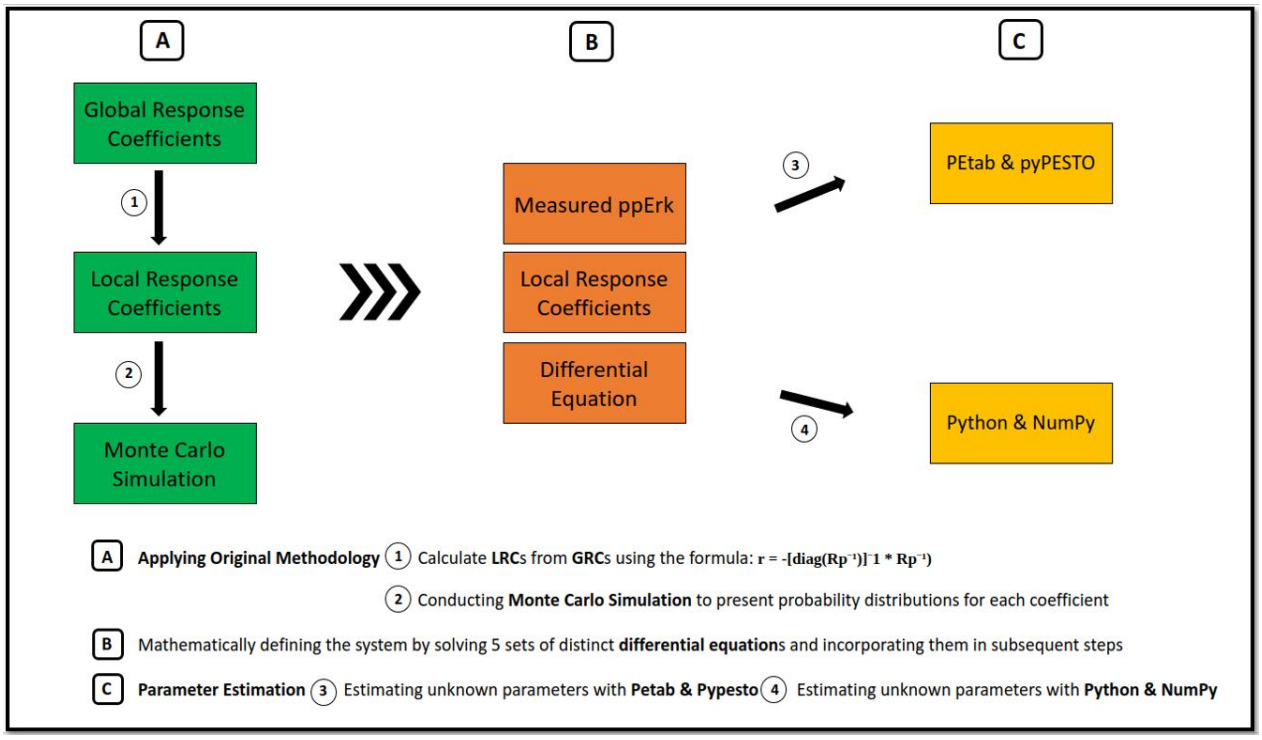
Unraveling the Complexity of MAPK Signaling Pathway

Abstract

In the dynamic landscape of systems biology, a pivotal and challenging task is the precise estimation of parameters that drive the behavior of intricate signaling pathways. This article delves into the heart of this challenge, focusing on the parameter estimation aspect within the context of the mitogen-activated protein kinase (MAPK) signaling pathways. These pathways, with their critical role in cellular responses, from proliferation to immunity, present an ideal backdrop for such investigations. Initially, our study sought to reproduce the methods employed in the original work [1] using Modular Response Analysis [6]. Subsequently, we initiated the process of rewiring the signaling pathway, defining it through mathematical equations. These equations formed the basis for a parameter estimation study, where we endeavored to estimate unknown parameters introduced during the system rewiring. Employing two distinct methodologies, our approach involved implementing a parameter estimation model using Python and NumPy, featuring a core gradient descent algorithm. In parallel, we utilized specialized packages, PEtab and pyPESTO in Python, for a comparative analysis of parameter estimation.

Graphical Abstract



Introduction

The mitogen-activated protein kinases (MAPKs) are conserved proteins that constitute essential components of various signaling pathways (Figure 1) found in organisms ranging from yeast to mammals [3]. These pathways consist of membranal and cytoplasmic signaling molecules [2], including several protein kinases that regulate critical cellular processes such as proliferation, differentiation, apoptosis, survival, inflammation, and innate immunity [3]. Among these pathways, one involving the epidermal growth factor receptor (EGFR), the tyrosine kinase receptor (TrkA), and three key proteins, namely Rapidly Accelerated Fibrosarcoma (Raf), mitogen-activated protein kinase kinase (Mek or MAPKK), and Extracellular Signal-Regulated Kinase (Erk), stands out. This pathway is influenced by various stimuli. Of these, we are specifically interested in two: nerve growth factor (NGF) and epidermal growth factor (EGF). The former leads to neuronal differentiation, while the latter induces cellular proliferation[1].

