parameter θ_i and the remaining sources of variability,

$$\mathrm{Var}\{\mathbf{Y}\} = \underbrace{\mathbb{E}_{\theta_i}\left\{\mathrm{Var}_{\boldsymbol{\theta}_{-i}}\{\mathbf{Y}|\theta_i\}\right\}}_{\text{variability generated by } \theta_i} + \underbrace{\mathrm{Var}_{\theta_i}\left\{\mathbb{E}_{\boldsymbol{\theta}_{-i}}\{\mathbf{Y}|\theta_i\}\right\}}_{\text{other sources}} \tag{68}$$

and

$$\mathrm{Var}\{\mathbf{Y}\} = \underbrace{\mathbb{E}_{\boldsymbol{\theta}_{-i}}\{\mathrm{Var}_{\theta_i}\{\mathbf{Y}|\boldsymbol{\theta}_{-i}\}\}}_{\text{variability generated by } \theta_i} + \underbrace{\mathrm{Var}_{\boldsymbol{\theta}_{-i}}\{\mathbb{E}_{\theta_i}\{\mathbf{Y}|\boldsymbol{\theta}_{-i}\}\}}_{\text{other sources}}, \tag{69}$$

where $\boldsymbol{\theta}_{-i}$ is the vector of all model parameters except θ_i.

The ratio of the conditional variance $\mathrm{Var}_{\theta_i}\left\{\mathbb{E}_{\boldsymbol{\theta}_{-i}}\{\mathbf{Y}|\theta_i\}\right\}$, and the overall variance is used as the so called first-order sensitivity index of θ_i [169]

$$SG_i^S = \frac{\mathrm{Var}_{\theta_i}\left\{\mathbb{E}_{\boldsymbol{\theta}_{-i}}\{\mathbf{Y}|\theta_i\}\right\}}{\mathrm{Var}\{\mathbf{Y}\}}, \tag{70}$$

which estimates how much the total variance of the model output could be reduced if the i-th parameter was fixed. On the other hand, $\mathbb{E}_{\boldsymbol{\theta}_{-i}}\{\mathrm{Var}_{\theta_i}\{\mathbf{Y}|\boldsymbol{\theta}_{\text{-}i}\}\}$ includes the variability that is induced through interactions with other model parameters. It leads to the total sensitivity coefficient [170]

$$SG_i^T = \frac{\mathbb{E}_{\boldsymbol{\theta}_{-i}}\{\mathrm{Var}_{\theta_i}\{\mathbf{Y}|\boldsymbol{\theta}_{-i}\}\}}{\mathrm{Var}\{\mathbf{Y}\}}, \tag{71}$$

which estimates how much total variance would remain if all parameters except θ_i were fixed. Using both coefficients in parallel allows one to understand the total impact of each parameter better. For instance, low SG_i^S and high SG_i^T, in relation to other parameters, suggest that parameter θ_i is sensitive through its interactions with other parameters, but not directly.
Finally, we described the Monte Carlo Filtering (MCF) that quantifies the role of every single parameter in inducing a specifically defined property of model behavior. For instance, it is often of interest to understand the combination of parameter values responsible for generating oscillations or causing a specific response variable to exceed a defined threshold. The method aims to detect to what extent each parameter θ_i determines the analyzed property. This is achieved by sampling parameters from the distribution $\mathcal{P}(\boldsymbol{\theta})$ and obtaining empirical estimates of the probabilities $\mathcal{P}(\theta_i|\text{condition is satisfied})$ and $\mathcal{P}(\theta_i|\text{condition is not satisfied})$, separately for each single parameter θ_i. The result quantifies how likely the value of the condition is due to a specific value of the parameter θ_i. If both probability distributions are identical, then the value of θ_i carries no information about the value of the condition. On the other hand, if these distributions have different supports, then the value of the parameter θ_i solely determines the value of the condition. Therefore the degree to which the distributions differ is a sensitive indicator of the impact of the parameter θ_i on the specified condition.

4.1.3 Application to biological models

All presented methods were applied to two exemplary models: a simple model of gene expression and the minimalistic model of a signaling pathway in both deterministic and stochastic scenarios. The gene expression model enabled us to compare sensitivity analysis results with intuition. The signaling pathway model allowed us to present the usefulness of sensitivity analysis methods and their potential applications in practice.
The chapter introduced the reader to the sensitivity methods that are adjusted to the quantitative biology dynamical models. Utilizing the potential of sensitivity analysis in providing a deeper understanding of biological phenomena relies mainly on developing new techniques that will account for the specificity of biological models and also will be able to provide definite answers to specific biological questions. The development of new tools that provide more direct answers to particular questions would make them more attractive for quantitative biologists and better understand modeled biological processes.

4.2 UTILIZATION OF IDENTIFIABILITY ANALYSIS APPROACHES TO EXAMINE MULTIPARAMETER MODELS OF SIGNALING PATHWAYS

Despite the usefulness of methodological developments, identification of identifiability parameters in a multi-parameter biochemical model still constitutes a substantial challenge. In the article, [2] **we have proposed a natural, universal and simple measure to quantify the similarity between groups of model parameters that combines canonical correlation analysis (CCA) with mutual information (MI)**. Based on this measure, **we have introduced a novel definition of asymptotic parameter identifiability that corresponds to well-known definitions based on profile likelihoods**. Moreover, **we applied hierarchical clustering to identify the maximal number of identifiable parameters**. In the article, we have shown its usefulness in meaningful experimental planning.

4.2.1 Similarity measure between groups of parameters

In the artile, we have proposed to use the canonical correlation analysis to measure similarity between groups of parameters (see Methods for details Section 3.1 and Figure 6A). Therein, the similarity between two groups of parameters $\theta_A = \{\theta_{i_1}, \ldots, \theta_{i_a}\}$, $\theta_B = \{\theta_{j_1}, \ldots, \theta_{j_b}\}$ is described as a canonical correlation (CC) between two sets of sensitivity vectors $\Omega_A = \{S_{i_1}, \ldots, S_{i_a}\}$, $\Omega_B = \{S_{i_1}, \ldots, S_{i_b}\}$. CCs form a set of correlation coefficients defined recursively. The first CC, ρ_1, is a maximal cosine between a linear combination, u_1, in Ω_A and a linear combination, v_1, in Ω_B, $\rho_1 = \cos(u_1, v_1)$. Each next CC is found in the same way under the constraint that the next linear combination must be orthogonal to these found in the previous steps. Repeating the procedure $m = \min(i_a, j_b)$ times provides a set of CCs $1 \geqslant \rho_1 \geqslant \ldots \geqslant \rho_m \geqslant 0$. The value of 1 indicates that there exists a linear combination of parameters in θ_A and θ_B having an identical impact, whereas 0 indicates existence of an orthogonal parameter combination. The CCs therefore provide an m-dimensional similarity measure between θ_A and θ_B. With specific assumption the CCs are closely related with the mutual information between the maximal likelihood estimates $\hat{\theta_A}$ and $\hat{\theta_B}$ (Eq. 31) (see Figure 6B), i.e.,

$$I(\theta_A, \theta_B) = H(\hat{\theta}_A) - H(\hat{\theta}_A|\hat{\theta}_B) = -\frac{1}{m}\sum_i^m \log(1-\rho_i^2). \tag{72}$$

The mutual information $I(\theta_A, \theta_B)$ describes the reduction of the entropy, $H(\hat{\theta}_A)$, of the estimate $\hat{\theta}_A$ resulting from knowledge of $\hat{\theta}_B$. The more similar θ_A and θ_B are, the more knowledge one will help determine the value of the other. Therefore, the mutual information $I(\theta_A, \theta_B)$ is an adequate measure for quantifying the overall similarity between parameter groups. In the article, the similarity measures $I(\theta_A, \theta_B)$ is called *MI-CCA*, as it combines two interpretations: canonical correlation analysis of groups of parameters with mutual information their estimates.

4.2.2 (δ, ζ)-identifiability

Based on the similarity measure MI-CCA we have proposed a novel definition of identifiability of individual parameters in multi-parameters models. This definition takes into account different sources of non-identifiability of parameters, i.e., the lack of their sensitivity, and structural or practical compensation effects of a parameter by remaining model parameters [171, 172, 173, 174, 175, 176, 177, 178, 179].

In this definition, the parameter θ_i is said to be (δ, ζ)-*identifiable* if $\rho(\theta_i, \theta_{-i}) < \delta$ and $\|S_i\| > \zeta$, where θ_{-i} denotes the remaining model parameters. Correlation ρ is used here in the canonical sense (as in CCA). If θ_i was estimated as a single model parameter, ζ-condition requires its asymptotic variance to be smaller than $1/\zeta$. The δ-condition requires the parameter not to be correlated with any linear combination of the remaining parameters by more than δ. Moreover, this condition translates into demanding that the variance of estimates does not increase by more than $1/(1-\delta^2)$ when the single parameter and multi-parameter scenarios are compared.

The (δ, ζ)-identifiability can be seen as an asymptotic version to the profile likelihoods based identifiability (see Eq.12). Let us denote, $APL(\theta_i)$ as the asymptotic profile likelihood, $APL(\theta_i) = \max \theta_{-i} (L(\theta))$. Assuming general reularity conditons it can be prooved that $PL_n(\theta_i) \to APL(\theta_i)$. In the article [2], we have shown that asymptotic profile likelihood, $APL(\theta_i)$, can be reformulated as

$$APL(\theta_i) = L(\theta^*) - \frac{1}{2} FI_{ii}(1-\rho^2)(\theta_i - \theta_i^*). \tag{73}$$

The assumption of (δ, ζ)-identifiability imposes conditions on the curvature of APL as the signless, second order derivative with respect to θ_i

$$\kappa = \left| \frac{\partial^2 APL(\theta_i)}{\partial \theta_i^2} \right| = FI_{ii}(\theta^*) \left(1 - \rho^2\right). \tag{74}$$

The contribution resulting from FI_{ii} must be greater than ζ, and the contribution coming from correlation must be more significant than $(1-\delta^2)$. The actual value of FI_{ii} can be neglected as it can be artificially inflated by scaling of the parameter or repeating the same experiment several times. The correlation ρ is affected by neither of these factors.

The identifiability based on profile likelihoods [126] is determined by confidence intervals $\{\theta_i : |\chi^2_{PL,i}(\theta_i) - \chi^2_{PL,i}(\hat{\theta}_i)| < \Delta\}$, where $\hat{\theta}$ is the argument maximising $L_n(\theta)$, and Δ is a constant selected based on χ^2 statistics. If confidence interval extends infinitely parameter is non-identifiable.

The two methods are applicable in different practical situations. If experimental data, X, is available and $L_n(\theta)$ can be calculated, profile likelihood is the superior method. Our method, however, is tailored to deal with a situation where only $L(\theta)$ is available, whereas $L_n(\theta)$ is not. For instance, if data X is not available or $L_n(\theta)$ cannot be evaluated, analyzing APL is the best one can do.

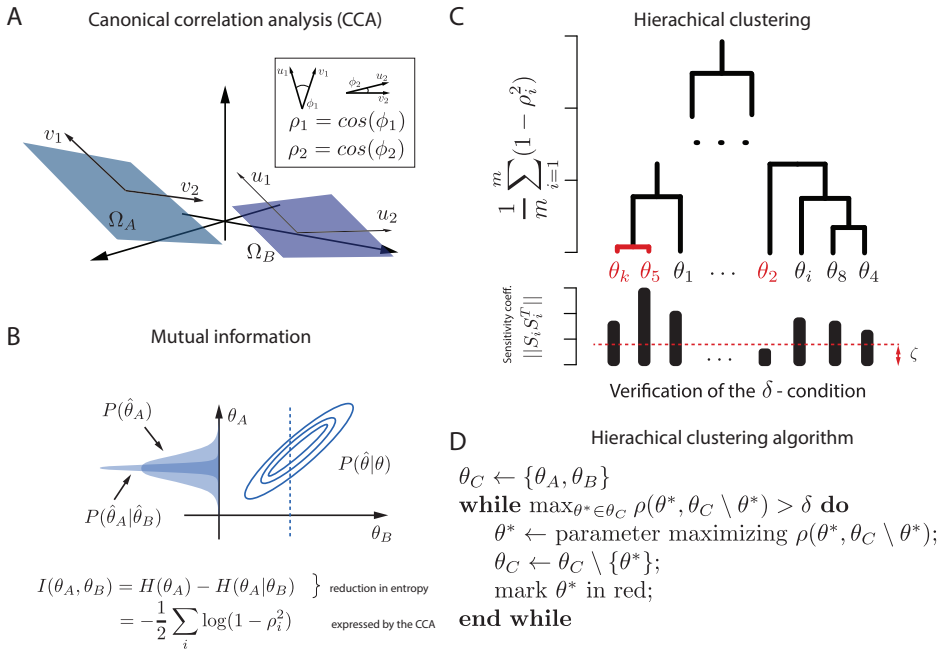


Figure 6: Canonical correlations and identifiability analysis.
a Conceptual illustration of the canonical correlations.Two subsets of sensitivity vectors represented as linear subspaces (planes Ω_A and Ω_B). Canonical vectors on the planes are found to yield maximum cosine. In a two-dimensional subspace case, the second canonical vectors, u_2, v_2, are required to be perpendicular to the first ones.
b Mutual information as a measure of similarity between two parameter sets θ_A and θ_B, which span linear subspaces Ω_A and Ω_B interpreted in terms of the asymptotic posterior $P(\hat{\theta}|\theta)$.
c Agglomerative hierarchical clustering of model parameters represented as dendrogram.
d The pseudo-code describing the heuristic procedure for maximising number of identifiability parameters, in the sense of (δ, ζ)-identifiability . Recursively, at each level, a pair ofmost similar clusters is merged into a single cluster and δ-condition is verified. Linkages between clusters, at each stage of clustering, areplotted at high $\frac{1}{m}\sum_{i=1}^{m}(1-\rho^2)$, where m is the size of a new cluster, compared to a previous linkage. Identifiability results from violation ofeither of the δ-condition or ζ-condition therefore even parameters that have sensitivities above a threshold can be non-identifiable. Non-identifiable parameters are marked in red.

4.2.3 Identifiaction of maximal number of (δ, ζ)-identifiable parameters

We proposed to use a hierarchical clustering algorithm with an MI-CCA as an aggregation measure to elucidate similarity between groups of model parameters. Results of the clustering can be presented in the form of dendrogram (see Figure 6C). The standard algorithm is enriched with the identifiability verification of clusters at each level of the procedure. Parameters that are identified as non-identifiable are assumed as fixed, which reduces the total compensation effects. This heuristic procedure leads to the maximization of the set of identifiable parameters in the model. Clustering algorithm is described in the Figure 6D. The clustering procedure was implemented and shared in the form of the R-package [180].

4.2.4 Application to biological systems

Apart from methodological development, the article shows that this tool provides relevant insight into how experimental designing contributes to the identifiability of model parameters.
The tool was applied to analyze the NF-κB pathway, one of the key components controlling the innate immune response. We used a model that describes a dynamic activation of NF-κB induced genes in response to the stimulation by a pro-inflammatory cytokine, TNF-α [181, 182]. The model is represented as a set of ODEs that contain 39 parameters and 19 variables.
The dendrogram obtained for the NF-κB system reveals that correlated parameters are grouped into clusters that essentially correspond to the network structure. This property provides relevant practical information that demonstrates how our method can help understand the functional role of individual parameters. If we want to change the model dynamical response, we should manipulate parameters values across different modules rather than within the same module.
Moreover, we have demonstrated the applicability of this tool to design experiments. Firstly, we checked which parameters of the NF-κB model can be estimated from the published experiments reported in 9 papers [183, 184, 185, 186, 187, 188, 189, 190, 182]. We showed that 18 out of 39 model parameters could not be estimated as they fail to satisfy the δ-condition. We have also analyzed how each of the analyzed papers increased the number of identifiable parameters. Chronologically first two papers, [183, 184], rendered 13 parameters identifiable. Subsequent seven papers provided information that enabled the estimation of only five new parameters. This indicates that to obtainining identifiable parameters requires specifically tailored experiments different from these performed to address conventional biological questions. Using this tool, we have also proposed two tailored experiments that would estimate additional six parameters.
The proposed tool establishes a theoretically grounded approach to examine the connection between correlations and non-identifiability systematically. This knowledge helps understand how non-identifiability arises and provides guidelines on whether experimental perturbations can remove detected correlations.

4.3 DEVELOPMENT OF NON-PARAMETRIC METHODS SUITED FOR THE COMPLEXITY OF CELLULAR SIGNALING PROCESSES

Multivariate, high-throughput single-cell signaling responses are typically represented in a standard form of dose-response curves. However, such representations do not capture the inherent complexity of single-cell high-throughput data, neglect the dynamics of cellular processes, and foreclose the analysis of multiple signaling effectors. These limitations can lead to simplified or even incorrect conclusions from experimental data [81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91]. Therefore, researchers applied Shannon information to quantify information transmission in signaling systems. However, the Shannon approach is tailored to specific communication and engineering conditions unsuitable for biological systems.

In the article [3] we have presented a novel analytic framework, fractional response analysis (FRA), that uses probabilistic modeling and information theory to deconvolute behaviors of cellular populations. **The main idea behind this work is to demonstrate that outcomes of physiological processes depend on the number of cells with specific responses rather than on mean or median, which constitutes the fraction of cells with a given response as a biologically relevant variable**. Therefore, we proposed to quantify what fractions of the cellular population exhibit different responses for different doses.

4.3.1 Fractional response curve

We introduced the fractional response curve (FRC) that quantifies fractions of cells that exhibit different responses to a change in dose. Formally, signaling system is represented as a input-output channel (similarly to 1.3). Briefly, we consider a series of doses $x_1, \ldots, x_i, \ldots, x_m$ and denote a response as y. The relation between response, y, and dose, x_i, is represented as the probability distribution, $P(Y|X = x_i)$. For the lowest dose, we set the value of FRC to 1 as the whole population has responses specific to the first dose when higher doses are not considered, i.e., $r(x_1) = 1$. For each subsequent dose, the FRC increases by the fraction of cells stimulated with the subsequent dose that exhibit responses different from lower doses, which counts the number of distinct distributions a cellular population generates for increasing doses. In other words, the value of the FRC for dose x_i is the sum of the previous value and the fraction of cells exhibiting different responses,

$$r(x_i) = r(x_{i-1}) + \Delta r, \tag{75}$$

where Δr is equivalent to the overall increase in the frequency of responses. For an illustration of FRC, we considered a simple hypothetical example involving one signaling effector and three doses, x_1, x_2 and x_3, which can be interpreted as control, intermediate and high dose (see Figure 7a-e).

Formally, FRC is defined as

$$r(x_i) = \int_{y \in \mathcal{Y}} \max_{x_j \geqslant x_i} P(y|x_j)dy, \tag{76}$$

where integration takes place over space $\mathcal{Y}$, the set of all possible responses, y. The integral in the above formula quantifies the area under the curve (or under surface for multivariate data) $max_{x_j \geqslant x_i} P(y|x_j)$. The curve describes the maximal frequency of the response y among doses $x_1, \ldots, x_i$. As subsequent doses are taken in the account, the change from $r(x_i)$ to $r(x_{i+1})$, Δr, quantifies the overall increase in the frequencies of responses due to considering the dose x_{i+1}.

Adding subsequent fractional increments, δr, leads to the value of FRC expressed in terms of the cumulative fraction of cells that exhibit different responses due to dose change.

4.3.2 Fractional cell-to-cell heterogeneity

The FRC quantifies fractions of cells that exhibit different responses due to dose change but does not quantify overall cell-to-cell heterogeneity: it does not show what fraction of cells exposed to one dose exhibits responses in the range characteristic for other doses. Therefore, within FRA, we propose augmenting the FRC by quantifying the overlaps between distributions corresponding to different doses. We call a given response as typical for a given dose if it is most likely, i.e., most frequent, to arise for this specific dose than all other doses. The fractional analysis is conveniently presented for the hypothetical examples of three doses, Figure 7f-i. The results of fractional analysis, presented as pie-charts, can be shown in a matrix as the fraction of cells stimulated with one dose (rows) that has responses typical for other doses (columns) Figure 7h. The pie-chart partitioning can be plotted along with the FRC (Figure 7.i) so that the fractional increments, Δr, and fractional cell-to-cell heterogeneity are concisely presented.

Formally, we define the response, y, to be typical for dose x_j if it is most likely to arise for this dose, which writes as

$$P(y|x_i) \geqslant P(y|x_j) \text{ for all } j \text{ other than } i. \tag{77}$$

The above condition allows assigning any response y to a dose for which it is typical. Further, for each stimulation level, x_i, we can quantify fractions of cells that exhibit response typical to each of the considered doses, x_j. Formally, it is defined as

$$v_{i,j} = \int_{y \in \mathcal{Y}} \mathbb{I}_{\{P(y|x_j) \geqslant P(y|x_k)|:k \neq j\}} P(y|x_i)dy. \tag{78}$$

description of the complex multivariate responses into a simple quantitative

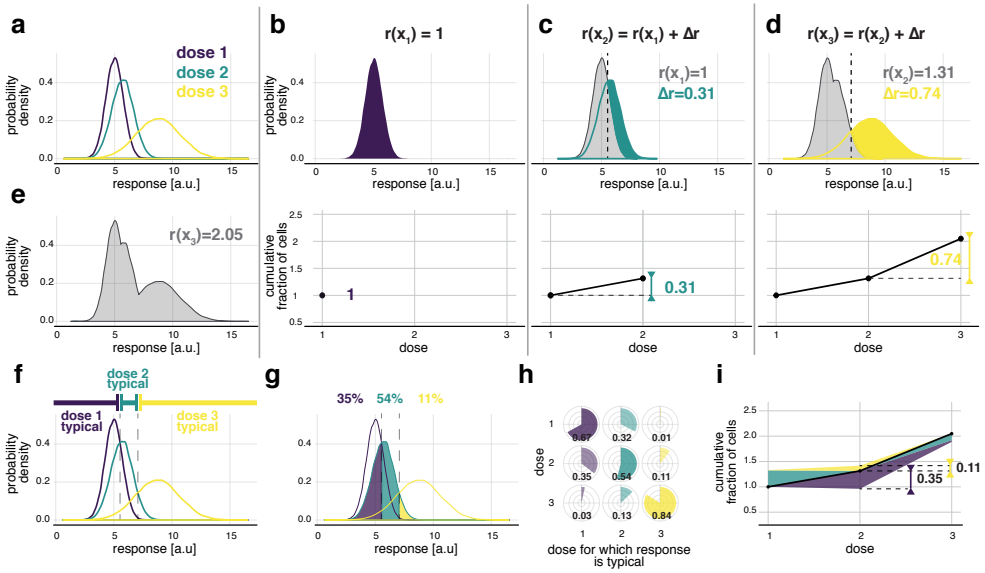


Figure 7: Fractional response analysis.

a Hypothetical response distributions to three different doses encoded by colors. Distributions are represented as a probability density, which is proportional to the frequency of cells with a given response level.

b-d Quantification of the fraction of cells that exhibit different responses due to dose increase, Δr, and constriction of FRC, for responses presented in **a**. Each panel from **b** to **d** corresponds to subsequent changes in dose. The color regions mark the overall increase in frequency due to considering the dose marked by the color. The area of the colored region quantifies Δr. The value of the FRC for each dose is obtained by adding the increment, Δr.

e Quantification of the number of distinct distributions induced by the three considered doses.

f Dose-typical responses for the response distributions of **a**.

g Dissection of the responses to dose 2 into responses typical to any of the three doses. The fraction of cells typical to a given dose is marked with the corresponding color. The surface area of each color quantifies the typical fraction.

h The fractions of cells stimulated with one dose (rows) with responses typical to any of the doses (columns).

i The FRC together with the bands representing cell-to-cell heterogeneity as quantified in **h**. For each reference dose (x-axis), the fractions of cells stimulated with the reference dose that exhibit responses typical to other doses can be plotted in the form of color bands around the curve. The color encodes the dose a given fraction refers to. The height of the band marks the size of the fraction (y-axis). Fractions corresponding to doses higher than the reference dose are plotted above the curve, whereas to doses lower than the reference dose below the curve.

In the article we show that FRC can be reforumlated as the fraction of cells that exhibit response typical to its own dose. Formally,

$$r(x_i) = \sum_{k=1}^{i} v_{kk} \tag{79}$$

This formula gives the novel interpretation of FRC as the number of distinct response distributions that were experimentally observed.

FRA compresses complex dose-response data into a comprehensive quantitative description accounting for cell-to-cell heterogeneity and multivariate measurements. We showed that FRA is a convenient and meaningful way of representing single-cell data. Moreover, FRA enables to uncover of hidden patterns in single-cell data.

4.3.3 Fractional analysis of biological systems

To show the advantages of FRA for discrimination between different response modalities, we can consider *in silico* models of binary and graded response in which the same mean response results from different response distributions in cellular populations (see Figure 8a-c). The two response modalities are represented differently within FRA, highlighting differences masked by the population average. For both cases, FRC shifts gradually, however for the graded system FRC reaches the value of over 3 Figure 8d, compared to nearly 2 for the binary case Figure 8e. FRC of 3 reflects the property of the considered graded system to generate responses in the range typical for basal, intermediate, and saturating doses. In contrast, the binary system can only generate responses in two ranges, basal and saturating, which leads to a maximal FRC of 2. The structure of the responses of the two systems is further reflected in the bands around FRC and the pie-charts, Figure 8f-g. The bands and pie charts indicate that response distribution corresponding to a given dose overlaps with distributions corresponding to several other doses for the graded system. For the binary system, on the other hand, responses to different doses overlap only with the distributions corresponding to the minimal and the highest dose; therefore, the bands have only two colors, violet, below the FRC, which corresponds to the minimal dose, and green, above the FRC, corresponding to the highest dose. The two response modalities are represented differently within FRA, highlighting differences masked by the population average. Similar differences in the functioning of a more complex system represented by multivariate data could not be detectable with a simple visual inspection as for simple univariate examples. In such cases, the different response modalities could be uncovered with FRA.

In the article [3], we have confirmed those properties by applying FRA to experimental data. We used FRA to experimental measures of responses to cytokines signaling, i.e., IFN-γ and IL-10 in the human monocyte cell line U937 and IFN-γ in cancer cell lines A549 and CALU1. Moreover, to show the application of FRA to multivariate data, we examined time-series TNF-α dose responses with live confocal imaging of a murine embryonic fibroblasts cell line stably expressing fluorescent NF-κB complex.

Moreover, we applied FRA to analyze type I interferon signaling in human peripheral blood mononuclear cells (PBMCs), a system involving multiple signaling effectors, cell-to-cell heterogeneity, and several cell types. Dose-responses to the type I interferon variant IFN-α2a were analyzed via whole-cell tyrosine phosphorylation levels of effector proteins STAT1, STAT3, STAT4, STAT5, and STAT6 (pSTATs) measured jointly in individual cells using mass cytometry (CyTOF). Mean levels and distributions of pSTATs in B-cells, CD4+ T-cells, CD8+ T-cells, NK cells, and CD14+ monocytes were calculated and

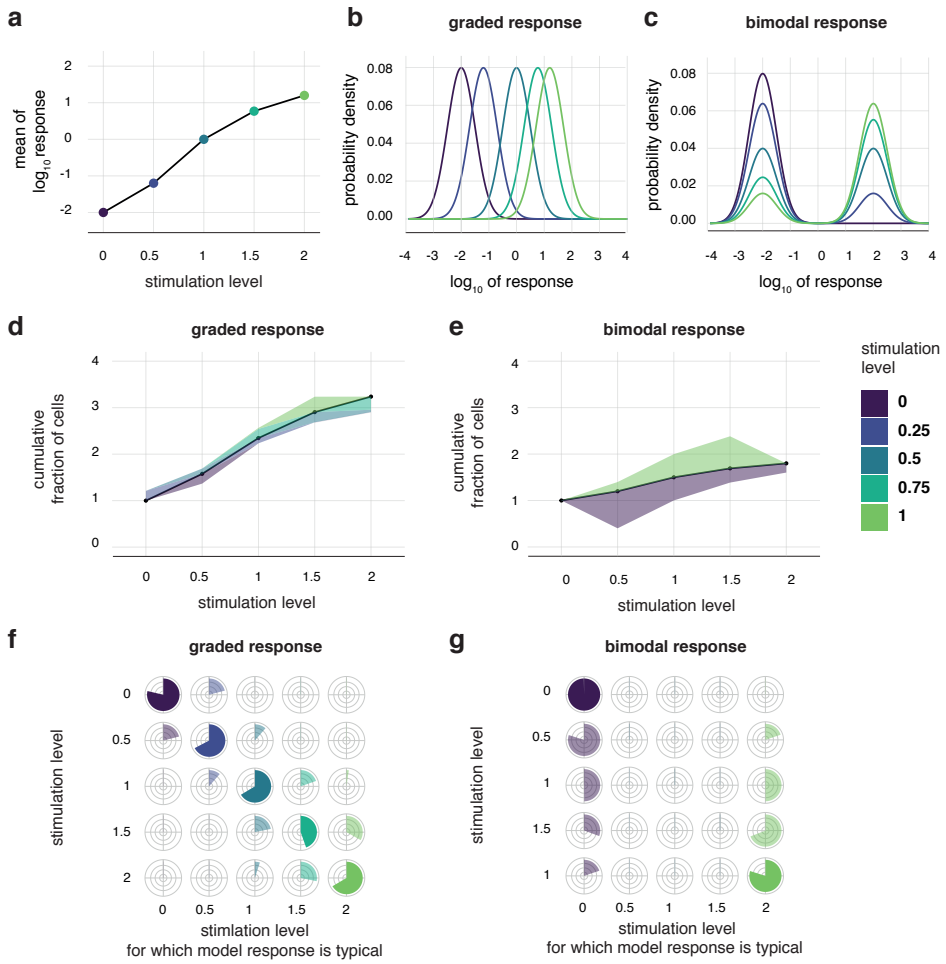


Figure 8: FRA of binary vs graded response for the in silico model.
a Mean of log responses in the binary and graded in silico models.
b Distribution of log responses in the graded models.
c Distribution of log responses in the binary models.
d FRA of distribution of log responses in the graded models shown in b.
e FRA of distribution of log responses in the binary models shown in c.
f Pie-charts corresponding to d, representing cell-to-cell heterogeneity structure in the graded model.
g Pie-charts corresponding to d, representing cell-to-cell heterogeneity structure in the binary model.

revealed that different STATs reached different maximal phosphorylation levels for different doses in different cell types. Overall, however, no apparent pattern in the functioning of the signaling system was revealed.

The FRA enabled to condenses the description of the complex multivariate responses to a simple quantitative form. The FRA results show entirely different conclusions than standard techniques. Both the FRC and fractional cell-to-cell heterogeneity are very similar for all cell types. Characteristically, the fractional dose responses in different cell types follow the same logarithmic pattern identifying a phenomenon that governs the behavior of multivariate cellular responses of our system, which remains hidden when

conventionally inspecting data. Formally, FRC scales as the logarithm of the dose

$$r(x) \propto \log(x), \tag{80}$$

which given incremental approximation, $\Delta\log(x) = \log(x + \Delta x) - \log(x) \sim \Delta x/x$, implies fold-change sensitivity in the population

$$\Delta r \propto \Delta x/x \tag{81}$$

which in the studied system universally describes dose-responses in populations across different cell types.

Interestingly, similarly to IFN-α2a in PBMC, in other experiments analyzed in this article, dose responses are sensitive to fold-changes in the dose instead of additive changes. Such property resembles the empirical Weber-Fechner law that characterizes the performance of many psycho-physiological sensory systems [191]. Minimal detectable stimulus change, Δx, in the sense of weight, hearing, vision, and smell, has been observed to be of fold-type. So far, in the signaling pathways, similar observations have been seen on a level of the single signaling effector in a representative cell, or population-average [192, 193, 194, 195, 196]. However, the application of FRA demonstrated that within heterogeneous populations of cells of a given type and across types, a number of cells that exhibit a different response is proportional to the fold-change in the dose to the state of the heterogeneous population described by multivariate data. However, a mechanistic explanation of cellular populations' fold-change sensitivity is unclear and remains to be determined.

Overall, FRA delivers a concise representation of complex single-cell data, which is particularly relevant for high-throughput techniques, which are increasingly allowing the measurement of a high number of parameters per cell, generating extensive, high-dimensional datasets [197].

4.3.4 Interpretation of FRA in terms of information-theory

Apart from the fact that the method has undeniable advantages when applied to the description of single-cell data, we have shown that it has a rigorous mathematical interpretation in terms of Rényi min-information. In section 3.3 we have to make a brief introduction to the generalization of Shannon information in Rényi sense.

Formally, we have noticed that the logarithm of the FRC for the highest dose, x_m is equal to the Csiszár's-Rényi min-capacity Eq.55, i.e.,

$$C^c_{min} = \log_2(r(x_m)), \tag{82}$$

or equivalently

$$r(x_m) = 2^{C^c_{min}}. \tag{83}$$

It means that FRA is a measure of information flow in the biochemical systems in a Rényi min-information sense. Analogously to Shannon capacity, Rényi min-information measures how many distinguishable states the signaling systems sense, which provides the overall signaling fidelity. In the article, we have shown that the properties of Rényi min-information make

this measure more appropriable to describe information flow in biochemical systems.

