

UNIVERSITY OF CALIFORNIA SAN DIEGO

Uncertainty quantification methods improve mechanistic predictability in systems biology

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Engineering Sciences with a Specialization in Multi-Scale Biology

by

Nathaniel James Linden-Santangeli

Committee in charge:

Professor Boris M. Krämer, Chair
Professor Padmini Rangamani, Co-Chair
Professor Matthias Morzfeld
Professor Jin Zhang

2025

Copyright

Nathaniel James Linden-Santangeli, 2025

All rights reserved.

The Dissertation of Nathaniel James Linden-Santangeli is approved,
and it is acceptable in quality and form for publication on microfilm
and electronically.

University of California San Diego

2025

DEDICATION

To my late grandfather Dr. Donald Linden, for inspiring my curiosity and planting the seed to pursue science. To my parents, Kerri and Marc Linden, for their endless support and encouragement. To my sister, Samantha Linden, for challenging me to always be creative. And to my partner, Dr. Grace Linden-Santangeli, for being my best friend and pushing me to achieve my goals.

EPIGRAPH

All models are approximations. Essentially, all models are wrong, but some are useful. However, the approximate nature of the model must always be borne in mind.

—George Box

TABLE OF CONTENTS

Dissertation Approval Page	iii
Dedication	iv
Epigraph	iv
Table of Contents	vi
List of Figures	viii
List of Tables	xii
Acknowledgements	xiv
Vita	xvi
Abstract of the Dissertation	xviii
Chapter 1 Introduction—Why do we need uncertainty quantification in systems biology? . . .	1
1.1 Computational systems biology—the study biological systems with mathematical models	2
1.2 Recent experimental innovations enable near-real-time observations of intracellular signaling	6
1.3 Uncertainty in systems biology	6
1.4 Uncertainty quantification—A Bayesian perspective	8
1.5 Mathematical preliminaries	10
1.6 Summary of Thesis	10
Chapter 2 A Bayesian Framework for Parameter Estimation of Dynamical Models in Systems Biology	11
2.1 Introduction	11
2.2 Methods	16
2.3 Results	50
2.4 Discussion	76
2.5 Code availability statement	81
2.6 Acknowledgements	82
Chapter 3 Increasing Certainty in Systems Biology Models using Bayesian Multimodel Inference	83
3.1 Introduction	83
3.2 Materials and Methods	86
3.3 Results	100
3.4 Discussion	122
3.5 Data availability	126
3.6 Code availability	126
3.7 Acknowledgments	126

Chapter 4	Systems modeling of AMP-activated protein kinase signaling using uncertainty quantification reveals exercise-type and duration-dependent input integration . . .	127
4.1	Introduction	128
4.2	Methods and Materials	131
4.3	Results	138
Chapter 5	Miscellanea: Quantifying heterogeneity in predictions of beta-cell insulin sensitivity	163
5.1	Introduction	163
5.2	Methods	166
5.3	Oral glucose tolerance test: C-peptide kinetics	166
5.4	Results	172
5.5	Discussion	177
Chapter 6	Conclusion—An expanded UQ workflow and the future of UQ in systems biology .	178
6.1	Summary and future directions	178
6.2	Closing Remarks	184
Appendix A	Supplemental Materials for Chapter 2	186
A.1	Supplemental Text	186
A.2	Supplemental Tables	187
A.3	Supplemental Figures	190
Appendix B	Supplemental Materials for Chapter 3	201
B.1	Supplemental Tables	202
B.2	Supplemental Figures	207
B.3	Supplemental Files	218
Appendix C	Supplemental Materials for Chapter 4	219
C.1	Supplemental Text	219
C.2	Supplemental Figures	226
C.3	Supplemental Tables	226
Bibliography		245

LIST OF FIGURES

Figure 1.1:	Timescales of signaling phenomena investigated in this thesis.	3
Figure 2.1:	A comprehensive Bayesian parameter estimation and uncertainty quantification framework for dynamical models in systems biology.	17
Figure 2.2:	Examples of point estimates depicted on an arbitrary probability density function (black line).	47
Figure 2.3:	Parameter estimation for a simple two-state model.	52
Figure 2.4:	Parameter estimation for a simplified MAPK cascade that exhibits multistability.	57
Figure 2.5:	Varying levels of uncertainty in the parameters associated with the MAPK model impact steady state prediction.	59
Figure 2.6:	Parameter estimation results for the MAPK model in the limit cycle regime with data only sampled from the oscillations.	62
Figure 2.7:	Parameter estimation results for the MAPK model in the limit cycle regime with varied sampling strategies.	64
Figure 2.8:	Marginal posterior distributions for k_2, k_4, k_5, and k_6 highlight the effects of excluding one or both of the structural identifiability and global sensitivity analyses.	68
Figure 2.9:	Parameter estimation for a coupled kinase-phosphatase switch for long-term potentiation and long-term depression in neurons as a function of calcium input.	71
Figure 2.10:	Comprehensive parameter estimation and uncertainty quantification reveal failures to predict the correct long-term model behavior.	74
Figure 3.1:	Bayesian multimodel inference (MMI) accounts for model uncertainty by incorporating contributions from intracellular signaling models specified by the user.	101
Figure 3.2:	Set of ERK signaling models highlights model form uncertainty due to differences in biological assumptions, model complexity, and mathematical formulation.	105
Figure 3.3:	Multimodel inference reduces uncertainty in cytoplasmic ERK activity predictions.	106
Figure 3.4:	Bayesian multimodel inference is robust to perturbations in the set of plausible models for cytoplasmic ERK activity trajectory predictions.	109

Figure 3.5:	Bayesian multitmodel inference predictions of cytoplasmic ERK activity are robust to uncertainties due to decreasing data length and quality.	112
Figure 3.6:	Out-of-sample predictions of the EGF-ERK trajectory with models trained on dose-response data.	114
Figure 3.7:	Out-of-sample predictions of the EGF-ERK dose-response curve with models trained on trajectory data.	116
Figure 3.8:	Bayesian multitmodel inference suggests that subcellular differences in ERK activity depend on both Rap1 activity and ERK negative feedback.	120
Figure 4.1:	AMPK is activated by decreasing cellular energy availability that leads to falling intracellular ATP.	130
Figure 4.2:	Uncertainty Quantification workflow used for data-informed model development.	132
Figure 4.3:	Identifiability varies between mass action, Michaelis-Menten, and Hill-type models of a single enzyme-catalyzed reaction.	139
Figure 4.4:	UQ workflow selects models that treat AMP and ADP as nonessential activators of AMPK.	147
Figure 4.5:	AMPKAR sensor and pAMPK activity vary proportionally with input strength.	151
Figure 4.6:	Stimulus duration induces greater changes in AMPK activity than stimulus strength	152
Figure 4.7:	AMPK activity varies as a function of input frequency.	154
Figure 4.8:	AMPK is robust to frequency differences during resistance and running exercise stimuli during short time-scale muscle activation.	157
Figure 4.9:	AMPK activation during long time-scale exercise stimuli depends on stimulus frequency and duration.	159
Figure 5.1:	Population level predictions of β-cell sensitivity with the Hovorka 1998 dataset.	174
Figure 5.2:	Population level predictions of β-cell sensitivity with the Hall 2018 dataset. . . .	175
Figure 5.3:	Effects of OGTT duration and sampling strategy on predictions of β-cell sensitivity.	176
Figure 6.1:	Expanded uncertainty quantification workflow for systems biology enables data-informed modeling with quantified uncertainty.	181
Figure A.1:	Posterior distributions of $x_2(t)$ in the two-state model experiments.	190

Figure A.2:	Markov chains for the two-state model parameters at increasing noise levels in the measurement noise experiment, Fig 2.3B-C.	191
Figure A.3:	Markov chains for the two-state model parameters with decreasing samples in the measurement data sparsity experiment, Fig 2.3D-E.	192
Figure A.4:	Posterior distributions of $x_1(t)$ and $x_2(t)$ for the MAPK model.	192
Figure A.5:	Posterior distributions of $x_1(t)$, $x_2(t)$, and $x_3(t)$ for the MAPK model with equidistant and non-equidistant sampled data.	193
Figure A.6:	Histograms of limit cycle amplitude and period for the oscillating trajectories for the equidistant sampling (A) and non-equidistant sampling (B) data.	194
Figure A.7:	Two-dimensional scatter plots between all unique pairings of model parameters estimated from data with oscillations only.	194
Figure A.8:	Two-dimensional scatter plots between all unique pairings of model parameters estimated from data with equidistant sampling.	195
Figure A.9:	Two-dimensional scatter plots between all unique pairings of model parameters estimated from data with non-equidistant sampling.	196
Figure A.10:	Markov chains for the MAPK model parameters and UKF-MCMC noise covariances for bistability.	197
Figure A.11:	Markov chains for the MAPK model parameters and UKF-MCMC noise covariance parameters for limit cycle oscillations with data from oscillations only; corresponds to Fig 2.6 and Fig 2.7E.	198
Figure A.12:	Markov chains for the MAPK model parameters and UKF-MCMC noise covariance parameters for limit cycle oscillations estimated from data with equidistant sampling; corresponds to Fig 2.7C.	198
Figure A.13:	Markov chains for the MAPK model parameters and UKF-MCMC noise covariances for limit cycle oscillations estimated from data with non-equidistant sampling; corresponds to Fig 2.7D.	199
Figure A.14:	Markov chains for the synaptic plasticity model parameters, Fig 2.10.	200
Figure B.1:	Morris sensitivity analysis reveals the parameters of ten ERK signaling models that most strongly influence ERK activity.	208
Figure B.2:	Posterior densities cytoplasmic ERK activity trajectory predictions.	209