PC-12 cells, derived from rat adrenal pheochromocytoma, serve as a model for neuronal differentiation and have been instrumental in studying the MAPK signaling pathway [4]. The differences in how Erk is activated in response to NGF and EGF stimuli play a crucial role in shaping the cellular responses [5]. In other words, the specific timing, magnitude, and duration of Erk activation can lead to a wide array of downstream effects, as depicted in (Figure 1).

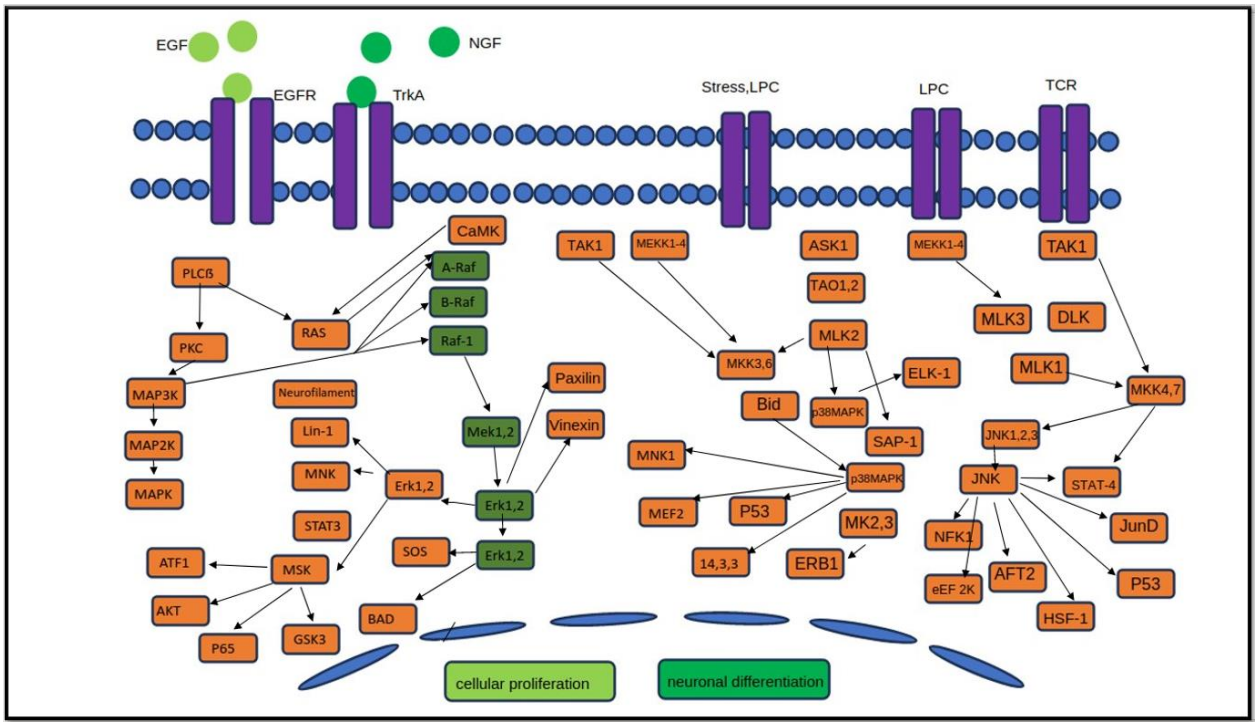


Figure 1: MAP Kinase Pathways. This diagram illustrates the intricate network of MAP Kinase Pathways, comprising multiple interconnected signaling pathways, each with its unique dependence on the others. The core proteins, including Raf, Mek, and Erk (illustrated in dark green), play pivotal roles in most of these pathways. Zooming into a specific segment of the MAPK signaling pathways, we focus our attention on this selected region (dark green) of the complex network for our research. The stimuli EGF (light green) and NGF (medium green) influence EGFR and TrkA

receptors, respectively. Both stimuli impact isoforms of the same proteins—Raf, Mek, and Erk—within the same pathway, yielding distinct cellular outcomes: EGF prompts cellular proliferation (light green), while NGF induces neuronal differentiation (medium green).

The central question at hand is how variations in Erk dynamics are influenced by the processes occurring upstream in the signaling pathway. The dynamics of Erk activation are primarily governed by the interconnected components within the MAPK signaling module, configuring themselves in response to different stimuli, such as EGF or NGF. The specific arrangement of these connections can give rise to distinct patterns of Erk activation dynamics [1].