First of all, Rényi-min information does not depend on the input distributions. Therefore, its application does not require any assumptions regarding the optimality of the system. Moreover, it does not involve long sequences of discrete symbols for rigorous interpretation.

Secondly, the Rényi-min approach provides the best decoding strategy based on the maximum-a-posteriori decision rule (MAP), i.e., assign the signaling response to the dose for which it is most frequent [198]. Then, among cells stimulated with the dose x_k, the faction that could be decoded correctly is the fraction that exhibits responses most frequent for x_k, which is referred to as *typical* and quantified as $v_{k,k}$ (see Figure 7.f). The Rényi-min channel capacity equals the logarithm of the sum of fractions of cells that can be decoded correctly, i.e.,

$$C^*_{min} = \log_2 \left(\sum_{k=1}^{m} v_{k,k} \right) \tag{84}$$

Therefore, Rényi-min capacity has a precise biological interpretation as the logarithm of the sum of fractions of cells simulated with a given dose with responses typical for this dose. Hence, in the Rényi-min sense number of distinguishable states based on the response equals the sum of fractions of cells decoded correctly. In contrast to the Shannon information, here we could indicate which states are distinguishable. Moreover, fractional interpretation of Rényi min-capacity offers the qualitative and quantitative decomposition into fractions of cells with non-typical responses levels, which can be presented as pie-charts (see Figure 7.h). Hence, we get information on why some of the states are not distinguishable.

Moreover, if we assume the equiprobability of the input distribution, then based on the equivalence of the Rényi min-capacity in Csiszár's and Arimoto's sense (see Methods 64), this measure has another interpretation in the probability of the correct discrimination, $\varepsilon_{Y|X}$. Precisely,

$$\varepsilon_{Y|X} = \frac{1}{m} 2^{C^*_{min}} \tag{85}$$

is the probability that an observer would correctly guess the dose of a randomly selected cell among cells stimulated with any m doses.

Therefore, Rényi-min capacity can be interpreted as the logarithm of the number of different doses that can be decoded correctly on average by the observer, according to the maximum a posteriori decision rule.

For comparison of the Rényi min-information capacity and Shannon information capacity, we considered a data set published in the [80], where authors quantify information transfer by G protein-coupled receptor signaling (see details in the introduction 1.3). Authors performed experiments that enabled comparison of information flow on single-cell and population. Dose-response distributions of the population are shown as violin plots in Figure 9.a, whereas dose-responses of three individual cells are presented in Figure 9.b. The authors first used the population distributions to calculate Shannon information capacity. Here, calculation of Shannon capac-

ity using R-package SLEMI [5] yielded 0.89 bits ($\log_2(1.85)$). Calculation of Shannon capacity for population distributions implicitly assumes that cell-to-cell heterogeneity constitutes a noise source leading to information loss. Therefore, subsequently, the authors calculated capacity using response distributions of individual cells, which quantified information transfer of individual cells, i.e., under noise arising inside of single-cells, and not from cell-to-cell heterogeneity. The average capacity among individual cells was 1.65 bits ($\log_2(3.14)$). The latter scenario's considerably higher capacity indicated that not all cell-to-cell heterogeneity could be interpreted as noise leading to information loss inside individual cells. Individual cells might be in different states that determine their dose-response levels, and different cells in the population are indifferent states[80].
Here, for comparison Rényi-min with Shannon mutual information, we firstly analyzed population response distributions, Figure 9.c. The value of FRC for the highest dose, 2.03, indicates that overall cellular populations can generate two distinct distributions. Comparing FRC with Shannon capacity, we have that FRC is more than 2^{C^*} , which is in line with the inequality of Eq.39. In addition to the value of Rényi min-information, FRA reveals the overlaps between distributions are substantial for low and high doses and moderate for intermediate doses. As pointed out by the authors of the paper presenting the data, Shannon's information cannot be interpreted here as the capacity of cells to discriminate between doses due to cell-to-cell heterogeneity, which is not equivalent to noise. In contrast, the interpretation of Rényi min-information capacity presented above is valid, as it does not refer directly to the capacity of individual cells to discriminate between the doses. Furthermore, FRA is designed to quantify dose responses of a heterogeneous population of cells. As the considered data set involves response distributions of individual cells, the question arises whether FRA could be deployed to analyze these in addition to population response distributions. From a technical perspective, using single-cell response distributions, as opposed to population response distributions, is irrelevant. FRA can be easily performed for each cell; nevertheless, interpretation of FRC and overlaps between response distributions need to be established. By analogy with population response distributions, we describe two possible interpretations:

(i) an increase in FRC, $\Delta r = r(x_i) - r(x_{i-1})$, can be interpreted as the probability of a given cell to generate a response to the dose x_i that is distinct from all lower doses (more likely for xx_i than for any lower dose). The value of $r(x_i)$ quantifies that the overall number of distinct response distributions specific to an examined cell. Value of FRC for the maximal dose, $r(x_m)$, is the maximal number of distinct response distributions that a given cell generates. Overlaps between response distributions, as quantified by pie-charts, and color bands around FRC, show probabilities of a given cell stimulated with a given does to generate a response that is typical (most frequent) for any of the considered doses;

(ii) considering the best possible strategy to decode the dose from the output of individual cells, i.e., maximum likelihood decoding, FRA can be interpreted in terms of correct and incorrect decoding probabilities.

The increase in FRC shows the probability of correctly discriminating the dose x_i from lower doses. Values of the FRC, $r(x_i)$, is the sum of probabilities of correct decoding of the doses not greater than x_i. Therefore, $r(x_m)/m$, is the overall probability of correct decoding. Finally, pie charts, and color bands around FRC, show probabilities of decoding a response to a given dose (rows) as another dose (columns).

FRA plots for three individual representative cells are shown in Fig.9.d. The highest value of FRC for individual cells is higher than for the population responses. An individual cell can generate, therefore, a more distinct response distribution than a population of cells, which is in line with findings presented in [80]. Analogously, an individual cell can resolve between higher numbers of doses than predicted by the population response distributions. In addition, FRA plots quantify that cells differ in terms of which concentrations can be resolved; some are better in resolving small doses and other high doses.

The analyses above imply several insightful conclusions regarding the applicability of information theory to quantify information transfer in single-cells. As pointed out by [80], responses of individual cells are much less variable than the overall cell-to-cell heterogeneity of responses. Cell-to-cell heterogeneity cannot be interpreted as noise leading to information loss about the level of signaling input. Shannon information capacity, therefore, calculated based on population response distributions, cannot be interpreted in terms of the potency of individual cells to resolve between different concentrations. Each cell is different, and therefore to calculate its potential to resolve between different concentrations, its own response distribution should be determined. On the other hand, Rényi information can be calculated and meaningfully interpreted in terms of fractions of cellular populations. Calculation of Shannon information capacity based on distributions of individual cells can be done and interpreted to the extent provided by the Shannon framework, including the Shannon coding theorem. On the other hand, the interpretation of Rényi is straightforward and does not require any advanced formalism like the Shannon coding theorem.

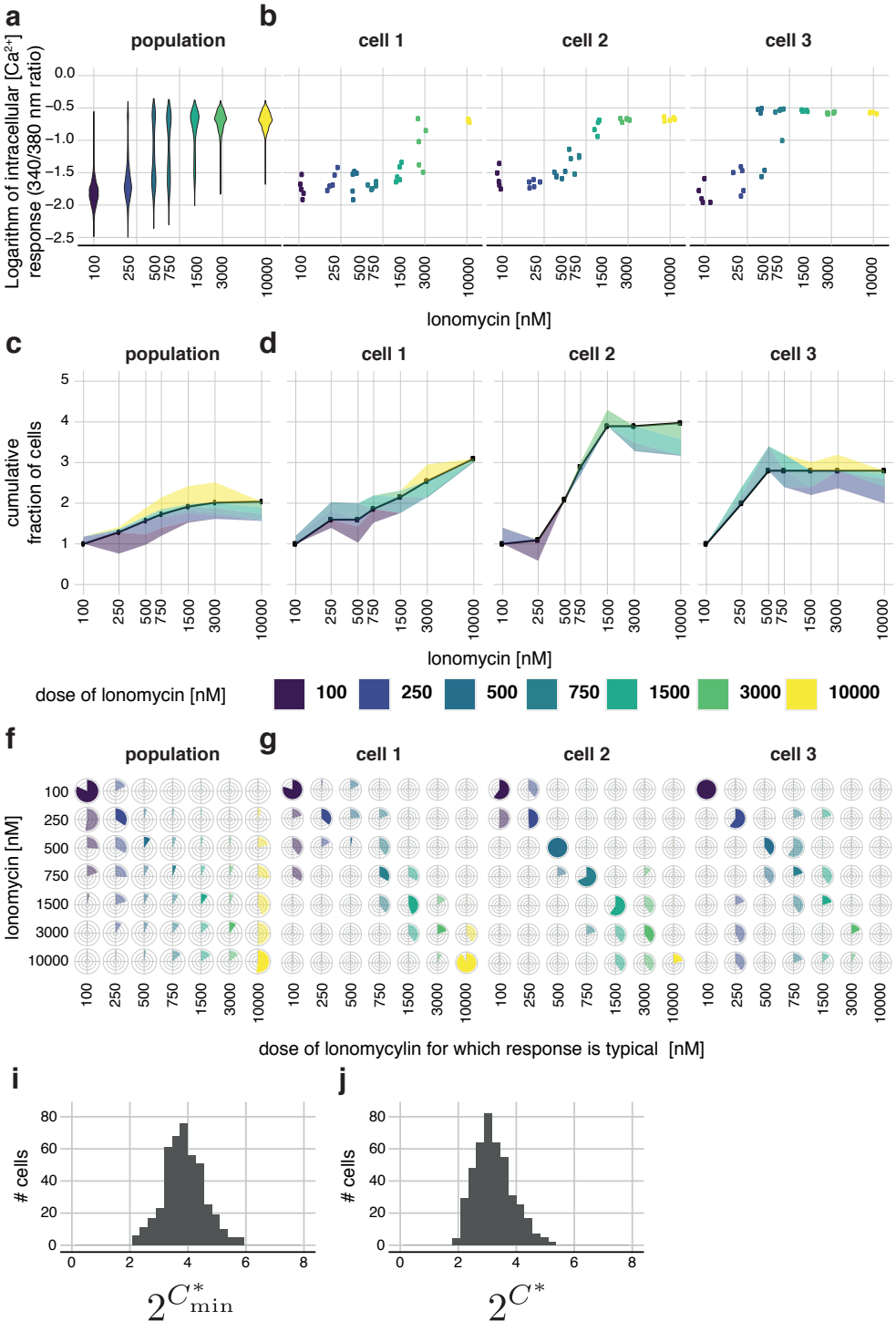


Figure 9: Comparison of Rényi min-capacity and Shannon capacity for GPCR signaling published in [80].

Figure 9: **a** Violin plot is representing population response distributions to different doses of lonomycin. At least 433 cells were measured per condition, whereas each cell was measured 35 times.
b Responses of three representative cells to different doses of lonomycin. Points at each panel correspond to measurements performed in the same cell.
c FRA of the population responses shown in a.
d FRA of the responses of three individual cells shown in b.
e Pie-chart representing cell-to-cell heterogeneity structure of population responses shown in a. **f** Pie-charts representing dose-response variability of responses shown in b corresponding to three individual cells.
i Variability of Shannon capacity in the cellular population. Shannon capacity, represented here as the power of 2, was calculated for each cell and presented as a histogram.
j Same as in i but for Rényi min-capacity

5 CONCLUSIONS

The development of experimental high-throughput techniques provides enormous data sets that describe molecular mechanisms of signaling pathways at the level of single cells. For instance, with automated confocal microscopy, immunocytochemistry enables to read the fluorescence signal from thousands of cells. On the other hand, more comprehensive mass cytometry (CyTOF) allows the analysis of multiple factors in thousands of cells at once. Despite the source of experimental data, its study requires quantitative support due to the complexity and stochasticity of cellular signaling.

During the last decades, there have been developed multiple mathematical and computational methods that are used to analyze biological systems. Most of these methods have been adapted directly from engineering problems. In my Ph.D., I have demonstrated that biological systems significantly differ from engineering models; hence they require specific techniques tailored to the complexity and stochasticity of the cellular signaling pathways. During my work, I have described current and new solutions to work with the multi-parameters models of signaling pathways. Moreover, I have proposed novel non-parametric approaches that suit the complexity of cellular mechanisms.

In the chapter "Sensitivity Analysis in quantitative biology models" [1] of the book "Quantitative Biology: Theory, Computational Methods and Examples of Mod- els" [9], we have presented an overview of sensitivity analysis methods in the context of quantitative biology. The chapter contains exemplary applications of those methods in the deterministic and stochastic regimes. Moreover, the chapter provides a list of exercises that help readers explore the problems of sensitivity analysis.

In the article[2], we have provided a novel technique for analyzing similarity between groups of model parameters and introduced the definition of asymptotic parameter identifiability. Both methods are helpful to describe the impact of parameters on the behavior of complex multi-parameter models. The proposed definition of identifiability is related to the well-known profile likelihood but is far more efficient to calculate in the local scenario. The concept, however, is very general and can be easily extended to the global case at the price of more intensive computations. The article provides relevant insight into how experimental designs contribute to the identifiability of model parameters. Moreover, we have illustrated how this tool can be used to design meaningful experiments. Created techniques are implemented and shared in an R-package. A potential extension of the work is making an automatic tool for verifying experiments and developing new ones.

Finally, in the article [3], we have presented drawbacks of the conventional representation of single-cell dose-response and propose to quantify changes of cellular fractions rather than mean or median. The introduced tool, frac-

tional response analysis, delivers a concise representation of complex single-cell data, particularly relevant for high-throughput techniques, increasingly allowing the measurement of a high number of parameters per cell, generating massive, high-dimensional datasets. Moreover, the fractional response analysis has a rigorous mathematical definition in terms of Reńyi-min information. We have described multiple advantages of Reńyi-min information over commonly used Shannon information in the context of biological systems. Using theoretical, model examples, and biological data, we have shown that the Reńyi-min measure detects patterns hidden for the Shannon measure. Based on the reviewers' opinions, I am firmly convinced that Reńyi-min information will be commonly used to measure information flow in the signal pathways. However, there are still many questions regarding the usage of Reńyi-min, for instance, (i) what type of response distributions are distinguishable by Reńyi-min, (ii) does Reńyi-min information preserve the information inequality in the biological systems, (iii) what the sensitivity of the measure is to the number of input signals. The application of fractional response analysis to several datasets gave promising biological results. Fractional response analysis demonstrated that the number of cells that exhibit a distinct response is proportional to the fold change in the dose within heterogeneous populations of cells. This phenomenon resembles the empirical Weber-Fechner law. However, a mechanistic explanation of the sensitivity to dose fold changes in cell populations is unclear and remains to be determined.

I firmly believe that the results of this thesis will be helpful for researchers to face the difficulties with the complexity and stochasticity of cellular signaling.

ACKNOWLEDGMENTS

Results presented in this thesis were obtained in an excellent and friendly research community of the Department of Biosystems and Soft Matter of the Institute of Fundemental Technological Research of the Polish Academy of Sciences (IPPT PAN). I graciously thank all my colleagues and friends that provided help, counsel, inspiration and motivation during my research.

Support from the Polish National Science Centre under grant PRELUDIUM 2016/23/N/ST6/03505 and the Foundation for Polish Science within the First TEAM (First TEAM/2017-3/21) program co-financed by the European Union under the European Regional Development Fund.

REFERENCES

[1] Karol Nienałtowski, Tomasz Jetka, and Michał Komorowski. Sensitivity analysis in quantitative biology models. In B Munsky, W S Hlavacek, and L S Tsimring, editors, *Quantitative Biology: Theory, Computational Methods and Examples of Models*, chapter 13, pages 245–270. MIT Press, 2018.

[2] Karol Nienałtowski, Michał Włodarczyk, Tomasz Lipniacki, and Michał Komorowski. Clustering reveals limits of parameter identifiability in multi-parameter models of biochemical dynamics. *BMC systems biology*, 9(1):1–9, 2015.

[3] Karol Nienałtowski, Rachel E Rigby, Jarosław Walczak, Karolina E Zakrzewska, Edyta Głów, Jan Rehwinkel, and Michał Komorowski. Fractional response analysis reveals logarithmic cytokine responses in cellular populations. *Nature Communications*, 12(1):1–10, 2021.

[4] Zahra Vahdat, Karol Nienałtowski, Zia Farooq, Michał Komorowski, and Abhyudai Singh. Information processing in unregulated and autoregulated gene expression. In *2020 European Control Conference (ECC)*, pages 258–263. IEEE, 2020.

[5] Tomasz Jetka, Karol Nienałtowski, Tomasz Winarski, Sławomir Błoński, and Michał Komorowski. Information-theoretic analysis of multivariate single-cell signaling responses. *PLoS computational biology*, 15(7):e1007132, 2019.

[6] Tomasz Jetka, Karol Nienałtowski, Sarah Filippi, Michael PH Stumpf, and Michał Komorowski. An information-theoretic framework for deciphering pleiotropic and noisy biochemical signaling. *Nature communications*, 9(1):4591, 2018.

[7] Karol Nienałtowski, Tomasz Jetka, and Michał Komorowski. Experimental design quantitative biology models. In B Munsky, W S Hlavacek, and L S Tsimring, editors, *Quantitative Biology: Theory, Computational Methods and Examples of Models*, chapter 13, pages 245–270. MIT Press, 2018.

[8] Weronika Wronowska, Agata Charzyńska, Karol Nienałtowski, and Anna Gambin. Computational modeling of sphingolipid metabolism. *BMC systems biology*, 9(1):1–16, 2015.

[9] B Munsky, W S Hlavacek, and L S Tsimring. *Quantitative Biology: Theory, Computational Methods and Examples of Models*. MIT Press, 2018.

[10] Eva Bianconi, Allison Piovesan, Federica Facchin, Alina Beraudi, Raffaella Casadei, Flavia Frabetti, Lorenza Vitale, Maria Chiara Pelleri, Simone Tassani, Francesco Piva, et al. An estimation of the number of cells in the human body. *Annals of human biology*, 40(6):463–471, 2013.

[11] Luiz Carlos Uchôa Junqueira, José Carneiro, and Robert O Kelley. Basic histology. 1995.

[12] JN Livingston, KM Lerea, J Bolinder, L Kager, L Backman, and P Arner. Binding and molecular weight properties of the insulin receptor from omental and subcutaneous adipocytes in human obesity. *Diabetologia*, 27(4):447–453, 1984.

[13] VJ Goyanes, A Ron-Corzo, E Costas, and E Maneiro. Morphometric categorization of the human oocyte and early conceptus. *Human Reproduction*, 5(5):613–618, 1990.

[14] Samir K Ballas. Erythrocyte concentration and volume are inversely related. *Clinica chimica acta; international journal of clinical chemistry*, 164(2):243–244, 1987.

[15] Hiroaki Kitano. Computational systems biology. *Nature*, 420(6912):206, 2002.

[16] Dan S Tawfik. Messy biology and the origins of evolutionary innovations. *Nature Chemical Biology*, 6(10):692, 2010.

[17] Olaf Wolkenhauer and Mihajlo Mesarović. Feedback dynamics and cell function: Why systems biology is called systems biology. *Molecular BioSystems*, 1(1):14–16, 2005.

[18] Hiroaki Kitano. Systems Biology: A Brief Overview. *Science*, 295:1662–1664, 2002.

[19] E N and Norbert Wiener. Cybernetics. Or Control and Communication in the Animal and the Machine. *The Journal of Philosophy*, 46:736, 1949.

[20] Harvey Lodish, Arnold Berk, S Lawrence Zipursky, Paul Matsudaira, David Baltimore, and James Darnell. Molecular cell biology 4th edition. *National Center for Biotechnology Information, Bookshelf*, 9, 2000.

[21] Alberto Mantovani and Elisabetta Dejana. Cytokines as communication signals between leukocytes and endothelial cells. *Immunology Today*, 10(11):370–375, 1989.

[22] David Alexander and Gary Goldberg. Transfer of Biologically Important Molecules Between Cells Through Gap Junction Channels. *Current Medicinal Chemistry*, 10(19):2045–2058, 2003.

[23] Lionel B Ivashkiv and Laura T Donlin. Regulation of type i interferon responses. *Nature reviews Immunology*, 14(1):36–49, 2014.

[24] Franck J Barrat, Mary K Crow, and Lionel B Ivashkiv. Interferon target-gene expression and epigenomic signatures in health and disease. *Nature immunology*, 20(12):1574–1583, 2019.

[25] Edouard M DeMaeyer, De Maeyer-Guignard, et al. *Interferons and other regulatory cytokines*. Wiley, 1988.

[26] Minoru Sakatsume, Ken-ichi Igarashi, Karen D Winestock, Gianni Garotta, Andrew C Larner, and David S Finbloom. The jak kinases differentially associate with the and (accessory factor) chains of the interferon receptor to form a functional receptor unit capable of activating stat transcription factors. *Journal of Biological Chemistry*, 270(29):17528–17534, 1995.

[27] AHH van Boxel-Dezaire and GR Stark. Cell type-specific signaling in response to interferon-γ. In *Interferon: The 50th Anniversary*, pages 119–154. Springer, 2007.

[28] U Boehm, T Klamp, M Groot, and JC Howard. Cellular responses to interferon-γ. *Annual review of immunology*, 15(1):749–795, 1997.

[29] Xiaoyu Hu, Carmen Herrero, Wai-Ping Li, Taras T Antoniv, Erik Falck-Pedersen, Alisa E Koch, James M Woods, G Kenneth Haines, and Lionel B Ivashkiv. Sensitization of ifn-γ jak-stat signaling during macrophage activation. *Nature immunology*, 3(9):859–866, 2002.

[30] David Olagnier and John Hiscott. Type i and type iii interferon-induced immune response: It's a matter of kinetics and magnitude, 2014.

[31] Nina Ank, Hans West, Christina Bartholdy, Kristina Eriksson, Allan R Thomsen, and Søren R Paludan. Lambda interferon (ifn-λ), a type iii ifn, is induced by viruses and ifns and displays potent antiviral activity against select virus infections in vivo. *Journal of virology*, 80(9):4501–4509, 2006.

[32] Tobias Marcello, Arash Grakoui, Giovanna Barba-Spaeth, Erica S Machlin, Sergei V Kotenko, Margaret R Macdonald, and Charles M Rice. Interferons α and λ inhibit hepatitis c virus replication with distinct signal transduction and gene regulation kinetics. *Gastroenterology*, 131(6):1887–1898, 2006.

[33] Anthony Meager, Kumuthini Visvalingam, Paula Dilger, Donna Bryan, and Meenu Wadhwa. Biological activity of interleukins-28 and-29: comparison with type i interferons. *Cytokine*, 31(2):109–118, 2005.

[34] Nicole E Pagliaccetti, Esther N Chu, Christopher R Bolen, Steven H Kleinstein, and Michael D Robek. Lambda and alpha interferons inhibit hepatitis b virus replication through a common molecular mechanism but with different in vivo activities. *Virology*, 401(2):197–206, 2010.

[35] Belinda S. Parker, Jai Rautela, and Paul J. Hertzog. Antitumour actions of interferons: implications for cancer therapy. *Nature Reviews Cancer*, 16:nrc.2016.14, 2016.

[36] Pau Rué and Alfonso Martinez Arias. Cell dynamics and gene expression control in tissue homeostasis and development. *Molecular Systems Biology*, 11(2):792, 2015.

[37] W A Lim. Designing customized cell signalling circuits. *Nature Reviews Molecular Cell Biology*, 11:393–403, 2010.

[38] Marcelo Behar, Derren Barken, Shannon L Werner, and Alexander Hoffmann. The Dynamics of Signaling as a Pharmacological Target. *Cell*, 155(2):448–461, 2013.

[39] I Amit, R Wides, Y Yarden Molecular systems biology, and 2007. Evolvable signaling networks of receptor tyrosine kinases: relevance of robustness to malignancy and to cancer therapy. *msb.embopress.org*.

[40] Akinori Takaoka, Yukiko Mitani, Hirofumi Suemori, Mitsuharu Sato, Taeko Yokochi, Shigeru Noguchi, Nobuyuki Tanaka, and Tadatsugu Taniguchi. Cross Talk Between Interferon-γ and -α/β Signaling Components in Caveolar Membrane Domains. *Science*, 288:2357–2360, 2000.

[41] Daniel A Haber, Nathanael S Gray, and Jose Baselga. The Evolving War on Cancer. *Cell*, 145(1):19–24, April 2011.

[42] Stephen B Baylin and Joyce E Ohm. Epigenetic gene silencing in cancer – a mechanism for early oncogenic pathway addiction? *Nature Reviews Cancer*, 6(2):107–116, February 2006.

[43] G Manning, D B Whyte, R Martinez, T Hunter, and S Sudarsanam. The Protein Kinase Complement of the Human Genome. *Science*, 298(5600):1912–1934, December 2002.

[44] D S Krause, R A Van Etten New England Journal of Medicine, and 2005. Tyrosine Kinases as Targets for Cancer Therapy | NEJM. *Mass Medical Soc*.

[45] Sean C Bendall and Garry P Nolan. From single cells to deep phenotypes in cancer. *Nature Biotechnology*, 30(7):639–647, July 2012.

[46] Takuya Fukazawa, Minzhe Guo, Naomasa Ishida, Tomoki Yamatsuji, Munenori Takaoka, Etsuko Yokota, Minoru Haisa, Noriko Miyake, Tomoko Ikeda, Tatsuo Okui, Nagio Takigawa, Yutaka Maeda, and Yoshio Naomoto. SOX2 suppresses CDKN1A to sustain growth of lung squamous cell carcinoma. *Scientific Reports*, 6(1):20113, February 2016.

[47] Alex Eccleston and Ritu Dhand. Signalling in cancer. *Nature*, 441(7092):423, 2006.

[48] Muhammad Furqan, Nikhil Mukhi, Byung Lee, and Delong Liu. Dysregulation of jak-stat pathway in hematological malignancies and jak inhibitors for clinical application. *Biomarker research*, 1(1):1–10, 2013.

[49] Chloé James, Valérie Ugo, Jean-Pierre Le Couédic, Judith Staerk, François Delhommeau, Catherine Lacout, Loïc Garçon, Hana Raslova, Roland Berger, Annelise Bennaceur-Griscelli, et al. A unique clonal jak2 mutation leading to constitutive signalling causes polycythaemia vera. *nature*, 434(7037):1144–1148, 2005.

[50] Ross L Levine and D Gary Gilliland. Myeloproliferative disorders. *Blood*, 112(6):2190–2198, 2008.

[51] Marina Bousquet, Cathy Quelen, Veronique De Mas, Eliane Duchayne, Blandine Roquefeuil, Georges Delsol, Guy Laurent, Nicole Dastugue, and Pierre Brousset. The t (8; 9)(p22; p24) translocation in atypical chronic myeloid leukaemia yields a new pcm1-jak2 fusion gene. *Oncogene*, 24(48):7248–7252, 2005.

[52] Pieter Peeters, Sophie D Raynaud, Jan Cools, Iwona Wlodarska, Josiane Grosgeorge, Patrick Philip, Fabrice Monpoux, Luc Van Rompaey, Mathijs Baens, Herman Van den Berghe, et al. Fusion of tel, the ets-variant gene 6 (etv6), to the receptor-associated kinase jak2 as a result of t (9; 12) in a lymphoid and t (9; 15; 12) in a myeloid leukemia. *Blood, The Journal of the American Society of Hematology*, 90(7):2535–2540, 1997.

[53] Frank Griesinger, Heike Hennig, Frauke Hillmer, Martina Podleschny, Rainer Steffens, Andreas Pies, Bernhard Wörmann, Detlef Haase, and Stefan K Bohlander. A bcr–jak2 fusion gene as the result of at (9; 22)(p24; q11. 2) translocation in a patient with a clinically typical chronic myeloid leukemia. *Genes, Chromosomes and Cancer*, 44(3):329–333, 2005.

[54] Andreas Reiter, Christoph Walz, Ann Watmore, Claudia Schoch, Ilona Blau, Brigitte Schlegelberger, Ute Berger, Nick Telford, Shilani Aruliah, John A Yin, et al. The t (8; 9)(p22; p24) is a recurrent abnormality in chronic and acute leukemia that fuses pcm1 to jak2. *Cancer research*, 65(7):2662–2667, 2005.

[55] A Murati, V Gelsi-Boyer, J Adelaide, C Perot, P Talmant, S Giraudier, L Lode, A Letessier, B Delaval, V Brunel, et al. Pcm1-jak2 fusion in myeloproliferative disorders and acute erythroid leukemia with t (8; 9) translocation. *Leukemia*, 19(9):1692–1696, 2005.

[56] Elisabetta Flex, Valentina Petrangeli, Lorenzo Stella, Sabina Chiaretti, Tekla Hornakova, Laurent Knoops, Cristina Ariola, Valentina Fodale, Emmanuelle Clappier, Francesca Paoloni, et al. Somatically acquired jak1 mutations in adult acute lymphoblastic leukemia. *Journal of Experimental Medicine*, 205(4):751–758, 2008.

[57] Qian Zhang, Irena Nowak, Eric C Vonderheid, Alain H Rook, Marshall E Kadin, Peter C Nowell, Leslie M Shaw, and Mariusz A Wasik. Activation of jak/stat proteins involved in signal transduction pathway mediated by receptor for interleukin 2 in malignant t lymphocytes derived from cutaneous anaplastic large t-cell lymphoma and sezary syndrome. *Proceedings of the National Academy of Sciences*, 93(17):9148–9153, 1996.

[58] Cecile Meier, Sylvia Hoeller, Caroline Bourgau, Petra Hirschmann, Juerg Schwaller, Philip Went, Stefano A Pileri, Andreas Reiter, Stephan Dirnhofer, and Alexandar Tzankov. Recurrent numerical aberrations

of jak2 and deregulation of the jak2-stat cascade in lymphomas. *Modern Pathology*, 22(3):476–487, 2009.

[59] Andrew E Schade, Marcin W Wlodarski, and Jaroslaw P Maciejewski. Pathophysiology defined by altered signal transduction pathways: the role of jak-stat and pi3k signaling in leukemic large granular lymphocytes. *Cell Cycle*, 5(22):2571–2574, 2006.

[60] Ralph J Greenspan. The flexible genome. *Nature Reviews Genetics*, 2(5):383–387, May 2001.

[61] Shay Stern, Tali Dror, Elad Stolovicki, Naama Brenner, and Erez Braun. Genome-wide transcriptional plasticity underlies cellular adaptation to novel challenge. *Molecular Systems Biology*, 3(1):106, January 2007.

[62] Zhihua Zhang and Jianzhi Zhang. A Big World Inside Small-World Networks. *PLoS ONE*, 4(5):e5686, May 2009.

[63] John M. Kirkwood, Thomas Richards, Hassane M. Zarour, Jeffrey Sosman, Marc Ernstoff, Theresa L. Whiteside, Joseph Ibrahim, Ronald Blum, Samuel Wieand, and Ruth Mascari. Immunomodulatory effects of high-dose and low-dose interferon α2b in patients with high-risk resected melanoma. *Cancer*, 95:1101–1112, 2002.

[64] Stergios J. Moschos, Howard D. Edington, Stephanie R. Land, Uma N. Rao, Drazen Jukic, Janice Shipe-Spotloe, and John M. Kirkwood. Neoadjuvant Treatment of Regional Stage IIIB Melanoma With High-Dose Interferon Alfa-2b Induces Objective Tumor Regression in Association With Modulation of Tumor Infiltrating Host Cellular Immune Responses. *Journal of Clinical Oncology*, 24:3164–3171, 2006.

[65] Laurence Zitvogel, Lorenzo Galluzzi, Oliver Kepp, Mark J. Smyth, and Guido Kroemer. Type I interferons in anticancer immunity. *Nature Reviews Immunology*, 15:nri3845, 2015.

[66] Ana Maria Gonzalez-Angulo, Bryan T J Hennessy, and Gordon B Mills. Future of Personalized Medicine in Oncology: A Systems Biology Approach. *Journal of Clinical Oncology*, 28(16):2777–2783, June 2010.

[67] Yaron E Antebi, Nagarajan Nandagopal, and Michael B Elowitz. An operational view of intercellular signaling pathways. *Current opinion in systems biology*, 1:16–24, 2017.

[68] Francisco Sanchez-Vega, Marco Mina, Joshua Armenia, Walid K Chatila, Augustin Luna, Konnor C La, Sofia Dimitriadoy, David L Liu, Havish S Kantheti, Sadegh Saghafinia, et al. Oncogenic signaling pathways in the cancer genome atlas. *Cell*, 173(2):321–337, 2018.

[69] Paul Nurse. Life, logic and information. *Nature*, 454(7203):424–426, 2008.

[70] Michael B Elowitz, Arnold J Levine, Eric D Siggia, and Peter S Swain. Stochastic gene expression in a single cell. *Science*, 297(5584):1183–1186, 2002.