Figure B.3:	Multimodel inference reduces uncertainty in plasma membrane ERK activity predictions.	210
Figure B.4:	Multimodel inference of the EGF-ERK dose-response curve using synthetic training data.	211
Figure B.5:	Bayesian multimodel inference is robust to perturbations in the set of plausible models for plasma membrane ERK activity trajectory predictions.	212
Figure B.6:	Bayesian multimodel inference is robust to perturbations in the set of plausible models for EGF-ERK dose-response predictions.	213
Figure B.7:	Posterior predictive densities of cytoplasmic ERK activity with shortened training data.	214
Figure B.8:	Multimodel inference predictions of plasma membrane ERK activity are robust to uncertainties due to decreasing data length and quality.	215
Figure B.9:	Differences in ERK model parameters can drive sub-cellular location-specific differences in ERK activity.	216
Figure B.10:	Posterior predictive densities of ERK activity under different location-specific hypotheses.	217
Figure C.1:	Simulated energy stress with decreased glycolytic ATP production increases the AMP/ATP and ADP/ATP ratios.	226
Figure C.2:	Prior refinement and nominal value selection for Model 2.	227
Figure C.3:	Prior predictive simulation samples with the original (left) and refined (right) priors for all models.	228
Figure C.4:	Posterior predictions for all models with parameters estimated with only wild-type data.	229
Figure C.5:	Estimated marginal posterior (orange) and prior (black dashed) densities of models with mass action kinetics, Models 1, 3, and 5.	230
Figure C.6:	Estimated marginal posterior (orange) and prior (black dashed) densities of models with Michaelis-Menten kinetics, Models 2, 4, and 6.	231
Figure C.7:	Posterior predictions for all models with parameters estimated with data from all three conditions simultaneously.	232
Figure C.8:	Area under the curve as a function of stimulus duration.	232

LIST OF TABLES

Table 4.1:	Set of AMPK models vary in assumed reaction mechanisms.	144
Table A.1:	Two-compartment model [149] model parameters and relevant ranges.	187
Table A.2:	MAPK model parameters and relevant ranges from [150].	188
Table A.3:	Initial conditions used for simulations of the MAPK model from [150].	188
Table A.4:	Long-term potentiation/depression model parameters from [151] and ranges.	189
Table B.1:	Set of extracellular signal-regulated kinase (ERK) signaling models used for Bayesian multimodel inference.	202
Table B.2:	ODE solver and steady-state methods and corresponding hyperparameters for each ERK model.	202
Table B.3:	Sampling runtimes and sample counts for all Sequential Monte Carlo sampling.	203
Table B.4:	Model predictive performance (ELPD) and MMI weights of the top and bottom ten models of the 94 total compartment-specific parameter combinations. Additional models are not shown. Bold entries indicate weight greater than 0.01 by any MMI method.	204
Table B.4:	Continued: Model predictive performance (ELPD) and MMI weights of the top and bottom ten models of the 94 total compartment-specific parameter combinations. Additional models are not shown. Bold entries indicate weight greater than 0.01 by any MMI method.	205
Table B.4:	Continued: Model predictive performance (ELPD) and MMI weights of the top and bottom ten models of the 94 total compartment-specific parameter combinations. Additional models are not shown. Bold entries indicate weight greater than 0.01 by any MMI method.	206
Table B.4:	Continued: Model predictive performance (ELPD) and MMI weights of the top and bottom ten models of the 94 total compartment-specific parameter combinations. Additional models are not shown. Bold entries indicate weight greater than 0.01 by any MMI method.	207
Table C.1:	Prior parameters for the single enzyme-mediated reaction models shown in Figure 4.3.	233
Table C.2:	Identifiability of two-reaction models.	234

Table C.3:	Fluxes for the metabolism module (Module 1).	235
Table C.4:	Model parameters for metabolism module (Module 1).	236
Table C.5:	AMPK model parameters.	237
Table C.6:	Local identifiability.	238
Table C.7:	Sobol total sensitivity indices for Model 1.	239
Table C.8:	Sobol total sensitivity indices for Model 2.	240
Table C.9:	Sobol total sensitivity indices for Model 3.	241
Table C.10:	Sobol total sensitivity indices for Model 4.	242
Table C.11:	Sobol total sensitivity indices for Model 5.	243
Table C.12:	Sobol total sensitivity indices for Model 6.	244

ACKNOWLEDGEMENTS

First, I want to thank my advisors, Professor Padmini Rangamani and Professor Boris Krämer, for their endless support over the past five years. This thesis was an experiment, the first of a new collaboration and a new line of research for all of us. I am deeply grateful for your patience while I learned to navigate two new fields and for your guidance during this process. I am very happy with what we have achieved.

Next, I want to thank my doctoral committee members, Professor Matthias Morzfeld and Jin Zhang, for your support over the last few years. Matti, thank you for providing me a second home at UCSD. The community we have built at CaCAO has been extremely welcoming and it is always a joy to come down the hill for our meetings. Jin, thank you for your thoughtful collaboration and for always providing an experimental perspective to our work. Additionally, I would like to thank my past mentors, Professors Bing and Steve Brunton, and Professor Andre Berndt, for laying the foundations of my research training.

I want to thank the UCSD Interfaces Graduate Training Program for making my interdisciplinary training possible. In particular, to Professor Andrew McCulloch and Kelley Raymond for leading and coordinating such a valuable program. Additionally, I would like to thank the Sloan Scholars program and the Alfred P. Sloan Foundation for their support and for the strong community I had the opportunity to join. I would also like to thank Dra. Sinai Cota and Shana Slebiada for their support and leadership of the Sloan Program at UCSD.

I want to thank all of the members of the Rangamani and Krämer labs for their friendship, support and for teaching my many new things. I want to especially thank Emmet, Danny, Kyle, Mayte, Kyle, Maria, Chris, Opal, Nick, and Harsh. I couldn't have done this without you. Thank you all for making my time at UCSD so special.

I want to thank all of my collaborators, Professor Amit Majithia, Emily Kobayashi, Johannes Reiner, Navid Zobeiry, Reza Vaziri, Jin Zhang, and Anne Lyons. Additionally, I want to thank the

funding sources that made this work possible, including the National Institute of Biomedical Imaging and Bioengineering, the Alfred P. Sloan Foundation, and the Wu Tsai Human Performance Alliance.

Finally, I want to thank my family, Marc, Kerri and Sami Linden, for your endless support. Most importantly, I want to thank my partner, Dr. Grace Linden-Santangeli, for supporting me through this journey. I couldn't have done it without you.

Chapter 2, in full, is a modified reprint of the material as it appears in Bayesian parameter estimation for dynamical models in systems biology. Nathaniel Linden, Boris Krämer, and Padmini Rangamani, PLoS Computational Biology, 2022. The dissertation author was the primary researcher and author of this paper.

Chapter 3, in full, is a modified reprint of the material as it appears in Increasing certainty in systems biology models using Bayesian multimodel inference. Nathaniel Linden-Santangeli, Jin Zhang, Boris Krämer, and Padmini Rangamani, *In review*, 2025.

Chapter 4, in full, is a modified reprint of the material as it appears in Systems modeling and uncertainty quantification of AMP-activated protein kinase signaling. Nathaniel Linden-Santangeli, Jin Zhang, Boris Krämer, and Padmini Rangamani, *Submitted*, 2025.