To delve into the intricacies of the MAPK signaling network and understand how variations in Erk dynamics result from upstream processes, researchers have employed the Modular Response Analysis (MRA) technique [6]. This technique involves analyzing network responses under steady-state conditions (at 5 minutes, representing a pseudo-steady state, and at 15 minutes after NGF stimulation, and steady-state condition at 5 minutes after EGF stimulation), following incremental perturbations introduced through small RNA interference [1]. While the measured global response coefficients (Table 1) demonstrate how perturbations propagate through the network, inferring the underlying network topology directly from these measurements is challenging. However, by calculating local response coefficients, which indicate the sensitivity of one module to another in isolation from the network, network connectivity maps can be generated (Figure 2).

Global Response Coefficients (GRCs)										
		EGF 5 minutes			NGF 5 minutes			NGF 15 minutes		
		Raf	Mek	Erk	Raf	Mek	Erk	Raf	Mek	Erk
Experiment 1	siRNA Raf	-0.686	-0.860	-1.201	-0.607	-0.878	-0.994	-0.506	-0.742	-1.154
	siRNA Mek	-0.128	-0.870	-0.952	-0.276	-0.476	-0.850	0.072	-0.553	-0.054
	siRNA Erk	-0.311	1.055	-0.714	-0.108	0.184	-0.908	-0.031	0.231	-0.778
Experiment 2	siRNA Raf	-1.190	-0.292	-1.593	-0.526	0.817	0.066	-0.501	0.098	-0.393
	siRNA Mek	0.272	0.228	-0.188	0.065	-0.435	0.066	0.088	-0.860	-0.625
	siRNA Erk	-0.963	0.224	-1.662	-0.115	0.430	-0.468	0.236	0.224	-0.823
Experiment 3	siRNA Raf	-0.395	-0.913	-0.281	-0.590	-0.488	-0.381	-0.835	-0.340	-0.166
	siRNA Mek	-0.098	-0.465	-0.265	-0.223	-0.276	-0.384	0.665	-0.249	-0.258
	siRNA Erk	-0.229	0.110	-1.241	-0.217	0.199	-0.322	0.344	0.592	-0.167
Experiment 4	siRNA Raf	-0.832	0.135	-0.184	-0.516	-0.099	-0.238	-0.835	-0.340	-0.166
	siRNA Mek	0.269	-0.048	0.125	-0.092	0.024	-0.150	0.665	-0.249	-0.258
	siRNA Erk	-0.159	0.416	-0.469	-0.146	0.087	-0.840	0.616	0.942	-1.140

Table1: Measured Global Response Coefficients. The tables show the measured global response coefficients derived from four sets of experiments, quantifying changes in the activity of specific modules, including Erk, Mek, and Raf proteins, before and after perturbation. Each experiment involved perturbations induced through RNA interference

(RNAi), resulting in the effective downregulation of protein levels. These response coefficients were obtained from quantitative western blot experiments and are sourced from the article [3], Note that the measured GRCs are rounded to three decimal places, the original table contains measurements with 8 decimal places.