[71] Jonathan M Raser and Erin K O'Shea. Control of stochasticity in eukaryotic gene expression. *science*, 304(5678):1811–1814, 2004.

[72] Yaron E Antebi, James M Linton, Heidi Klumpe, Bogdan Bintu, Mengsha Gong, Christina Su, Reed McCardell, and Michael B Elowitz. Combinatorial Signal Perception in the BMP Pathway. *Cell*, 170(6):1184–1196.e24, September 2017.

[73] Xiaoyu Hu and Lionel B Ivashkiv. Cross-regulation of signaling pathways by interferon-γ: implications for immune responses and autoimmune diseases. *Immunity*, 31(4):539–550, 2009.

[74] MS Ko, Hiromitsu Nakauchi, and Naomi Takahashi. The dose dependence of glucocorticoid-inducible gene expression results from changes in the number of transcriptionally active templates. *The EMBO journal*, 9(9):2835–2842, 1990.

[75] Mark C Walters, Steven Fiering, Jeff Eidemiller, Wendy Magis, Mark Groudine, and DI Martin. Enhancers increase the probability but not the level of gene expression. *Proceedings of the National Academy of Sciences*, 92(15):7125–7129, 1995.

[76] Andrew M Kringstein, Fabio MV Rossi, Andreas Hofmann, and Helen M Blau. Graded transcriptional response to different concentrations of a single transactivator. *Proceedings of the National Academy of Sciences*, 95(23):13670–13675, 1998.

[77] Attila Becskei, Bertrand Séraphin, and Luis Serrano. Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. *The EMBO journal*, 20(10):2528–2535, 2001.

[78] 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.

[79] Ryan Suderman, John A Bachman, Adam Smith, Peter K Sorger, and Eric J Deeds. Fundamental trade-offs between information flow in single cells and cellular populations. *Proceedings of the National Academy of Sciences*, 114(22):5755–5760, 2017.

[80] Amiran Keshelava, Gonzalo P Solis, Micha Hersch, Alexey Koval, Mikhail Kryuchkov, Sven Bergmann, and Vladimir L Katanaev. High capacity in g protein-coupled receptor signaling. *Nature communications*, 9(1):876, 2018.

[81] Rahul Satija and Alex K Shalek. Heterogeneity in immune responses: from populations to single cells. *Trends in immunology*, 35(5):219–229, 2014.

[82] Yao Lu, Qiong Xue, Markus R Eisele, Endah S Sulistijo, Kara Brower, Lin Han, El-ad David Amir, Dana Pe'er, Kathryn Miller-Jensen, and

Rong Fan. Highly multiplexed profiling of single-cell effector functions reveals deep functional heterogeneity in response to pathogenic ligands. *Proceedings of the National Academy of Sciences*, 112(7):E607–E615, 2015.

[83] Clive G Bowsher and Peter S Swain. Identifying sources of variation and the flow of information in biochemical networks. *Proceedings of the National Academy of Sciences*, 109(20):E1320–E1328, 2012.

[84] Robin EC Lee, Sarah R Walker, Kate Savery, David A Frank, and Suzanne Gaudet. Fold change of nuclear nf-κb determines tnf-induced transcription in single cells. *Molecular cell*, 53(6):867–879, 2014.

[85] Orsolya Symmons and Arjun Raj. What?s luck got to do with it: single cells, multiple fates, and biological nondeterminism. *Molecular Cell*, 62(5):788–802, 2016.

[86] Ermelinda Porpiglia, Daniel Hidalgo, Miroslav Koulnis, Abraham R Tzafriri, and Merav Socolovsky. Stat5 signaling specifies basal versus stress erythropoietic responses through distinct binary and graded dynamic modalities. *PLoS biology*, 10(8):e1001383, 2012.

[87] Michael A Rowland, Brian Harrison, and Eric J Deeds. Phosphatase specificity and pathway insulation in signaling networks. *Biophysical journal*, 108(4):986–996, 2015.

[88] Michael A Rowland, Joseph M Greenbaum, and Eric J Deeds. Crosstalk and the evolvability of intracellular communication. *Nature communications*, 8:16009, 2017.

[89] Ido Amit, Ron Wides, and Yosef Yarden. Evolvable signaling networks of receptor tyrosine kinases: relevance of robustness to malignancy and to cancer therapy. *Molecular systems biology*, 3(1), 2007.

[90] Isabella Rauch, Mathias Müller, and Thomas Decker. The regulation of inflammation by interferons and their STATs. *JAK-STAT*, 2(1):e23820, January 2013.

[91] M L Schmitz, A Weber, T Roxlau, and M Gaestel. Signal integration, crosstalk mechanisms and networks in the function of inflammatory cytokines. *Biochimica et Biophysica . . .*, 2011.

[92] Fritz Horn and Roy Jackson. General mass action kinetics. *Archive for rational mechanics and analysis*, 47(2):81–116, 1972.

[93] Leonor Michaelis and Maud L Menten. Die kinetik der invertinwirkung. *Biochem. z*, 49(333-369):352, 1913.

[94] Alan C Burton. The properties of the steady state compared to those of equilibrium as shown in characteristic biological behavior. *Journal of Cellular and Comparative Physiology*, 14(3):327–349, 1939.

[95] Athanasios Polynikis, SJ Hogan, and Mario di Bernardo. Comparing different ode modelling approaches for gene regulatory networks. *Journal of theoretical biology*, 261(4):511–530, 2009.

[96] Boris N Kholodenko. Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades. *European journal of biochemistry*, 267(6):1583–1588, 2000.

[97] Nicolaas Godfried Van Kampen. *Stochastic processes in physics and chemistry*, volume 1. Elsevier, 1992.

[98] Daniel T Gillespie. Exact stochastic simulation of coupled chemical reactions. *The journal of physical chemistry*, 81(25):2340–2361, 1977.

[99] Daniel T Gillespie. Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of Chemical Physics*, 115(4):1716–1733, 2001.

[100] Brian Munsky and Mustafa Khammash. The finite state projection algorithm for the solution of the chemical master equation. *The Journal of chemical physics*, 124(4):044104, 2006.

[101] Michał Komorowski, Bärbel Finkenstädt, Claire V Harper, and David A Rand. Bayesian inference of biochemical kinetic parameters using the linear noise approximation. *BMC bioinformatics*, 10(1):343, 2009.

[102] Tina Toni and Bruce Tidor. Combined model of intrinsic and extrinsic variability for computational network design with application to synthetic biology. *PLoS computational biology*, 9(3):e1002960, 2013.

[103] Johan Paulsson. Summing up the noise in gene networks. *Nature*, 427(6973):415, 2004.

[104] Peter S Swain, Michael B Elowitz, and Eric D Siggia. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proceedings of the National Academy of Sciences*, 99(20):12795–12800, 2002.

[105] Andreas Hilfinger and Johan Paulsson. Separating intrinsic from extrinsic fluctuations in dynamic biological systems. *Proceedings of the National Academy of Sciences*, 108(29):12167–12172, 2011.

[106] Vahid Shahrezaei, Julien F Ollivier, and Peter S Swain. Colored extrinsic fluctuations and stochastic gene expression. *Molecular systems biology*, 4(1), 2008.

[107] Christoph Zechner, Jakob Ruess, Peter Krenn, Serge Pelet, Matthias Peter, John Lygeros, and Heinz Koeppl. Moment-based inference predicts bimodality in transient gene expression. *Proceedings of the National Academy of Sciences*, 109(21):8340–8345, 2012.

[108] Matthew Scott, Brian Ingalls, and Mads Kaern. Estimations of intrinsic and extrinsic noise in models of nonlinear genetic networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 16(2):026107, 2006.

[109] M D Brennan, R Cheong, and A Levchenko. How Information Theory Handles Cell Signaling and Uncertainty. *Science*, 338(6105):334–335, 2012.

[110] R Cheong, A Rhee, C J Wang, I Nemenman, and A Levchenko. Information Transduction Capacity of Noisy Biochemical Signaling Networks. *Science*, 334(6054):354–358, 2011.

[111] Thomas M Cover and Joy A Thomas. *Elements of Information Theory*. John Wiley & Sons, 2012.

[112] Claude Elwood Shannon. A mathematical theory of communication. *Bell system technical journal*, 27(3):379–423, 1948.

[113] Gašper Tkačik, Curtis G Callan, and William Bialek. Information flow and optimization in transcriptional regulation. *Proceedings of the National Academy of Sciences*, 105(34):12265–12270, 2008.

[114] Jason A Papin, Tony Hunter, Bernhard O Palsson, and Shankar Subramaniam. Reconstruction of cellular signalling networks and analysis of their properties. *Nature reviews Molecular cell biology*, 6(2):99–111, 2005.

[115] Reinhart Heinrich, Benjamin G Neel, and Tom A Rapoport. Mathematical models of protein kinase signal transduction. *Molecular cell*, 9(5):957–970, 2002.

[116] Jason A Papin and Bernhard O Palsson. Topological analysis of mass-balanced signaling networks: a framework to obtain network properties including crosstalk. *Journal of theoretical biology*, 227(2):283–297, 2004.

[117] Sam Vaseghi, Franz Macherhammer, Susanne Zibek, and Matthias Reuss. Signal transduction dynamics of the protein kinase-a/phosphofructokinase-2 system in saccharomyces cerevisiae. *Metabolic engineering*, 3(2):163–172, 2001.

[118] Ira Swameye, TG Müller, Jens Timmer, O Sandra, and Ursula Klingmüller. Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by databased modeling. *Proceedings of the National Academy of Sciences*, 100(3):1028–1033, 2003.

[119] Boris N Kholodenko, John F Hancock, and Walter Kolch. Signalling ballet in space and time. *Nature reviews Molecular cell biology*, 11(6):414, 2010.

[120] Nikolay Borisov, Edita Aksamitiene, Anatoly Kiyatkin, Stefan Legewie, Jan Berkhout, Thomas Maiwald, Nikolai P Kaimachnikov, Jens Timmer, Jan B Hoek, and Boris N Kholodenko. Systems-level interactions between insulin–egf networks amplify mitogenic signaling. *Molecular systems biology*, 5(1), 2009.

[121] Chun-Chao Wang, Murat Cirit, and Jason M Haugh. Pi3k-dependent cross-talk interactions converge with ras as quantifiable inputs integrated by erk. *Molecular systems biology*, 5(1), 2009.

[122] Edda Klipp, Wolfram Liebermeister, Christoph Wierling, and Axel Kowald. *Systems biology: a textbook*. John Wiley & Sons, 2016.

[123] Guy C. Brown and Boris N. Kholodenko. Spatial gradients of cellular phospho-proteins. *FEBS Letters*, 457:452–454, 1999.

[124] Herbert M. Sauro, Michael Hucka, Andrew Finney, Cameron Wellock, Hamid Bolouri, John Doyle, and Hiroaki Kitano. Next Generation Simulation Tools: The Systems Biology Workbench and BioSPICE Integration. *OMICS: A Journal of Integrative Biology*, 7:355–372, 2003.

[125] Pablo Meyer, Thomas Cokelaer, Deepak Chandran, Kyung H Kim, Po-Ru Loh, George Tucker, Mark Lipson, Bonnie Berger, Clemens Kreutz, Andreas Raue, et al. Network topology and parameter estimation: from experimental design methods to gene regulatory network kinetics using a community based approach. *BMC Systems Biology*, 8(1):13, 2014.

[126] A. Raue, C. Kreutz, T. Maiwald, J. Bachmann, M. Schilling, U. Klingmüller, and J. Timmer. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. *Bioinformatics*, 25(15):1923–1929, 8 2009.

[127] Erich Leo Lehmann, George Casella, and George Casella. Theory of point estimation. wadsworth & brooks. *Cole Advanced Books and Software, Pacific Grove, California*, 16:24–25, 1991.

[128] William H Press, Saul A Teukolsky, William T Vetterling, and Brian P Flannery. *Numerical recipes in Fortran 90*, volume 2. Cambridge university press Cambridge, 1996.

[129] William Q Meeker and Luis A Escobar. Teaching about approximate confidence regions based on maximum likelihood estimation. *The American Statistician*, 49(1):48–53, 1995.

[130] Andrea Saltelli, Marco Ratto, Stefano Tarantola, and Francesca Campolongo. Sensitivity analysis for chemical models. *Chemical Reviews*, 105(7):2811–2828, 2005.

[131] Bryan C Daniels, Yan-Jiun Chen, James P Sethna, Ryan N Gutenkunst, and Christopher R Myers. Sloppiness, robustness, and evolvability in systems biology. *Current Opinion in Biotechnology*, 19(4):389–395, 2008.

[132] Jeremy Gunawardena. Models in systems biology: the parameter problem and the meanings of robustness. *Elements Of Computational Systems Biology*, 1, 2010.

[133] D Degenring, C Froemel, G Dikta, and R Takors. Sensitivity analysis for the reduction of complex metabolism models. *Journal of Process Control*, 14(7):729–745, 2004.

[134] Andrea Saltelli, Marco Ratto, Terry Andres, Francesca Campolongo, Jessica Cariboni, Debora Gatelli, Michaela Saisana, and Stefano Tarantola. *Global sensitivity analysis: the primer*. John Wiley & Sons, 2008.

[135] Raymond Cheong, Alex Rhee, Chiaochun Joanne Wang, Ilya Nemenman, and Andre Levchenko. Information transduction capacity of noisy biochemical signaling networks. *science*, 334(6054):354–358, 2011.

[136] Thalia E Chan, Michael PH Stumpf, and Ann C Babtie. Gene regulatory network inference from single-cell data using multivariate information measures. *Cell systems*, 5(3):251–267, 2017.

[137] Raul Vicente, Michael Wibral, Michael Lindner, and Gordon Pipa. Transfer entropy—a model-free measure of effective connectivity for the neurosciences. *Journal of computational neuroscience*, 30(1):45–67, 2011.

[138] Yu Terada, Tomoyuki Obuchi, Takuya Isomura, and Yoshiyuki Kabashima. Objective and efficient inference for couplings in neuronal network. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(12):124010, dec 2019.

[139] Justin B Kinney and Gurinder S Atwal. Equitability, mutual information, and the maximal information coefficient. *Proceedings of the National Academy of Sciences*, 111(9):3354–3359, 2014.

[140] Lin Song, Peter Langfelder, and Steve Horvath. Comparison of co-expression measures: mutual information, correlation, and model based indices. *BMC bioinformatics*, 13(1):1–21, 2012.

[141] Juan Zhao, Yiwei Zhou, Xiujun Zhang, and Luonan Chen. Part mutual information for quantifying direct associations in networks. *Proceedings of the National Academy of Sciences*, 113(18):5130–5135, 2016.

[142] Wentian Li. Mutual information functions versus correlation functions. *Journal of statistical physics*, 60(5):823–837, 1990.

[143] H C Berg and E M Purcell. Physics of chemoreception. *Biophysical Journal*, 20(2):193–219, 1977.

[144] Anders S Hansen and Erin K O'Shea. Limits on information transduction through amplitude and frequency regulation of transcription factor activity. *Elife*, 4:e06559, 2015.

[145] Michał Komorowski and Dan S Tawfik. The limited information capacity of cross-reactive sensors drives the evolutionary expansion of signaling. *Cell systems*, 8(1):76–85, 2019.

[146] Jacob Cohen, Patricia Cohen, Stephen G West, and Leona S Aiken. *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge, 2013.

[147] Harold Hotelling. Relations between two sets of variates. In *Breakthroughs in statistics*, pages 162–190. Springer, 1992.

[148] Antonietta Guagliardi, Cinzia Giannini, Massimo Ladisa, Antonio Lamura, Teresa Laudadio, Alessia Cedola, Stefano Lagomarsino, and

Ranieri Cancedda. Canonical correlation and quantitative phase analysis of microdiffraction patterns in bone-tissue engineering. *Journal of Applied Crystallography*, 40(5):865–873, 2007.

[149] Ola Friman, Jonny Cedefamn, Peter Lundberg, Magnus Borga, and Hans Knutsson. Detection of neural activity in functional mri using canonical correlation analysis. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 45(2):323–330, 2001.

[150] C Bach, D Ceglia, L Song, and F Duddeck. Randomized low-rank approximation methods for projection-based model order reduction of large nonlinear dynamical problems. *International Journal for Numerical Methods in Engineering*, 118(4):209–241, 2019.

[151] Jim Kay. Feature discovery under contextual supervision using mutual information. In *[Proceedings 1992] IJCNN International Joint Conference on Neural Networks*, volume 4, pages 79–84. IEEE, 1992.

[152] Alfréd Rényi et al. On measures of entropy and information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. The Regents of the University of California, 1961.

[153] Tim Van Erven and Peter Harremos. Rényi divergence and kullback-leibler divergence. *IEEE Transactions on Information Theory*, 60(7):3797–3820, 2014.

[154] Imre Csiszár. Generalized cutoff rates and rényi's information measures. *IEEE Transactions on information theory*, 41(1):26–34, 1995.

[155] Igal Sason and Sergio Verdú. Arimoto–rényi conditional entropy and bayesian m-ary hypothesis testing. *IEEE Transactions on Information Theory*, 64(1):4–25, 2017.

[156] Moshe Ben-Bassat and Josef Raviv. Rényi's entropy and the probability of error. *IEEE Transactions on Information Theory*, 24(3):324–331, 1978.

[157] Sergio Verdú. α-mutual information. In *2015 Information Theory and Applications Workshop (ITA)*, pages 1–6. IEEE, 2015.

[158] Suguru Arimoto. Computation of random coding exponent functions. *IEEE Transactions on Information Theory*, 22(6):665–671, 1976.

[159] Robert G Gallager. Source coding with side information and universal coding. 1979.

[160] Boris Ryabko. Comments on" a source matching approach to finding minimax codes" by davisson, ld and leon-garcia, a. *IEEE Transactions on Information Theory*, 27(6):780–781, 1981.

[161] Semih Yagli, Yücel Altuğ, and Sergio Verdú. Minimax rényi redundancy. *IEEE Transactions on Information Theory*, 64(5):3715–3733, 2018.

[162] Andrew Barron, Jorma Rissanen, and Bin Yu. The minimum description length principle in coding and modeling. *IEEE Transactions on Information Theory*, 44(6):2743–2760, 1998.

[163] Pedro Domingos. The role of occam's razor in knowledge discovery. *Data mining and knowledge discovery*, 3(4):409–425, 1999.

[164] Jorma Rissanen. Modeling by shortest data description. *Automatica*, 14(5):465–471, 1978.

[165] Z Zi. Sensitivity analysis approaches applied to systems biology models. *Systems Biology, IET*, 5(6):336–346, 2011.

[166] Rudiyanto Gunawan, Yang Cao, Linda Petzold, and Francis J Doyle. Sensitivity analysis of discrete stochastic systems. *Biophysical Journal*, 88(4):2530–2540, 2005.

[167] Maria Rodriguez-Fernandez, Julio R Banga, and Francis J Doyle. Novel global sensitivity analysis methodology accounting for the crucial role of the distribution of input parameters: application to systems biology models. *International Journal of Robust and Nonlinear Control*, 22(10):1082–1102, 2012.

[168] GEB Archer, A Saltelli, and IM Sobol. Sensitivity measures, anova-like techniques and the use of bootstrap. *Journal of Statistical Computation and Simulation*, 58(2):99–120, 1997.

[169] I.M. Sobol'. On sensitivity estimation for nonlinear mathematical models. *Mathematical Modelling*, 2(1):112–118, 1990.

[170] Toshimitsu Homma and Andrea Saltelli. Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering & System Safety*, 52(1):1 – 17, 1996.

[171] Maksat Ashyraliyev, Yves Fomekong-Nanfack, Jaap A Kaandorp, and Joke G Blom. Systems biology: parameter estimation for biochemical models. *Febs Journal*, 276(4):886–902, 2009.

[172] D. A. Rand. Mapping the global sensitivity of cellular network dynamics. *Journal of The Royal Society Interface*, 5:S59, 2008.

[173] 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 computational biology*, 3(10):e189, 2007.

[174] K. Erguler and M.P.H. Stumpf. Practical limits for reverse engineering of dynamical systems: a statistical analysis of sensitivity and parameter inferability in systems biology models. *Mol. BioSyst.*, 7(5):1593–1602, 2011.

[175] Roland Brun, Peter Reichert, and Hans R Künsch. Practical identifiability analysis of large environmental simulation models. *Water Resources Research*, 37(4):1015–1030, 2001.

[176] Yunfei Chu and Juergen Hahn. Parameter set selection for estimation of nonlinear dynamic systems. *AIChE journal*, 53(11):2858–2870, 2007.

[177] Yunfei Chu and Juergen Hahn. Parameter set selection via clustering of parameters into pairwise indistinguishable groups of parameters. *Industrial & Engineering Chemistry Research*, 48(13):6000–6009, 2008.

[178] Yunfei Chu and Juergen Hahn. Generalization of a parameter set selection procedure based on orthogonal projections and the d-optimality criterion. *AIChE Journal*, 58(7):2085–2096, 2012.

[179] Andreas Raue, Johan Karlsson, Maria Pia Saccomani, Mats Jirstrand, and Jens Timmer. Comparison of approaches for parameter identifiability analysis of biological systems. *Bioinformatics*, page btt006, 2014.

[180] stork119. stork119/ClusteringIdentifiability: Clustering reveals limits of parameter identifiability in multi-parameter models of biochemical dynamics, May 2021.

[181] T. Lipniacki, P. Paszek, A.R. Brasier, B. Luxon, and M. Kimmel. Mathematical model of NF-κB regulatory module. *Journal of theoretical biology*, 228(2):195–215, 2004.

[182] S. Tay, J.J. Hughey, T.K. Lee, T. Lipniacki, S.R. Quake, and M.W. Covert. Single-cell nf-κb dynamics reveal digital activation and analogue information processing. *Nature*, 466(7303):267–271, 2010.

[183] M. Delhase, M. Hayakawa, Y. Chen, and M. Karin. Positive and negative regulation of iκb kinase activity through ikkβ subunit phosphorylation. *Science*, 284(5412):309–313, 1999.

[184] E.G. Lee, D.L. Boone, S. Chai, S.L. Libby, M. Chien, J.P. Lodolce, and A. Ma. Failure to regulate tnf-induced nf-κb and cell death responses in a20-deficient mice. *Science Signalling*, 289(5488):2350, 2000.

[185] A. Hoffmann, A. Levchenko, M. L. Scott, and D. Baltimore. The IκB-NF-κB Signaling Module: Temporal Control and Selective Gene Activation. *Science*, 298(5596):1241–1245, 2002.

[186] DE Nelson, AEC Ihekwaba, M. Elliott, JR Johnson, CA Gibney, BE Foreman, G. Nelson, V. See, CA Horton, DG Spiller, et al. Oscillations in nf-κb signaling control the dynamics of gene expression. *Science Signalling*, 306(5696):704, 2004.

[187] S.L. Werner, D. Barken, and A. Hoffmann. Stimulus specificity of gene expression programs determined by temporal control of ikk activity. *Science Signalling*, 309(5742):1857, 2005.

[188] T. Lipniacki, K. Puszynski, P. Paszek, A.R. Brasier, and M. Kimmel. Single tnfα trimers mediating nf-κb activation: stochastic robustness of nf-κb signaling. *BMC bioinformatics*, 8(1):376, 2007.

[189] S.L. Werner, J.D. Kearns, V. Zadorozhnaya, C. Lynch, E. O'Dea, M.P. Boldin, A. Ma, D. Baltimore, and A. Hoffmann. Encoding nf-κb temporal control in response to tnf: distinct roles for the negative regulators iκbα and a20. *Genes & Development*, 22(15):2093–2101, 2008.

[190] L. Ashall, C.A. Horton, D.E. Nelson, P. Paszek, C.V. Harper, K. Sillitoe, S. Ryan, D.G. Spiller, J.F. Unitt, D.S. Broomhead, et al. Pulsatile stimulation determines timing and specificity of nf-{kappa} b-dependent transcription. *Science Signalling*, 324(5924):242, 2009.

[191] Stanislas Dehaene. The neural basis of the weber–fechner law: a logarithmic mental number line. *Trends in cognitive sciences*, 7(4):145–147, 2003.

[192] Christopher L Frick, Clare Yarka, Harry Nunns, and Lea Goentoro. Sensing relative signal in the tgf-β/smad pathway. *Proceedings of the National Academy of Sciences*, 114(14):E2975–E2982, 2017.

[193] Lea Goentoro, Oren Shoval, Marc W Kirschner, and Uri Alon. The incoherent feedforward loop can provide fold-change detection in gene regulation. *Molecular cell*, 36(5):894–899, 2009.

[194] Lea Goentoro and Marc W Kirschner. Evidence that fold-change, and not absolute level, of β-catenin dictates wnt signaling. *Molecular cell*, 36(5):872–884, 2009.

[195] James E Ferrell Jr. Signaling motifs and weber's law. *Molecular cell*, 36(5):724–727, 2009.

[196] Oren Shoval, Lea Goentoro, Yuval Hart, Avi Mayo, Eduardo Sontag, and Uri Alon. Fold-change detection and scalar symmetry of sensory input fields. *Proceedings of the National Academy of Sciences*, 107(36):15995–16000, 2010.

[197] Stéphane Chevrier, Helena L Crowell, Vito RT Zanotelli, Stefanie Engler, Mark D Robinson, and Bernd Bodenmiller. Compensation of signal spillover in suspension and imaging mass cytometry. *Cell Systems*, 6(5):612–620, 2018.

[198] Raymond W Yeung. *Information Theory and Network Coding*. Springer Science & Business Media, 2008.

APPENDIX 1: DECLARATIONS OF AUTHOR CONTRIBUTIONS

CO-AUTHORS LIST

1. Tomasz Jetka,
   Institute of Fundamental Technological Research, Polish Academy of Sciences, Warsaw, Poland,
   Contact: t.jetka@sysbiosig.org
2. Michał Komorowski,
   Institute of Fundamental Technological Research, Polish Academy of Sciences, Warsaw, Poland,
   Contact: m.komorowski@sysbiosig.org
3. Karolina Zakrzewska,
   Institute of Fundamental Technological Research, Polish Academy of Sciences, Warsaw, Poland,
   Contact: k.zakrzewska@sysbiosig.org
4. Jarosław Walczak,
   Institute of Fundamental Technological Research, Polish Academy of Sciences, Warsaw, Poland,
   Contact: j.walczak@sysbiosig.org
5. Edyta Głów,
   Institute of Fundamental Technological Research, Polish Academy of Sciences, Warsaw, Poland,
   Contact: e.glow@sysbiosig.org
6. Tomasz Lipniack,
   Institute of Fundamental Technological Research, Polish Academy of Sciences, Warsaw, Poland,
   Contact: tlipnia@ippt.pan.pl
7. Michał Włodarski,
   Faculty of Mathematics Informatics and Mechanics,
   University of Warsaw, Warsaw, Poland
8. Rachel Rigby,
   Medical Research Council Human Immunology Unit,
   Medical Research Council Weatherall Institute of Molecular Medicine,
   Radcliffe Department of Medicine,
   University of Oxford, Oxford, United Kingdom,
   Contact: rachel.rigby@rdm.ox.ac.uk
9. Jan Rehwinkel,
   Medical Research Council Human Immunology Unit,
   Medical Research Council Weatherall Institute of Molecular Medicine,
   Radcliffe Department of Medicine,
   University of Oxford, Oxford, United Kingdom,
   Contact: jan.rehwinkel@ndm.ox.ac.uk
10. Zahra Vahdat,
    Department of Electrical and Computer Engineering,
    University of Delaware, Newark, DE, USA,
    Contact: zahravd@udel.edu
11. Abhyudai Singh,
    Department of Electrical and Computer Engineering,
    Department of Biomedical Engineering and Department of Mathematical Sciences
    University of Delaware, Newark, DE, USA,
    Contact: absingh@udel.edu
12. Zia Farooq,
    Department of Public Health and Clinical Medicine, Umea University,

Umea, Sweden,
Contact: zia.farooq@umu.se

13. Sarah Filippi,
Department of Mathematics, Imperial College London, United Kingdom,
Contact: s.filippi@imperial.ac.uk

14. Michael P.H. Stumpf,
Melbourne Integrative Genomics, University of Melbourne, Australia
Contact: m.stumpf@imperial.ac.uk

15. Tomasz Winarski,
Medical University of Warsaw, Poland
Contact: twinarski@gmail.com

16. Sławomir Błoński,
Institute of Fundamental Technological Research, Polish Academy of Sciences, Warsaw, Poland,
Contact: sblonski@ippt.gov.pl

17. Agata Charzyńska,
Faculty of Mathematics Informatics and Mechanics, University of Warsaw, Poland,
Contact: a.charzynska@nencki.gov.pl

18. Anna Gambin,
Faculty of Mathematics Informatics and Mechanics, University of Warsaw, Poland,
Contact: aniag@mimuw.edu.pl

19. Weronika Wronowska,
Faculty of Mathematics Informatics and Mechanics, University of Warsaw, Poland,

APPENDIX 2: PUBLICATIONS INCLUDED IN THE THESIS

Nienaltowski 2016/6/16 19:58 Page 1 #1

1 Sensitivity Analysis in Quantitative Biology Models

Karol Nienałtowski[1], **Tomasz Jetka**[1], **and Michał Komorowski**[1]

[1]*Institute of Fundamental Technological Research, Polish Academy of Sciences*

Parameters constitute an inherent component of quantitative biology models. Model dependence on parameters is a source of valuable insights as well as a cause of numerous modelling difficulties. Sensitivity analysis provides rigorous tools to quantify how parameter values determine model behaviour. It allows to obtain a deeper understanding of studied phenomena as well as helps in manipulating complex models. In the chapter basic methods of sensitivity analysis are described in the context of quantitative biology dynamical modelling. Specific tools that analyse parameter dependence locally and globally in parameters space are presented in both deterministic and stochastic regimes. Theoretical concepts are followed by illustratory examples.

Nienaltowski 2016/6/16 19:58 Page 2 #2

1.1 Introduction

Quantitative models are intrinsically dependent on parameters. Moreover, compared to engineering or physics problems, dynamical models in quantitative biology typically depend on a relatively large number of parameters [1, 2]. Therefore, understanding how parameter values determine model behaviour is highly intricate and requires a methodological support [3, 4]. Sensitivity analysis provides rigorous tools to investigate how parameter values impact quantitative characteristics of model behaviour [5].

A key concept in sensitivity analysis is the 'sensitivity coefficient', which quantifies how sensitive the model output is to changes in a given parameter. The sensitivity coefficient can be calculated for small changes around nominal values (*local sensitivity*) or arbitrary changes within a range (*global sensitivity*) [6].

Learning which parameters have stronger and weaker relationships to model behaviour is valuable in itself. However, it provides other insights about model properties. For instance, sensitivity analysis shows how robust the system's behaviour is to external perturbations of parameters [7, 8]. Moreover, if experiments exhibit unexpected variability, then sensitivity analysis helps to verify if such observations can be explained by fluctuations in parameter values. Sensitive parameters can be inferred from data with higher accuracy than insensitive parameters. Therefore, calculating sensitivities is helpful to select experiments that yield more informative data to constrain unknown parameter values. Finally, knowledge of parameter sensitivities can also be used to identify and eliminate insensitive parameters, which can lead to simplified models [9].

This chapter introduces methods of sensitivity analysis in the context of quantitative biology dynamical models. We present a small number of the most relevant methods without superfluous technical details. The focus on major methodological concepts is aimed to help in gaining intuition how sensitivity analysis can address specific questions arising in modeling practice. We present both local and global methods in both deterministic and stochastic regimes. The theoretical concepts are presented first and are followed by an illustratory analysis of two example models. The chapter concludes with a small set of exercises to help the student to better understand the presented methods and concepts.

1.2 Methods of sensitivity analysis

In general, a mathematical model of a biological system can be seen as a mapping between a vector of model parameters, $\theta = (\theta_1, \ldots, \theta_l) \in \Theta$, to model responses, $Y = (Y_1, \ldots, Y_k) \in \mathbb{R}^k$. In other words, θ and Y can be thought as the model input (e.g. kinetic rates and initial conditions) and the model output (e.g. mRNA, protein levels nad receptor states). The output, Y, and the parameter vector, θ, are related to each other by a

functional dependence,

$$Y = G(\theta), \tag{1.1}$$

in the deterministic regime or by a probabilistic dependence,

$$Y \sim \mathcal{P}(Y|\theta), \tag{1.2}$$

in the stochastic setting.

The main aim of the sensitivity analysis is to systematically and quantitatively characterise how changes in θ effect changes in Y.

1.2.1 Local methods

Local methods are concerned with small changes (perturbations) in parameter values, Δ, with respect to some nominal value, θ^*. For deterministic models, a natural way to evaluate the local sensitivity is to estimate the derivative using a Taylor expansion for small perturbations

$$Y(\theta^* + \Delta) \approx Y(\theta^*) + \frac{\partial Y}{\partial \theta} \cdot \Delta. \tag{1.3}$$

For stochastic models, the derivative of the output cannot be defined. Instead, one can quantify sensitivity of the distribution of outputs given the inputs, $\mathcal{P}(Y|\theta)$. A statistically meaningful way of such quantification is the use of the Fisher information,

$$\mathbb{E}\left\{\left(\frac{\partial \log(\mathcal{P}(Y|\theta))}{\partial \theta_i}\right)^2\right\}. \tag{1.4}$$

Below we present how local sensitivities can be calculated for quantitative biology dynamical models in the deterministic and stochastic setting.