VITA

2017-2019	Undergraduate Research Assistant Berndt Lab, Department of Bioengineering, University of Washington
2019-2020	Washington Research Foundation Innovation Undergraduate Fellow in Neuro-engineering Brunton Lab, Department of Biology, University of Washington
2020	Bachelor of Science in Bioengineering University of Washington
2020	Minor in Applied Mathematics University of Washington
2020-2025	Graduate Research Fellow Rangamani and Kramer Labs, University of California San Diego
2023	Master of Science in Engineering Sciences (Mechanical Engineering) University of California San Diego
2025	Doctor of Philosophy in Engineering Sciences with a Specialization in Multiscale Biology University of California San Diego

PUBLICATIONS

Linden, N. J.; Kramer, B.; Rangamani, P. *Bayesian Parameter Estimation for Dynamical Models in Systems Biology*. PLoS Comput. Biol. 2022, 18 (10), e1010651.

Reiner, J.; **Linden, N.**; Vaziri, R.; Zobeiry, N.; Kramer, B. *Bayesian Parameter Estimation for the Inclusion of Uncertainty in Progressive Damage Simulation of Composites*. Compos. Struct. 2023. 321, 117257.

- Contributed Bayesian parameter estimation and global sensitivity analysis.

Qiao L.*; Khalilimeybodi A.*; **Linden-Santangeli N. J.**, Rangamani P. *The evolution of systems biology and systems medicine: From mechanistic models to uncertainty quantification* Annual Reviews in Biomedical Engineering. 2024.

- Contributed to sections on uncertainty quantification.

Linden-Santangeli N.J., Zhang J., Kramer B, and Rangamani P. *Increasing certainty in systems biology models using Bayesian multimodel inference*. In Review: Nature Communications. 2025.

Kobayashi E., **Linden-Santangeli N.J.**, Chan N., Toomey C.B., Diabetes Prevention Program Research Group, Mudaliar S., Temprosa M., Edelstein S., Goyal R., Rangamani P., Majithia A.R.. *Longitudinal metabolic trajectories in Diabetes Prevention Program participants reveal subgroups with varying micro and macrovascular complication risk*. In Review: Diabetes Care. 2025.

- Contributed to the development of the dimensionality reduction approach using tensor decomposition and the overall study design.

Linden-Santangeli N.J., Zhang J., Kramer B, and Rangamani P. *Systems modeling of AMP-activated protein kinase signaling using uncertainty quantification reveals exercise-type and duration-dependent input integration*. In Preparation. 2025.

ABSTRACT OF THE DISSERTATION

Uncertainty quantification methods improve mechanistic predictability in systems biology

by

Nathaniel James Linden-Santangeli

Doctor of Philosophy in Engineering Sciences with a Specialization in Multi-Scale Biology

University of California San Diego, 2025

Professor Boris M. Krämer, Chair
Professor Padmini Rangamani, Co-Chair

Mathematical models are indispensable for understanding the architecture and behavior of biological systems using engineering science approaches. Systems biology, analogous to reaction network theory in chemical engineering, enables researchers to investigate the dynamics and behaviors of biological systems beyond what is experimentally feasible. However, despite the successful application of systems biology to studying numerous biological systems, three key challenges limit the predictability of models, including epistemic uncertainty, uncertain model parameters, and incorporation of noisy and sparse experimental data. In this thesis, we propose that uncertainty quantification (UQ)—the field of computational

mathematics that models uncertainties—can address these challenges. Recent advances in fluorescent microscopy and protein engineering, which enable near-real-time biological measurements, provide data with the resolution necessary to apply UQ methods.

First, we address the estimation of model parameters from noisy and sparse experimental data. To investigate this, we examined the effects of data noise, data quality, parameter identifiability, and parameter sensitivity on our ability to estimate model parameters. In doing so, we developed a framework for parameter estimation that leverages identifiability and sensitivity analysis to pre-process the parameter space and then applies Bayesian inference to learn model parameters.

Second, we address the problem of epistemic uncertainty, which commonly arises because our understanding of biology is still evolving. This often leads to the availability of multiple models of the same system. To make predictions with multiple models, we applied a methodology called Bayesian multimodel inference, which combines the specified models to yield a single consensus estimator. We found that multimodel inference increases the certainty of systems biology models by reducing predictive variance, increasing robustness to poor-quality data, and selecting models that best explain the data.

Third, we demonstrate how UQ enables the modeling of a system for which we have experimental data, but few published models exist. Specifically, we leveraged high-quality experimental measurements from the Jin Zhang lab (UCSD Dept. of Pharmacology) and applied the methods established in the first two sections to develop a new model of the AMP-activated protein kinase signaling pathway.

Through these interdisciplinary studies, we find that UQ addresses key challenges in systems biology, enabling data-informed modeling. Together, these methods constitute a broadly applicable workflow for UQ in systems biology and computational modeling applications. Uncertainty quantification has the potential to drive new biological discoveries by enhancing mechanistic predictability through principled incorporation of experimental data and systematic accounting of uncertainties.

Chapter 1

Introduction—Why do we need uncertainty quantification in systems biology?

Engineering science has a long history of applying mathematical and computational models to study complex physical systems. In this thesis, I apply an engineering sciences approach to studying a complex system in biology, that is intracellular signaling. Mathematical and computational modeling has long been an essential component of the systems biology toolkit to comprehend and explain observed biological phenomena [1]–[3]. However, as experimental techniques become ever more advanced and we enter into a world of *big-data biology*, mathematical biologists face the challenge of integrating this data into their models [4]. This thesis explores a workflow to begin to address this challenge. Here, we take the viewpoint of data-informed modeling, where we combine mechanistic representations with available data. We leverage tools from uncertainty quantification, a field of computational science and engineering [5], to enable data-informed systems biology. The work is further enabled by the availability of high-quality data, which has only recently been made available by ongoing developments in microscopy, biosensors, and

molecular biology [6]–[8]. In this chapter, we first introduce computational systems biology models of intracellular signaling in Section 1.1 and then briefly review modern experimental approaches to study intracellular signaling in Section 1.2. Next, we discuss uncertainty in systems biology in Section 1.3 and introduce the Bayesian perspective to UQ in Section 1.4. We end with a brief note on mathematical preliminaries in Section 1.5 and summarize the remainder of this thesis in Section 1.6.

1.1 Computational systems biology—the study biological systems with mathematical models

Systems biology is the study of the structure and function of biological systems at the level of an entire biological system [1], [9]. Importantly, experimental systems biology approaches probe the entire system while it is functioning and have revealed the complex interactions and regulatory networks underlying cellular and tissue functions [10], [11]. While the system-level approach has been crucial to understanding how organisms function, it also complicates our ability to determine a mechanism for observed phenomena [1], [10], [12]. To address this challenge, computational models enable researchers to hypothesize a mechanism of the observed phenomena and run *in silico* experiments to validate the proposed hypothesis against experimental data [13]. System biology models leverage a variety of mathematical and computational frameworks, including mechanistic ordinary differential equations (ODEs), rule-based ODEs, delay-differential equations, stochastic differential equations, agent-based models, and partial differential equations [14]–[17]. Often these equations include functions that are nonlinear in both the state variables and parameters to capture complex biochemical interactions [2]. After a proposed model, and with it the proposed mechanism, have been validated against experimental data, that model can then be used to run additional simulations which generate experimentally testable predictions [1], [10], [13].

Computational models of biological systems span a wide range of biological systems and therefore a range of spatial and temporal scales [14]. Because of the ability to interpret and explain experimental findings, models have been widely used in the field of systems biology to reveal the underlying mechanisms of how cells execute various biological functions [1], [10], [18]–[20].

1.1.1 Intracellular signal transduction processes information at multiple time-scales

In this work, we focus on intracellular signal transduction systems that enable cells to sense, process, and react to external cues [21], [22]. Cells contain extracellular-facing, membrane-bound receptors that detect extracellular biochemical signals by binding a matching ligand, which induces a biochemical change in the receptor [21]. Ligand-binding induces a cascade of intracellular biochemical reactions that act to amplify [23]–[25], process [26]–[29], and respond to signals [30]–[32]. These cascades of biochemical reactions inherently span multiple temporal scales with individual reactions occurring over micro-to-millisecond timescales and pathway responses over second-to-hour timescales [33]–[36]. In this work, we model signaling phenomena that span second timescales (CaMKK2 and AMPK signaling) to hour timescales (ERK and AMPK; Figure 1.1). Mathematical models allow us to simulate multi-scale systems and probe components that evolve across scales.