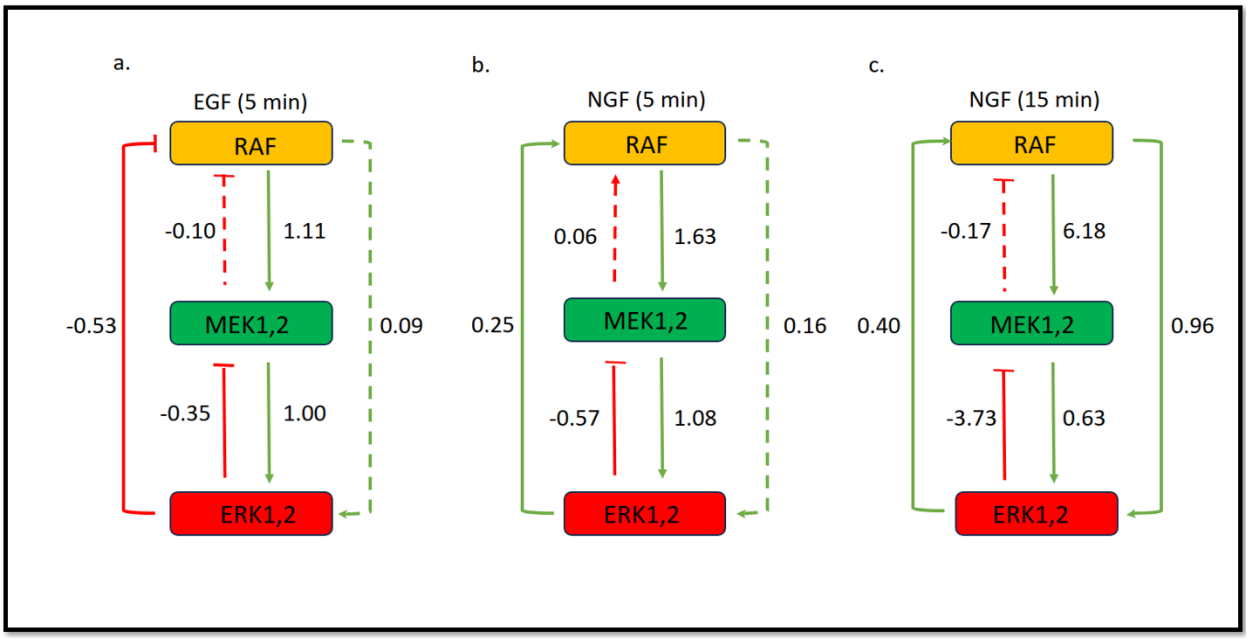


Figure 2: Wiring Diagrams of Local Response Coefficients. a) Five minutes after stimulating the system with EGF, showcasing the network's dynamic responses under this specific condition. b) Five minutes after stimulating the system with NGF, representing the pseudo-steady state response. c) Fifteen minutes after stimulating the system with NGF, representing the steady state response. Arrows indicate coefficients, with a minus coefficient indicating an inhibitory effect, and dashed lines representing minimal effects. Highlighting the network's adaptation over time. The boxes represent the proteins involved in the pathways.

In our research, we initially aimed to replicate the work[7] from Santos et al. [1], using the MRA method to estimate local response coefficients from global response coefficients obtained in experiments. Subsequently, we extended our investigation by refining the network, introducing EGF and NGF stimuli and linking them to the main proteins - Raf, Mek and Erk. We then generated mathematical models for this extended network [8], which were derived from various circuit diagrams (Figure 3). Our next challenge was to determine the unknown coefficients in these models as well as to validate the coefficients whose values were calculated using the MRA method. This was done using two different methods: First, we attempted to write Python programs based on the differential equations [8] (one program for each set of equations) that describe the system in different ways to estimate the unknown coefficients. Each of these programs has a gradient descent algorithm at its core and can be used in both cases (NGF or EGF as stimulus). In the second method, we used special Python packages (PEtab [14] and pyPESTO [15]) to achieve the same result, namely estimating the unknown parameters and validating the already calculated parameters.