1.2.1.1 Deterministic models

The approach to compute the derivative of the model output depends on an available form of the functional dependence $G(\theta)$. Deterministic models most often have the form of ordinary differential equations (ODEs) that typically describe how abundances of a set of k entities, $y = (y_1, ..., y_k)$, change with time t given an initial condition $y(t_0) = y_0$

$$\frac{dy}{dt} = F(y, \theta), \tag{1.5}$$

where $F(\cdot)$ is a law that determines the temporal evolution of the system. The output of the model can be then defined as a concatenated vector of values of y at specified times, $(t_1, \ldots, t_n)$,

$$Y(\theta) = (y(t_1), \ldots, y(t_n))^T. \tag{1.6}$$

Alternative definitions of the output may be more appropriate depending on the purpose of the analysis. For instance, the output may contain only selected components of y (e.g. those observed experimentally) or could be a specific function of y (e.g. linear combination or ratio of variables). For simplicity of notation, we assume that the output is defined by Eq. 1.6.

The derivative $\frac{\partial Y}{\partial \theta_i}$ defines the sensitivity vector for the parameter θ_i

$$SV_i = \frac{\partial Y}{\partial \theta_i} = \left(\frac{\partial y(t_1)}{\partial \theta_i}, \ldots, \frac{\partial y(t_n)}{\partial \theta_i} \right)^T, \tag{1.7}$$

where each of $\frac{\partial y(t_j)}{\partial \theta_i}, j = 1, ..., n$, is a k-dimensional row vector. A collection of sensitivity vectors for $i = 1, \ldots, l$ constitutes the sensitivity matrix

$$SM = (SV_1, \ldots, SV_l) \tag{1.8}$$

that contains all information needed to describe local changes in the model output resulting from the inifinitesimal parameter perturbation, $\Delta\theta$,

$$\Delta Y = SM\ \Delta\theta. \tag{1.9}$$

The squared length of ΔY can be then written as

$$\|\Delta Y\|^2 = \Delta Y^T \Delta Y = \Delta\theta^T SM^2\ \Delta\theta, \tag{1.10}$$

where $\|\cdot\|$ is Euclidean norm and SM^2 is the squared sensitivity matrix defined as

$$SM^2 = SM^T\ SM. \tag{1.11}$$

By definition, the i-th diagonal element of SM^2 is $\|SV_i\|^2$. This provides a basic measure of the overall sensitivity to the parameter θ_i. The *local sensitivity coefficient of the i-th parameter* is defined as [10, 11]

$$\boxed{SL_i = \sqrt{SM^2_{i,i}} = \|SV_i\|.} \tag{1.12}$$

The off-diagonal elements of SM^2 correspond to the angles between sensitivity vectors as

$$\cos\left(\sphericalangle(\overrightarrow{SV_i}, \overrightarrow{SV_j}) \right) = \frac{SM^2{}_{i,j}}{\sqrt{SM^2{}_{i,i} \cdot SM^2{}_{j,j}}}. \tag{1.13}$$

and as such represent the ability of one parameter to compensate for changes in the values of the other [4]. The cosine of 1 represents parallel sensitivity vectors, where a change in one parameter can be compensated fully by a change in the other. Conversely, the value of 0 represents orthogonal sensitivity vectors, where no compensation is possible.

For ODE models, the derivative $\frac{\partial Y}{\partial \theta_i}$ can be calculated relatively easily. It can be shown (see Question 1.A) that the derivative with respect to the parameter θ_i of a solution of the Eq. 1.5, $z^{(i)}(t) = \frac{\partial y(t)}{\partial \theta_i}$, is described by [12]

$$\frac{dz^{(i)}(t)}{dt} = \nabla_y F(y(t),\theta)\, z^{(i)}(t) + \frac{\partial F(y(t),\theta)}{\partial \theta_i}, \tag{1.14}$$

where $\nabla_y F(y(t),\theta)$ is the Jacobian matrix of $F(\cdot)$ with respect to y.

Calculation of S requires solving Eq. 1.5 and 1.14, which can be done efficiently with available ODE solvers. Derivation of Eq. 1.14 directly from Eq. 1.5 requires symbolic calculations, which can be effectively achieved with available toolboxes (e.g. [13]).

1.2.1.2 Stochastic models

In the stochastic setting, the model output Y is described by the probability distribution, $Y \sim \mathcal{P}(Y|\theta)$. Therefore conventional derivative of Y cannot be defined. The mean output, $\mu(\theta) = \mathbb{E}\{Y\}$, on the other hand, is deterministic and its derivative can be considered as a way to evaluate local sensitivity [14]

$$SV_i = \frac{\partial \mu(\theta)}{\partial \theta_i}. \tag{1.15}$$

The mean however does not contain entire information about the impact of a parameter on the output [15, 16]. A parameter can for instance change the variability of the output without affecting the mean (compare with Question 4). Therefore one may also consider sensitivity of the variance

$$\frac{\partial \Sigma}{\partial \theta_i} = \frac{\partial \mathbb{E}\left\{(Y-\mu(\theta))^2\right\}}{\partial \theta_i}. \tag{1.16}$$

The derivatives of higher moments could also be taken into account. Yet, to summarise the overall impact of a parameter, the sensitivity of the entire distribution, $\mathcal{P}(Y|\theta)$, should be calculated. The theory of statistical inference [17] showed that a natural candidate is the *the Fisher information matrix (FIM)* defined as

$$\boxed{FIM_{i,j}(\theta) = \mathbb{E}\left\{\frac{\partial \log(\mathcal{P}(Y|\theta))}{\partial \theta_i} \cdot \frac{\partial \log(\mathcal{P}(Y|\theta))}{\partial \theta_j}\right\}} \tag{1.17}$$

Nienaltowski 2016/6/16 19:58 Page 6 #6

It quantifies changes in the probability distribution due to joint changes of all model parameters. The diagonal elements of the *FIM* are used to calculate *local sensitivity coefficients in the stochastic setting*

$$\boxed{SL_i = \sqrt{FIM_{i,i}}.} \tag{1.18}$$

The interpretation of the off-diagonal elements of the *FIM* is not as straightforward as in the deterministic regime [15]. Although the local sensitivity coefficients in the deterministic and stochastic scenarios exhibit many analogies, important differences exist. Matrix SM^2 (Eq. 1.11) gives sensitivities in units of the output, whereas *FIM* measures the relative change of unitless probability density values. As a result, the deterministic and stochastic scenarios cannot be compared directly with each other [18].

In general, calculation of the Fisher information matrix is difficult. For a small number of analytically tractable probability distributions a closed form can be derived. For instance, if a random variable Y follows a multivariate normal distribution (MVN),

$$Y \sim \text{MVN}(\mu(\theta), \Sigma(\theta)), \tag{1.19}$$

with mean, μ, and covariance matrix, Σ, the FIM is given as [19]

$$FIM(\theta)_{i,j} = \frac{\partial \mu}{\partial \theta_i}^T \Sigma^{-1}(\theta) \frac{\partial \mu}{\partial \theta_j} + \frac{1}{2}\text{tr}\left(\Sigma^{-1}\frac{\partial \Sigma}{\partial \theta_i}\Sigma^{-1}\frac{\partial \Sigma}{\partial \theta_j}\right). \tag{1.20}$$

An important example of models with outputs expressed as the MVN distribution are certain stochastic models of biochemical reactions. A primary tool to model stochastic biochemial systems is the Chemical Master Equation (CME) (see Chapter XXX-CME above). One of CME approximations that provides a sound balance between accuracy and insight is the linear noise approximation (LNA) [20] (see Chapter YYY-LNA). It is valid under the assumptions of large copy number of molecules in the system [21]. In LNA model variables have normal distribution, therefore it can be easily employed to quantify sensitivities.

In order to describe how the FIM matrix can be computed for the LNA model, we extend notation and introduce the stoichiometric matrix $S = \{s_{ij}\}_{i=1,\ldots,k}^{j=1,\ldots,r}$ that describes changes in the population sizes due to r different chemical reactions, where s_{ij} is the change in the number of type i molecules after a reaction of type j. The probability that the reaction of type j occurs in time interval $[t, t+dt)$ equals $w_j(y,\theta)dt$, where functions $w_j(y,\theta)$ are parameter dependent transition rates.

With the LNA, if we assume that initially model output has normal distribution, then for each t

$$y(t) \sim \text{MVN}\left(\mu(t,\theta), \Sigma(t,\theta)\right). \tag{1.21}$$

Nienaltowski 2016/6/16 19:58 Page 7 #7

The mean, $\mu(t,\theta)$ and variance, $\Sigma(t,\theta)$, of the system state $y(t)$ can be calculated from ODEs [15, 20]

$$\frac{d\mu(t,\theta)}{dt} = F(\mu,\theta), \tag{1.22}$$

$$\frac{d\Sigma(t,\theta)}{dt} = A(\mu,\theta)\Sigma(t,\theta) + \Sigma(t,\theta)A(\mu,\theta)^T + E(\mu,\theta)E(\mu,\theta)^T, \tag{1.23}$$

where

$$F(\mu,\theta) = S \cdot (w_1(\mu,\theta),...,w_l(\mu,\theta)),$$

$$\{A(\mu,\theta)\}_{i,k} = \sum_{j=1}^{r} s_{ij}\frac{\partial w_j}{\partial \mu_k}, \tag{1.24}$$

$$E(\mu,\theta) = S\sqrt{diag(F(\mu,\theta))}.$$

If the output is defined as a trajectory measured at times $t_1,\ldots,t_n$

$$Y = (y(t_1),\ldots,y(t_n))^T$$

then it can also be shown that

$$Y \sim \text{MVN}\left(\mu((t_1,\ldots,t_n),\theta), \Sigma((t_1,\ldots,t_n),\theta)\right). \tag{1.25}$$

The matrix

$$\Sigma((t_1,\ldots,t_n),\theta) = \begin{pmatrix} \Sigma(t_1,\theta) & \text{Cov}\{y(t_1),y(t_2)\} & \ldots & \text{Cov}\{y(t_1),y(t_n)\} \\ \text{Cov}\{y(t_1),y(t_2)\} & \Sigma(t_2,\theta) & \ldots & \text{Cov}\{y(t_2),y(t_n)\} \\ \vdots & \vdots & \ldots & \vdots \\ \text{Cov}\{y(t_1),y(t_n)\} & \text{Cov}\{y(t_2),y(t_n)\} & \ldots & \Sigma(t_n,\theta) \end{pmatrix} \tag{1.26}$$

is $nk \times nk$ dimensional and requires covariances between time points t_i and t_j, $\text{Cov}\left\{y(t_i),y(t_j)\right\}$, to be calculated. These are determined by the following equation

$$\text{Cov}\left\{y(t_i),y(t_j);\theta\right\} = \Sigma(t_i,\theta)\Phi(t_i,t_j)^T \quad \text{for} \quad t_i \le t_j, \tag{1.27}$$

where $\Phi(t_i,t_j)$ is the fundamental matrix of the non-autonomous system of ODEs

$$\frac{dX}{dt_j} = A(\mu,\theta)X(t_j), \tag{1.28}$$

such that $\Phi(t_i,t_i) = I$ with I being the identity matrix.

Therefore in the LNA both mean μ and covariance matrix Σ are given by solutions of ODEs and the Eq. 1.14 provides a method to calculate their derivatives with respect to parameters. Hence, the formula of the Eq. 1.20 can be used to calculate the FIM. A detailed description of this procedure can be found in [15] and is implemented in Matlab package – StochSens [13].

Nienaltowski 2016/6/16 19:58 Page 8 #8

1.2.2 Global methods

The sensitivities provided by the local methods are valid around nominal parameter values, $\theta^{\star}$, and may be substantially different from sensitivities calculated for other points in the parameter space, $\theta \in \Theta$. Methods for global sensitivity analysis (GSA) aim to provide a remedy to this drawback [6]. Instead of using fixed nominal values of the parameters GSA methods specify a probability distribution, $\mathcal{P}(\theta)$, of plausible parameter values [22]. With this distribution, one can sample parameter space and assemble a representative collection of model outputs in order to analyse impact of parameters on the output. How the impact of parameters is assessed varies between methods and depends on the specific aim of the analysis.

Here we focus on three commonly used GSA techniques. The so-called *screening* technique is the conceptually simplest and computationally most efficient method, therefore is used as a first-line tool before more advanced tools are employed [23, 24]. Next, *variance based* methods quantify sensitivity of parameters using variance of the output resulting from sampling the model parameters from the distribution $\mathcal{P}(\theta)$. Finally, *filtering* allows users to measure the impact of parameters on specifically defined properties of model behaviour. All GSA methods described here rely on sampling parameters and corresponding model outputs. Therefore both deterministic and stochastic models can be analysed using these methods without substantial modifications [25].

1.2.2.1 Screening

Screening can be seen as a simple generalisation of the local methods. Sensitivity coefficients are not calculated for nominal values only but for values on a grid, Ω, that represent the considered parameter space, Θ. A simple grid is usually created by setting a minimal and a maximal value for each of the model parameters, and uniformly distributing intermediate values. Formally, the i-th parameter θ_i can take h values in the set Ω_i

$$\theta_i \in \Omega_i = \left\{ \theta_i^{min}, \theta_i^{min} + \frac{(\theta_i^{max} - \theta_i^{min})}{h-1}, \ldots, \theta_i^{min} + \frac{(h-2)(\theta_i^{max} - \theta_i^{min})}{h-1}, \theta_i^{max} \right\}, \quad (1.29)$$

where θ_i^{min} and θ_i^{max} are lower and upper bounds, respectively. The grid is then defined as the product $\Omega = \Omega_1 \times \ldots \times \Omega_l$, which are usually called elementary effects (EEs). *The elementary effect* of the i-th parameter between the point $\theta = (\theta_1, \ldots, \theta_i, \ldots, \theta_l) \in \Omega$ and its incremental change $\theta + \Delta_i = (\theta_1, \ldots, \theta_i + \Delta, \ldots, \theta_l) \in \Omega$ is defined as

$$EE_i = \frac{dist\,(Y(\theta), Y(\theta + \Delta_i))}{\Delta}, \quad (1.30)$$

where $dist(\cdot)$ is a distance function between two corresponding outputs, e.g. Euclidian distance. Definitions of the EEs may differ depending on the application [24] . The definition presented above is presented as it is suitable for analysis of models with multi-dimensional output.

Quantification of sensitivities of individual parameters requires computations of EEs across the entire grid. This can be performed systematically using a variety of approaches including the Morris Trajectory algorithm (Box 1) selected here for simplicity [23]. In each run, $j = 1, \ldots, r$, the algorithm generates a vector of EEs, $EE^{(j)} = (EE_1, ..., EE_l)$, such that i-the element of the vector contains EE corresponding to i-th parameter . Repeated runs produce a representative sample of elementary effects $(EE^{(1)}, EE^{(2)}, \ldots, EE^{(r)})$. The sample is intended to cover the entire grid and is used to estimate *the mean*, μ_i, and *the standard deviation*, σ_i, of EEs

$$\mu_i = \frac{1}{r}\sum_{j=1}^{r} EE_i^{(j)}, \qquad \sigma_i = \sqrt{\frac{1}{r-1}\sum_{j=1}^{r}\left(EE_i^{(j)} - \mu_i\right)^2}, \tag{1.31}$$

where $EE_i^{(j)}$ is the elementary effect of the i-th parameter calculated in the j-th run.

The estimate of μ_i measures the overall influence of the i-th parameter on the output. The standard deviation, σ_i, quantifies the extent to which the influence of the parameter varies across the considered grid. The EE of the i-th parameter depends on values of other parameters, therefore σ_i can be also interpreted as a measure of the interactions of the i-th parameter with other parameters.

The screening method can be applied in both deterministic and stochastic scenario with appropriate modifications of the distance function, $dist\,(Y(\theta), Y(\theta + \Delta_i))$. In the deterministic case, the Euclidian distance is most often used. In stochastic scenario, the distance must quantify differences between the probability distributions $\mathcal{P}(\cdot|\theta)$ and $\mathcal{P}(\cdot|\theta + \Delta_i)$ [26]. This can be achieved using one of the available measures e.g. Kolmogorov distance, Hellinger distance or Kullback-Leibler divergence [27, 28], of which the Kolmogorov distance is prevalently used to ease computations. The Kolmogorov distance between the random variables U and V is defined as

$$dist(U, V) = \sup_{u\in\mathbb{R}}\{|F_U(u) - F_V(u)|\}, \tag{1.32}$$

where $F_U(\cdot)$ and $F_V(\cdot)$ are empirical cumulative distribution functions of U and V, respectively. Within the screening method, the Kolmogorov distance can be therefore used to calculate the distance between distributions $\mathcal{P}(\cdot|\theta)$ and $\mathcal{P}(\cdot|\theta + \Delta_i)$, based on samples collected for parameter values on the Morris trajectories.

Nienaltowski 2016/6/16 19:58 Page 10 #10

Box 1

function MORRIS TRAJECTORY
  $\mathcal{I} = \{1, 2, \ldots, l\}$ ▷ Set of parameters indexes
  $\theta^{(0)} = \mathcal{P}(\Omega_1 \times \Omega_2 \times \ldots \times \Omega_l)$ ▷ Sampling initial parameters values
  $Y^{(0)} = G(\theta^{(0)})$ ▷ Initial model output
  for q = $(1, 2, \ldots, l)$ **do** ▷ Index of subsequent algorithm steps – q
    $\theta^{(q)} = \theta^{(q-1)}$ ▷ Assigning parameters values from the previous step
    $i = \mathcal{U}(\mathcal{I})$ ▷ Sampling parameter index i from uniform distribution
    $\theta_i^{(q)} = \mathcal{P}(\Omega_i)$ ▷ Sampling value of parameter θ_i
    $\Delta = |\theta_i^{(q)} - \theta_i^{(q-1)}|$
    $Y^{(q)} = G(\theta^{(q)})$ ▷ Model output
    $EE_i = \frac{dist(Y^{(q)}, Y^{(q-1)})}{\Delta}$
    $\mathcal{I} = \mathcal{I} \setminus \{i\}$ ▷ Removing index of the already computed elementary effect
  end for
  return $(EE_1, EE_2, \ldots, EE_l)$
end function

1.2.2.2 Variance based methods

Variance based methods offer an alternative approach to quantify the global sensitivities of model parameters. Instead of using differences in outputs generated by perturbations of parameters, as in the local methods and screening, the overall variability of the output is used to extract information about parameter sensitivities. Precisely, the distribution of parameters, $\mathcal{P}(\theta)$, induces a corresponding distribution of outputs, $\mathcal{P}(Y) = \int_\Theta \mathcal{P}(Y|\theta)\mathcal{P}(\theta)d\theta$, where in the deterministic scenario $\mathcal{P}(Y|\theta) = \delta_{G(\theta)}(Y)$, with δ being Dirac delta, and in the stochastic regime $\mathcal{P}(Y|\theta)$ is the analysed model itself. The variance of the distribution $\mathcal{P}(Y)$ can be decomposed into contributions resulting from individual parameters. Using the variance decomposition theorem [29] the overall variance can be written as the sum of the components corresponding to the parameter θ_i and the remaining sources of variability,

$$\mathrm{Var}\{Y\} = \underbrace{\mathbb{E}_{\theta_i}\left\{\mathrm{Var}_{\theta_{-i}}\{Y|\theta_i\}\right\}}_{\text{variability generated by } \theta_i} + \underbrace{\mathrm{Var}_{\theta_i}\left\{\mathbb{E}_{\theta_{-i}}\{Y|\theta_i\}\right\}}_{\text{other sources}}, \tag{1.33}$$

where θ_{-i} is the vector of all model parameters except θ_i, i.e. $\theta_{-i} = (\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_l)$; $\mathbb{E}_{\theta_i}\{\cdot\}$ and $\mathrm{Var}_{\theta_i}\{\cdot\}$ are the expected value and variance with respect to the distribution $\mathcal{P}(\theta_i) = \int_\Theta \mathcal{P}(\theta)d\theta_{-i}$; $\mathbb{E}_{\theta_{-i}}\{Y|\theta_i\}$ and $\mathrm{Var}_{\theta_{-i}}\{Y|\theta_i\}$ are the expected value and the variance with respect to the distribution $\int_\Theta P(Y|\theta)P(\theta_{-i}|\theta_i)d\theta_{-i}$.

Nienaltowski 2016/6/16 19:58 Page 11 #11

Intuitively, a sensitive parameter is expected to have a large contribution to the overall variance. The decomposition of the Eq. 1.33 demonstrates that if the parameter θ_i has a substantial contribution to the variability of the model output the component $\mathrm{Var}_{\theta_i}\left\{\mathbb{E}_{\theta_{-i}}\{Y|\theta_i\}\right\}$ is large. Therefore the ratio of the conditional variance $\mathrm{Var}_{\theta_i}\left\{\mathbb{E}_{\theta_{-i}}\{Y|\theta_i\}\right\}$ and the overall variance is used as the so called *first-order sensitivity index* of θ_i [30]

$$SG_i^S = \frac{\mathrm{Var}_{\theta_i}\left\{\mathbb{E}_{\theta_{-i}}\{Y|\theta_i\}\right\}}{\mathrm{Var}\{Y\}} \tag{1.34}$$

The major caveat of the above coefficient is that it does not quantify the entire variability of the output Y induced by the parameter θ_i. Precisely, it quantifies only the direct contribution of each parameter but neglects contribution exerted by interactions with other parameters. This can be illustrated by comparing the sum of the first-order contributions with the total output variability. Precisely, it can be shown (see Question 1.B) that

$$\mathrm{Var}\{Y\} \geq \sum_{i=1}^{l} \mathrm{Var}_{\theta_i}\left\{\mathbb{E}_{\theta_{-i}}\{Y|\theta_i\}\right\}. \tag{1.35}$$

In order to measure the indirect effects the decomposition of the Eq. 1.33 can be written in the reversed order

$$\mathrm{Var}\{Y\} = \underbrace{\mathbb{E}_{\theta_{-i}}\left\{\mathrm{Var}_{\theta_i}\{Y|\theta_{-i}\}\right\}}_{\text{variability generated by } \theta_i} + \underbrace{\mathrm{Var}_{\theta_{-i}}\left\{\mathbb{E}_{\theta_i}\{Y|\theta_{-i}\}\right\}}_{\text{other sources}}. \tag{1.36}$$

Now the contribution of the parameter θ_i is expressed as $\mathbb{E}_{\theta_{-i}}\left\{\mathrm{Var}_{\theta_i}\{Y|\theta_{-i}\}\right\}$. Compared to the $\mathbb{E}_{\theta_i}\left\{\mathrm{Var}_{\theta_{-i}}\{Y|\theta_i\}\right\}$, it includes the variability that is induced through interactions with other model parameters and leads to *the total sensitivity coefficient* [31] (see Question 1.B)

$$SG_i^T = \frac{\mathbb{E}_{\theta_{-i}}\left\{\mathrm{Var}_{\theta_i}\{Y|\theta_{-i}\}\right\}}{\mathrm{Var}\{Y\}}. \tag{1.37}$$

Sensitivity coefficients SG_i^S and SG_i^T have a straightforward interpretation. The value of SG_i^S estimates how much the total variance of the model output could be reduced if the i-th parameter was fixed. Conversely, coefficients SG_i^T estimate how much total variance would remain if all parameters except θ_i were fixed. Using both coefficients in parallel allows to better understand the total impact of each parameter. For instance, a low SG_i^S and high SG_i^T, in relation to other parameters, suggests that parameter θ_i is sensitive through its interactions with other parameters, but not directly. When considering parameters to

Nienaltowski 2016/6/16 19:58 Page 12 #12

be removed from complex models, the total-effect coefficient is a more reliable tool as it provides a more comprehensive quantification of parameter impact.

Calculation of the variances and expected values used to express the sensitivity coefficients in Eq. 1.37 and 1.34 can be achieved relatively easily using Monte Carlo approaches by generating a sample of parameters, from the distribution $\mathcal{P}(\theta)$, and corresponding outputs. Unfortunately, obtaining precise estimation using direct sampling is inefficient [25] as a large sample size is required. Therefore, quasi-Monte Carlo sequences instead of Monte Carlo sampling or Fourier Amplitude Sensitivity Test (FAST) are often used to improve accuracy of the estimates [32].

1.2.2.3 Monte Carlo Filtering

In comparison to the previous techniques, Monte Carlo Filtering (MCF) does not provide sensitivity coefficients, but rather intends to quantify the role of each single parameter in inducing a specifically defined property of model behaviour. For instance, it is often of interest to understand what combination of parameter values is responsible for generating oscillations or causing a specific response variable to exceed a defined threshold.

Formally, MCF requires the condition to be defined as a binary function of the model output $cond$: $\mathbb{R}^k \rightarrow \{0, 1\}$,

$$cond(Y) = \begin{cases} 1 & \text{if condition is satisfied} \\ 0 & \text{otherwise} \end{cases}, \tag{1.38}$$

which takes 1 if the condition is satisfied (e.g. oscillations exist) and 0 otherwise.

The aim of the method is to detect to what extent the value of the condition is determined by each parameter θ_i. This is achieved by sampling parameters from the distribution $\mathcal{P}(\theta)$ and obtaining empirical estimates of the probabilities $\mathcal{P}(\theta_i|\ cond(Y) = 0)$ and $\mathcal{P}(\theta_i|\ cond(Y) = 1)$, separately for each single parameter θ_i, that describe how likely the value of the condition is due to a specific value of the parameter θ_i. If both probability distributions are identical, then the value of θ_i carries no information about the value of the condition. On the other hand, if these distributions have distinct supports then value of the parameter θ_i solely determines value of the condition. Therefore the degree to which the distributions differ is a sensible indicator of the impact of the parameter θ_i on the specified condition. Differences are usually quantified using Kolmogorov distance (Eq. 1.32), as it is relatively straightforward to estimate, which is interpreted as a measure of the impact of the single parameter θ_i on value of the condition.

1.2.3 Logarithmic parametrisation

Differences in order of magnitudes or units often make both variables and parameters incomparable within each other. Therefore it is advisable to consider relative rather than

Nienaltowski 2016/6/16 19:58 Page 13 #13

absolute changes when performing sensitivity analysis. A simple way to introduce relative sensitivities is to use logarithmic transformation. In case of local methods, derivatives can be modified as follows

$$\frac{\partial \log Y}{\partial \log \theta} = \frac{\partial Y}{\partial \theta}\frac{\theta}{Y} = \frac{\frac{\partial Y}{Y}}{\frac{\partial \theta}{\theta}}. \tag{1.39}$$

This simple relation gives a straightforward transformation from the relative to absolute sensitivities so that the vectors of relative sensitivities,

$$\widetilde{SV}_i = \left(\frac{\partial \log y_m(t_1)}{\partial \log \theta_i}, \ldots, \frac{\partial \log y_m(t_n)}{\partial \log \theta_i}\right), \tag{1.40}$$

can be computed directly from the previously defined vectors SV_i (Eq. 1.7).

In case of stochastic models, the definition of the Fisher information matrix already includes the logarithm of the density function. The transformation of parameters is performed analogously as in the Eq. 1.39.

In order to account for relative changes when using GSA methods, two modifications can be introduced: (i) the output can be defined in terms of logarithms of endogenous variables; (ii) the distribution of parameters should be appropriately selected. For instance, instead of assuming that the parameter θ_i is sampled from a uniform distribution at the interval $[\theta_i^{min}, \theta_i^{max}]$, i.e. $\mathcal{U}\left(\theta_i^{min}, \theta_i^{max}\right)$, one can assume that the parameter is sampled from the distribution $10^{\mathcal{U}(\log_{10}(\theta_i^{min}), \log_{10}(\theta_i^{max}))}$ [22].

1.3 Applications of the sensitivity analysis

In order to illustrate the presented methods, we analyse two generic examples of dynamical models in quantitative biology. First, simple model of gene expression (GE) is sufficiently simple to allow a clear comparison of results provided by the sensitivity analysis with intuition. The second example represents a minimal model of a signalling pathway (SP) activating a transcription factor.

1.3.1 Model examples

1.3.1.1 Gene expression (GE)

We assume that the gene expression process begins with the production of mRNA molecules at rate θ_1. Each mRNA molecule, y_1, may be independently translated into protein molecules, y_2, at rate θ_2. Both mRNA and protein molecules are degraded at rates θ_3 and θ_4, respectively. Therefore, the vector of model parameters is $\theta = (\theta_1, \theta_2, \theta_3, \theta_4)$, and the system of ODEs describing the model can be written as follows

$$\text{(mRNA)} \quad \frac{dy_1}{dt} = \theta_1 - \theta_3 y_1, \qquad \text{(protein)} \quad \frac{dy_2}{dt} = \theta_2 y_1 - \theta_4 y_2, \tag{1.41}$$

Nienaltowski 2016/6/16 19:58 Page 14 #14

and has the following stationary solution $y^* = \left(y_1^*,\ y_2^*\right)$

$$y_1^* = \frac{\theta_1}{\theta_3}, \quad y_2^* = \frac{\theta_1}{\theta_3}\frac{\theta_2}{\theta_4}. \tag{1.42}$$

The stationary solution of the stochastic version of the model can be obtained using the linear noise approximation by solving the Eq. 1.22 and Eq. 1.23. Then, $y^* \sim \text{MVN}(\mu, \Sigma)$ with

$$\mu = \left(\frac{\theta_1}{\theta_3}, \frac{\theta_1\theta_2}{\theta_3\theta_4}\right), \quad \Sigma = \begin{pmatrix} \frac{\theta_1}{\theta_3} & \frac{\theta_1\theta_2}{\theta_3(\theta_3+\theta_4)} \\ \frac{\theta_1\theta_2}{\theta_3(\theta_3+\theta_4)} & \frac{\theta_1\theta_2}{\theta_3\theta_4}\cdot\left(\frac{\theta_2+\theta_3+\theta_4}{\theta_3+\theta_4}\right) \end{pmatrix}. \tag{1.43}$$

The schematic of the model together with trajectories corresponding to the deterministic and stochastic versions of the model are presented in the Fig. 1.1.A-B.

1.3.1.2 Signaling pathway (SP)

The second example represents a minimal model that aims to incorporate major components of a signaling pathway [33, 34]. An extracellular ligand, with a time dependent concentration denoted as $u(t)$, leads to the activation of kinase, y_1. The activated kinase induces activation of a transcription factor, y_2. Subsequently, the TF molecules translocate to the nucleus, y_3, where they modulate expression of specific genes. In particular, they induce transcription of repressor molecules, y_4, that inhibit the activation of the kinase and spontaneously degrade. For simplicity, we assume that the produced repressor molecules are immediately present in the cytoplasm. The process is depicted in the Fig. 1.1.C and can be described by the following ODEs

$$\begin{aligned}
\text{(active kinase)} \quad & \frac{dy_1}{dt} = \frac{u(t)}{1+u(t)}(\theta_1 - y_1)\frac{1}{1+y_4/\theta_2} - \theta_3 y_1, \\
\text{(cytoplasmic activated TF)} \quad & \frac{dy_2}{dt} = \frac{y_1/\theta_4}{1+y_1/\theta_4}(\theta_5 - y_2 - y_3) - \theta_6 y_2 - \theta_7 y_2, \\
\text{(nuclear activated TF)} \quad & \frac{dy_3}{dt} = \theta_7 y_2 - \theta_8 y_3, \\
\text{(repressor of signaling)} \quad & \frac{dy_4}{dt} = \theta_9 \frac{(y_3/\theta_{10})^{\theta_{12}}}{1+(y_3/\theta_{10})^{\theta_{12}}} - \theta_{11} y_4,
\end{aligned}$$

where θ_1 is the total number of kinase molecules; θ_2 is the dissociation constant for the repressor activity; θ_3 is the kinase deactivation rate; θ_4 is the dissociation constant for the TF activation by the kinase; θ_5 is the total number of the TF molecules; θ_6 is the TF deactivation rate; θ_7 is the TF nuclear translocation rate; θ_8 is the TF cytoplasmic translocation rate; θ_9 is the maximal repressor transcription rate; θ_{10} is the dissociation constant for repressor transcription activation by TF; and θ_{11} is the repressor degradation rate. In order

to keep the example as simple as possible, we assume a constant stimulation profile for time $t > 0$ i.e. $u(t) = 100$ and that at time $t = 0$ the system reached the stationary state with $u(t) = 0$ for $t \leq 0$. Knowing the stoichiometry and reaction rates involved in the process allows us to write a stochastic version of the model and simulate it in the CME framework. A schematic of the model together with trajectories corresponding to the deterministic and stochastic scenarios are presented in the Fig. 1.1.C-D.

1.3.2 Results

Below we apply the introduced local and global methods to analyse the deterministic and stochastic versions of the GE and SP models.

1.3.2.1 Gene expression model

The first step in the analysis of any model is to choose the model output. In the case of the GE model, the protein level is a natural candidate. For simplicity, we also assume that only the stationary state is of interest. Therefore we set $Y = y_2^s$.