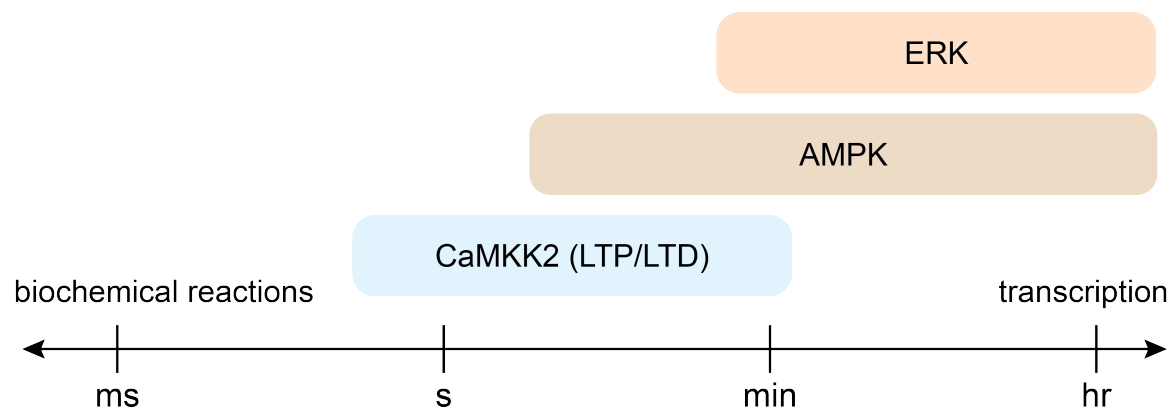
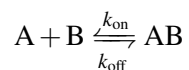


Figure 1.1: We investigate ERK signaling in Chapters 2 and 3, CaMKK2 signaling in 2 and 4, and AMPK signaling in Chapter 4.

Researchers have developed models for various well-known signaling pathways. For example, signaling pathways that regulate cell proliferation, apoptosis, inflammation response, metastasis, differentiation [37], including, EGFR [38], NF κ B [39], [40], Ca²⁺ [41], cAMP [42], PI3K [43], ERK [38], [44], [45], YAP/TAZ [46]–[48] pathways. Models have helped elucidate mechanisms in pathways that regulate cellular metabolism, including AMPK [49]–[51], mTOR [51], and insulin signaling [24]. Additionally, mathematical models have been used to study aberrant signaling in a variety of disease states, including cancer [52], [53] and diabetes [54]–[56]. Finally, models have been used to understand the coupling between cell mechanics and biochemistry [57] with applications in mechanosensing and mechanotransduction [48] and the interplay between signaling and cell shape [58]–[60]. With these models, researchers have revealed the underlying mechanisms that give rise to physiological and pathological intracellular signaling.

1.1.2 A continuum perspective to modeling biochemical reactions

The models we focus on in this work aim to mathematically represent the mechanisms of biochemical reactions within a signaling network. Mechanistic models leverage parametric nonlinear ODEs to describe biochemical reactions through time [61], [62]. We will use an illustrative example to introduce the basic concepts of constructing a mechanistic model. Consider the elementary reaction where species A binds reversibly to species B with a rate of k_{on} to form complex AB, and AB dissociates with a rate of k_{off} . The reaction



describes the formation of complex AB. We assume that species A, B, and AB are abundant enough to be represented using concentrations (quantity/volume) as opposed to individual copy numbers (quantity). This allows us to take a continuum approach, which assumes that the rates of change vary continuously with time and leverages ordinary differential equations. To build a model for this example, we can assume

the Law of Mass Action ¹, to prescribe the flux $k_{\text{on}}[A][B]$ for the forward reaction and the flux $k_{\text{off}}[AB]$ for the reverse ². The model is then defined as

$$\frac{d[A]}{dt} = -k_{\text{on}}[A][B] + k_{\text{off}}[AB] \quad (1.1a)$$

$$\frac{d[B]}{dt} = -k_{\text{on}}[A][B] + k_{\text{off}}[AB] \quad (1.1b)$$

$$\frac{d[AB]}{dt} = k_{\text{on}}[A][B] - k_{\text{off}}[AB]. \quad (1.1c)$$

Applying such strategies to larger reaction networks allows one to develop models of various biochemical phenomena. We refer the reader to [1], [2], [61]–[63] for further information on model construction.

We typically compose models of individual reactions, such as Eq. (1.1a), to mathematically represent an entire signaling pathway. In general, systems biology models can be defined as

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{f}(\mathbf{x}(t); \boldsymbol{\theta}), \quad (1.2)$$

where $\mathbf{x} \in \mathbb{R}^d$ is the vector of states and $\boldsymbol{\theta} \in \mathbb{R}^p$ is the vector of model parameter. The states $\mathbf{x}$ often represent physical quantities such as chemical concentrations or copy numbers, so they are restricted to $\mathbb{R}_{\geq 0}^d$. The function $\mathbf{f}(\cdot; \cdot) : \mathbb{R}^d \times \mathbb{R}^p \rightarrow \mathbb{R}^d$ is the vector field that governs the dynamics of the states. Due to the nature of possible reaction mechanisms, $\mathbf{f}$ is often nonlinear in both the states and model parameters and often contains polynomial and rational nonlinearities; see, e.g., Eq (1.1a). Additional examples include Michaelis-Menten and Hill-type kinetics, which both contain rational nonlinearities [64].

¹The Law of Mass-Action states that the rate of a chemical reaction is proportional to the product of the concentrations of the reactants.

²The notation $[S]$ denotes the concentration of species S

1.2 Recent experimental innovations enable near-real-time observations of intracellular signaling

Intracellular signaling phenomena involve numerous coupled biochemical events that evolve over a range of timescales and occur across both cellular and subcellular scales. Traditional experimental approaches for studying intracellular biochemistry sacrifice resolution across all three axes [7]. For example, the Western blot allows researchers to quantify the relative abundance of a target protein (e.g., a phosphorylated protein in a signaling pathway), but it requires large quantities of substrate from an entire batch of cells and is extremely time-consuming. To that end, the best observations using traditional approaches can be made on the order of every 10 minutes of a reaction, only capture a few target molecules, and do not have single-cell resolution. Our ability to observe more biochemical species has vastly improved over the past few decades with the rapid advancement of mass-spec and sequencing methods for -omics [65]. Recently, these methods have even been extended to single-cell resolutions [66]. However, -omics-based approaches typically cannot observe cell signaling in real-time [7]. Concurrent advancements in protein engineering and molecular biology stemming from the discovery of green fluorescent protein [67], have led to the introduction of modern fluorescent protein biosensors [68]. These tools can be expressed within a cell and used to quantify signaling events in near-real-time at the single-cell [6], [7], [69], [70] and subcellular resolutions [8], [71], [72]. This high-resolution data provides the temporal resolution necessary to combine with mathematical models [73], [74].

1.3 Uncertainty in systems biology

Handling uncertainties is a key challenge of systems biology modeling [75], [76]. There are many sources of uncertainty in mechanistic modeling of signal transduction, including the model structure itself,

the distribution of the model parameters, and the quality of the data used for model calibration [75]–[77]. Uncertainty in the model structure arises because models are built from incomplete knowledge of a biological system using a set of assumptions. As discussed above, reaction fluxes for many biochemical reactions can be established in terms of different choices of rate equations [61]–[63], but they could also be formulated using alternative phenomenological representations such as generalized Hill equations [78]. This uncertainty in the model equations is known as model form or topological uncertainty and can drive significant uncertainty in model predictions [79]–[82]. Once a model form has been established, one must determine suitable model parameters, including the kinetic rate constants, for various flux terms [75], [83]. While model parameters represent biological quantities such as reaction rates, direct experimental measurement of these parameters often occurs in isolated reaction systems and does not capture the complexity of the entire network of reactions represented by the models. Therefore, we must rely on methods to infer and estimate parameter values from data. This is complicated by the strong nonlinearity in the states and parameters, as well as the large parameter count relative to the number of states, which collectively result in parameter spaces that are difficult to explore with estimation algorithms. Estimation and identification of the kinetic parameters associated with these reactions and quantifying the associated uncertainty requires selecting a statistical model and then learning the distribution of these parameters from available data [5], [83]–[87]. However, addressing model form uncertainty and parameter estimation is complicated by the *noisy, temporally sparse, and incomplete nature of data* found in systems biology (few or select readouts due to experimental limitations) that introduce uncertainties in the biological parameters [87]–[90]. Uncertainty exists in experimental measurements for several reasons, including physical limitations such as the diffraction limit of light [91], inability to directly measure concentrations of biological species [6], [7], and intrinsic and extrinsic noise [92]. Reducing this uncertainty can increase trustworthiness in model predictions and help to drive model-based predictions.

1.4 Uncertainty quantification—A Bayesian perspective

Uncertainty quantification determines how model assumptions, inferred quantities, and data uncertainties impact model predictions. UQ methods and applications embrace the reality that model predictions and experimental observations are inherently uncertain and aim to include more than just deterministic equations with point estimates. Uncertainties are naturally represented using probabilities [84]. Therefore, UQ can be characterized by the bridging of computational mathematics with probability and statistics [5]. UQ was originally developed for traditional engineering and national security applications, where models are heavily relied upon for decision making [5]. A complete handling of UQ goes far beyond this thesis, but some key applications include inverse problems [93]–[97], risk estimation [98], rare events modeling [99], sensitivity analysis [100], [101], model form uncertainty [79], [102], and surrogate/reduced-order modeling [79], [103]–[107]. Due to the nature of systems biology models and the problems that arise, we focus on three of these UQ tasks: inverse problems (parameter estimation) and sensitivity analysis.