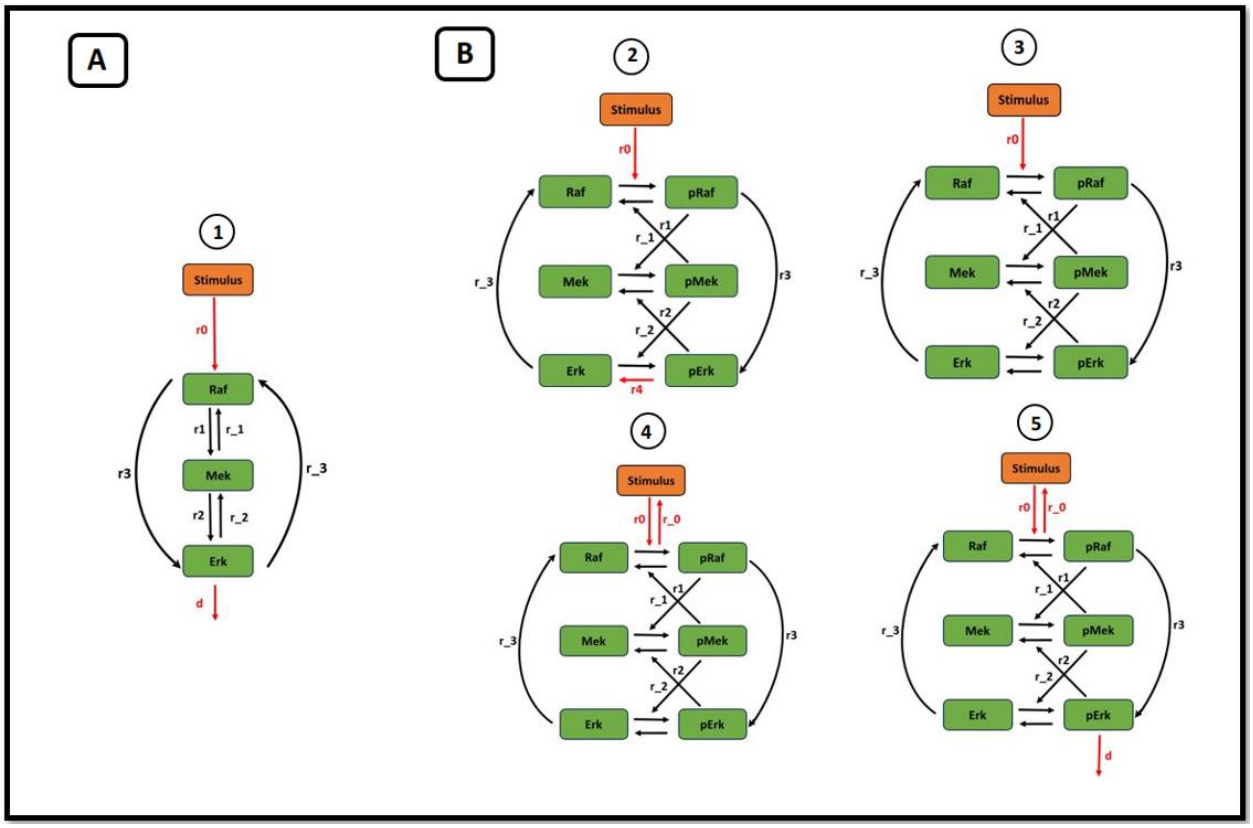


Figure 3: Diagrams Illustrating the Pathway. Five distinct diagrams are presented to elucidate the pathway in different manners. A) Single-State Diagram: In this case, a single diagram defines one state for each protein. B) Multi-State Diagrams: Four different diagrams offer varied representations of the model. In these instances, proteins exhibit two states (phosphorylated and unphosphorylated). Calculated coefficients are denoted in black (r1, r_1, r2, r_2, r3, and r_3), while unknown coefficients are highlighted in red (r0, r_0, r4, and d).

Methods and Results

To initiate our investigation, we visually represented the Global Response Coefficients (GRCs) - mean of the coefficient sets from four experiments - and Local Response Coefficients (LRCs) through bar plots (Figure 4).

To calculate LRCs from GRCs, we developed a Python program [9], grounded in the core formula ($r = - [diag(Rp^{-1})]^{-}1 * Rp^{-1}$) introduced in the original MRA-method article [6]. This process aligns with Modular Response Analysis (MRA), facilitating a comprehensive understanding of the system's dynamics. Subsequently, we executed Monte Carlo Simulations [10] using the NumPy [19] library in Python, incorporating normal (Gaussian) distributions to generate probability distributions for each coefficient under EGF 5 minutes, NGF 5 minutes, and NGF 15 minutes conditions [11] (Figure 5a).

During this process, we realized that MRA, with its matrix calculations (utilizing the formula: r = -[diag(Rp^{-1})]$^{-}$1 * Rp^{-1}), could be computationally intensive. To address this, we created a Python program [12] to measure computation times for different input matrix shapes (Figure 5 b).