Deterministic scenario (local). The simplicity of the example allows local sensitivity analysis of the deterministic version of the model to be performed analytically. With the deterministic solution for the stationary state (Eq. 1.42) the sensitivity (Eq. 1.40) can be expressed as

$$\widetilde{SM} = \left(\frac{\partial \log \frac{\theta_1\theta_2}{\theta_3\theta_4}}{\partial \log \theta_1}, \frac{\partial \log \frac{\theta_1\theta_2}{\theta_3\theta_4}}{\partial \log \theta_2}, \frac{\partial \log \frac{\theta_1\theta_2}{\theta_3\theta_4}}{\partial \log \theta_3}, \frac{\partial \log \frac{\theta_1\theta_2}{\theta_3\theta_4}}{\partial \log \theta_4}\right) = (1, 1, -1, -1). \tag{1.44}$$

Sensitivities SL_i can be read from matrix $SM^2 = \widetilde{SM}^T \widetilde{SM}$ (Eq. 1.11). Therefore, we obtain $SL_1 = SL_2 = SL_3 = SL_4 = 1$. This is consistent with intuition. Each parameter has exactly the same role in changing the stationary level of the protein y_2^s regardless of the nominal parameter values. Change of 1% in value of each parameter leads to the 1% change of the protein abundance.

Stochastic scenario (local). Analytical calculation of the SL_i coefficients for the stochastic scenario although in principle possible, using the Eq. 1.43, could be troublesome (see Question 4). Therefore, we calculate these numerically using the software package StochSens [13]. In order to calculate the coefficients, we first set the nominal parameter values, as presented in the Table 1.1. Parameters are scaled so that time is measured in minutes, whereas species are expressed as total number of molecules. The analysis demonstrates that all parameters exhibit almost identical importance similarly as in the deterministic scenario (see Fig. 1.2.A). Values of sensitivity coefficients can be compared only between parameters. Comparison between the stochastic and deterministic regimes is not justified as coeffecients for the two cases have different units.

Nienaltowski 2016/6/16 19:58 Page 16 #16

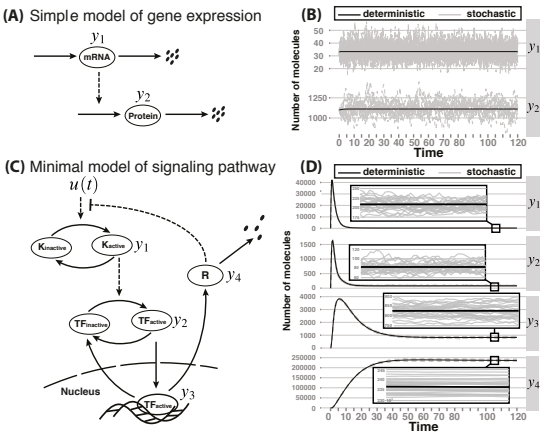


Figure 1.1

Illustration of the models. (**A**) Gene expression model (GE); (**B**) Signaling pathway model (SP); (**C**) Simulated trajectories of the GE model for a nominal parameter set in the deterministic (solid line) and the stochastic (dashed lines) regime; (**D**) Simulated trajectories of the SP model or nominal parameter set in the deterministic (solid line) and the stochastic (dashed lines) regime. A time interval between 105 and 110 minute is zoomed for each species. Nominal parameters are presented in the Table 1.1.

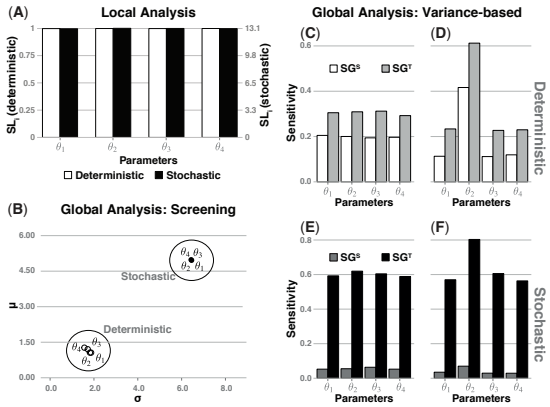


Figure 1.2

Sensitivity analysis of the deterministic (white colouring) and stochastic (black colouring) versions of the GE model using local and global methods. (**A**) Local sensitivity coefficients, SL_i, for the deterministic (white boxes, left axis) and stochastic (black boxes, right axis) regimes. Boxes are scaled so that they have the same relative magnitudes for each regime. (**B**) Mean elementary effects and their standard deviations (Eq. 1.31) for each model parameter obtained by the screening method using $r = 10^4$ and $h = 10^3$ in deterministic (white circles, left and bottom axis) and stochastic (black circles, right and top axis) cases. (**C-F**) Comparison of the first-order, SG_i^S, and total sensitivity coefficients, SG_i^T within the deterministic (**C, D**) and stochastic (**E, F**) regimes. (**D, F**) Effects of altering of the parameter distribution, $\mathcal{P}(\theta)$ on the first-order and total sensitivity coefficients. For plots(**C, E**) θ_2 is sampled from the different distribution $\theta_2^* \cdot 10^{\mathcal{U}(-1,1)}$ whereas for plots (**D, F**)from a wider distribution $\theta_2^* \cdot 10^{\mathcal{U}(-2,2)}$, where θ^* is defined in the Table 1.1 and $\mathcal{U}(-a,a)$ is the uniform on the interval $(-a,a)$.

Nienaltowski 2016/6/16 19:58 Page 18 #18

Deterministic scenario (global). In order to compare local methods with global analysis, we have first analysed the deterministic version of the model with the screening and variance based methods. Results of the screening illustrated in the Fig. 1.2.B depict the relation between means and standard deviations of elementary effects (Eq. 1.30). The approximately equal means of elementary effects of all parameters show that, as expected, the parameters have a similar impact. The standard deviations of the elementary effects are also similar for all parameters indicating that the impact of each parameter is similar across the considered parameter grid.

Similar conclusions are obtained using the variance based methods (Fig. 1.2.C-F). To calculate the variance based coefficients, we assumed that the probability distribution describing plausible parameters, $\mathcal{P}(\theta)$, corresponds to the following sampling procedure: each parameter θ_i is sampled by generating an auxiliary random variable, u_i, from the uniform distribution on the interval $(-1, 1)$, i.e. $u_i \sim \mathcal{U}(-1, 1)$, and setting $\theta_i = \theta_i^* \cdot 10^{u_i}$, where θ_i^* is a fixed nominal value. Under this assumption according to the Eq. 1.34 and the form of the output, $Y = \frac{\theta_1\theta_2}{\theta_3\theta_4}$, the first order coefficient of the parameter θ_1 is equal to

$$SG_1^S = \frac{\theta_1^*\theta_2^*}{\theta_3^*\theta_4^*} Var\left\{10^{u_1}\right\} \mathbb{E}\left\{\frac{10^{u_2}}{10^{u_3}10^{u_4}}\right\} = \frac{\theta_1^*\theta_2^*}{\theta_3^*\theta_4^*} Var\left\{10^{\mathcal{U}(-1,1)}\right\} \left(\mathbb{E}\left\{10^{\mathcal{U}(-1,1)}\right\}\right)^3. \quad (1.45)$$

Performing the same calculation for the remaining parameters be get

$$SG_1^S = SG_2^S = SG_3^S = SG_4^S. \quad (1.46)$$

The above derivation demonstrates why the numerically calculated values of the variance based coefficients (Fig. 1.2.C) are similar for all parameters. It also shows why assumptions regarding the probability distribution, $\mathcal{P}(\theta)$, may have a significant impact on results of the analysis. To demonstrate the latter, we considered the case where θ_2 is sampled from a wider space (Fig. 1.2.D), $u_2 \sim \mathcal{U}(-2, 2)$ as opposed to the previous case with $u_2 \sim \mathcal{U}(-1, 1)$. Altering the distribution results in a significant relative increase of the first order sensitivity coefficient of the parameter θ_2 (Fig. 1.2.D). A similar property is exhibited by the total sensitivity coefficients (Fig. 1.2.C and Fig. 1.2.D).

Stochastic scenario (global). The stochastic version of the model was analysed with the screening and variance based methods yielding very similar results as the deterministic version. The obtained sensitivity coefficients are presented in the Fig. 1.2.B,E & F.

1.3.2.2 Signaling pathway model

The simplicity of the GE model made it possible to explain the results of sensitivity analysis with analytical considerations and confront them with intuition. However that is rarely possible. Here we analyse the SP model to demonstrate that sensitivity analysis can provide insight that is beyond intuition. We define the output, Y, as the temporal profile of

Nienaltowski 2016/6/16 19:58 Page 19 #19

the nuclear transcription factor level. Thus we assume that the value of the variable y_3 at multiples of 5 minutes ranging from 0 to 120 minutes is the output of the system

$$Y = (y_3(0), y_3(5), \ldots, y_3(115), y_3(120)). \tag{1.47}$$

As we assume that the system is activated at time $t = 0$ by a constant stimuli, the system is out of the stationary state. Moreover, it contains nonlinear reaction rates and therefore numerical computations need to be used to calculate the sensitivity coefficients.

Local analysis. The local sensitivity coefficients for both deterministic and stochastic versions of the model are presented in the Fig. 1.3.A. Two parameters appear to be significantly more sensitive compared to the rest: θ_5, describing the total number of the TF molecules; and θ_8, describing the translocation rate of the TF from the nucleus to the cytoplasm.

In contrast, θ_6, the rate of the deactivation of the TF in the cytoplasm, hardly influences the output of the model. The analysis highlights the role of the total number of the TF molecules and translocation rate in shaping the response of the system and suggests the possibility to neglecting the parameter θ_6 in further analysis. This holds for both stochastic and deterministic regimes. In addition, a direct comparison of the coefficients obtained for the deterministic and stochastic regimes shows that the parameter θ_5 is not only the most sensitive but it has also the highest impact on stochastic effects (Fig. 1.3.A), as its sensitivity coefficient is much higher in the stochastic than in the deterministic scenario, compared to other parameters.

Global analysis. Global methods can verify to what extend findings of the local methods hold throughout the parameter space. Results of the screening method and variance based method are presented in the Fig. 1.3.B and the Fig. 1.3.C-F, respectively. The results are consistent with the local analysis and allow to conclude that previous findings hold throughout the considered parameter space. A unique component provided by the screening method is the high value of the standard deviation of elementary effects of the parameter θ_5 that implies that the impact of this parameter exhibits a substantial dependence on other parameter values.

In addition, variance-based methods allow for an alternative interpretation of the sensitivity coefficients. Variance of the output reflects lack of robustness to perturbations of parameter values. From this perspective the computed sensitivities for the SP model imply that the output is not robust to perturbations in some parameters, particularly θ_5, i.e. fluctuations of the total concentration of the TF.

The MCF method allows to quantify impact of parameters on whether a condition of interest is satisfied. Here we intended to identify parameters that have the strongest impact on whether trajectories of the SP model contain a time widow of at least 20 minutes length with over 50% of the total amount of the TF located in the nucleus. If his condition is satisfied then $cond(Y) = 1$ and 0 otherwise. Fig. 1.4.A presents estimates of the Kolmogorov distance between distributions $\mathcal{P}(\theta_i|\ cond(Y) = 1)$ and $\mathcal{P}(\theta_i|\ cond(Y) = 0)$,

Nienaltowski 2016/6/16 19:58 Page 20 #20

for each parameter θ_i. According to these coefficients, rates of TF translocation from the nucleus to the cytoplasm, θ_8, and deactivation of the kinase, θ_3 are mostly responsible for inducing model behaviour that satisfies the condition. Moreover, Fig. 1.4.B shows that condition is mostly satisfied for small values of parameters θ_3 and/or θ_8. Parameters θ_6 and θ_7, on the other hand, have been identified as insignificant, which is also confirmed by the scatter plot at Fig. 1.4.C.

1.3.3 Parameters collinearity

Conventional sensitivity analysis aims to quantify the impact of individual parameters. Although mutual relationships are implicitly taken into account while calculating the sensitivity coefficients, no direct insight is provided on how joint changes in parameter values impact the model output. One of the essential observations that highlights how mutual interactions between parameters shape model output is the possibility to compensate a change in value of one parameter with a change in values of other parameters [35, 36].

To illustrate this phenomenon consider the GE model with the output defined as the stationary abundance of mRNA and protein, $Y = (y_1^*, y_2^*)$. Assume we consider changes of the two parameters, θ_1, θ_2, inside a unit ball $(\partial \log\theta_1)^2 + (\partial \log\theta_2)^2 \leq 1$ in the log-parameter space. As a result of mutual compensation, these are mapped onto output ellipsoid $\widetilde{SV}_1 \partial \log\theta_1 + \widetilde{SV}_2 \partial \log\theta_2$ in the log-output space, where $\widetilde{SV}_i$ are sensitivity vectors (Eq. 1.40). Equation 1.10 shows that the principal axes of this ellipsoid can be calculated as eigenvectors and their radiuses are given as the square roots of the reciprocal eigenvalues of the matrix SM^2 (Fig. 1.5.A).

Similar insight can be gained by analysing which parameter combinations result in changes of the output within a unit ball $(\partial \log Y_1)^2 + (\partial \log Y_2)^2 \leq 1$ (Fig. 1.5.B). Then, in the log-parameter space there is a corresponding ellipsoid, whose axes are also defined by eigenvectors of matrix SM^2 with radiuses as inverses of the square roots of their eigenvalues. The minor axis of this ellipsoid defines a direction that leads to the biggest change in the model output. This direction is often referred to as 'stiff'. On the other hand, 'sloppy' direction have little effect on the model output. Figure 1.5.C shows a significant difference between 'stiff' and 'sloppy' directions in GE model.

In the context of the above considerations, the cosine between sensitivity vectors (Eq. 1.13) is a sensible measure of similarity between parameters. The value of 1 indicates that sensitivity vectors are parallel, which implies that corresponding parameters have (locally) an equivalent effect on the output. On the other hand, the value of 0 means that these vectors are orthogonal and the corresponding parameters have unique impacts on the output.

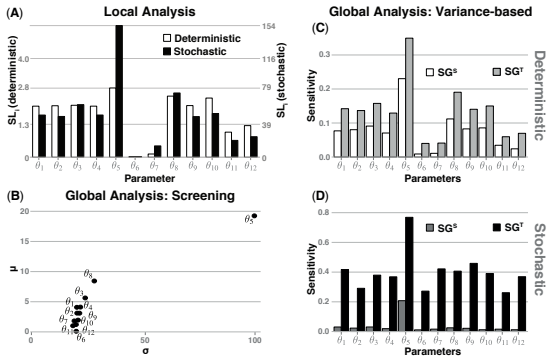


Figure 1.3

Sensitivity analysis of the deterministic (white colouring) and stochastic (black colouring) versions of the SP model using local and global methods. (**A**) Local sensitivity coefficients, SL_i, for the deterministic (white boxes, left axis) and stochastic (black boxes,right axis) regimes. Boxes are scaled so that they have the same relative magnitudes for each regime. (**B**) Mean elementary effects and their standard deviations (Eq. 1.31) for each model parameter obtained by the screening method using $r = 10^4$ and $h = 10^3$. (**C-D**) Comparison of the first-order, SG_i^S, and total sensitivity coefficients, SG_i^T within the deterministic (**C**) and stochastic (**D**) regimes.

Nienaltowski 2016/6/16 19:58 Page 22 #22

GE			SP					
Nominal parameters values								
	value	units		value	units		value	units
θ_1	100	$\frac{molec.}{min.}$	θ_1	100000	$molec.$	θ_7	1	$\frac{1}{min.}$
θ_2	20	$molec.$	θ_2	250	$\frac{1}{min.}$	θ_8	0.1	$\frac{1}{min.}$
θ_3	3	$\frac{1}{min.}$	θ_3	0.5	$\frac{1}{min.}$	θ_9	16000	$\frac{molec.}{min.}$
θ_4	0.6	$molec.$	θ_4	10000	$\frac{1}{min.}$	θ_{10}	2000	$molec.$
			θ_5	5000	$molec.$	θ_{11}	0.01	$\frac{1}{min.}$
			θ_6	0.01	$\frac{1}{min.}$	θ_{12}	2	1
Initial conditions								
	value	units		value	units		value	units
y_1	33	$molec.$	y_1	0	$molec.$	y_3	0	$molec.$
y_2	1111	$molec.$	y_2	0	$molec.$	y_4	0	$molec.$

molec.=molecules, min.=minutes

Table 1.1

Nominal parameters values and initial conditions used in the analysis of the GE and SP models.

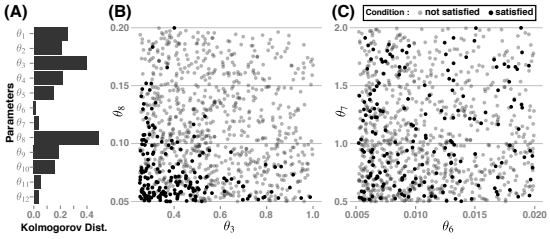


Figure 1.4

Results of the MCF analysis applied to the SP model. The parameter values were tested if the corresponding trajectories contained a time window of 20 minutes with at least 50% of the transcription factor molecules present in the nucleus. In the analysis 10000 parameters sets were sampled. (**A**) Kolmogorov distance between the estimates of the distribution $\mathcal{P}(\theta_i|\, cond(Y) = 1)$ and $\mathcal{P}(\theta_i|\, cond(Y) = 0)$. (**B,C**) Scatter plots corresponding to pairs of parameters with the highest (**B**) and the lowest (**C**) Kolmogorov distance with marking describing the value of the tested condition.

Nienaltowski 2016/6/16 19:58 Page 23 #23

The analysis can be extended to quantify similarity between sub-groups of model parameters $\theta_A = (\theta_{a_1}, \ldots, \theta_{a_{l_a}})$ and $\theta_B = (\theta_{b_1}, ..., \theta_{b_{l_b}})$. As for correlations between multidimensional random variables, a concept of canonical correlations (CCs) [37] can provide valuable insight. The canonical correlation between θ_A and θ_B is a vector $\rho = [\rho_1, \ldots, \rho_m]$, which contains maximal cosines between mutually orthogonal linear combinations of parameters within θ_A and θ_B, where $m = \min(l_a, l_b)$. CCs lead to one dimensional measure of similarity between θ_A and θ_B [4]

$$I(\theta_A, \theta_B) = -\frac{1}{m}\sum_{i}^{m} \log(1 - \rho_i^2). \tag{1.48}$$

Quantification of similarity between groups of parameters combined with a hierarchical clustering algorithm can be used to depict similarity between model parameters in form of a dendrogram [4].

A dendrogram with parameters similarity for the GE model is plotted in the Fig. 1.5.D. The linkage between individual parameters θ_1 and θ_3 as well as between θ_2 and θ_4 is plotted at zero height, which indicates that the parameters can be mutually compensated. Contrary, linkage between the pairs (θ_1, θ_3) and (θ_2, θ_4) is at a non-zero height, as the pairs cannot be mutually entirely compensated. A similar analysis can be performed for more complex models where mutual relationships between parameters are intricate. The dendrogram plotted for the SP model reveals which parameters and to what extend can be mutually compensated. The possibility of mutual compensation leads to dominance of 'sloppy' directions presented in the Fig. 1.5.E-F.

1.4 Summary

Parameters are inherent components of quantitative biology models. Model dependence on parameters is a source of valuable insights into studied processes as well as a cause of numerous difficulties.

For instance, dependence on parameters allows to examine impact of underlying processes on studied phenomena. Having a model of a signaling pathway it is in principle possible to predict an impact of up-regulation or down-regulation of receptor number on a pathway's function. However, dissecting an impact individual parameters make on model behaviour is usually nontrivial to quantify as it is exerted via a complex set of interactions with other parameters and systems's components. For instance, the effect of up-regulating the number of receptors may be different in case of high and low rate of receptor's repressor expression.

In addition, being able to compare model behaviour with experimental data provides a mean to estimate unknown parameter values. In a typical quantitative biology model, determination of values of all parameter based on available data is difficult or impossible

Nienaltowski 2016/6/16 19:58 Page 24 #24

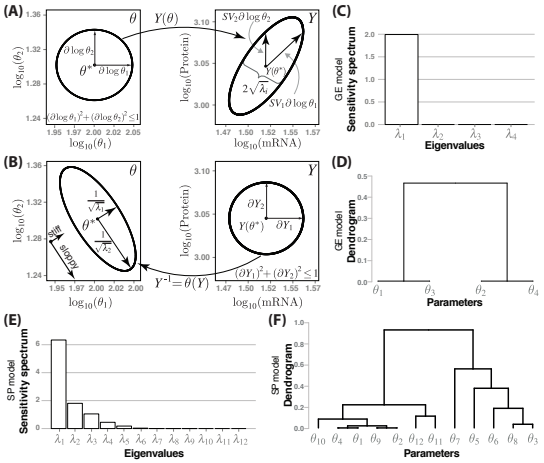


Figure 1.5

Collinearity and sloppiness in the GE and SP models **(A-B)** A mapping between parameter and output spaces for the GE model. **(A)** The image of the unit ball is the ellipsoid, whose principal axes have the length of the square roots of the eigenvalues λ_i of the matrix **S**. **(B)** Inverse image of the output onto parameter space. Eigenvectors of the matrix **S** are the principal axes of the ellipsoid, whereas $\lambda_i^{-0.5}$ are lengths of these axes. The direction corresponding to the first eigenvalue is called 'stiff' as a small change in parameter combination is needed to induce a substantial change in the output, the second direction is called 'sloppy' as a significantly larger change in the parameters is needed to induce the same change in the output. **(D)** Parameters collinearity in the GE model presented as the dendrogram.**(E)** Spectral decomposition of the *S* matrix of the SP model. **(F)** Parameters collinearity in the SP model presented as the dendrogram. In all calculations nominal parameters were used, presented in the Table 1.1.

Nienaltowski 2016/6/16 19:58 Page 25 #25

[8, 35] due to a large number of parameters compared to data size [38]. Lack of sufficient information to constrain parameter values makes estimates to be associated with high uncertainties.

Sensitivity analysis allows to extract insights about modelled processes as well as to avoid some difficulties resulting from parameters dependance. Local and global sensitivities enable comparison of parameters impact across parameter space. Knowing sensitivities is useful for understanding effects of uncertainties [16] in parameter estimates. In particular sensitivities enable to verify to what extent conclusions drawn from a model depend on parameter values [7]. As sensitive parameters can be estimated from experimental data with higher accuracy than insensitive parameters knowing parmater sensitivities is helpful when designing experiments aimed at parameter inference. Specifically, it is desired to select experiments that yield model parameters sensitive [39]. Moreover, sensitivity analysis helps to identify and eliminate parameters that have a negligible impact on model behaviour [2].

Despite their usefulness existing methods suffer from certain drawbacks. As being largely inherited from engineering and theoretical statistics, they are not entirely suited to the specificity of the quantitative biology dynamical models. Utilising the potential of sensitivity analysis in providing deeper understanding of biological phenomena relies largely on developing new techniques that will account for specificity of biological models and also will be able to provide tangible answers to specific biological questions. In particular, tools to disentangle complexity of models resulting from non-linearities and large number of parameters remain to be developed. Development of new tools that provide more explicit answers to specific questions would make them more attractive for quantitative biologists and therefore lead to better understanding of modelled biological processes.

Acknowledgment

Authors were supported by the Polish Ministry of Science and Higher Education under IUVENTUS PLUS grant IP2012016572 as well as by the European Commission Research Executive Agency under grant CIG PCIG12-GA-2012-334298.

Nienaltowski 2016/6/16 19:58 Page 27 #27

Exercises

1. **Premilimaries**

   a. Prove the Eq. 1.14. Let's assume notation as in the chapter: y is model's output, governed by ODE equation

   $$\frac{dy}{dt} = F(y,\theta), \quad y(t_0) = y_0.$$

   Prove that if we introduce $z^j(t) = \frac{\partial y(t)}{\partial \theta_j}$, $i = 1,\ldots,l$, then it fulfills the Eq. 1.14

   $$\frac{dz^j(t)}{dt} = \nabla_y F(y(t),\theta) z^j(t) + \frac{\partial F(y(t),\theta)}{\partial \theta_j}.$$

   b. Let $Y = G(\theta)$, then variance of the Y can be decomposed according to

   $$\mathrm{Var}\{Y\} = \sum_{i=1}^{l} V_i + \sum_{i<j}^{l} V_{ij} + \ldots + V_{1\ldots l}, \tag{1.49}$$

   where

   $$V_i = \mathrm{Var}_{\theta_i}\left\{\mathbb{E}_{\theta_{\sim i}}\{Y|\theta_i\}\right\}, \quad V_{ij} = \mathrm{Var}_{\theta_{ij}}\left\{\mathbb{E}_{\theta_{\sim ij}}\{Y|\theta_i,\theta_j\}\right\} - V_i - V_j, \quad \ldots \quad V_{12\ldots l} = \mathrm{Var}\{Y\} - \sum_{u\subset\{1,\ldots,l\}} V_u \ .$$

   The Equation 1.49 is usually called 'Sobol variance decomposition'. Prove the Eq. 1.49 for $l = 3$.

   Notice that V_J (where $J = \{j_1,\ldots,j_k\} \subseteq \{1,2,\ldots,l\}$) describes the contribution of interactions of parameters θ_J into overall variance. Indexes above normalised by total variance of the output lead to definition of higher order-indexes $SG_J^S = \frac{V_J}{\mathrm{Var}\{Y\}}$. Using them, total effects coefficients can be explained as a total contribution to the output variation due to individual parameter

   $$\theta_i{:}SG_i^T = SG_i^S + \sum_{i<j} SG_{ij}^S + \sum_{i\neq j, i\neq k, j<k} SG_{ijk}^S + \ldots + SG_{12\ldots l}^S.$$

2. **Fisher Information Matrix**

   a. Prove the Eq. 1.20 that describes the FIM for the normal distribution in the one dimensional scenario, i.e. if $Y \sim (N)(\mu(\theta),\sigma^2(\theta)), \quad Y \in \mathbb{R}$.

   b. Consider a deterministic system subject to a measurement error, e.g.

   $$Y(t,\theta) = y(t,\theta) + \varepsilon,$$

   where $y(t,\theta)$ is a solution of some ODE system and $\varepsilon \sim \mathcal{N}(0,\Sigma)$, Σ – measurement error, independent of θ. Express *FIM* of this system using sensitivity vectors and matrix (Eq. 1.7, Eq. 1.11).

   c. Calculate coefficients SL_j (Eq. 1.18) for the gene expression model in the stochastic scenario assuming that the system is in the stationary state,$Y = y_1$ (RNA) and knowing that stationary distribution of y_1 is a Poisson distribution with an appropriate coefficient.

3. **Simple models** Consider two simple models with two parametrs $\theta = (\theta_1,\theta_2)$:

   (Model I:) $G^I(\theta) = \theta_1 + \theta_2$ *(Model II:)* $G^{II}(\theta) = \theta_1 \cdot \theta_2$

   a. Calculate local the sensitivity coefficients with and without logarithmic parametrisation for both models. Use nominal vector $\theta^* = (10,10)$. Which approach is more appropriate in each case?

   b. Calculate the global sensitivity coefficients – first order and total effects for both models. Assume that $\theta_i \in (1,100)$ and use both direct and logarithmic approach to space sampling. How the sampling procedure influence results for each of models?

   c. Repeat b), assuming that $\theta_1 \in \left(\frac{1}{10},10\right)$ and $\theta_2 \in (10,1000)$.

Nienaltowski 2016/6/16 19:58 Page 28 #28

d. Repeat b), assuming that $\theta_1 \in (1, 10)$ and $\theta_2 \in (1, 1000)$.

e. Compare total variance of the model output Var $\{Y\}$ computed directly and using Eq. 1.33 and Eq. 1.36 in cases b), c) and d).

4. **Gene expression model**

a. In the conventional definition of the sensitivity coefficient, Euclidean distance is used to aggregate all entries of output vector. However, any other norm can be used in its place. Let's consider GE model and assume that the output is $Y = (y_1(0.5), y_1(1), y_1(2))$ and $y_1(0) = 0$. Repeat calculation of sensitivity coefficient for GE model (see Section. 1.3.2), but instead of a standard norm, use a weighted norm defined as below

$$\|x\|_D = \|x^T D x\|_2, \quad \text{where} \quad D = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad SL_i = \sqrt{W_i^T D W_i}$$

b. Expand parameters set for GE model by adding initial condition, i.e. $\theta = (\theta_1, \theta_2, \theta_3, \theta_4, y_1(0))$. Repeat b) for added parameter.

c. Calculate Fisher Information Matrix analytically for gene expression model in stochastic setting.

d. In the local analysis of the stochastic GE model presented in the chapter all parameters turned out to be similarly sensitive. Find such nominal parameter values that generate significant differences between parameters in stochastic scenario.

e. Derivatives of momements is one of the strategies to quantify sensitivity in the stochastic scenario. Calculate $\frac{\partial \mathbb{E}[Y(t)]}{\partial \theta_i}$ as well as $\frac{\partial \mathrm{Var}[Y(t)]}{\partial \theta_i}$ and compare with the analysis presented in the chapter.

Nienaltowski 2016/6/16 19:58 Page 29 #29

Index

Bibliography

[1] K. Erguler, M. P. Stumpf. 'Practical limits for reverse engineering of dynamical systems: a statistical analysis of sensitivity and parameter inferability in systems biology models'. *Molecular BioSystems*, **7**(5):1593–1602, 2011.

[2] A. Ihekwaba, D. Broomhead, R. Grimley, N. Benson, D. Kell. 'Sensitivity analysis of parameters controlling oscillatory signalling in the NF-κB pathway: the roles of IKK and IκBα'. *Systems Biology*, **1**(1):93–103, 2004.

[3] K. S. Brown, J. P. Sethna. 'Statistical mechanical approaches to models with many poorly known parameters'. *Physical Review E*, **68**(2):021904, 2003.

[4] K. Nienałtowski, M. Włodarczyk, T. Lipniacki, M. Komorowski. 'Clustering reveals limits of parameter identifiability in multi-parameter models of biochemical dynamics'. *BMC Systems Biology*, **9**(1):65, 2015.

[5] Z. Zi. 'Sensitivity analysis approaches applied to systems biology models'. *Systems Biology, IET*, **5**(6):336–346, 2011.

[6] A. Saltelli, M. Ratto, S. Tarantola, F. Campolongo. 'Sensitivity analysis for chemical models'. *Chemical Reviews*, **105**(7):2811–2828, 2005.

[7] B. C. Daniels, Y.-J. Chen, J. P. Sethna, R. N. Gutenkunst, C. R. Myers. 'Sloppiness, robustness, and evolvability in systems biology'. *Current Opinion in Biotechnology*, **19**(4):389–395, 2008.

[8] J. Gunawardena. 'Models in systems biology: the parameter problem and the meanings of robustness'. *Elements Of Computational Systems Biology*, **1**, 2010.

[9] D. Degenring, C. Froemel, G. Dikta, R. Takors. 'Sensitivity analysis for the reduction of complex metabolism models'. *Journal of Process Control*, **14**(7):729–745, 2004.

[10] D. Rand, B. Shulgin, D. Salazar, A. Millar. 'Design principles underlying circadian clocks'. *Journal of the Royal Society Interface*, **1**(1):119–130, 2004.

[11] H. Yue, M. Brown, J. Knowles, H. Wang, D. S. Broomhead, D. B. Kell. 'Insights into the behaviour of systems biology models from dynamic sensitivity and identifiability analysis: a case study of an NF-κB signalling pathway'. *Molecular BioSystems*, **2**(12):640–649, 2006.

[12] H. Rabitz, M. Kramer, D. Dacol. 'Sensitivity analysis in chemical kinetics'. *Annual Review of Physical Chemistry*, **34**(1):419–461, 1983.

[13] M. Komorowski, J. Žurauskiene, M. P. Stumpf. 'StochSens matlab package for sensitivity analysis of stochastic chemical systems'. *Bioinformatics*, **28**(5):731–733, 2012.

[14] A. Gupta, M. Khammash. 'Unbiased estimation of parameter sensitivities for stochastic chemical reaction networks'. *SIAM Journal on Scientific Computing*, **35**(6):A2598–A2620, 2013.

[15] M. Komorowski, M. J. Costa, D. A. Rand, M. P. Stumpf. 'Sensitivity, robustness, and identifiability in stochastic chemical kinetics models'. *Proceedings of the National Academy of Sciences*, **108**(21):8645–8650, 2011.

[16] D. J. Wilkinson. 'Stochastic modelling for quantitative description of heterogeneous biological systems'. *Nature Reviews Genetics*, **10**(2):122–133, 2009.

[17] L. Le Cam, G. L. Yang. *Asymptotics in Statistics: Some Basic Concepts*. Springer Science & Business Media, 2012.

[18] R. Gunawan, Y. Cao, L. Petzold, F. J. Doyle. 'Sensitivity analysis of discrete stochastic systems'. *Biophysical Journal*, **88**(4):2530–2540, 2005.

[19] L. L. Scharf. *Statistical Signal Processing*. Addison-Wesley Reading, MA, 1991.

[20] M. Komorowski, B. Finkenstädt, C. V. Harper, D. A. Rand. 'Bayesian inference of biochemical kinetic parameters using the linear noise approximation'. *BMC Bioinformatics*, **10**(1):343, 2009.