A brief introduction to inference: Frequentist vs. Bayes Here, we make an effort to describe the fundamentals of statistical inference for a general model $\hat{y} = g(\theta)$ with model parameter θ and available data $\mathcal{Y}$. Statistical inference aims to solve the inverse problem that answers: *Which value (or values) of θ best recapitulate the data?* [84], [93]. The frequentist approach treats this as an optimization problem, which is defined as

$$\hat{\theta} = \arg \min_{\theta} J(\theta, \mathcal{Y}), \quad (1.3)$$

where $J(\theta, \mathcal{Y})$ is an objective function that measures the discrepancy between model predictions and data, and $\hat{\theta}$ is the best-fit point estimate for the unknown parameters. For example, the objective function could be the sum of squared errors, which is defined as $J(\theta, \mathcal{Y}) := \sum_i^N (g(\theta) - y_i)^2$, for data $\mathcal{Y} = (y_1, \dots, y_N)$.

Alternatively, Bayesian inference aims to learn a probability density for the unknown parameters conditioned on the data, $p(\theta|\mathcal{Y})$, called the posterior density. Bayes' Rule allows one to equate the posterior density to a product of the prior density $p(\theta)$ and the likelihood function $p(\mathcal{Y}|\theta)$ normalized by the evidence $p(\mathcal{Y})$, and is defined as

$$p(\theta|\mathcal{Y}) := \frac{p(\theta)p(\mathcal{Y}|\theta)}{p(\mathcal{Y})}. \quad (1.4)$$

Here, the prior density encodes assumptions about the parameters before observing any data, and the likelihood function measures the probability of correctly predicting data. The evidence is often intractable and is therefore dropped, replacing equality for proportionality in Eq (1.4). There are several excellent texts on Bayesian inference that provide in-depth theory beyond what is covered here, including [108]. Importantly, Bayesian methods are always subjective, with beliefs encoded in the prior and likelihood densities; these beliefs will only be overcome in the limit of infinite data [108].

A key difference between traditional modeling and modeling with UQ is how unknown quantities are handled and inferred from data. More traditionally, inference is approached using frequentist statistics, where unknowns are thought to have a single true value that we aim to infer from data [5]. Here, inference is typically posed as an optimization problem where the objective is to find the values for unknown parameters that minimize the discrepancy between model predictions and data. In the frequentist setting, it is possible to estimate predictive uncertainties using quantities such as confidence intervals. On the contrary, UQ comes more naturally when unknown quantities are treated as random variables where the goal of inference is to characterize their probability densities conditioned on data [84]. From the Bayesian perspective, the resulting probability density conveys all of the necessary information to quantify predictive uncertainty. In the remainder of this section, we briefly introduce both points of view.

1.5 Mathematical preliminaries

This work assumes the reader has a working knowledge of differential equations and dynamical systems, which are typically covered in a standard undergraduate degree in engineering or applied mathematics. The reader should be familiar with numerical schemes for solving nonlinear differential equations (for reference, see [109]). Additionally, the reader should be comfortable with concepts in probability and statistics, including random variables and Monte Carlo estimation. There are countless books on these topics; however, I found the following resources particularly valuable [84], [110]–[112]. A strong foundation in a scripting language with a strong statistics community, such as R, Python, or Julia, is essential.

1.6 Summary of Thesis

This thesis addresses uncertainties in biology by answering three questions. First in Chapter 2, we ask: *Given a system of equations, how do we best estimate uncertain model parameters from available data?* Next, in Chapter 3, we ask: *How do we make predictions if we have multiple models of the same system?* Finally, in Chapter 4, we ask: *How do we build a new model for a system where we have experimental data, but few published models exist?* We also highlight how UQ methods can be applied beyond systems biology to diabetes care (Chapter 5).

Chapter 2

A Bayesian Framework for Parameter Estimation of Dynamical Models in Systems Biology

2.1 Introduction

Mathematical modeling is an integral part of systems biology; indeed, the use of approaches from dynamical systems analyses resulted in a paradigmatic shift in our understanding of biochemical signal transduction and enabled the identification of the emergent properties of a signaling network [29], [113]–[115]. Additionally, mathematical models allow us to investigate the dynamics of biological systems beyond what is experimentally possible [116]–[119]. A classical approach to modeling the dynamics of signal transduction is the use of systems of ordinary differential equations (ODEs) [61], [62]. Often these equations include nonlinear functions to capture complex biochemical interactions using Michaelis-Menten

kinetics and Hill functions for cooperative binding [2]. One of the ongoing challenges in developing and constraining predictive models of signal transduction has been the estimation and identification of the kinetic parameters associated with these reactions and quantifying the associated uncertainty [85]–[87]. The use of rigorous, quantitative approaches to estimate kinetic parameters and their uncertainties is in its early stages in systems biology [83], [85], [120]–[122] even though such methods are far more prevalent in the greater computational science community under the field of uncertainty quantification (UQ) [5], [93], [123], [124].

There are many sources of uncertainty in dynamical systems modeling of signal transduction, including the model structure itself, the values of model parameters, and the quality of the data used for model calibration. Uncertainty in the model equations, known as model form or topological uncertainty [79]–[82] often arises during model development. However, the reaction fluxes for many biochemical reactions (ODE formulations) can be established in terms of classical rate equations [2], [61], [62]. The more significant challenge is establishing suitable model parameters, including the kinetic rate constants, for various flux terms [83]. Direct measurement of these parameters often occurs in isolated reaction systems and does not capture the complexity of the entire network of reactions represented by the models. Estimating these biological parameters and identifying any remaining uncertainties requires selecting a statistical model and then learning the distribution of these parameters from available data [5], [84]. From this viewpoint, the biological parameters are random variables that either have parametric or nonparametric distributions. However, parameter estimation is complicated by the noisy, sparse (few time points), and incomplete nature of data found in systems biology (few or select readouts due to experimental limitations) that introduce uncertainties in the biological parameters [87]–[90], [125]. In the face of these complicating factors, there is a need for statistical modeling of parameters that enables uncertainty quantification.

A comprehensive parameter estimation and UQ framework should consider the impact of structural

parameter identifiability and parameter sensitivity [83], [89], [126]–[128]. Structural parameter identifiability analysis reveals which of the parameters can be estimated given a specific dynamical systems model and a set of measurable outputs [88], [89], [129], [130]. A parameter is globally structurally identifiable if there is only one unique model output for each value of that parameter [89]. Parameters that do not meet this criterion are deemed structurally nonidentifiable and cannot successfully be estimated from the specified model outputs. Structural nonidentifiabilities can arise due to complex nonlinear equations and incomplete experimental data that only measures a subset of the system’s states. Additionally, parametric sensitivity analysis [5], [131], [132] quantifies how sensitive a model output is to variations in the model parameters. Gutenkunst *et al.* [88] found that most models in systems biology contain parameters with a wide range of sensitivities, which they termed ‘sloppy’. Despite this challenge, sensitivity analysis helps rank the set of identifiable parameters by their contributions to specified model outputs, [133] as was done in Mortlock *et al.* [134] for the prolactin-mediated JAK-STAT signaling pathway. This analysis enabled them to select a subset of 33 out of 60 total parameters that significantly contribute to variations in the model outputs.

Commonly used methods to estimate parameters for systems biology models include frequentist [121] and Bayesian approaches [122]. In the frequentist setting, parameter estimation is formulated as an optimization problem, and the solution to the parameter estimation problem is the set of parameters that best recapitulates the data [5]. Additionally, frequentist approaches quantify uncertainty via estimated confidence intervals around the optimal parameters [5], [135], [136]. In contrast, in the Bayesian perspective, parameters are assumed to be random variables whose unknown probability distributions, called *posterior distributions*, quantify the probability of assuming any value in the parameter space [5], [84], [93]. The advantage of Bayesian approaches comes from their ability to characterize the entire posterior distribution and quantify the uncertainty in parameter estimates via credible intervals [84]. Many methods have been

developed for Bayesian parameter estimation [122], [137]–[143] that all aim to characterize the posterior distribution, by leveraging Bayes’ rule [5], [84]. For example, Mortlock et al. [134] successfully used Bayesian estimation to study the uncertainty in the model predictions and assess the statistical significance of their modeling results.