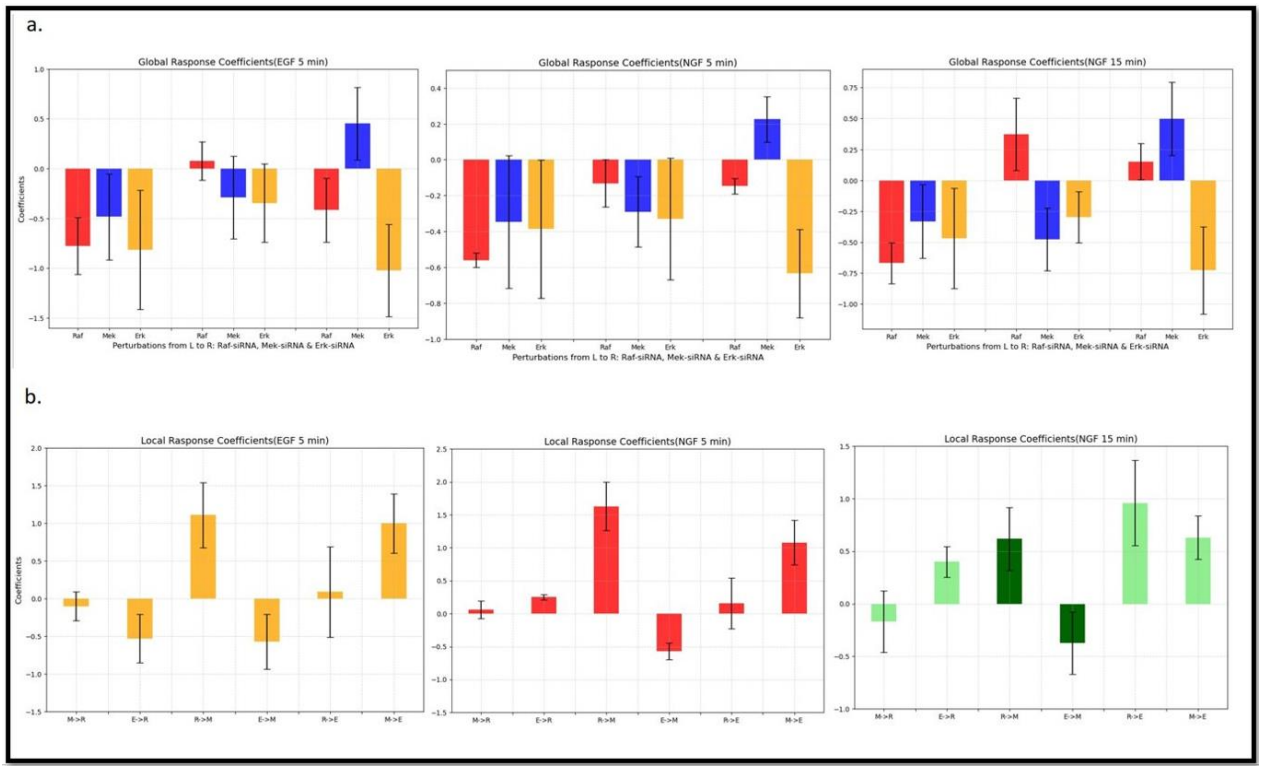


Figure 4: Visualized Bar Plots of Global- and Local-Response Coefficients. a) Average global response coefficients (GRC) for EGF and NGF (5 min) and NGF (15 min), The perturbations are represented on the *x* axis (*Raf* siRNA, *Mek* siRNA and *Erk* siRNA). Average GRC are depicted for pRaf (red), pMek (blue) and pErk(orange). b) Computed local response coefficients. in both case (GRCs and LRCs) error bars represent the standard deviation.