[21] N. G. Van Kampen. *Stochastic Processes in Physics and Chemistry*, volume 1. Elsevier, 2006.

[22] M. Rodriguez-Fernandez, J. R. Banga, F. J. Doyle. 'Novel global sensitivity analysis methodology accounting for the crucial role of the distribution of input parameters: application to systems biology models'. *International Journal of Robust and Nonlinear Control*, **22**(10):1082–1102, 2012.

[23] M. D. Morris. 'Factorial sampling plans for preliminary computational experiments'. *Technometrics*, **33**(2):161–174, 1991.

[24] F. Campolongo, J. Cariboni, A. Saltelli. 'An effective screening design for sensitivity analysis of large models'. *Environmental Modelling & Software*, **22**(10):1509–1518, 2007.

[25] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, S. Tarantola. *Global Sensitivity Analysis: the Primer*. John Wiley & Sons, 2008.

[26] A. Degasperi, S. Gilmore. 'Sensitivity Analysis of Stochastic Models of Bistable Biochemical Reactions'. In *Formal Methods for Computational Systems Biology*, pages 1–20. Springer, 2008.

[27] R. Kass, P. Vos. *Geometrical Foundations of Asymptotic Inference*. John Wiley & Sons, 1997.

[28] F. Pianosi, T. Wagener. 'A simple and efficient method for global sensitivity analysis based on cumulative distribution functions'. *Environmental Modelling & Software*, **67**:1–11, 2015.

[29] G. Archer, A. Saltelli, I. Sobol. 'Sensitivity measures, ANOVA-like techniques and the use of bootstrap'. *Journal of Statistical Computation and Simulation*, **58**(2):99–120, 1997.

[30] I. Sobol'. 'On sensitivity estimation for nonlinear mathematical models.' *Mathematical Modelling*, **2**(1):112–118, 1990.

[31] T. Homma, A. Saltelli. 'Importance measures in global sensitivity analysis of nonlinear models'. *Reliability Engineering & System Safety*, **52**(1):1 – 17, 1996.

[32] A. Saltelli. 'Making best use of model evaluations to compute sensitivity indices'. *Computer Physics Communications*, **145**(2):280 – 297, 2002.

[33] S. Krishna, M. H. Jensen, K. Sneppen. 'Minimal model of spiky oscillations in NF-κB signaling'. *Proceedings of the National Academy of Sciences*, **103**(29):10840–10845, 2006.

[34] T. Lipniacki, P. Paszek, A. R. Brasier, B. Luxon, M. Kimmel. 'Mathematical model of NF-κB regulatory module'. *Journal of Theoretical Biology*, **228**(2):195–215, 2004.

[35] R. N. Gutenkunst, J. J. Waterfall, F. P. Casey, K. S. Brown, C. R. Myers, J. P. Sethna. 'Universally sloppy parameter sensitivities in systems biology models'. *PLoS Computational Biology*, **3**(10):e189, 2007.

[36] B. K. Mannakee, A. P. Ragsdale, M. K. Transtrum, R. N. Gutenkunst. 'Sloppiness and the geometry of parameter space'. In *Uncertainty in Biology*, pages 271–299. Springer, 2016.

[37] W. Härdle, L. Simar. *Applied multivariate statistical analysis*. Springer, 2007.

[38] P. Meyer, T. Cokelaer, D. Chandran, K. H. Kim, P.-R. Loh, G. Tucker, M. Lipson, B. Berger, C. Kreutz, A. Raue, *et al.* 'Network topology and parameter estimation: from experimental design methods to gene regulatory network kinetics using a community based approach'. *BMC Systems Biology*, **8**(1):13, 2014.

[39] K.-H. Cho, S.-Y. Shin, W. Kolch, O. Wolkenhauer. 'Experimental design in systems biology, based on parameter sensitivity analysis using a monte carlo method: A case study for the TNF-α-mediated NF-κB signal transduction pathway'. *Simulation*, **79**(12):726–739, 2003.

Nienałtowski *et al. BMC Systems Biology* (2015) 9:65
DOI 10.1186/s12918-015-0205-8

BMC Systems Biology

RESEARCH ARTICLE Open Access

Clustering reveals limits of parameter identifiability in multi-parameter models of biochemical dynamics

Karol Nienałtowski[1], Michał Włodarczyk[2], Tomasz Lipniacki[1] and Michał Komorowski[1*]

Abstract

Background: Compared to engineering or physics problems, dynamical models in quantitative biology typically depend on a relatively large number of parameters. Progress in developing mathematics to manipulate such multi-parameter models and so enable their efficient interplay with experiments has been slow. Existing solutions are significantly limited by model size.

Results: In order to simplify analysis of multi-parameter models a method for clustering of model parameters is proposed. It is based on a derived statistically meaningful measure of similarity between groups of parameters. The measure quantifies to what extend changes in values of some parameters can be compensated by changes in values of other parameters. The proposed methodology provides a natural mathematical language to precisely communicate and visualise effects resulting from compensatory changes in values of parameters. As a results, a relevant insight into identifiability analysis and experimental planning can be obtained. Analysis of NF-κB and MAPK pathway models shows that highly compensative parameters constitute clusters consistent with the network topology. The method applied to examine an exceptionally rich set of published experiments on the NF-κB dynamics reveals that the experiments jointly ensure identifiability of only 60 % of model parameters. The method indicates which further experiments should be performed in order to increase the number of identifiable parameters.

Conclusions: We currently lack methods that simplify broadly understood analysis of multi-parameter models. The introduced tools depict mutually compensative effects between parameters to provide insight regarding role of individual parameters, identifiability and experimental design. The method can also find applications in related methodological areas of model simplification and parameters estimation.

Background

Methods to understand the relationship between parameters (input) and model properties (output) are of particular interest in the context of biochemical dynamics and related phenomena. Sensitivity analysis and statistical inference have proven their importance for utilising modelling in physics and engineering. Models of biochemical dynamics, however, are different from conventional models in a number of ways. Primarily they involve a substantially larger number of parameters compared to available data. The high number of parameters and sparse data in ordinary differential equation (ODE) models make a conventional sensitivity analysis and statistical inference methods often prohibitively difficult to apply. This challenge has given rise to a number of approaches aimed at improving our ability to develop, verify and manipulate multi-parameter mechanistic models of such systems. These methods can be vaguely grouped into those aiming at: 1) improved description of parameter sensitivities; 2) detection of parameters that cannot be inferred from experimental data (identifiability analysis) and 3) guided experimental design to improve parameter identifiability and inference accuracy. Within the first group a number of studies have reported an intrinsic feature of dynamic multi-parameter models of biochemical dynamics to be sensitive only to a small number of linear combinations

*Correspondence: m.komorowski@sysbiosig.org
[1]Institute of Fundamental Technological Research, Polish Academy of Sciences, Warsaw, Poland
Full list of author information is available at the end of the article

of parameters [1–5]. The conventional identifiability analysis verifies whether local distinct changes in parameter values imply distinct changes in model behaviour. A priori methods focus on determining whether this condition is satisfied prior to data collection. This can be done either based on model structure, often by attempting to find functional relationships between parameters [6], or by analysing model responses to local perturbations in parameter values. The latter is achieved by examining the Fisher information matrix (FIM). Two natural sources of non-identifiability have been recognised: insensitivity of individual parameters and compensative effects of parameter changes, also known as collinearity. Both problems have gained substantial attention. As a remedy, most approaches aim to select an optimal subset of parameters that is both sufficiently sensitive and has lowest collinearity. The identifiable subset can be then estimated jointly with the remaining parameters assumed fixed. The determinant of the FIM and its least eigenvalue are used to measure optimality [7–10] of the selected set. Pairwise clustering has also been proposed to reduce the number of parameters [9]. A posteriori methods focus on finding identifiable parameters when experimental data are available. The likelihood surface around its maximum is then examined by means of the Hessian matrix [11, 12]. A statistical concept of profile likelihoods is particularly helpful [13] in this case. Identifiability analysis is closely related to experimental design. It has been used to show how the information content in experimental measurements can be maximised [13–16]. Despite useful methodological developments performing routine modelling tasks with a multi-parameter model still constitutes a substantial challenge. Here, we introduce a natural, universal and simple measure to quantify similarity between groups of model parameters. The measure links canonical correlation analysis (CCA) with Shannon's mutual information (MI) and is called MI-CCA throughout the paper. Similarity between model parameters has been previously addressed (e.g. [9, 10, 17]). However a precise, statistically interpretable similarity measure has not been proposed. MI-CCA, when employed in a hierarchical clustering, provides statistically meaningful and precise information about mutual compensability of parameters. It can also be used as an assistance tool to validate parameters identifiability in experimental planning. Apart from its simplicity and rigorous statistical interpretation, the main advantage of our tool is that it can be applied to large models, for which other, well established, approaches are computationally infeasible. We demonstrate the power of our framework by analysis of the NF-κB and MAPK signalling models. We find that highly similar parameters constitute groups consistent with the network topology. For the NF-κB model we analyse the majority of published experimental protocols [18–26] and examine parameters identifiability. We show how the method can be used to guide further experiments.

Methods

A typical model of biochemical dynamics describes how abundances of a set of k molecular entities, $y = (y_1, \ldots, y_q, \ldots, y_k)$, change with time t. Deterministically it is usually written as an ordinary differential equation (ODE)

$$\frac{dy}{dt} = F(y, \theta), \tag{1}$$

where $F()$ is a law that determines the temporal evolution of y and implicitly contains a control signal. The vector $\theta = (\theta_1, \ldots, \theta_l)$ is a vector of model parameters. To numerically simulate the model, parameter values and initial condition, $(y_1(0), \ldots, y_k(0))$, must be set. The method proposed in this paper is a priori in nature, therefore the parameter values and initial conditions are not inferred from data and must be assumed in advance based on the modellers knowledge.

Often only certain components of y, for instance first q, $y^{(q)} = (y_1, \ldots, y_q)$, at specified times, $(t_1, \ldots, t_n)$, are of interest. These components, which may correspond to experimentally measured variables, are denoted here as $Y = \left(y^{(q)}(t_1), \ldots, y^{(q)}(t_n)\right)$.

Conventional sensitivity analysis fails to capture collective interactions between model parameters

Sensitivity analysis provides a prediction how Y will change, ∂Y, in response to small changes in a single parameter, $\partial\theta_i$, or all parameters, $\partial\theta = (\partial\theta_1, \ldots, \partial\theta_l)$. If changes in parameters are small, the problem is solved by finding the derivative of a solution of the equation (1), $y(t)$, with respect to the parameter θ_i, $z_i(t) = \frac{\partial y(t)}{\partial \theta_i}$. This derivative can be easily calculated by solving another ODE (see Additional file 1). Evaluation of $z_i(t)$ at the times and components of interests defines the sensitivity vector $S_i = \left(z_i^{(q)}(t_1), \ldots, z_i^{(q)}(t_n)\right)$ of the parameter θ_i. The sensitivity vector describes the shift in Y in response to perturbation in the parameter θ_i, $\partial Y = S_i \partial\theta_i$. A collection of the sensitivity vectors for all $i = 1, \ldots, l$ constitutes the sensitivity matrix $S = (S_1, \ldots, S_l)$, which summarises the change in Y in response to perturbation of all of the model parameters $\partial Y = S\partial\theta$. The sensitivity matrix, S, is directly linked with the concept of Fisher information. Given that Y is observed with the Gaussian unit variance error the FIM can be written as (see Additional file 1)

$$FI(\theta) = S^T S. \tag{2}$$

Therefore the FIM contains information regarding the size of a perturbation, $||\partial Y|| = \sqrt{\partial\theta^T FI(\theta)\partial\theta}$. The pairwise

similarity between parameters, quantified as the cosine between the S_i and S_j vectors, is also given by elements of the FIM, $\cos(S_i, S_j) = S_i^T S_j / ||S_i|| ||S_j||$. It is not clear, however, how the FIM can serve as a tool to analyse mutual relations between groups of parameters. Below we provide a rigorous and practical solution to this problem.

Measuring similarity between parameters groups

Canonical correlations. The canonical correlation analysis (CCA) is a simple extension of the Pearson correlation. With CCs it is possible to measure correlations between multidimensional covariates. We modify the well established definition to suit the considered context. Assume, we measure similarity between two subsets of parameters $\theta_A = \{\theta_{i_1}, \ldots, \theta_{i_a}\}$ and $\theta_B = \{\theta_{j_1}, \ldots, \theta_{j_b}\}$ that correspond to the two subsets of sensitivity vectors, $\Omega_A = \{S_{i_1}, \ldots, S_{i_a}\}$ and $\Omega_B = \{S_{j_1}, \ldots, S_{j_b}\}$. The latter can be interpreted as hyper-planes. CCs form a set of correlation coefficients defined recursively. The first CC, ρ_1, is a maximal cosine between a linear combination, u_1, in Ω_A and a linear combination, v_1, in Ω_B, $\rho_1 = \cos(u_1, v_1)$. Each next CC is found in the same way under the constraint that the next linear combination must be orthogonal to these found in the previous steps (see Additional file 1). Repeating the procedure $m = \min(i_a, j_b)$ times provides a set of CCs $1 \geq \rho_1 \geq \ldots \geq \rho_m \geq 0$ (see Fig. 1a–b). The value of 1

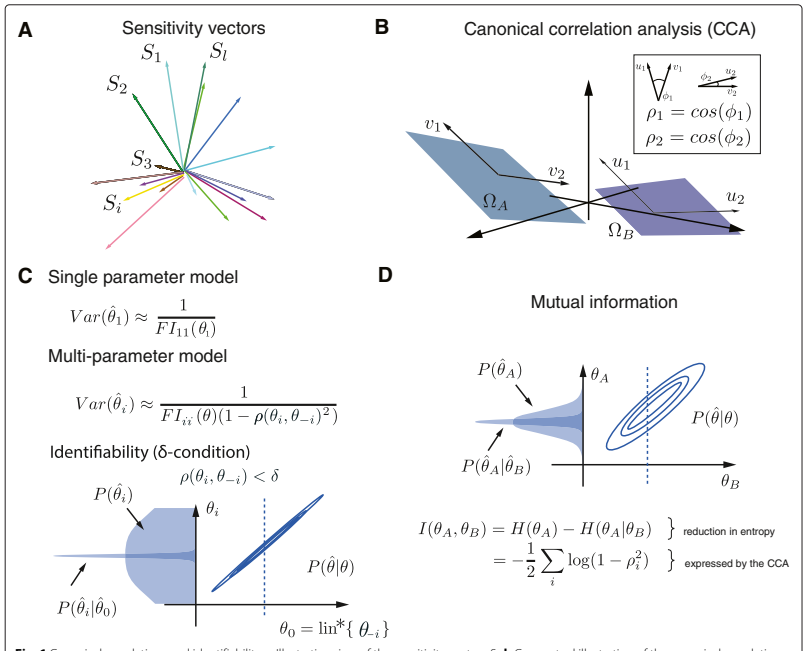


Fig. 1 Canonical correlations and identifiability. **a** Illustrative view of the sensitivity vectors S_i. **b** Conceptual illustration of the canonical correlations. Two subsets of sensitivity vectors represented as linear subspaces (planes Ω_A and Ω_B). Canonical vectors on the planes are found to yield maximum cosine. In a two-dimensional subspace case, the second canonical vectors u_2, v_2 are required to be perpendicular to the first ones. **c** The introduced δ-condition requires that each parameter θ_i is correlated less than δ with the remaining parameters $\theta_{-i} = (\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_l)$. It can be interpreted in terms of how variance of the estimates changes when a single parameter and all model parameters are estimated. Parameter θ_0 denotes the linear combination of θ_{-i} maximally correlated with θ_i, i.e. $\theta_0 = \text{lin}^*\{\theta_{-i}\}$. **d** Mutual information as a measure of similarity between two parameter sets θ_A, θ_B, which span linear subspaces Ω_A, Ω_B interpreted in terms of the asymptotic posterior $P(\hat{\theta}|\theta)$

indicates that there exists a linear combination of parameters in θ_A and θ_B having an identical impact, whereas 0 indicates existence of an orthogonal parameter combination. The CCs therefore provide an m-dimensional similarity measure between θ_A and θ_B.

Mutual information. The above geometric view has a natural probabilistic interpretation that provides a natural, one-dimensional similarity measure. Assume, we estimate the parameter vector θ using the maximal likelihood estimate $\hat{\theta}$ (equivalently Bayesian posterior estimate) from data $X = Y + \xi$, where ξ is a measurement error. Asymptotically (for large number of independent copies of X, denoted here by N) the distribution of the estimate $\hat{\theta}$ given a true value θ is asymptotically multivariate normal

$$P(\hat{\theta}|\theta) \propto \exp(-\frac{1}{2N}(\hat{\theta}-\theta)FI(\theta)(\hat{\theta}-\theta)^T). \quad (3)$$

Consider the entropy, $H(\hat{\theta}_A)$, of the estimate $\hat{\theta}_A$, and the average conditional entropy of $\hat{\theta}_A$ given $\hat{\theta}_B$, $H(\hat{\theta}_A|\hat{\theta}_B)$. The reduction in entropy of $\hat{\theta}_A$ resulting from knowledge of $\hat{\theta}_B$ is given by Shannon's mutual information between $\hat{\theta}_A$ and $\hat{\theta}_B$, denoted here by $I(\theta_A, \theta_B)$. We propose to use $I(\theta_A, \theta_B)$ as the natural measure of similarity. The more similar θ_A and θ_B are, the more knowing one will help in determining the value of the other. In Additional file 1 we show that the mutual information between estimates $\hat{\theta}_A$ and $\hat{\theta}_B$ and CCs are closely related

$$I(\theta_A, \theta_B) = H(\hat{\theta}_A) - H(\hat{\theta}_A|\hat{\theta}_B) = -\frac{1}{m}\sum_i^m \log\left(1-\rho_i^2\right), \quad (4)$$

where $H(\hat{\theta}_A|\hat{\theta}_B)$ is the condition entropy of $\hat{\theta}_A$ given $\hat{\theta}_B$. The above measure, which throughout the paper is called MI-CCA, provides a novel and efficient way to quantify overall similarity between parameter groups via mutual information and CCs.

We use the constructed measures to propose a natural definition of parameters identifiability in the multi-parameter scenario.

(δ, ζ)-identifiability

Conventionally, parameters of a statistical model $P(Y|\theta)$ are said to be identifiable if there exists a neighbourhood of θ such that for all parameter values in that neighbourhood $P(Y|\theta)$ represents a different distribution. Equivalently the FIM must have the full rank. This definition refers simultaneously to the entire vector of model parameters θ. The definition of [13] introduces a notion of practical non-identifiability by examining the flatness of the likelihood surface. We propose a novel definition of identifiability of individual parameters in multi-parameters models. It is widely recognised that lack of identifiability can arise from two sources: lack of sensitivity, or compensation of a parameter by remaining model parameters [7–10, 12, 27–30]. A definition that quantifies this intuition has been missing. Therefore, we propose a natural criterion of whether the parameter θ_i can be identified along with the remaining model parameters, θ_{-i}. The parameter θ_i is said to be (δ, ζ)-identifiable if $\rho(\theta_i, \theta_{-i}) < \delta$ and $||S_i|| > \zeta$. Correlation ρ is used here in the canonical sense. If θ_i was estimated as a single parameter of the model ζ-condition requires its asymptotic variance to be smaller than $1/\zeta$. The δ-condition requires the parameter not to be correlated with any linear combination of the remaining parameters by more than δ. In variance terms, it translates into demanding that the variance does not increase by more than $1/(1-\delta^2)$ when the single parameter and multi-parameter scenarios are compared (Fig. 1c). The above definition is conceptually similar to the profile likelihood approach. However it uses asymptotic likelihood instead of actual likelihood and therefore does not require any numerical optimisation. Based on the FIM, solutions are given analytically by CCs. As a result identifiability can be determined for models of virtually any size. In practical applications values of δ and ζ must be selected. The above interpretation of δ and ζ values provides a theoretical ground to guide how these thresholds can be set. For instance, in the logarithmic parametrisation setting $\zeta = 1$ requires a parameter to be learned with at most an order of magnitude error. Parameter δ controls how the estimate's variance increases when the parameter is estimated as a single parameter and jointly with remaining model parameters. Setting stricter values (lower δ and higher ζ) will result in lower variance of parameter estimates. Efficiency of the method enables the analysis to be performed for a range δ and ζ values that correspond to different levels of stringency. In the applications considered in this paper we used $\zeta = 1$ and $\delta = 0.95$. The latter corresponds to approximately 10-fold increase of variance (Fig. 1c). In Additional file 1 we use one of the analysed experiments to show that these thresholds provide results consistent with the profile likelihood approach. In general, profile likelihoods can also be used to validate method's predictions as experimental data become available (see Sections 4.3 and 6.6 of the Additional file 1).

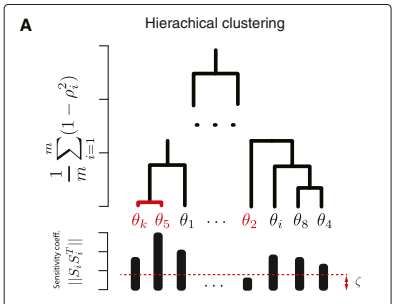
Clustering reveals similarity structure and identifiability

Using the constructed similarity measure we can meaningfully group model parameters. We provide a modification of the conventional hierarchical clustering algorithm. At each level of the hierarchy, clusters are created by merging clusters at the next lower level. At the lowest level, each cluster contains a single parameter. The pair chosen for merging consists of the two groups with the highest mutual information, $I(\theta_A, \theta_B)$. When a new cluster is formed we verify if each of the parameters

within the newly created cluster satisfies the δ-condition. The parameters of the clusters most correlated with the remaining parameters of the cluster are removed until all satisfy the δ-condition. We use average canonical correlation between the clusters, $\frac{1}{m}\sum_{i=1}^{m}(1-\rho^2)$, which is normalised opposed to $I(\theta_A,\theta_B)$, to determine the height of linkages. A set of identifiable parameters is not guaranteed to be maximal. Finding the maximal set would require testing each of the subsets of the parameter set, which is computationally infeasible. As the output of the algorithm, we obtain the visualisation of similarity structure and a set of identifiable parameters (see Fig. 2). The pseudocode describing the clustering algorithm in details is presented in Section 3 of the Additional file 1 and an R-implementation (Additional file 2) is available as an online supplement.

Example: a simple gene expression model

To clarify the principles behind the method, we use a simplistic gene expression model. We assume that the process begins with the production of mRNA molecules at rate k_r. Each mRNA molecule r may be independently translated into protein molecules at rate k_p. Both mRNA and protein molecules are degraded at rates γ_r and γ_p, respectively. Therefore, we have the vector of model parameters $\theta = (k_r, k_p, \gamma_r, \gamma_p)$ and ODEs presented in Figure 1A in Additional file 1. Consider the steady state $Y = \left(\frac{k_r}{\gamma_r}, \frac{k_r k_p}{\gamma_r \gamma_p}\right)$. We address the following questions: 1) Which model parameters are most similar?; 2) Which parameters are identifiable?; 3) What consequence does the similarity structure have for the model robustness?; 4) How can the steady state experiment be modified to reduce parameter correlations? The similarity of the parameters is determined entirely by the response of the model to changes in parameter values. The steady state formula implies that perturbations in k_r and γ_r have the same impact i.e. they increase or decrease the RNA and protein level. The same holds for perturbations in k_p and γ_p. On the contrary, a perturbation in (k_r, γ_r) does not have the same impact as one in (k_p, γ_p). The first pair affects the level of both RNA and protein; the latter only the level of protein. This intuition is formalised and visualised by the method. The linkage between parameters k_r, γ_r and k_p, γ_p is plotted at zero height, and the non-identifiable parameters are marked red (Figure 1B in Additional file 1). Linkage between the pairs is at a non-zero height, as they are not entirely correlated. As for model robustness, the dendrogram depicts that mutually compensative perturbations occur within pairs (k_r, γ_r) and (k_p, γ_p). The analysis highlights the sources of non-identifiability and therefore helps to find experiments that render more parameters identifiable. For instance, in this example, pushing the initial condition $r(t_0), p(t_0)$ above the steady state levels changes the model dynamics (Figure 1C in Additional file 1). The resulting exponential decay is not invariant with respect to parameter changes. As a result all parameters can be identified (Figure 1C in Additional file 1).



B Verification of the δ - condition

$\theta_C \leftarrow \{\theta_A, \theta_B\}$
while $\max_{\theta^* \in \theta_C} \rho(\theta^*, \theta_C \setminus \theta^*) > \delta$ **do**
  $\theta^* \leftarrow$ parameter maximizing $\rho(\theta^*, \theta_C \setminus \theta^*)$;
  $\theta_C \leftarrow \theta_C \setminus \{\theta^*\}$;
  mark θ^* in red;
end while

Fig. 2 a Agglomerative hierarchical clustering of model parameters. **b** Verification of the δ-condition. Recursively, at each level, a pair of most similar clusters is merged into a single cluster and δ-condition is verified. Linkages between clusters, at each stage of clustering, are plotted at high $\frac{1}{m}\sum_{i=1}^{m}(1-\rho^2)$, where m is the size of a new cluster, compared to a previous linkage. Identifiability results from violation of either of the δ-condition or ζ-condition therefore even parameters that have sensitivities above a threshold can be non-identifiable. Non-identifiable parameters are marked red

Results

The NF-κB pathway is one of the key components controlling the innate immune response. The model considered (see Additional file 1) was first proposed in [3] and further developed in [26]. For the simulations we have used parameter values and initial conditions introduced therein and reproduced in the Table 1 of the Additional file 1. The model represents a dynamic activation of NF-κB induced genes in response to stimulation by a pro-inflammatory cytokine, TNF-α. It involves 39 parameters and 19 variables and encapsulates typical features of systems biology models. We address three questions: 1) What can we learn from the structure of parameter similarities? 2) Which parameters of the network can be estimated from the experiments published in the literature? 3) What

experiments can be performed to increase the number of identifiable parameters?

Correspondence between parameter correlations and topology of the NF-κB system. The dendrogram obtained for the NF-κB system reveals that correlated parameters are grouped into clusters that largely correspond to the network structure (Fig. 3b). The cluster *C*1 contains parameters describing IKK kinase post-translational modifications and its interactions with the IκBα-NF-κB complex; *C*2: TNF-α receptor activation and signalling; *C*3: IKKK kinase post-translational modifications and its interactions with A20 and IKK; *C*4: nuclear shuttling of NF-κB and IκBα - NF-κB binding; *C*4: A20 transcription and mRNA degradation; *C*6: IκBα transcription, translation, degradation and post-translational modifications *C*7: NF-κB - DNA interactions and nuclear shuttling of IκBα.

The correspondence of the correlation structure with the network topology is one of the main findings of the paper. After that is explicitly stated it may seem intuitive. Although it provides relevant practical information, it has not been reported before. When aiming to change model dynamical response, parameters of various network modules should be manipulated rather than those within the same module. Regarding parameter inference, knowing a priori some parameters within various modules is more likely to help in estimating the remaining parameters than knowing the same number of parameters within a single module. The analogous conclusion holds for the system robustness. In the above analysis, we assumed that all model variables define model behaviour, i.e. $q = n$, and considered a response of the system to a physiological stimulation: gradual increase, plateau and gradual decrease of TNF-α. In a later subsection we present analogous observation for a MAPK signalling model. Earlier work of Huang *et al.* [31] reported similar fining using pairwise correlations. Moreover, the authors demonstrated that parameter correlations can be effectively used for systematic model reduction.

Experiments examining the NF-κB dynamics jointly exhibit highly correlated parameters. It is debatable how much data is needed to ensure parameters' identifiability in systems biology models, and whether it is realistically achievable. Here we examined collectively all experiments reported in 9 papers [18–26] that contain rich data sets on the dynamics of the NF-κB system. We asked which parameters of the NF-κB model can be estimated from the published experiments (see Table 1 in Additional file 1). We found that 18 out of 39 model parameters cannot

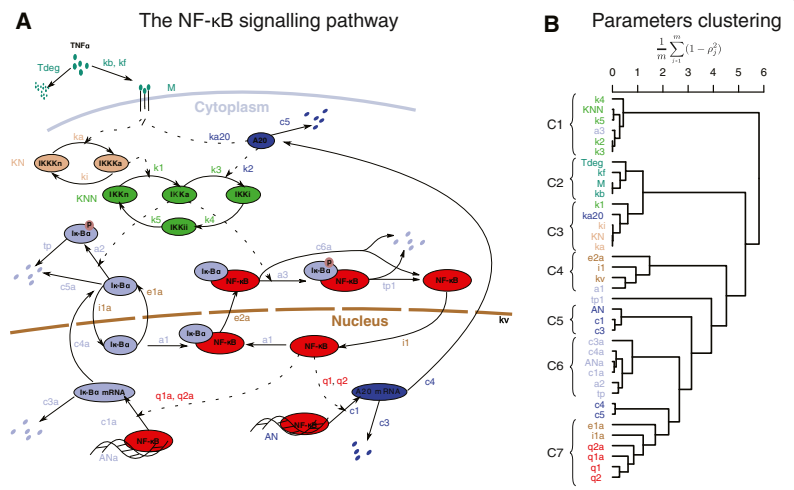


Fig. 3 The NFκB signalling pathway. **a** Schematic representation of the reactions involved. **b** Clustering of the model parameters reveals that parameters which characterise modules close in the network topology are generally more correlated than those far apart. Sensitivity vectors were calculated for a specific TNF-α stimulation to reflect physiological conditions (increase, plateau, decrease): see Figure 6B in Additional file 1

be estimated as they fail to satisfy the δ-condition (red parameters in Fig. 4a). The huge amount of literature available data, providing a comprehensive knowledge on the dynamics of the NF-κB system, was not sufficient to ensure identifiability of all model parameters. The identifiability problem is widely reported. Here we demonstrate that it is not mitigated by a huge number of experiments performed to obtain insights other than values of kinetic rates. To draw our conclusions we have initially set $\delta = 0.95$ and $\zeta = 1$. As we used logarithmic parameterisation, i.e. $\log(\theta_i)$ instead of θ_i the latter corresponds to learning a parameter more accurately than with an order of magnitude error if the remaining model parameters were known. Value $\delta = 0.95$ requires the estimate's variance not to increase by more than approximately 10 times when the single parameter and multi-parameter scenarios are compared. Thereafter we have verified that our main findings remain robust to assumptions regarding specific values of δ and ζ (Figure 3 in Additional file 1). We have also analysed how each of the analysed papers increased the number of identifiable parameters (Figure 2 in Additional file 1) . Chronologically first two papers [18, 19], rendered 13 parameters identifiable. Subsequent 7 papers provided information to estimate 8 new parameters, which gives approximately 1 parameter per paper. This indicates that making more parameters identifiable requires specifically tailored experiments different to these performed to address conventional biological questions.

Given the size of the model analysed and the size of the data included in the aforementioned papers a posterior identifiability analysis would be hardly feasible. Identifiability studies available so far analyse single or small number of experiments. Importantly the dendrogram in Fig. 4a identifies which parameters are most correlated and therefore non-identifiable. This information can be effectively used to design experimental perturbations that decrease parameter correlations and enhance parameters identifiability.

Tailored experiments can decrease parameter correlations and increase the number of identifiable parameters. In order to find experiments that can provide information about non-identifiable parameters, we first randomly searched a space of potential new TNF-α stimulation time-profiles that together with available data would make new parameters identifiable. Details of considered protocols are presented in Additional file 1. We have assumed that only variables proven before to be measurable could be quantified. After having generated 1000 random TNF-α stimulation time-profiles we surprisingly found that none of the generated protocols can make more parameters to satisfy (δ, ζ)-condition. The underling cause is show in Figure 5 in Additional file 1: in all such protocols certain parameters have close to 1 correlation with the remaining parameters. This finding indicates that a successful strategy to obtain new identifiable parameters in multi-parameter models may require more careful design of new experiments. Correlation structure (Fig. 4a) revealed the underlying cause of non-identifiability and therefore we can select some of the highly correlated parameters to be estimated in additional experiments. We propose a small number of experiments that lead to identifiability of *ki*, *KN*, *ka*, *c*3, *c*4, and *c*3*a*. Here we describe how *ki*, *KN*, *ka* can be estimated whereas experiments to estimate *c*3, *c*4 and *c*3*a* are described in Additional file 1. Parameters *ki*, *KN*, *ka* and *ka*20 describe dynamics of phosphorylated IKKK (y_1).