Despite the successes of Bayesian parameter estimation in systems biology [83], [85], [120], [122], failure to account for all sources of uncertainty in a model can significantly inhibit parameter estimation and uncertainty quantification [82], [83], [88]. Thus, a comprehensive framework for UQ in systems biology should include rigorous accounting of uncertainties in the model structure, nonidentifiable parameters, mixed parameter sensitivities, and noisy, sparse, or incomplete experimental data. While identifiability and sensitivity analyses are typically performed prior to parameter estimation [83], [89], accounting for model form uncertainty requires us to consider a stochastic model instead of a deterministic one [79], [82], [90], [144]. One promising approach to account for model form uncertainty is the Unscented Kalman filter Markov chain Monte Carlo (UKF-MCMC) method [82], [145]. This method includes statistical models for noisy data and model form uncertainty simultaneously; however, it has not been adapted for dealing with the unique challenges in system biology. Similarly, the parameter estimation and model selection method in [90] accounts for model form uncertainty with extended Kalman filtering but does not provide complete uncertainty estimates because it takes a frequentist approach for parameter estimation. Thus, there is a need for a framework that combines structural identifiability analysis, global sensitivity analysis, a statistical model for data and model form uncertainty, and Bayesian parameter estimation for comprehensive UQ of dynamical models in systems biology.

The novelty of this work lies in the comprehensive workflow we have developed to advance the current state of the art for uncertainty quantification in systems biology. Additionally, the proposed framework accounts for uncertainty in the data and model structure by building on the Bayesian framework

and the UKF-MCMC method [82], [145]. Our proposed workflow begins with structural identifiability analysis [130], [146] and global sensitivity analysis [131], [147] to eliminate parameters that cannot be learned from the measurements or that are not affecting the model outputs and then extends UKF-MCMC for systems biology by leveraging the constrained interval unscented Kalman filter (CIUKF) [148]. Taken together, each of these steps quantitatively addresses uncertainties encountered during model development and calibration to improve predictive modeling in systems biology.

The remainder of this paper details our comprehensive workflow for uncertainty quantification in systems biology and presents several examples to highlight this analysis. First, in Section 4.2 we introduce the proposed framework and provide the mathematical details needed to understand and apply the approach. Next, in Section 2.3 we apply this framework to three systems biology models of increasing complexity, including a simple two-state model [149], a model of the core mitogen-activated protein kinase (MAPK) signaling pathway [150], and a phenomenological model of synaptic plasticity to capture long-term potentiation/depression [151]. We found that even in simple models, estimation of parameters depends on the level of data noise and data sparsity. Finally, the framework enables uncertainty quantification for model structures that include non-linearities and multistability. In all of these cases, we leverage identifiability and sensitivity analyses to narrow the subset of parameters for estimation and then use Bayesian estimation to determine the role of model structure in parameter estimation. These results establish an uncertainty quantification-focused approach to systems biology that can enable rigorous parameter estimation and analysis. Lastly, in Section 4.3.5 we discuss these findings in the context of the three examples, address challenges in applying Bayesian methods, and provide directions for future UQ in systems biology.

2.2 Methods

This section describes the technical details of the comprehensive framework for uncertainty quantification (see Fig 2.1) proposed in Section 2.2.1. Next, Section 2.2.2 introduces dynamical systems biology models and the parameter estimation problem. Sections 2.2.3 and 2.2.4 respectively overview structural identifiability and global sensitivity analyses to reduce the dimension of the parameter space. Then Section 2.2.5 introduces Bayesian estimation, Section 2.2.6 outlines the CIUKF-MCMC algorithm, and Section 2.2.7 describes the constrained interval unscented Kalman filter in detail. Following this, Section 2.2.8 discusses how to construct prior distributions and Section 2.2.9 details Markov chain Monte Carlo sampling. Lastly, Section 2.2.10 discusses output uncertainty propagation with ensemble modeling, Section 2.2.11 highlights choosing point estimators for the parameters, Section 2.2.12 delineates synthetic data generation and Section 2.2.13 outlines limit cycle analysis.

2.2.1 Framework for comprehensive uncertainty quantification for dynamical models in systems biology

This section previews the proposed comprehensive framework for parameter estimation and uncertainty quantification of dynamical models in systems biology. Figure 2.1 outlines the framework and its components, which are then described in much more detail in the subsequent sections. The proposed framework follows three main steps. First, we assume that dynamical models of intracellular signal transduction (Fig 2.1A1) use classical biochemical rate laws, such as mass action kinetics, Michaelis-Menten kinetics, and Hill functions [2], [61], [62] (Fig 2.1A2). The key challenge to applying these models is estimating the associated parameters, such as the rate constants k and $V_{\max}$, equilibrium coefficients K_m and K_A , and Hill coefficients n in Fig 2.1A2, from available experimental data (Fig 2.1A4). The comprehensive framework uses Bayesian inference to estimate a statistical model (a probability distribution;

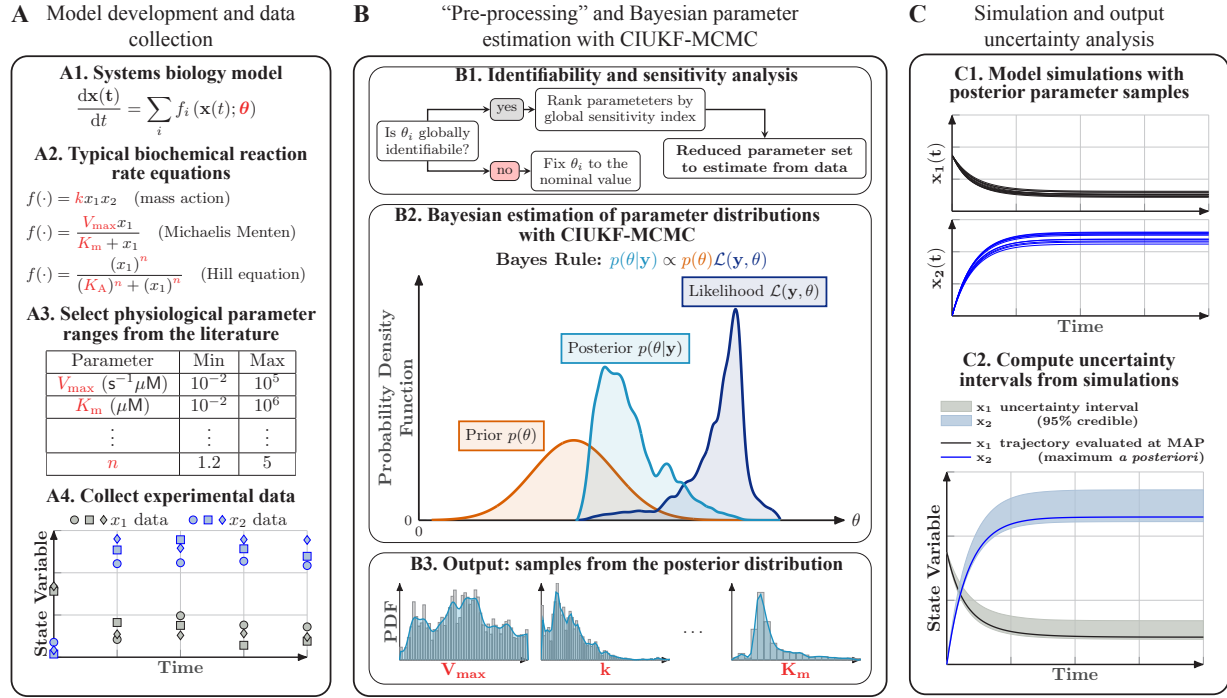


Figure 2.1: (A) Model development in systems biology begins with model construction and data collection. Dynamical models in systems biology typically involve a system of ODEs that capture the dynamics of the concentrations of different chemical species in the system (A1). The reaction rates associated with these concentration changes are usually mass action, Michaelis Menten kinetics, or cooperative kinetics represented by the Hill equation (A2). The free parameters in these models include kinetic rate constants, e.g. k , $V_{\max}$, equilibrium constants, e.g. K_m , K_A , and Hill coefficients, e.g. n . These parameters are first constrained by *best guess* values based on physiological ranges and typical values of model parameters from the literature (A3). Finally, the model needs experimental data for validation; this data can either be from published work or new experiments. (B) Parameter preprocessing and Bayesian parameter estimation with the CIUKF-MCMC algorithm. First structural identifiability and global sensitivity analyses on the entire parameter set reduce the set of free parameters that can be estimated (B1). Next, we perform Bayesian parameter estimation for this reduced set of parameters to learn their posterior distributions. The posterior distribution is the parameter distribution conditioned on the data (B2). Bayes’ rule relates the posterior distribution to the product of the prior distribution and the likelihood function (which requires simulating the model) measures the misfit between predictions and the data. A state-constrained Unscented Kalman filter approximates the likelihood function to account for uncertainty in the model equations. Although Bayes’ rule provides a means to evaluate the posterior distribution at specific points in the parameter space, we use Markov chain Monte Carlo (MCMC) sampling (B3) to characterize the entire distribution. (C) The posterior distributions enable uncertainty analysis of model outputs through ensemble simulation. We perform simulations using the posterior parameter samples to propagate parameter uncertainty through the model (C1). Statistical analysis of the ensemble enables us to compute uncertainty intervals and study various system behaviors, for example, the statistics of the steady state values (C2).

see Fig 2.1B) for the model parameters given a set of noisy measurement data and a specific model form.