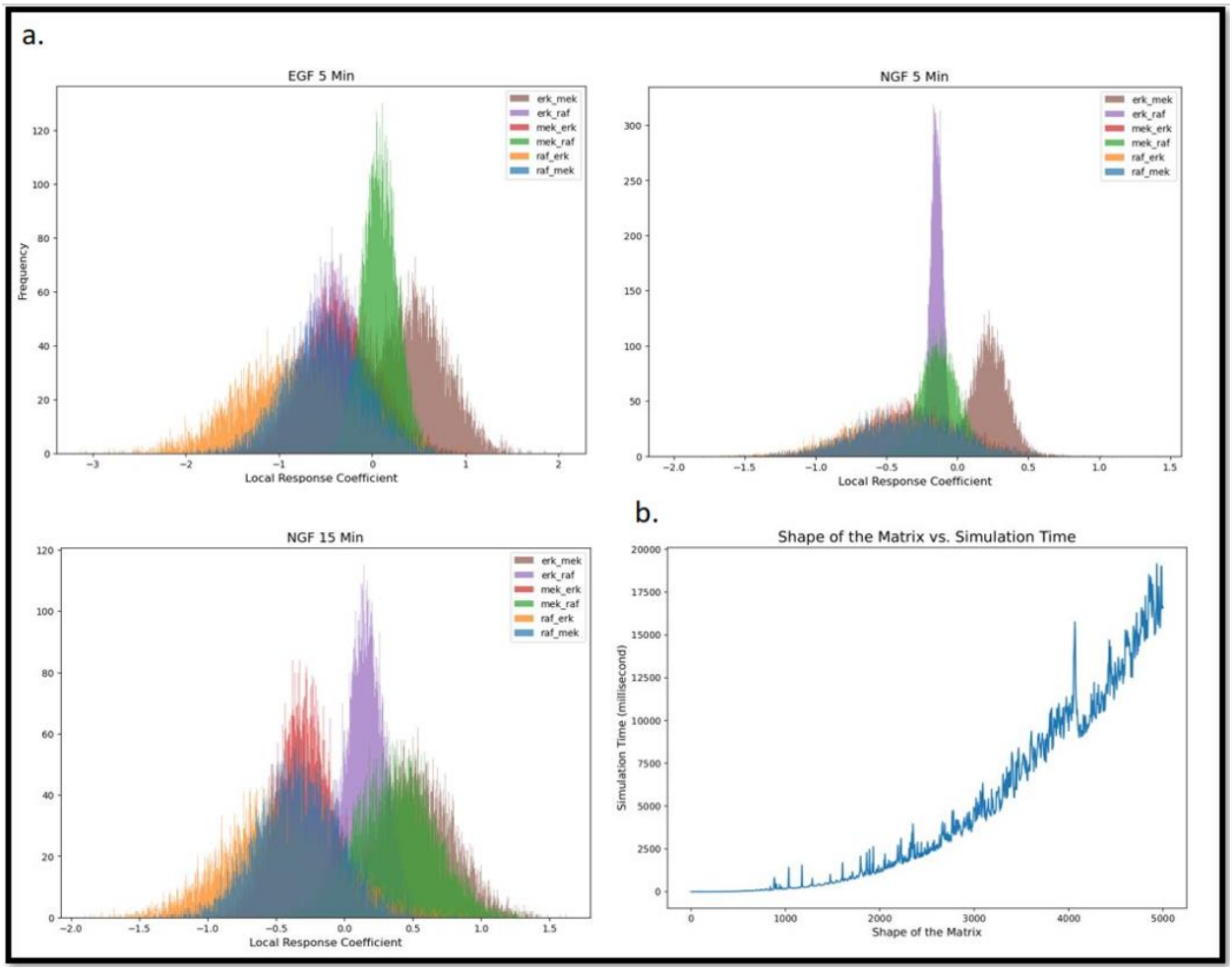


Figure 5: a) Monte Carlo Simulations Histograms: Histograms depict Monte Carlo simulations for EGF (5 mins), NGF (5 mins), and NGF (15 mins). In each simulation, the standard deviation of Global Response Coefficients (GRCs) was utilized as the spread parameter for the normal distribution (numpy.random.normal()). b) Matrix Size vs. Computation Time Visualization: This visualization illustrates the correlation between the size of the coefficient matrix and the time required to execute local response matrix computations based on the global response matrix. The graphic enables observation of how simulation time fluctuates with increasing matrix size.

Following the replication of the original work's methods [1] to familiarize ourselves with the subject, we attempted to estimate the unknown parameters, particularly those connecting the stimulus and the Raf protein, which were not defined in the original article. To do this, we formulated five different sets of differential equations, each describing slightly different network configurations within the signaling pathway [8]. We then employed two distinct methods to estimate the unknown parameters.

Gradient Descent Optimization Approach

In our optimization approach, we created Python programs [13], each tailored to the gradient descent algorithm, with the aim of estimating unknown parameters from sets of differential equations. Within each program, three pivotal functions collaborate seamlessly to optimize the defined coefficients.

The first function calculates the concentration of each species (stimulus, Raf/pRaf, Mek/pMek, and Erk/pErk) based on the differential equations designed for rewiring the pathway. Subsequently, the second function, cost function, employs a mean squared error approach to quantify the disparity between predicted and observed concentrations. This function serves as a cornerstone in the optimization process. The third function embodies the gradient descent algorithm, a dynamic component updating coefficients by calculating derivatives of residuals concerning unknown parameters. This iterative process enhances the precision of coefficient optimization. The program orchestrates these functions iteratively, steadily refining coefficients and converging towards an optimized solution. This detailed approach ensures a robust and effective optimization strategy for the complex system described by the differential equations.

PEtab and pyPESTO

In our methodology, we harnessed specialized tools, specifically PEtab [14] and the Python Parameter Estimation Toolbox (pyPESTO) [15], to streamline the parameter estimation process. Initially, we employed Pandas DataFrame [20], a Python library, to construct tables that were subsequently transformed into PEtab-compatible tables essential for constructing a pyPESTO optimization problem. This conversion involved generating specific PEtab tables (experimental_condition.tsv, parameter.tsv, observables.tsv, measurement.tsv, and visualization.tsv) [16].

Following this, we utilized model information, encompassing reactions, initial species concentrations, and coefficients, to generate SBML files for each case. This step was facilitated by the online tool "make SBML" [21].

Subsequently, using the generated PEtab tables and SBML files, we crafted a YAML file necessary for conducting a pyPESTO problem. This file encapsulates crucial information for the optimization process.

The final step in the workflow involved the development of a pyPESTO model, with the standard optimization algorithm (ScipyOptimizer) employed to utilize the prepared data [17]. This comprehensive approach, integrating PEtab and pyPESTO, facilitated a robust parameter estimation methodology for the system, providing a thorough and efficient means of exploring parameter space and refining the model.

Challenges

Despite our meticulous approach, we encountered challenges that hindered the attainment of desired outcomes in both the Gradient Descent Optimization Approach and the PEtab & pyPESTO approach. Notably, the predicted concentrations of pErk in each case diverged significantly from the experimentally derived values, either converging to zero or diverging to infinity [22]. This discrepancy underscored potential inaccuracies in our approaches to estimate unknown parameters within the pathway.

Several factors contributed to this setback. Insufficient information about the system's topology and structure, coupled with a lack of raw data in a suitable format (the pErk concentration was normalized by an unknown method and we obtained it from a graph defining the change in pErk concentration over time), posed significant obstacles. Furthermore, the intricate nature of the signaling pathway, illustrated in Figure 1, reflected only a fraction of a larger, complex network, involving numerous proteins with intricate interdependencies. Furthermore, neither we nor the original article have defined the membrane proteins (EGFR and TrkA) in the system.

Additionally, computational limitations emerged as a substantial challenge. The demands of parameter estimation, when compounded by computational constraints, limited our ability to explore a broader range of unknown parameters. Overcoming these limitations could potentially enhance the success of our approach.