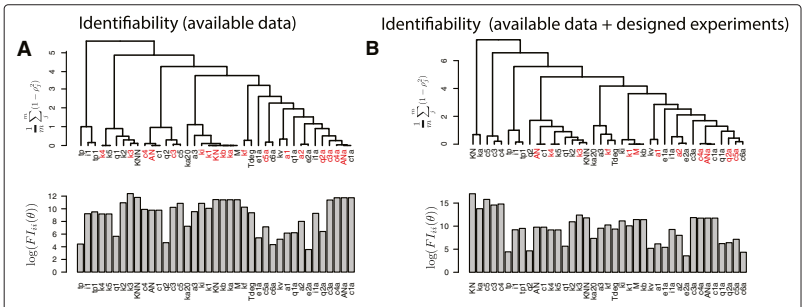


Fig. 4 Identifiability study of the NF-κB system. **a** Clustering results together with the identifiability analysis computed based on all the major published experiments. Non-identifiable parameters are marked in red. We used $\delta = 0.95$, $\zeta = 1$ to verify the identifiability condition. Sensitivity coefficients, i.e diagonal elements of the FIM, are shown below the dendrogram. **b** Clustering results as in (**a**) but for the published experiments together with the suggested experimental protocols

$$\dot{y}_1 = ka\, y_{16}\, (KN - y_1)\, ka20/(ka20 + y_9) - ki\, y_1, \quad (5)$$

where y_{16} and y_9 denote activated TNF-α receptors and cytoplasmic A20 protein, respectively (see also equation (31) in Additional file 1). We assume phosphorylated IKKK, phosphorylated TNF-α receptors and cytoplasmic A20 protein can be measured by means of immunchemistry and we are able to evaluate the equation and compare it to a data. As identified by the dendrogram (Fig 4a), structure of the equation (5) also indicates that considered three parameters have very similar impact on y_1. Figure 7A in Additional file 1 shows that indeed in a TNF-α stimulation experiment in wild type cells all parameters are highly correlated and non-identifiable. However, combining the dynamics in wild type cells, in A20 knockout cells and in A20 knockout cells with blocked phosphatase activity provides information to make *ki*, *KN* and *ka* identifiable (Figure 7C in Additional file 1). We verified that these identifiability predictions are correct using profile likelihood approach (Figure 7 B,D in Additional file 1). Identifiability also does not depend on specific parameter values used (Figure 7E in Additional file 1).

Analysis of the MAPK signalling model. In order to verify whether other biochemical models exhibit similar properties regarding correspondence between parameters similarity and network topology we have performed analysis of a MAPK signalling model [32]. The dendrogram of this model reflects the network topology (Figure 9 and 10 in Additional file 1). Our observations, therefore, might have a more general character. The model of [32] incorporates over 200 parameters and 100 equations. Computations required to plot dendrogram take several minutes on a standard desktop computer. The computational time scales with the cube of number of parameters. Therefore, the method can be applied to much larger models.

Discussion and conclusions

The mutually compensative effects of parameters changes in mathematical models have gained substantial attention in recent years [1, 4, 5, 27, 28, 33]. Methods to better understand origins and consequences of parameter correlations have began to emerge. Particularly, authors of [7] defined identifiability of parameter subsets using the smallest eigenvalue of corresponding sub-matrices of the FIM. Selection of an identifiable set of parameters based on orthogonalisation of sensitivity vectors was proposed in [8, 10]. In [9, 10, 17] authors used pairwise correlations to better understand parametric sensitivity. In addition, the method introduced in [17] allows to detect existence of an explicit functional relationship between parameters but, in contrary to our method, it does not quantify the degree of collinearity. The existing methods are largely based on the determinant, the eigenvalues of the FIM or the pairwise correlations, and do not reveal the complexity of mutual relationships between parameters in multi-parameter models. Pairwise correlations cannot reflect similarity between groups of parameters. For instance, three parameters that have low pairwise correlations can be jointly non-identifiable. This is detected by CCA. MI-CCA allowed us to phrase intuitions about the impact of parameter correlations on parameter sensitivity and identifiability in a natural, statistically justified framework. In addition efficiency of the method makes it ideally suitable for large ODE models.

In the setting of this paper the mutual information $I(\theta_A, \theta_B)$ is calculated based on the asymptotic posterior (3), which makes it exceptionally efficient to calculate in the local scenario. The concept however is very general and can be easily extended to the global case at the price of more intensive computations (see Additional file 1).

Apart from methodological development, the paper provides relevant insight into how experiments designed for purposes other than parameter estimation contribute to identifiability of model parameters. Non-identifiability problem may not be easily mitigated by collecting large number of measurements in experiments aimed at biological insight other than parameter estimation. Despite exceptionally rich data on the NF-κB dynamics, a large fraction of model parameters remains non-identifiable. Experimental design strategies to be used in the multi-parameter scenario have not been developed yet. Systematic improvement of experimental design requires origins of non-identifiability to be pinpointed and removed. Our method constitutes a theoretically grounded approach to examine link between correlations and non-identifiability in a systematic way. Having a precise picture how correlations translate into non-identifiability allows targeted and rational design of further experiments. However it does not provide any automated or systematic approach to indicate a sequence of experiments leading to a full identifiable model. It only provides information to the modeller regarding sources of non-identifiability. It only helps to understand how non-identifiability arrises and provides guidelines whether considered experimental perturbations can remove detected correlations.

Additional files

Additional file 1: Supplementary information containing detailed description of the method and reported results. (PDF 2467 kb)

Additional file 2: Open source R implementation of the clustering algorithm is available at http://sysbiosig.org/start/resources/. (PDF 2.88 kb)

Competing interests

The authors declare that they have no competing interests.

Authors' contributions
KN designed research, analyzed data, wrote the paper. MW designed research, analyzed data. TL and MK designed research, analyzed data, wrote the paper. All authors read and approved the final manuscript.

Acknowledgements
MW and MK were supported by the Foundation for Polish Science under the program Homing Plus HOMING 2011-3/4. KN and MK were funded by the Polish Ministry of Science and Higher Education under IUVENTUS PLUS grant IP2012016572. TL acknowledges support from National Science Center (Poland) grant no. 2011/03/B/NZ2/00281. MK is EMBO Installation Grantee.

Author details
[1]Institute of Fundamental Technological Research, Polish Academy of Sciences, Warsaw, Poland. [2]Faculty of Mathematics Informatics and Mechanics, University of Warsaw, Warsaw, Poland.

Received: 24 March 2015 Accepted: 28 August 2015
Published online: 29 September 2015

References

1. Brown KS, Sethna JP. Statistical mechanical approaches to models with many poorly known parameters. Phys Rev E Phys Rev E. 2003;68:021904.
2. Brown KS, Hill CC, Calero GA, Myers CR, Lee KH, Sethna JP, et al. The statistical mechanics of complex signaling networks; nerve growth factor signaling. Phys Biol. 2004;1:185–95.
3. Lipniacki T, Paszek P, Brasier AR, Luxon B, Kimmel M. Mathematical model of NF-κB regulatory module. J Theor Biol. 2004;228(2):195–215.
4. Rand DA, Shulgin BV, Salazar D, Millar AJ. Design principles underlying circadian clocks. J R Soc, Interface. 2004;1(1):119–30.
5. Komorowski M, Costa MJ, Rand DA, Stumpf MPH. Sensitivity, robustness, and identifiability in stochastic chemical kinetics models. Proc Natl Acad Sci. 2011;108(21):8645.
6. Chis OT, Banga JR, Balsa-Canto E. Structural identifiability of systems biology models: a critical comparison of methods. PloS one. 2011;6(11):27755.
7. Brun R, Reichert P, Künsch HR. Practical identifiability analysis of large environmental simulation models. Water Resour Res. 2001;37(4):1015–30.
8. Chu Y, Hahn J. Parameter set selection for estimation of nonlinear dynamic systems. AIChE J. 2007;53(11):2858–870.
9. Chu Y, Hahn J. Parameter set selection via clustering of parameters into pairwise indistinguishable groups of parameters. Ind Eng Chem Res. 2008;48(13):6000–009.
10. Chu Y, Hahn J. Generalization of a parameter set selection procedure based on orthogonal projections and the d-optimality criterion. AIChE J. 2012;58(7):2085–096.
11. Hidalgo M, Ayesa E. Numerical and graphical description of the information matrix in calibration experiments for state-space models. Water res. 2001;35(13):3206–214.
12. Ashyraliyev M, Fomekong-Nanfack Y, Kaandorp JA, Blom JG. Systems biology: parameter estimation for biochemical models. Febs J. 2009;276(4):886–902.
13. Raue A, Kreutz C, Maiwald T, Bachmann J, Schilling M, Klingmüller U, et al. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood. Bioinformatics. 2009;25(15):1923–9.
14. Kreutz C, Timmer J. Systems biology: experimental design. FEBS J. 2009;276(4):923–42.
15. Vanlier J, Tiemann C, Hilbers P, van Riel N. A Bayesian approach to targeted experiment design. Bioinformatics. 2012;28:1136–1142.
16. Liepe J, Filippi S, Komorowski M, Stumpf MP. Maximizing the information content of experiments in systems biology. PLoS Comput Biol. 2013;9(1):1002888.
17. Li P, Vu DQ. Identification of parameter correlations for parameter estimation in dynamic biological models. BMC Syst Biol. 2013;7(1):91.
18. Delhase M, Hayakawa M, Chen Y, Karin M. Positive and negative regulation of iκb kinase activity through ikkβ subunit phosphorylation. Science. 1999;284(5412):309–13.
19. Lee EG, Boone DL, Chai S, Libby SL, Chien M, Lodolce JP, et al. Failure to regulate tnf-induced nf-κb and cell death responses in a20-deficient mice. Sci Signal. 2000;289(5488):2350.
20. Hoffmann A, Levchenko A, Scott ML, Baltimore D. The Iκb-NF-κB signaling module: temporal control and selective gene activation. Science. 2002;298(5596):1241–5. doi:10.1126/science.1071914.
21. Nelson D, Ihekwaba A, Elliott M, Johnson J, Gibney C, Foreman B, et al. Oscillations in nf-κb signaling control the dynamics of gene expression. Sci Signal. 2004;306(5696):704.
22. Werner SL, Barken D, Hoffmann A. Stimulus specificity of gene expression programs determined by temporal control of ikk activity. Sci Signal. 2005;309(5742):1857.
23. Lipniacki T, Puszynski K, Paszek P, Brasier AR, Kimmel M. Single tnfα trimers mediating nf-κb activation: stochastic robustness of nf-κb signaling. BMC Bioinformatics. 2007;8(1):376.
24. Werner SL, Kearns JD, Zadorozhnaya V, Lynch C, O'Dea E, Boldin MP, et al. Encoding nf-κb temporal control in response to tnf: distinct roles for the negative regulators iκbα and a20. Genes Dev. 2008;22(15):2093–101.
25. Ashall L, Horton CA, Nelson DE, Paszek P, Harper CV, Sillitoe K, et al. Pulsatile stimulation determines timing and specificity of nf-{*kappa*}b-dependent transcription. Sci Signal. 2009;324(5924):242.
26. Tay S, Hughey JJ, Lee TK, Lipniacki T, Quake SR, Covert MW. Single-cell nf-κb dynamics reveal digital activation and analogue information processing. Nature. 2010;466(7303):267–71.
27. Rand DA. Mapping the global sensitivity of cellular network dynamics. J R Soc Interface. 2008;5:59.
28. Gutenkunst RN, Waterfall JJ, Casey FP, Brown KS, Myers CR, Sethna JP. Universally sloppy parameter sensitivities in systems biology models. PLoS Comput Biol. 2007;3(10):189.
29. Erguler K, Stumpf MPH. Practical limits for reverse engineering of dynamical systems: a statistical analysis of sensitivity and parameter inferability in systems biology models. Mol BioSyst. 2011;7(5):1593–1602.
30. Raue A, Karlsson J, Saccomani MP, Jirstrand M, Timmer J. Comparison of approaches for parameter identifiability analysis of biological systems. Bioinformatics. 2014;30:1440–1448.
31. Huang ZJ, Chu Y, Hahn J. Model simplification procedure for signal transduction pathway models: An application to il-6 signaling. Chem Eng Sci. 2010;65(6):1964–75.
32. Schoeberl B, Eichler-Jonsson C, Gilles ED, Müller G. Computational modeling of the dynamics of the map kinase cascade activated by surface and internalized egf receptors. Nat Biotechnol. 2002;20(4):370–5.
33. Komorowski M, Zurauskiene J, Stumpf MPH. StochSens - matlab package for sensitivity analysis of stochastic chemical systems. Bioinformatics. 2012;28(5):731–3.

nature COMMUNICATIONS

ARTICLE

https://doi.org/10.1038/s41467-021-24449-2 OPEN

Fractional response analysis reveals logarithmic cytokine responses in cellular populations

Karol Nienałtowski[1], Rachel E. Rigby[2], Jarosław Walczak[1], Karolina E. Zakrzewska[1], Edyta Głów[1], Jan Rehwinkel[2] & Michał Komorowski[1✉]

Although we can now measure single-cell signaling responses with multivariate, high-throughput techniques our ability to interpret such measurements is still limited. Even interpretation of dose–response based on single-cell data is not straightforward: signaling responses can differ significantly between cells, encompass multiple signaling effectors, and have dynamic character. Here, we use probabilistic modeling and information-theory to introduce fractional response analysis (FRA), which quantifies changes in fractions of cells with given response levels. FRA can be universally performed for heterogeneous, multivariate, and dynamic measurements and, as we demonstrate, quantifies otherwise hidden patterns in single-cell data. In particular, we show that fractional responses to type I interferon in human peripheral blood mononuclear cells are very similar across different cell types, despite significant differences in mean or median responses and degrees of cell-to-cell heterogeneity. Further, we demonstrate that fractional responses to cytokines scale linearly with the log of the cytokine dose, which uncovers that heterogeneous cellular populations are sensitive to fold-changes in the dose, as opposed to additive changes.

[1] Institute of Fundamental Technological Research, Polish Academy of Sciences, Warsaw, Poland. [2] Medical Research Council Human Immunology Unit, Medical Research Council Weatherall Institute of Molecular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, UK. ✉email: m.komorowski@sysbiosig.org

Many studies of signaling systems involve examining how the intensity of a stimulus, e.g., cytokine dose, translates into the activity of signaling effectors, e.g., transcription factors[1–7]. This is usually done by exposing cells to a range of doses and measuring responses either in bulk or at the single-cell level. Results of such experiments are then represented and interpreted in terms of dose–response curves. The standard dose–response curve depicts how the mean, median, or a characteristic of choice, changes with the increasing dose, and provides a basic, first-order model of how a signaling system operates. Several aspects of cellular signaling are difficult to analyze with mean/median dose–responses. For example, signaling responses can differ significantly between cells, encompass multiple signaling effectors, and are dynamic. First, outwardly very similar cells exposed to the same stimulus exhibit substantial cell-to-cell heterogeneity[8–11] (see refs. [12–14] for review). Therefore, the same mean/median response can result from a small fraction of strongly responding cells or a significant fraction of weakly responding cells[1,2,15]. Second, the highly interconnected architecture typical for mammalian signaling usually results in a single stimulus activating several primary signaling effectors or downstream genes[16–21]. For example, effectors of type I interferons include six members of the signal transducer and activator of transcription (STAT) family[22], which are activated with different sensitivities at different doses. Therefore, the description of dose–response in terms of an individual signaling effector is incomplete[23]. Third, live-cell imaging experiments demonstrated that the dose may not only alter the response at a single timepoint but can control temporal profiles of signaling responses[24,25]. For instance, low doses of tumor necrosis factor-alpha (TNF-α) may induce one peak of nuclear factor-κB signaling activity, whereas higher doses may induce additional peaks[7,26]. Besides, the dose may control the onset, shut off, amplitude, or, in principle, any other characteristics of the responses[27–30]. Overall, mean/median dose–response curves do not capture the inherent complexity of single-cell high-throughput data, and an alternative approach is required. We have used probabilistic modeling and information-theory to develop a different analytic framework, fractional response analysis, involving fractional cell counting, which is capable of deconvoluting behaviors of heterogeneous cellular populations.

Results

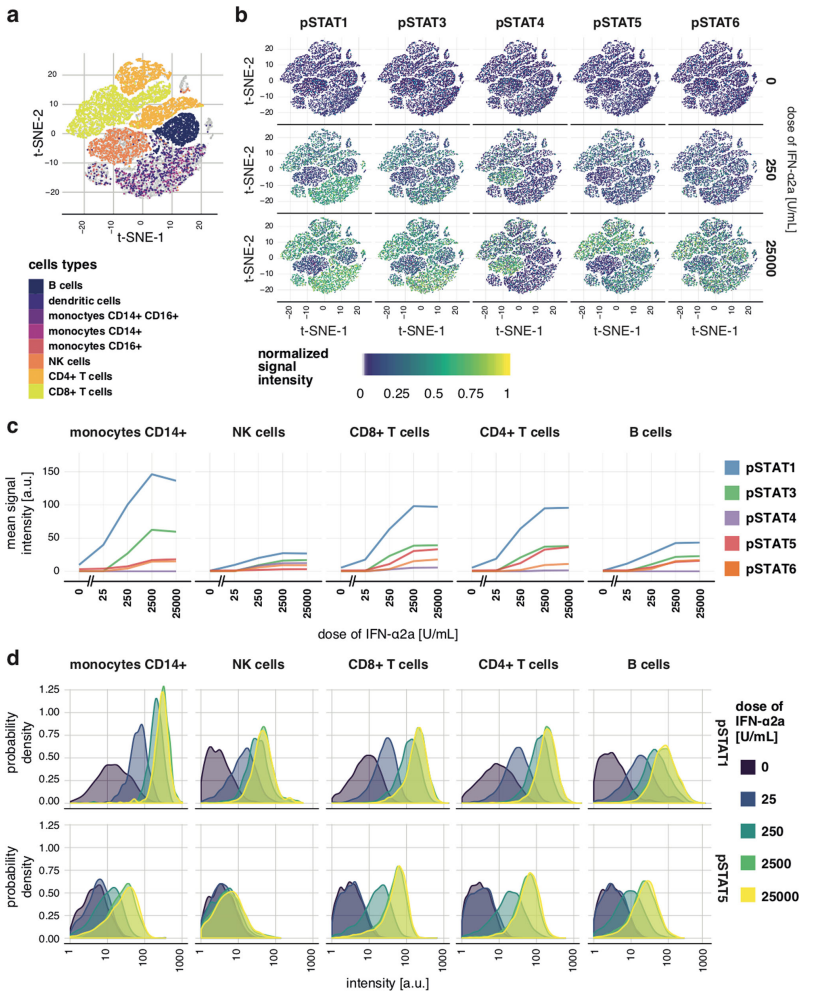
Conventional dose–response analysis does not capture complex data. To demonstrate the need and utility of FRA, we studied type I interferon signaling in human peripheral blood mononuclear cells (PBMCs), a system involving multiple signaling effectors, cell-to-cell heterogeneity, and several cell types. Dose–responses to the type I interferon variant IFN-α2a were analyzed via whole-cell tyrosine phosphorylation levels of effector proteins STAT1, STAT3, STAT4, STAT5, and STAT6 (pSTATs) measured jointly in individual cells using mass cytometry (CyTOF). Cells were collected from a healthy donor, and measurements were performed 15 min after IFN-α2a stimulation, the time of maximal response (Supplementary Fig. 1). Along with signaling effectors, 26 phenotypic markers (Supplementary Table 1), such as CD3 that marks T cells, were measured to allow for identification of several cell types including B cells, CD4+ T cells, CD8+ T cells, natural killer (NK) cells, and CD14+ monocytes[31–33] (Supplementary Fig. 2). Such multivariate data are often analyzed using t-SNE plots to visualize multiple cell types and signaling effectors[32,33] (Fig. 1a, b, Supplementary Fig. 3), which is a prerequisite for a more-detailed quantitative analysis usually involving mean/median dose–responses and population response distributions of individual signaling effectors. Following this strategy, mean levels and distributions of pSTATs in B cells, CD4+ T cells, CD8+ T cells, NK cells, and CD14+ monocytes were calculated (Fig. 1c, d) and revealed that each STAT reached different maximal phosphorylation level for different doses in a particular cell type. Medians and means of the log-data (Supplementary Fig. 4a, b) yielded similar conclusions. Plotting distributions of individual signaling effectors (Fig. 1d, Supplementary Fig. 4c) exposed considerable differences in terms of cell-to-cell heterogeneity between cell types and STATs. Nonetheless, no pattern in the functioning of the signaling system was apparent. However, the data involved five signaling effectors measured in single cells of five different types resulting in a tangible complexity possibly covering any existent regularities, which highlights the need for comprehensive approaches capable of handling complex data.

Fractional response curves. Outcomes of physiological processes, e.g., of inflammation or stress responses, depend on the number of cells with specific responses, rather than on their mean or median, which constitutes the fraction of cells with a given response as a biologically relevant variable. We proposed, therefore, to quantify dose–responses in terms of cellular fractions and show here how this can be achieved for multivariate data.

We first introduced the fractional response curve (FRC) that quantifies fractions of cells that exhibit different responses to a change in dose, or in fact any other experimental condition. For each subsequent dose, the increase of FRC reflects the fraction of cells that exhibit responses different from lower doses. Adding cumulatively distinct fractions results in counting the number of distinct response distributions.

For an illustration of FRC, in addition to the formal definition derived in Methods, we considered a simple hypothetical example involving one signaling effector and three doses, although the approach extends to a general multivariate scenario. Response distributions to three doses, x_1, x_2, x_3, which can be interpreted as control, intermediate, and high dose, are shown in Fig. 2a. When dose 1 was considered alone, fractions of cells with all possible responses sum up to 1 (Fig. 2b). Therefore, we defined the value of the FRC for dose 1 to be 1, and write $r(x_1) = 1$. We then asked what fraction of the cellular population exhibits different responses after the change from dose 1 to dose 2. The fraction of cells exhibiting different responses is equivalent to the overall increase in the frequency of responses (Fig. 2c, green region). The overall fractional increase, denoted as Δr, is calculated as the area of the green region, and $\Delta r = 0.31$, represents the 31% of the cellular population exhibiting different responses due to dose increase. Therefore, we defined the value of the FRC for dose 2 to be the sum of the previous value and the fractional increment, $r(x_2) = r(x_1) + \Delta r = 1.31$. When dose 3 was considered, the fraction of cells that exhibited different responses is again equivalent to the overall increase in the frequency of different responses, now compared with the two lower doses (Fig. 2d). As before, the overall increase, Δr, is equivalent to the area of the yellow region (Fig. 2d), with $\Delta r = 0.74$, representing 74% of cells stimulated with dose 3 exhibiting responses different to populations stimulated with lower doses. Again, the value of the FRC for dose 3 was defined as the sum of the previous value and the fractional increment, $r(x_3) = r(x_2) + \Delta r = 2.05$. Changes in the FRC show what fraction of cells exhibit different responses owing to the dose increase. Adding subsequent fractional increments, Δr, leads to the value of FRC expressed in terms of the cumulative fraction of cells that exhibit different responses due to dose change.

The sum of the dose-to-dose increments, also, records the number of distinct response distributions that were



experimentally observed, which provides the second interpretation of the FRC. Precisely, for dose 1 considered alone, a single response distribution was observed, $r(x_1) = 1$. Dose 2 added 31% of a distinct distribution, and $r(x_2) = 1.31$ (the gray area, Fig. 2d). Similarly, accounting for all three doses we had 2.05 distinct response distributions (the gray area, Fig. 2e). The number of distinct response distributions induced by changing dose quantifies the number of programmed responses of a cellular population, which appears to provide relevantly, yet, so far, unexplored, quantitative characteristics of signaling systems (Supplementary Fig. 5).

The FRC can be universally calculated for any type of signaling data, i.e., an arbitrary number of signaling effectors, time points of measurements, doses, or other experimentally varied

Fig. 1 Dose–responses to IFN-α2a in PBMCs. a t-SNE plots constructed based on phenotypic markers. Cell types are encoded by color and each dot represents a single cell. **b** t-SNE plots of whole-cell pSTATs levels 15 min after stimulation with two selected doses of IFN-α2a as well as in unstimulated cells. Positions of dots corresponding to single cells are the same as in **a** allowing cell type identification. The color of each dot represents normalized (0 for minimum and 1 for maximum) mass cytometry signal. Analogous t-SNE plots for all considered doses are shown in Supplementary Fig. 3. **c** Mean pSTATs levels in five cell types as a function of the dose calculated from mass cytometry signals of single cells. The mean of log-data and medians are shown in Supplementary Fig. 4a, b. **d** Distributions of responses in five cell types after stimulation with different doses of IFN-α2a in terms of pSTAT1 (top row) and pSTAT5 (bottom row) as measured with mass cytometry. The shown probability density is proportional to the frequency of cells with a given level of the pSTAT. The value of the probability density is proportional to the frequency of cells with given response levels. Distributions of other pSTATs are shown in Supplementary Fig. 4c. Different doses correspond to different colors. Technical details: at least 2500 cells were measured per condition. The plot shows one representative of two biological replicates.

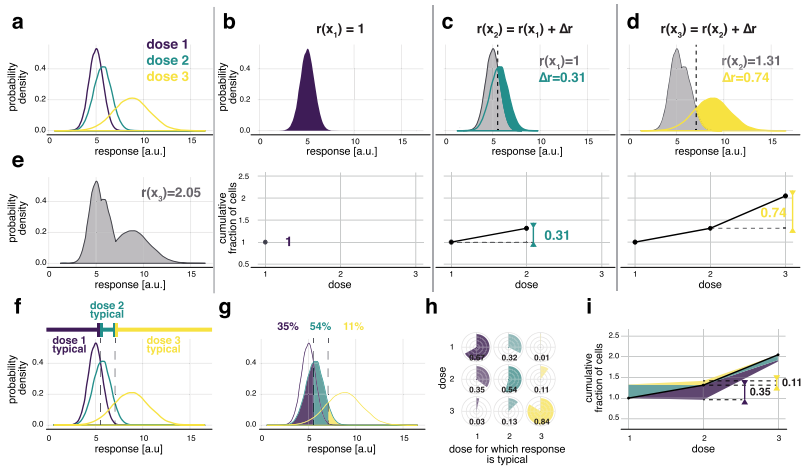


Fig. 2 Fractional response analysis. a Hypothetical response distributions to three different doses encoded by colors. Distributions are represented as a probability density, which is proportional to the frequency of cells with a given response level. **b**–**d** Quantification of the fraction of cells that exhibit different responses due to dose increase, Δ*r*, and constriction of FRC, for responses presented in **a**. Each panel from **b** to **d** corresponds to subsequent changes in dose. The color regions mark the overall increase in frequency due to considering the dose marked by the color. The area of the colored region quantifies Δ*r*. The value of the FRC for each dose is obtained by adding the increment, Δ*r*. **e** Quantification of the number of distinct distributions induced by the three considered doses. **f** Dose-typical responses for the response distributions of **a**. **g** Dissection of the responses to dose 2 into responses typical to any of the three doses. The fraction of cells typical to a given dose is marked with the corresponding color. The surface area of each color quantifies the typical fraction. **h** The fractions of cells stimulated with one dose (rows) with responses typical to any of the doses (columns). **i** The FRC together with the bands representing cell-to-cell heterogeneity as quantified in **h**. For each reference dose (*x*- axis), the fractions of cells stimulated with the reference dose that exhibit responses typical to other doses can be plotted in the form of color bands around the curve. The color encodes the dose a given fraction refers to. The height of the band marks the size of the fraction (*y*-axis). Fractions corresponding to doses higher than the reference dose are plotted above the curve, whereas to doses lower than the reference dose below the curve.

parameters, as long as sufficient data are available. The interpretation for univariate and multivariate data are the same (compare Fig. 2 and Supplementary Fig. 6). The response probability distributions do not need to be explicitly quantified, as the distinct fractions can be estimated with logistic regression (see Methods) or, in principle, other statistical classifiers, which is particularly relevant for multivariate data.

In addition to the interpretations in terms of fractions of cells, FRC has a rigorous mathematical definition in terms of Rényi information, which, broadly speaking, counts probability distributions corresponding to outputs of a communication system (see Supplementary Note 1 and Supplementary Table 2). It differs from more frequently used Shannon information as discussed in detail in Supplementary Note 2 using G protein-coupled receptors (GPCRs) signaling data[6] as well as theoretically.

Fractional cell-to-cell heterogeneity. The FRC quantifies fractions of cells that exhibit different responses due to dose change but does not quantify overall cell-to-cell heterogeneity: it does not show what fraction of cells exposed to one dose exhibits responses in the range characteristic for other doses. Therefore, within FRA, we propose to augment the FRC with quantification of the overlaps between distributions corresponding to different doses. We call a given response as typical for a given dose if it is most likely, i.e., most frequent, to arise for this specific dose compared

to all other doses. In the hypothetical example, low responses are most likely, and therefore typical, for dose 1, intermediate responses are typical for dose 2, and high responses for dose 3 (Fig. 2f). We can then divide responses to a given dose into responses typical for any dose. For instance, for dose 2, 35% of cells have responses typical for dose 1, 54% typical for dose 2, and 11% typical for dose 3 (Fig. 2g). The results, presented as pie-charts, can be shown in a matrix as the fraction of cells stimulated with one dose (rows) that has responses typical for other doses (columns) (Fig. 2h). This pie-chart partitioning can be plotted along with the FRC (Fig. 2i) so that the fractional increments, Δr, and fractional cell-to-cell heterogeneity are concisely presented. Similar to FRC, quantification of the fractional cell-to-cell heterogeneity structure can be performed for multivariate data without quantification of the response distributions (see Methods and Supplementary Fig. 6).

Populations of different types of PBMCs exhibit very similar logarithmic dose–responses. FRA compresses complex dose–response data into a simple quantitative description accounting for cell-to-cell heterogeneity and multivariate measurements. To determine the kinds of biological information that can be uncovered, we performed FRA for IFN-α2a multivariate dose–responses in different types of PBMCs, assuming that all five measured pSTATs jointly constitute a cell's response. The FRC and fractional cell-to-cell heterogeneity (Fig. 3a, b) are very similar for all cell types. Counter to the differences seen in the analysis presented in Fig. 1a–d and Supplementary Fig. 4, the dose–responses in different cell types follow the same logarithmic pattern identifying a phenomenon that governs the behavior of multivariate cellular responses of our system, which remains hidden when inspecting data in the conventional way.

For all cell types, the FRC is linear and increases at the same rate with respect to the log of the dose, which means that the fraction of cells showing different responses is proportional to the dose fold-change, over a broad range of doses, i.e., from 0 to 2500 U/mL. The linear increase of the FRC demonstrates that the fraction of cells that exhibit different responses are very similar from 0 to 25 U/mL, from 25 to 250 U/mL, and from 250 to 2500 U/mL. For each subsequent dose change, $\Delta r \approx 0.5$ so that 50% of cells have different responses. A given fold-change in the dose induces a different response in the fixed fraction of cells, across a broad range of doses. Therefore, cellular populations are sensitive to fold-changes in the dose as opposed to additive changes.

Formally, FRC scales as the log of the dose

$$r(x) \propto \log(x), \tag{1}$$

which given incremental approximation, $\Delta\log(x) = \log(x + \Delta x) - \log(x) \approx \Delta x/x$, implies fold-change sensitivity in the population

$$\Delta r \propto \Delta x/x, \tag{2}$$

which in the studied system universally describes dose–responses in populations across different cell types.

The FRA, therefore, condenses the description of the complex multivariate responses into a simple quantitative formula. Furthermore, FRA uncovered that the number of programmed response distributions, i.e., maximal value of FRC, and the fractional cell-to-cell heterogeneity structure are very similar for all cell types. This similarity indicates that the immune system may precisely control responses of fractions of cells rather than responses of individual cells. In multicellular organisms, a fraction of cells with a given response level is a biologically essential response variable. For example, the outcome of a viral infection in tissue depends on the number of NK cells with given

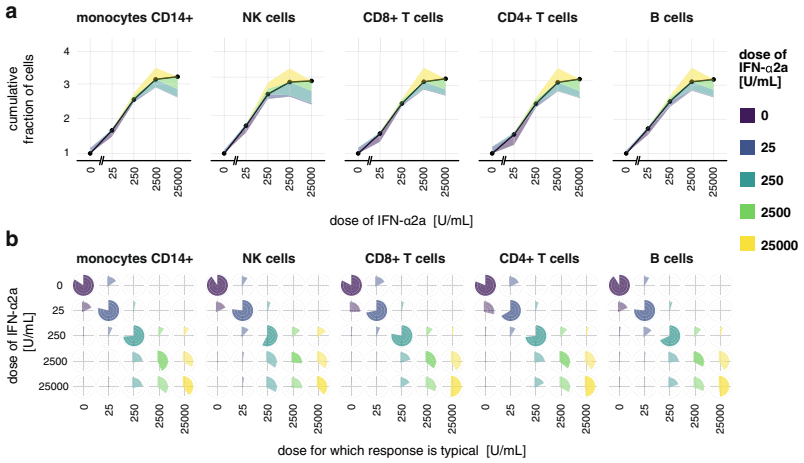


Fig. 3 Different types of PBMCs exhibit highly similar dose–responses to IFN-α2a. a FRA of IFN-α2a responses. Here, levels of all pSTATs were assumed to jointly constitute cell's response. Supplementary Fig. 7 shows FRA for individual STATs. **b** Pie-charts of the cell-to-cell heterogeneity structure used to plot color bands in **a**. Cell-to-cell heterogeneity is shown as pie-charts, in addition to **a**, in order to clearly visualize the similarity between the cell types. Technical details: The plot presents an analysis of one biological replicate. The analysis of the second biological replicate yielding very similar results is shown in Supplementary Fig. 8.

response levels and induced cytotoxic activity. Our analysis revealed that in the studied system the fraction of cells that have responses in a specific range is not only tightly controlled in the population of a given cell type but is controlled in the same way across different cell types, as opposed to responses of individual cells that are largely heterogeneous within one cell type and across cell types. The role for controlling the fractions of cells with specific responses can, in principle, be tested by perturbing cell-to-cell heterogeneity through genetic manipulation and observing the phenotypic effects on the performance of the immune system.