We argue that identifiability and sensitivity analysis are necessary steps to perform before parameter

estimation (Fig 2.1B1). To eliminate uncertainty due to nonidentifiable parameters, we perform global structural identifiability analysis using the Structural Identifiability Analyzer (SIAN) [130], [146] (see Section 2.2.3 for details). The nonidentifiable parameters are fixed to their nominal values from the literature or based on their physiological ranges. Next, variance-based global sensitivity analysis [5], [131] is performed to rank the identifiable parameters in order of their contributions to the variance of the model outputs (see section 2.2.4 for details). A subset of the identifiable parameters with the largest sensitivity indices is selected for parameter estimation. The remaining model parameters are then fixed to their nominal values in the same fashion as nonidentifiable parameters.

Bayesian parameter estimation completely characterizes uncertainty in the model parameters by estimating a nonparametric statistical model. Bayes' rule (see Fig 2.1B2) provides the *best guess* distribution, called the *posterior distribution*, for the parameters starting from an initial guess (the prior distribution) that is transformed by the available experimental data with the likelihood function. The likelihood function measures the mismatch between the data and the model predictions and returns higher probabilities for parameter sets that produce outputs *closer* to the data. We use the CIUKF-MCMC algorithm [82], [145] to approximate the likelihood function and account for uncertainty in the model formulation, data, and parameters. Markov chain Monte Carlo sampling, with either delayed rejection adaptive Metropolis [152] or the affine invariant ensemble sampler [153], generates a set of samples that represents the posterior distribution (Fig 2.1B3).

Lastly, we leverage the posterior distribution to quantify how uncertainty in the model parameters affects uncertainty in the model predictions (Fig 2.1C). An ensemble simulation with the parameter samples generates sets of trajectories (see Fig 2.1C1) that capture the uncertainty in the predicted dynamics. Computing uncertainty intervals such as the 95% credible intervals presented in Fig 2.1C2 provides a visualization of this uncertainty. Notably, credible intervals are different from confidence intervals

because credible intervals capture a specified percentage of the samples whereas confidence intervals are random variables that capture regions where estimators will lie at a specified probability level [84] (see Fig 2.2 for an example of a credible interval). Additionally, statistical analysis of the ensemble enables quantitative analysis of computational modeling results in the same way that running multiple experimental trials enables analysis of experimental results. Before discussing the methods that enable comprehensive uncertainty quantification in the following sections, the next section formally introduces parameter estimation for dynamical models in systems biology.

2.2.2 Parameter estimation for systems biology models in the form of partially-observed systems of ordinary differential equations

We consider nonlinear ordinary differential equation models of the form

$$\frac{d\mathbf{x}(t)}{dt} = \sum_i f_i(\mathbf{x}(t); \theta_f) \quad (2.1)$$

$$\mathbf{y}(t) = h(\mathbf{x}(t); \theta_h) + \eta(t), \quad \eta(t) \sim \mathcal{N}(0, \Gamma(\theta_\Gamma)), \quad (2.2)$$

where $\mathbf{x} \in \mathbb{R}_{\geq 0}^d$ is the state vector of nonnegative species concentrations and $\mathbf{y} \in \mathbb{R}^m$ is the vector of potentially incomplete, $m \leq d$, measurements of $\mathbf{x}$. The functions $f_i(\cdot; \cdot) : \mathbb{R}^d \times \mathbb{R}^p \rightarrow \mathbb{R}^d$ govern the rates of the involved biochemical reactions and are derived using biochemical theory (see Fig 2.1A2 for example terms). Further, $\theta_f \in \mathbb{R}^{p_f}$ is the vector of biological model parameters, including but not limited to rate constants, binding coefficients, and equilibrium coefficients. The function $h(\cdot; \cdot) : \mathbb{R}^d \times \mathbb{R}^p \rightarrow \mathbb{R}^m$ is the measurement function that maps from the states to the set of observables (experimental data), where θ_h is the vector of associated parameters. Lastly, the measurements $\mathbf{y}(t)$ are corrupted by independently

and identically distributed (iid) Gaussian measurement noise $\eta(t) \in \mathbb{R}^m$ with zero mean and covariance matrix $\Gamma \in \mathbb{R}^{m \times m}$. The covariance matrix is parameterized by $\theta_\Gamma \in \mathbb{R}^m$ such that $\Gamma(\theta_\Gamma) = \text{diag}(\theta_\Gamma)$. The parameter space is then defined as the multidimensional space of all possible values of $\theta = [\theta_f, \theta_\Gamma]$, e.g., $\theta \in \mathbb{R}^{p_f+m}$.

In this work, we make several simplifying assumptions to the model in Eqs (2.1)–(2.2). First, we assume that the measurement function $h(\cdot; \cdot)$ is linear and that all the parameters in θ_h are known, so Eq (2.2) becomes

$$\mathbf{y}(t) = \mathbf{H}\mathbf{x}(t) + \eta(t), \quad (2.3)$$

with $\mathbf{H} \in \mathbb{R}^{m \times d}$. Second, we assume that the initial condition $\mathbf{x}(t=0) = \mathbf{x}_0$ is known, so it is excluded from parameter estimation.

Although Eqs (2.1)–(2.2) define a continuous-time dynamical system, we mostly deal with discrete observations. The set of n measurements $\mathcal{Y}_n = \{\mathbf{y}_1, \dots, \mathbf{y}_n\}$ denotes the experimental data taken at time instances $t_1, \leq t_2, \dots, \leq t_n$, where $\mathbf{y}_k = \mathbf{H}\mathbf{x}_k + \eta_k$. In this discrete setting, $\mathbf{x}_k$ is the state vector at time t_k and η_k is an independent realization of the measurement noise. Additionally, the set of states at the discrete measurement times defined above is $\mathcal{X}_n = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$. Note that the full (internal) states may not be available for estimation, only the measurements.

We can now define the problem setting of the proposed parameter estimation framework. This work assumes that the model form is known and seeks to estimate the model parameters and associated uncertainties by learning a probability distribution for the parameters. The following problem statement formalizes the parameter estimation problem.

Problem 1. *Given a known model form as in Eq (2.1) and a set of noisy, sparse and incomplete, experimental measurements $\mathcal{Y}_n = \{\mathbf{y}_1, \dots, \mathbf{y}_n\}$ at time instances $t_1 \leq t_2, \dots, \leq t_n$, estimate the complete probability distribution, $p(\theta|\mathcal{Y}_n)$, for the model parameters $\theta = [\theta_f, \theta_\Gamma]^\top$.*

The next section outlines the structural identifiability and global sensitivity analyses performed to reduce the dimension of θ before estimating parameters.

2.2.3 Global structural identifiability analysis with the structural identifiability analyzer (SIAN)

Structural identifiability analysis determines if parameters can be uniquely estimated from the available measurement function [89]. Structural identifiability is a mathematical property of the model itself and does not consider the quality or quantity of the available experimental data. The following definition from [154] provides an intuitive condition for global structural identifiability and can be shown to be equivalent to alternative definitions [127], [130].

Definition 1 (Global structural identifiability [154]). *A parameter θ_i is globally structurally identifiable if, for almost all $\theta \in \mathbb{R}^p$ and for all times $t > 0$,*

$$\mathbf{y}(t, \theta') = \mathbf{y}(t, \theta) \quad \implies \quad \theta'_i = \theta_i,$$

where $\theta' = [\theta'_1, \dots, \theta'_i, \dots, \theta'_p]^\top$ and $\theta = [\theta_1, \dots, \theta_i, \dots, \theta_p]^\top$.

Global structural identifiability, as in Definition 1, implies that a parameter θ_i can be uniquely identified from data. If a parameter is globally structurally identifiable, then there is a single unique value of that parameter that gives each observed output value with the same initial conditions [130].