Conclusion

In the complex landscape of systems biology, precise methods for parameter estimation are fundamental and shape the course of research efforts [18]. Our exploration of MAPK signaling pathways using modular response analysis (MRA) and innovative computational approaches revealed valuable insights and encountered notable challenges. The observed deviation of predicted pErk concentrations from experimental values, ranging from convergence towards zero to deviation towards infinity, highlights the complexity inherent in parameter estimation in dynamic biological systems. In light of these challenges, we propose several avenues for future research to strengthen the robustness of such studies.

First, the use of high-performance computing is a possible means to overcome the computational limitations encountered in our study. Expanding computational resources could enable a more comprehensive exploration of the parameter space, potentially revealing previously unseen patterns and improving the accuracy of estimates. Additionally, employing different optimization algorithms may provide a viable solution to this matter, enhancing the robustness and efficiency

of the analyses. A better understanding of the structural intricacies of the system and the acquisition of additional raw data available in a format suitable for analysis are essential. This endeavor could provide a clearer map of the topology of the pathway and provide a more comprehensive basis for parameter estimation. Furthermore, expanding the parameter range for each variable proves to be a strategy to increase the effectiveness of parameter estimation. A broader exploration of the parameter space can reveal critical nuances in the system dynamics and thus contribute to a more accurate representation of the biological processes under investigation.

References

[1] Santos, S., Verveer, P. & Bastiaens, P. Growth factor-induced MAPK network topology shapes Erk response determining PC-12 cell fate. *Nat Cell Biol* **9**, 324–330 (2007). [nature cell biology] Accessed 12 November 2023

[2] Seger, Rony, and Edwin G. Krebs. "The MAPK signaling cascade." *The FASEB journal* 9.9 (1995): 726-735. [FASEB] Accessed 12 November 2023

[3] Kim EK, Choi EJ. Compromised MAPK signaling in human diseases: an update. Arch Toxicol. 2015 Jun;89(6):867-82. doi: 10.1007/s00204-015-1472-2. Epub 2015 Feb 18. PMID: 25690731. [Archives of Toxicology] Accessed 12 November 2023

[4] Greene, L. A. & Tischler, A. S. Establishment of a noradrenergic clonal line of ratadrenalpheochromocytoma cells which respond to nerve growth factor. Proc. Natl Acad. Sci.USA 73, 2424–2428 (1976). [PANS] Accessed 15 November 2023

[5] Specificity of Receptor Tyrosine Kinase Signaling: Transient versus Sustained ExtracellularSignal-Regulated Kinase Activation, C. J. Marshall. Cell, Vol, 80, 179-185, January27, 1995, Copyright© 1995by Cell Press. [Cell] Accessed 15 November 2023

[6] Kholodenko, B. N. et al. Untangling the wires: a strategy to trace functional interactions in signaling and gene networks. Proc. Natl Acad. Sci. USA 99, 12841–12846 (2002). [PANS] Accessed 12 November 2023

[7] GitHub – Replica Work. Accessed 15 December 2023

[8] GitHub – Differential Equations. Accessed 15 December 2023

[9] GitHub – Local Response Coefficients. Accessed 15 December 2023

[10] Bonate, P.L. A Brief Introduction to Monte Carlo Simulation. *Clin Pharmacokinet* **40**, 15–22 (2001). [Clinical Pharmacokinetics] Accessed 21 November 2023

[11] GitHub – Monte Carlo Simulation. Accessed 15 December 2023

[12] GitHub – Calculation Intensity. Accessed 15 December 2023

[13] GitHub – Parameter Estimation. Accessed 15 December 2023

[14] Schmiester L, Schälte Y, Bergmann FT, Camba T, Dudkin E, Egert J, et al. (2021) PEtab—Interoperable specification of parameter estimation problems in systems biology. PLoS Comput Biol 17(1): e1008646. [PLOS] Accessed 30 November 2023

[15] Leonard Schmiester, Daniel Weindl, Jan Hasenauer, Efficient gradient-based parameter estimation for dynamic models using qualitative data, *Bioinformatics*, Volume 37, Issue 23, December 2021, Pages 4493–4500. [Bioinformatics] Accessed 30 November 2023

[16] GitHub – PEtab. Accessed 15 December 2023

[17] GitHub – pyPESTO. Accessed 15 December 2023

[18] Lillacci G, Khammash M (2010) Parameter Estimation and Model Selection in Computational Biology. PLoS Comput Biol 6(3): e1000696. [PLOS] Accessed 18 November 2023

[19] Python Library - NumPy. Accessed 15 December 2023

[20] Python Library - pandas. Accessed 15 December 2023

[21] GitHub - Make SBML Models. Accessed 15 December 2023

[22] GitHub – pyPESTO Problem. Accessed 15 December 2023