Fold-change sensitivity of cellular populations is a recurrent property of cytokine signaling. To explore how responses that are qualitatively different translate into differences in FRA, we examined responses to the cytokines interferon-gamma (IFN-γ) and interleukin 10 (IL-10). As these are implicated in macrophage phenotypic diversity[34], we used the human monocyte cell line U937, differentiated into macrophage-like cells, and immunostained to measure responses via nuclear levels of the key signaling effectors, phosphorylated STAT1 for IFN-γ, and phosphorylated STAT3 for IL-10 at 30 min after stimulation (Fig. 4a, b). For IFN-γ, response distributions shift gradually towards higher values as the dose increases, which is referred to as the graded response[35–37]. For IL-10, the distributions flatten over a broad region as the dose increases reflecting the higher number of responding cells for high doses, with the dose having a limited impact on the level of the response, similar to a binary system[2,37] where responses aggregate in "on" and "off" regions. The qualitative differences in the responses to IFN-γ and IL-10 cytokines are mirrored by FRA (Fig. 4d, e and Supplementary Fig. 11a, b). Compared with IL-10, FRC for IFN-γ increases faster and reaches a higher value, which reflects the higher number of cells with distinct responses for increasing doses. Besides, bands around FRC for IFN-γ are narrower than for IL-10, indicating that the response distributions are more distinct. Furthermore, for IL-10 the bands below the curve are broader than above the curve, which reflects the large fractions of cells with response typical to doses lower than encountered, which points to the similarity with the binary system. In Supplementary Note 3, we used in silico generated data of exact binary and graded responses, as well as responses of lung cancer cell lines to IFN-γ to show in detail how FRA can discriminate between different response modalities. Overall, differences in the response distributions visible to the naked eye are adequately mapped onto the FRA plot, which

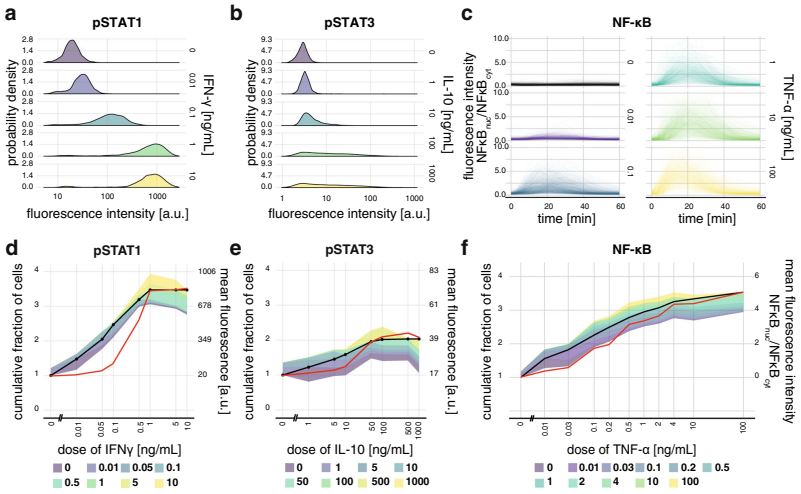


Fig. 4 IFN-γ, IL-10, and TNF-α exhibit logarithmic dose–responses in cell lines. a Distributions of responses 30 min after stimulation with different doses of IFN-γ in terms of nuclear pSTAT1 as measured with confocal microscopy imaging and immunostaining in the U937 cell line. Responses are expressed as mean fluorescence of nuclear pixels. Selected doses are shown. See Supplementary Fig. 9 for all doses. **b** The same as in **a** but for pSTAT3 after stimulation with IL-10. **c** Temporally resolved responses of individual murine embryonic fibroblasts to increasing concentrations of TNF-α. Each line corresponds to a single cell. Responses are expressed as the ratio of the nuclear pixels fluorescence average to the cytoplasmic-pixels fluorescence average. Selected doses are shown in the panel. See Supplementary Fig. 10 for all doses. Measurements were taken every 3 min in a murine embryonic fibroblast cell line stably expressing the p65 protein component of the NF-κB complex fused with the fluorescent dsRed protein. **d** FRA of IFN-γ responses. For comparison, the red line presents the mean response with *y*-axis on the right. The pie-chart corresponding to the plotted cell-to-cell heterogeneity structure is shown in Supplementary Fig. 11. Comparison with mean of log-data is shown in Supplementary Fig. 12. **e** Same as in **d** but for IL-10 responses. **f** FRA of the temporally resolved responses to TNF-α. The red line presents the mean response at the time of maximal response, i.e., at 18 min. Comparison with mean of log-data at 18 min is shown in Supplementary Fig. 12. Technical details: at least 892 cells were measured per condition. Experiments were performed in two biological repeats each consisting of two technical replicates. In **a** and **b**, the sum of two technical replicates of a representative biological replicate is shown. In **c** sum of biological replicates is shown[26].

indicates that similar phenomena hidden by the complexity of multivariate data could be uncovered in the same way.

To further explore how generally applicable FRA is, we examined time-series responses to cytokine stimulation. We measured TNF-α dose–responses with live confocal imaging of a murine embryonic fibroblast cell line stably expressing fluorescent NF-κB complex[4], a key TNF-α signaling effector[4,29], for 60 min (Fig. 4c). As the data constituting the time-series are multivariate, it is not feasible to assess the fraction of cells with distinct responses for increasing doses as well as overlaps between response distributions using visual inspection alone. FRA, on the other hand, enables the quantification and visualization of the cell-to-cell heterogeneity structure (Fig. 4f and Supplementary Fig. 11c). The overlaps between distributions are considerable. The rate of the increase of FRC, as well as width of the bands around FRC, are more similar to IFN-γ than for IL-10, which cannot be seen here directly from the time-series data (Fig. 4c). Primarily, however, FRC increases linearly with the log of TNF-α dose, which again discloses the fold-change sensitivity of the cellular population.

Despite differences in the responses to the above cytokines and type of data used, FRC increases almost linearly with respect to the log of the dose for all three cytokines. Therefore, similarly to IFN-α2a in PBMCs, IFN-γ, IL-10, and TNF-α responses are sensitive to fold-changes in the dose, as opposed to additive changes, suggesting that this mode of the response may be a more universal biological pattern that describes cytokine signaling in cellular populations.

Discussion

Sensitivity to dose fold-changes in populations of cells resembles the empirical Weber-Fechner law that characterizes the performance of many psycho-physiological sensory systems. Minimal detectable stimulus change, Δx, in the sense of weight, hearing, vision, and smell, has been observed to be of fold-type. So far, several pathways have been shown to follow some form of fold-change sensitivity either by observing representative individual cells or population averages. For instance, Wnt and TGF-β signaling exhibit desensitization to background ligand concentration and subsequent sensitivity to fold-changes[38,39] with an incoherent feed-forward loop motif being the explanatory mechanism[40–42]. Also, single-cell gene expression induced by the nuclear signaling effectors β-catenin, SMAD, or NF-κB, were shown to be sensitive to fold-changes in their nuclear levels[11,38,39]. Similarly, changes in inter-spike intervals in Ca^{2+} spike trains are proportional to baseline inter-spike intervals in GPCRs signaling activated with phospholipase C ligands[43].

Here, FRA allowed us to make a considerably different observation. We demonstrated that within heterogeneous populations of cells of a given type, and across types, the number of cells that exhibit a different response is proportional to the fold-change in the dose. We did not refer to a single signaling effector in a representative cell or population average but to the state of the heterogeneous population described by multivariate data. Ultimate outcomes of multicellular processes like immunity are not determined by individual cells alone or population averages but by a heterogeneous collective. By accounting for cell-to-cell heterogeneity, we showed that the distribution of the collective, which encodes stimulation levels in multicellular systems, shifts with the fold-changes of the dose. Therefore, the way in which heterogeneous cell populations encode signals is quantitatively similar to the way we perceive differences in certain sensations (weight/light).

Weber-Fechner law is a pattern that can arise from a range of different mechanisms[44,45] with the underlying neural implementations still being discovered[45,46]. Here also, a mechanistic explanation of the fold-change sensitivity of cellular populations is not clear and remains to be determined, possibly by relating cell-to-cell heterogeneity with ligand sensitivity, which might involve feed-forward loops, as in fold-change detection in individual cells over a long time scale[40–42].

Overall, FRA delivers a concise representation of complex single-cell data, which is particularly relevant for high-throughput techniques, which are increasingly allowing the measurement of a high number of parameters per cell, generating very large, high-dimensional datasets[47]. The high information content of multivariate, single-cell measurements makes biological discoveries more likely. On the other hand, however, insights may be difficult to extract due to data complexity. Therefore, making use of the increasing amount of single-cell high-throughput measurements requires approaches that can extract relevant insights in spite of complexity. FRA is not limited to cytokine signaling, proteomic data, or dose–responses, enabling the systematic investigation of single-cell high-throughput data in a wide range of situations, in which responses are measured in single cells at any "-omics" scale. Accurate estimation requires, however, the number of measured cells to considerably exceed the number of measures signaling effectors, and a representative selection of doses (see Supplementary Note 4 for caveats of FRA). Nonetheless, FRA should yield insights into the structure of signaling heterogeneity in immunology, developmental biology, cancer research, and diverse other fields in which response analysis in single cells is of relevance.

Methods

Software implementation. The methodology to perform and visualize FRA is provided as a user-friendly R-package available for download at http://github.com/sysbiosig/FRA. The package contains an installation guide and a brief user manual.

Formal definition of the FRC. Consider a series of doses $x_1,\ldots,x_i,\ldots,x_m$ and denote a single-cell response as y. Depending on the context, y, may be a number or a vector, e.g., the level of one or more measured signaling effectors. Suppose that responses to a given dose, x_i, are represented as the probability distribution,

$$P(Y|x_i). \tag{3}$$

The FRC is then formally defined as

$$r(x_i) = \int_{\mathcal{Y}} \max_{x_k \le x_i} P(y|x_k)dy, \tag{4}$$

where integration takes place over $\mathcal{Y}$, the set of all possible responses, y. The integral quantifies the area under the curve (or under surface for multivariate data), with respect to y, defined as $\max_{x_k \le x_i} P(y|x_k)$. For the calculations shown in Fig. 2 the integration corresponds to the calculation of the area of the gray regions in c–e. As explained in Supplementary Note 1, the FRC defined as above is closely related Rényi min-information capacity.

Formal definition of typical fractions. Having the responses represented in terms of the probability distribution, Eq. 3, we can define which responses, y, are typical to any of the doses. Precisely, we define the response, y, to be typical for dose x_j if it is most likely to arise for this dose, which writes as

$$P(y|x_j) > P(y|x_k) \text{ for all } k \text{ other than } j. \tag{5}$$

The above condition allows assigning any response, y, to a dose for which it is typical. Therefore, for a given dose, x_i, we can identify what fraction of cells stimulated with this dose exhibits responses typical to any dose, x_j, for j from 1 to m. These fractions, denoted as v_{ij}, can be practically computed as explained below.

Calculation of typical fractions. The fractions of cells stimulated with dose i that have responses typical to dose j, v_{ij}, can be easily calculated from data regardless of the number of doses and the type of experimental measurements. We have that

$$v_{ij} = \frac{\text{number of cells stimulated with } x_i \text{ with responses typical for } x_j}{\text{number of cells stimulated with } x_i}. \tag{6}$$

Calculation of typical fractions, v_{ij}, with the above formula requires the possibility to examine the condition $P(y\,|\,x_j) > P(y\,|\,x_k)$ for any experimentally observed response, y. The distributions $P(y\,|\,x_j)$ can be reconstructed from data using a variety of probability density estimators[48]. The use of the available estimators,

however, might be problematic for multivariate responses[26,49]. We, therefore, propose a more convenient strategy. We replace the condition $P(y \mid x_j) > P(y \mid x_k)$ with an equivalent condition that is computationally much simpler to evaluate. Precisely, we propose to use the Bayes formula

$$P(x_j|y) = \frac{P(y|x_j)P(x_j)}{\sum_{k=1}^{m} P(y|x_k)P(x_k)}. \quad (7)$$

If we set the equiprobable prior distribution, i.e., $P(x_j) = 1/m$, we have that $P(y|x_j)$ is proportional to $P(x_j|y)$ and the condition $P(y|x_j) > P(y|x_k)$ is equivalent to

$$P(x_j|y) > P(x_k|y). \quad (8)$$

The above strategy allows avoiding estimation of the response distributions, $P(y|x_j)$, from data. For continuous and multivariate variable y the estimation of $P(x_j|y)$ is generally simpler than estimation of $P(y|x_j)$[26,48]. Precisely, an estimator $\hat{P}\left(x_j|Y=y\right)$ of the distribution $P(x_j|y)$ can be built using a variety of Bayesian statistical learning methods. For simplicity and efficiency, here we propose to use logistic regression, which is known to work well in a range of applications[48]. In principle, however, other classifiers could also be considered. The logistic regression estimators of $P(x_j|Y=y)$ arise from a simplifying assumption that log-ratio of probabilities, $P(x_j|Y=y)$ and $P(x_m|Y=y)$ is linear. Precisely,

$$\log\left(\frac{P(x_j|Y=y)}{P(x_m|Y=y)}\right) \approx \alpha_j + \beta_j^T y. \quad (9)$$

The above formulation allows fitting the logistic regression equations to experimental data, i.e., finding values of the parameters, α_j and β_j that best represent the data. The fitted logistic regression model allows assigning cellular responses to typical doses based on conditions given by Eq. 8. Formally, the fractions v_{ij} defined by Eq. 6 are calculated as

$$v_{ij} = \frac{1}{n_i}\sum_{l=1}^{n_i} \mathbb{1}_{\left\{\hat{P}(x_j|Y=y) > \hat{P}(x_k|Y=y)\,\forall k \neq j\right\}}\left(y_i^l\right), \quad (10)$$

where n_i is the number of cells measured for the dose x_i, y_i^l denotes response of the l-th cell, and $\mathbb{1}_{\left\{\hat{P}(x_j|Y=y) > \hat{P}(x_k|Y=y)\,\forall k \neq j\right\}}(y_i^l)$ is equal 1 if $\hat{P}\left(x_j|Y=y\right) > \hat{P}\left(x_k|Y=y\right)$ for any $k \neq j$ and 0 otherwise.

Calculation of the FRC. Calculation of the FRC can be conveniently performed using the typical fractions, as defined above, rather than through integration of Eq. 4. Precisely, to calculate the FRC for the dose, x_i, consider doses $x_1,\ldots,x_i$ in isolation from higher doses. Then, the sum of typical fractions $v_{11},\ldots,v_{ii}$ is equivalent to FRC for the dose x_i

$$r(x_i) = \sum_{k=1}^{i} v_{ik}. \quad (11)$$

The equivalency of the above equation and Eq. 4 is derived in the Supplementary Methods.

Mass cytometry (CyTOF). PBMCs were isolated from the peripheral blood of healthy adult donors using Lymphoprep (Stemcell Technologies), according to the manufacturer's instructions. Cells were washed in serum-free Roswell Park Memorial Institute (RPMI) then resuspended at 10^7 cells/mL in serum-free RPMI containing 0.5 mM Cell-ID Cisplatin (Fluidigm) and incubated at 37 °C for 5 min. Cells were washed with RPMI containing 10% (v/v) FCS (Sigma) and 2 mM L-Glutamine (R10), centrifuging at 300 × *g* for 5 min before being resuspended to 6×10^7 cells/mL in R10 and rested at 37 °C for 15 min. 50 mL of cells (3×10^6 cells) were transferred to 15 mL falcon tubes for stimulation and antibody staining. Antibodies and their dilutions are listed in Supplementary Table 1. Staining for CD14, CCR6, CD56, CD45RO, CD27, CCR7, CCR4, and CXCR3 was done before stimulation/fixation for 30 min in R10 at 37 °C. Cells were stimulated with 0, 25, 250, 2500, or 25000 U/mL recombinant human IFN-α2a (PBL Assay Science, #11100-1) diluted in R10 for 15 min at 37 °C. After washing with 5 mL cold Maxpar PBS (Fluidigm), cells were fixed with 1× Maxpar Fix I Buffer (Fluidigm) for 10 min at RT before being washed with 1.5 mL Maxpar Cell Staining Buffer (CSB, Fluidigm). All centrifugation steps after this point were at 800 × *g* for 5 min. Cells were barcoded using Cell-ID 20-Plex Pd Barcoding Kit (Fluidigm), according to the manufacturer's instructions, and washed twice with CSB before samples were pooled and counted. All further steps were performed on the pooled cells. Fc receptors were blocked using Fc Receptor Binding Inhibitor Antibody (eBioscience, #14-9161-73) diluted 1:10 in CSB for 10 min at RT. Surface antibody staining mixture was added directly to the blocking solution and incubated for 30 min at RT. Cells were washed twice with CSB, resuspended in ice-cold methanol, and stored at −80 °C overnight. After washing twice with CSB, cells were stained with intracellular antibody staining mixture for 30 min at RT before two further washes in CSB. Cells were resuspended in 1.6% (v/v) formaldehyde (Pierce, #28906) diluted in Maxpar PBS and incubated for 10 min at RT. Cells were resuspended in 125 nM Cell-ID Intercalator (Fluidigm) diluted in Maxpar Fix and Perm Buffer (Fluidigm) and incubated overnight at 4 °C. Compensation beads (OneComp eBeads Compensation Beads, Invitrogen, #01-1111-42) stained with 1 mL of each antibody were also prepared. The next day, cells and compensation beads were washed twice with CSB and twice with Maxpar water (Fluidigm), mixed with a 1:10 volume EQ Four Element Calibration Beads (Fluidigm) before acquisition on a Helios Mass Cytometer (Fluidigm) using the HT injector. Data were normalized, randomized, and concatenated using Helios CyTOF Software v6.7 (Fluidigm), cytofCore v0.4, flowCore v1.46.2, and Cytobank v6.2. Compensation and de-barcoding were performed using the CATALYST v1.5.3.23 package[50]. Different immune cell subpopulations were gated in R v3.5.1 from single, live, CD45+ cells as shown in Supplementary Fig. 2.

Collection and analysis of PBMCs were carried out in accordance with the EU Directive 2004/23/EC and the UK Human Tissue Act 2004 (HTA), under the HTA licence (number 12433) of the Weatherall Institute of Molecular Medicine. Informed consent was obtained and the samples were fully anonymised.

U937 cells. U937 cells (CRL-1593.2, ATCC), a human monocyte cell line, were cultured under standard conditions at 37 °C in a humidified atmosphere of 5% CO_2/95% air in low glucose RPMI 1640 (Corning, #10-040-CV) medium supplemented with 10% fetal bovine serum (FBS, ThermoFisher, #10500064) and 1% penicillin–streptomycin solution (P/S, ThermoFisher, #15140122). For macrophage differentiation, U937 cells were suspended in a medium with 20 ng/mL phorbol 12-myristate 13-acetate (PMA, Sigma Aldrich, #P1585) and plated in 96-well microplates with μClear®flat bottom (Greiner, #655090) in density 2×10^4 cells per well. After 24 h medium with PMA was removed and fresh medium was added to cells. 72 h after seeding on 96-well microplates differentiated cells were incubated with recombinant human IFN-γ (ThermoFisher, #PHC4031) at concentrations 0DAPI10 ng/mL or recombinant human IL-10 (PeproTech, #200-10) at concentrations 0–1000 ng/mL for 30 min. Afterwards, cells were fixed with 3.7% paraformaldehyde (PFA, Sigma Aldrich, #P6148) for 10 min at room temperature, RT, then permeabilized with 90% ice-cold methanol (Sigma, #322415), for 30 min at −20 °C, blocked with 5% bovine serum albumin (BSA, Merck, #821006) and 0.3% Triton X-100 (Sigma Aldrich, #T9284) for 1 h at RT, and incubated with primary antibody–phospho-STAT1 (Tyr701) (pSTAT1, Cell Signaling, #9167) diluted 1:100 or phospho-STAT3 (Tyr705) (pSTAT3, Cell Signaling, #4113) diluted 1:200 in 1% BSA with 0.3% Triton X-100 for 18 h at 4 °C. Next day, cells were incubated with an appropriate secondary antibody–Alexa Fluor 488 (Life Technologies, #A-21206) or Alexa Fluor 555 (Life Technologies, #A-31570) diluted 1:500 in 1% BSA with 0.3% Triton X-100 for 1.5 h at RT and stained with 2 μg/mL 4′,6-diamidino-2-phenylindole (Sigma Aldrich, #D9542) for 10 min at RT. The fluorescence signal was acquired using an automated confocal microscope (Pathway 435, BD) and analyzed with Cell Profiler v2.1.1, R v3.5.1, and ImageJ v1.48.

Murine immortalized fibroblasts. Murine embryonic fibroblasts 3T3 cell line, previously used in several studies including[4,26], expressing fluorescent fusion proteins relA-dsRed as wells H2B-GFP for nuclei identification were cultured in an incubator under standard conditions at 37 °C in a humidified atmosphere of 5% CO2/95% air. The cell line was kindly provided by Professor S. Tay. The cells were cultured in high glucose Dulbecco's Modified Eagle's Medium without phenol red (ThermoFisher, #21063029) supplemented with 10% FBS (ThermoFisher, #10500064) and 1% penicillin–streptomycin solution (P/S, ThermoFisher, 15140122). Approximately 1.3×10^5 cells were plated on 35-mm confocal dish for imaging. After 48 h in the incubator, cells were transferred to the environmental chamber in a microscope. At time 0, medium was removed from cells and recombinant mouse TNF-α (Sigma Aldrich, #T7539) was added at concentrations 0-100 ng/mL as a 5-minute pulse. Live imaging was performed using a confocal microscope, Leica TCS SP5 X. During single experiment, images have been captured every 3 min over 1 h in two channels simultaneously at nine different positions on the plate. The experiment has been repeated at least four times to test reproducibility and to allow for a sufficient number of observations. Nuclear and cytoplasmic fluorescence (pixel mean) was then quantified from microscopic images using Cell Profiler v2.1.1, and R v3.5.1. The response of each cell was then represented as the ratio of nuclear to cytoplasmic fluorescence in order to ensure the robustness of measurements to changes in confocal plane over time. The data set is described in detail in ref. [26], where it was initially published.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Data generated during the study are available for download from https://github.com/sysbiosig/FRA/ and are also deposited in the open-access repository https://doi.org/10.5281/zenodo.4835622. The study also involves data published in ref. [26].

Code availability

FRA is made available as an R-package (see Supplementary Note 5) downloadable from https://github.com/sysbiosig/FRA/ that is also deposited in the open-access repository https://doi.org/10.5281/zenodo.4818586.

Received: 9 July 2020; Accepted: 17 June 2021;
Published online: 07 July 2021

References

1. Ko, M. S., Nakauchi, H. & Takahashi, N. The dose dependence of glucocorticoid-inducible gene expression results from changes in the number of transcriptionally active templates. *EMBO J.* **9**, 2835–2842 (1990).
2. Walters, M. C. et al. Enhancers increase the probability but not the level of gene expression. *Proc. Natl Acad. Sci. USA* **92**, 7125–7129 (1995).
3. Cheong, R., Rhee, A., Wang, C. J., Nemenman, I. & Levchenko, A. Information transduction capacity of noisy biochemical signaling networks. *Science* **334**, 354–358 (2011).
4. Tay, S. et al. Single-cell NF-kappaB dynamics reveal digital activation and analogue information processing. *Nature* **466**, 267–271 (2010).
5. Suderman, R., Bachman, J. A., Smith, A., Sorger, P. K. & Deeds, E. J. Fundamental trade-offs between information flow in single cells and cellular populations. *Proc. Natl Acad. Sci. USA* **114**, 5755–5760 (2017).
6. Keshelava, A. et al. High capacity in G protein-coupled receptor signaling. *Nat. Commun.* **9**, 876 (2018).
7. Selimkhanov, J. et al. Accurate information transmission through dynamic biochemical signaling networks. *Science* **346**, 1370–1373 (2014).
8. Satija, R. & Shalek, A. K. Heterogeneity in immune responses: from populations to single cells. *Trends Immunol.* **35**, 219–229 (2014).
9. Lu, Y. et al. Highly multiplexed profiling of single-cell effector functions reveals deep functional heterogeneity in response to pathogenic ligands. *Proc. Natl Acad. Sci. USA* **112**, E607–E615 (2015).
10. Bowsher, C. G. & Swain, P. S. Identifying sources of variation and the flow of information in biochemical networks. *Proc. Natl Acad. Sci. USA* **109**, E1320–E1328 (2012).
11. Hagai, T. et al. Gene expression variability across cells and species shapes innate immunity. *Nature* **563**, 197–202 (2018).
12. Lee, R. E. C., Walker, S. R., Savery, K., Frank, D. A. & Gaudet, S. Fold change of nuclear NF-κB determines TNF-induced transcription in single cells. *Mol. Cell* **53**, 867–879 (2014).
13. Eling, N., Morgan, M. D. & Marioni, J. C. Challenges in measuring and understanding biological noise. *Nat. Rev. Genet.* **20**, 536–548 (2019).
14. Symmons, O. & Raj, A. What's luck got to do with it: single cells, multiple fates, and biological nondeterminism. *Mol. Cell* **62**, 788–802 (2016).
15. Porpiglia, E., Hidalgo, D., Koulnis, M., Tzafriri, A. R. & Socolovsky, M. Stat5 signaling specifies basal versus stress erythropoietic responses through distinct binary and graded dynamic modalities. *PLoS Biol.* **10**, e1001383 (2012).
16. Rowland, M. A., Greenbaum, J. M. & Deeds, E. J. Crosstalk and the evolvability of intracellular communication. *Nat. Commun.* **8**, 16009 (2017).
17. Rowland, M. A., Harrison, B. & Deeds, E. J. Phosphatase specificity and pathway insulation in signaling networks. *Biophys. J.* **108**, 986–996 (2015).
18. Amit, I., Wides, R. & Yarden, Y. Evolvable signaling networks of receptor tyrosine kinases: relevance of robustness to malignancy and to cancer therapy. *Mol. Syst. Biol.* **3**, 151 (2007).
19. Housden, B. E. & Perrimon, N. Spatial and temporal organization of signaling pathways. *Trends Biochem. Sci.* **39**, 457–464 (2014).
20. Komorowski, M. & Tawfik, D. S. The limited information capacity of cross-reactive sensors drives the evolutionary expansion of signaling. *Cell Syst.* **8**, 76–85 (2019). e6.
21. Bridgham, J. T., Carroll, S. M. & Thornton, J. W. Evolution of hormone-receptor complexity by molecular exploitation. *Science* **312**, 97–101 (2006).
22. Rauch, I., Müller, M. & Decker, T. The regulation of inflammation by interferons and their STATs. *JAKSTAT* **2**, e23820 (2013).
23. Schmitz, M. L., Weber, A., Roxlau, T. & Gaestel, M. Signal integration, crosstalk mechanisms and networks in the function of inflammatory cytokines. *Biochim. Biophys. Acta* **1813**, 2165–2175 (2011).
24. Behar, M., Barken, D., Werner, S. L. & Hoffmann, A. The dynamics of signaling as a pharmacological target. *Cell* **155**, 448–461 (2013).
25. Purvis, J. E. & Lahav, G. Encoding and decoding cellular information through signaling dynamics. *Cell* **152**, 945–956 (2013).
26. Jetka, T., Nienałtowski, K., Winarski, T., Błoński, S. & Komorowski, M. Information-theoretic analysis of multivariate single-cell signaling responses. *PLoS Comput. Biol.* **15**, e1007132 (2019).
27. Sampattavanich, S. et al. Encoding growth factor identity in the temporal dynamics of FOXO3 under the combinatorial control of ERK and AKT kinases. *Cell Syst.* **6**, 664–678 (2018).
28. Nandagopal, N. et al. Dynamic ligand discrimination in the notch signaling pathway. *Cell* **172**, 869–880 (2018).
29. Lane, K., Andres-Terre, M., Kudo, T., Monack, D. M. & Covert, M. W. Escalating threat levels of bacterial infection can be discriminated by distinct MAPK and NF-κB signaling dynamics in single host cells. *Cell Syst.* **8**, 183–196 (2019).
30. Nunns, H. & Goentoro, L. Signaling pathways as linear transmitters. *Elife* **7**, e33617 (2018).
31. Bendall, S. C. et al. Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science* **332**, 687–696 (2011).
32. van der Maaten, L. & Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
33. Amir, E. D. et al. viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nat. Biotechnol.* **31**, 545–552 (2013).
34. Biswas, S. K. & Mantovani, A. Macrophage plasticity and interaction with lymphocyte subsets: cancer as a paradigm. *Nat. Immunol.* **11**, 889–896 (2010).
35. Kringstein, A. M., Rossi, F. M., Hofmann, A. & Blau, H. M. Graded transcriptional response to different concentrations of a single transactivator. *Proc. Natl Acad. Sci. USA* **95**, 13670–13675 (1998).
36. Becskei, A., Séraphin, B. & Serrano, L. Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. *EMBO J.* **20**, 2528–2535 (2001).
37. Biggar, S. R. & Crabtree, G. R. Cell signaling can direct either binary or graded transcriptional responses. *EMBO J.* **20**, 3167–3176 (2001).
38. Goentoro, L. & Kirschner, M. W. Evidence that fold-change, and not absolute level, of beta-catenin dictates Wnt signaling. *Mol. Cell* **36**, 872–884 (2009).
39. Frick, C. L., Yarka, C., Nunns, H. & Goentoro, L. Sensing relative signal in the Tgf-β/Smad pathway. *Proc. Natl Acad. Sci. USA* **114**, E2975–E2982 (2017).
40. Goentoro, L., Shoval, O., Kirschner, M. W. & Alon, U. The incoherent feedforward loop can provide fold-change detection in gene regulation. *Mol. Cell* **36**, 894–899 (2009).
41. Ferrell, J. E. Signaling motifs and Weber's law. *Mol. Cell* **36**, 724–727 (2009).
42. Shoval, O. et al. Fold-change detection and scalar symmetry of sensory input fields. *Proc. Natl Acad. Sci. USA* **107**, 15995–16000 (2010).
43. Thurley, K. et al. Reliable encoding of stimulus intensities within random sequences of intracellular Ca2+ spikes. *Sci. Signal.* **7**, ra59 (2014).
44. Gorea, A. & Sagi, D. Disentangling signal from noise in visual contrast discrimination. *Nat. Neurosci.* **4**, 1146–1150 (2001).
45. Dehaene, S. The neural basis of the Weber-Fechner law: a logarithmic mental number line. *Trends Cogn. Sci. (Regul. Ed.)* **7**, 145–147 (2003).
46. Pardo-Vazquez, J. L. et al. The mechanistic foundation of Weber's law. *Nat. Neurosci.* **22**, 1493–1502 (2019).
47. Sagner, A. & Briscoe, J. Morphogen interpretation: concentration, time, competence, and signaling dynamics. *Wiley Interdiscip. Rev. Dev. Biol.* **6**, e271 (2017).
48. Hastie, T., Friedman, J. & Tibshirani, R. The elements of statistical learning. (Springer New York, 2001).
49. Mack, Y. P. & Rosenblatt, M. Multivariate k-nearest neighbor density estimates. *J. Multivar. Anal.* **9**, 1–15 (1979).
50. Chevrier, S. et al. Compensation of signal spillover in suspension and imaging mass cytometry. *Cell Syst.* **6**, 612–620 (2018). e5.

Acknowledgements

K.N., J.W., K.E.Z., and M.K. were supported by Foundation for Polish Science within the First TEAM (First TEAM/2017-3/21) program co-financed by the European Union under the European Regional Development Fund. K.N. was supported by Polish National Science Centre under grant PRELUDIUM 2016/23/N/ST6/03505. R.E.R. and J.R. were funded by the UK Medical Research Council [MRC core funding of the MRC Human Immunology Unit]. We thank Guy Riddihough for helpful comments during the preparation of this manuscript. M.K. and J.R. also thank Matteo Iannacone, the organizer of an EMBO workshop on Immunology, Bergamo 2017, which had initiated our collaboration.

Author contributions

Conceptualization of FRA.: K.N., M.K.; interpretation of FRA.: K.N., J.W., M.K.; CyTOF study design and experimentation: R.E.R., J.R.; microscopy study design and experimentation: K.E.Z., E.G.; data interpretation and analysis: K.N., R.E.R., J.W., K.E.Z., J.R., M.K.; implementation of the R-software package and of the user manual: K.N.; figure preparation: K.N., M.K.; writing the manuscript: K.N., K.E.Z., R.E.R., M.K.; editing the manuscript: K.N., R.E.R., J.W., K.E.Z., J.R., M.K.

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/s41467-021-24449-2.

Correspondence and requests for materials should be addressed to M.K.

Peer review information *Nature Communications* thanks Luis Aguilera, Brian Munksy, Kevin Thurley and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.