Alternatively, a parameter may be *locally* structurally identifiable if the condition in Definition 1 only holds in a local neighborhood, $\mathcal{V}(\theta_i)$, of parameter space around θ_i , e.g., for $\theta_i \in \mathcal{V}(\theta_i)$ [89], [127], [130]. This condition implies that a finite number of values of a locally identifiable parameter can give the same output values [130]. Lastly, if a parameter is nonidentifiable, then infinitely many values of that parameter can give the same model output. While models with locally structurally identifiable and nonidentifiable parameters are still valid over the entire parameter space, the existence of such parameters can confound successful parameter estimation. Many computational methods have been developed to assess structural identifiability [89], [127], [130], [155] for ODE-based models. In this work, we use the differential algebra and power series-based approach presented in [130], [146], because the method specifically assesses global identifiability. The SIAN (Structural Identifiability ANalyser) software [146] provides a numerical implementation of the approach proposed in [130]. The SIAN algorithm for assessing global structural identifiability uses a combination of symbolic and probabilistic computation [130], [146]. First, SIAN uses Taylor expansions of the model equations to obtain a polynomial representation of the system. Second, the algorithm truncates the polynomial system to produce a minimal system containing all parameter identifiability information. Third, SIAN solves the identifiability problem for a single parameter set that is randomly selected to guarantee correctness up to a user-specified probability level p (see Theorem 5 in [130]). Fourth, the algorithm uses the results in the third step to separate the parameters into globally identifiable, locally identifiable, and nonidentifiable sets. SIAN is implemented in Maple (Maplesoft, Waterloo, ON) and Julia (The Julia Project [156]). We refer the reader to [130] for additional mathematical details on SIAN.

We use the Julia implementation of the SIAN algorithm with the default probability of correctness, $p = 0.99$ (available at <https://github.com/alexeyovchinnikov/SIAN-Julia>). Furthermore, we set the additional `p_mod` parameter to $2^{29} - 3$ to enable the algorithm to run faster [157]. Any parameters that

are not globally structurally identifiable are fixed to nominal values informed by the available literature following identifiability analysis. While these parameters may convey meaningful biological information, nonidentifiability implies that it is mathematically impossible to identify them from the available data. Next, global sensitivity analysis is used to further reduce the set of identifiable parameters.

2.2.4 Variance-based global sensitivity analysis

Parametric sensitivity analysis quantifies the contributions of the model parameters to variations in the model output [5], [131]. Specifically, global sensitivity analysis aims to quantify the effects of the model parameters on the output quantity of interest over the entire parameter space [5], [131] and is well suited for studying parameters in nonlinear models [126], [131]. This work applies Sobol's method [147] for variance-based sensitivity analysis because it provides a quantitative output to rank parameters and leverages the prior distributions defined for the parameters. Sobol sensitivity analysis decomposes the variance of model outputs based on contributions from individual parameters and interactions between parameters [131], [147]. The total variance of the output quantity $f(\theta)$ is

$$D = \int_{\mathbb{R}^p} f^2(\theta) d\theta - f_0^2, \quad (2.4)$$

where $f_0 := \int_{\mathbb{R}^p} f(\theta) d\theta$ is the mean of the output. The following definition for the analysis of variance (ANOVA) representation provides an expansion for the output variance in a high-dimensional representation (HDMR), also known as a Sobol representation [5], [147].

Definition 2 (Analysis of variance (ANOVA) representation [5], [131], [147]). *The ANOVA expansion states that the output function $f(\theta)$, for $\theta \in \mathbb{R}^p$ defined as $\theta = [\theta_1, \theta_2, \dots, \theta_p]^\top$, can be represented as*

$$f(\boldsymbol{\theta}) = \sum_i f_i(\boldsymbol{\theta}_i) + \sum_{i < j} f_{i,j}(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j) + \dots + f_{1,2,\dots,p}(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \dots, \boldsymbol{\theta}_p),$$

where the zero-, first-, and second-order terms are defined recursively as

$$\begin{aligned} f_0 &= \int_{\mathbb{R}^p} f(\boldsymbol{\theta}) d\boldsymbol{\theta} \\ f_i(\boldsymbol{\theta}_i) &= \int_{\mathbb{R}^p} f(\boldsymbol{\theta}) d\boldsymbol{\theta}_{\sim i} - f_0 \\ f_{i,j}(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j) &= \int_{\mathbb{R}^p} f(\boldsymbol{\theta}) d\boldsymbol{\theta}_{\sim \{i,j\}} - f_i(\boldsymbol{\theta}_i) - f_j(\boldsymbol{\theta}_j) - f_0, \end{aligned}$$

respectively. The recursion is extended further for increasing numbers of parameters to compute higher-order terms. This definition assumes that the contribution terms are orthogonal (see Def 1 in [147]), and the notation $\sim i$ refers to the set excluding index i , for example: $d\boldsymbol{\theta}_{\sim i} = \{d\boldsymbol{\theta}_1, \dots, d\boldsymbol{\theta}_{i-1}, d\boldsymbol{\theta}_{i+1}, \dots, d\boldsymbol{\theta}_p\}$.

The ANOVA representation (see Def 2) expands the total variance, Eq (2.4), as

$$D = \sum_{i=1}^p D_i + \sum_{1 \leq i < j \leq p} D_{i,j}, \quad (2.5)$$

where the variances D_i and $D_{i,j}$ are

$$D_i = \int_{\mathbb{R}^p} f_i^2(\boldsymbol{\theta}_i) d\boldsymbol{\theta}_i \quad \text{and} \quad D_{i,j} = \int \int_{\mathbb{R}^p} f_{i,j}^2(\boldsymbol{\theta}_i, \boldsymbol{\theta}_j) d\boldsymbol{\theta}_i d\boldsymbol{\theta}_j.$$

Note that it is possible to compute higher-order variances by increasing the dimension of the integral and following the recursion in Def 2, however we limit our discussion to second-order or lower variances for

brevity. The first and second-order Sobol sensitivity indices are then defined using the variance terms in Eq (2.5) as

$$S_i = \frac{D_i}{D} \quad \text{and} \quad S_{i,j} = \frac{D_{i,j}}{D}.$$

The first-order sensitivity index S_i quantifies the fraction of the total variance attributed to parameter θ_i , and the second-order sensitivity index $S_{i,j}$ quantifies this for the interactions between θ_i and θ_j . Lastly, the *total-order sensitivity* is

$$S_{T_i} = S_i + \sum_{j=1}^p S_{i,j},$$

which quantifies all contributions from parameter θ_i on the output variance.

This work uses the UQLab toolbox [158], [159] to perform Sobol sensitivity analysis in Matlab (MathWorks, Natick, MA) and the `DifferentialEquations.jl` package [160] for analysis in Julia. Both softwares use Monte Carlo estimators to compute Sobol sensitivity indices from parameter samples (see [5], [131] and references therein for details). In Matlab, unless otherwise specified, first and total sensitivity indices are computed using Sobol pseudo-random sampling (e.g. `SOpts.Sobol.Sampling = 'sobol'`) and the default estimator (`SOpts.Sobol.Sampling = 'janon'`), see [159] for details. Additionally, the number of samples, `SOpts.Sobol.SampleSize`, is set specifically for each problem in Section 2.3. In Julia, we perform similar Sobol sampling with `QuasiMonteCarlo.jl` (<https://github.com/SciML/QuasiMonteCarlo.jl>) with the `SobolSample()` sampler, and the default estimator for the sensitivity indices, e.g. `Ei_estimator` set to `:jansen1999`.

Sobol sensitivity analysis computes sensitivity indices for scalar output quantities rather than for an entire trajectory. Thus, we define quantities of interest (QoI) for each problem that capture the

biologically relevant information in a trajectory, such as the steady state value of a certain biochemical species. Using the computed sensitivity indices for the given QoI, we rank the parameters and aim to find the most influential parameters by selecting those with the greatest sensitivity indices. The critical challenge is selecting a cutoff point that separates the influential parameters from the complete set. To do so, we chose a threshold corresponding to a pronounced decline in the sensitivity index value or a logical value of that index, for example, $S_i \geq 0.1$. All parameters with sensitivity indices above the cutoff value are left free for estimation, and the remaining parameters are fixed to their nominal values. The following section recasts the parameter estimation problem (Problem 1) in the Bayesian framework to estimate the free, influential parameters.

2.2.5 Bayesian parameter estimation

Bayesian parameter estimation solves Problem 1 by considering the model parameters, θ , as random variables that do not have a pre-specified parametric distribution. Note that parametric distributions have probability density functions that are fully described by a closed-form parametric equation, such as a normal distribution that is characterized by its mean and variance. The approach characterizes the posterior probability distribution for the parameters $p(\theta|\mathcal{Y}_n)$ conditioned on a given dataset $\mathcal{Y}_n$ and provides the *best guess* probability distribution for θ *given the data*. We can use Bayes' rule to express the posterior distribution as

$$p(\theta|\mathcal{Y}_n) \propto p(\theta)\mathcal{L}(\theta;\mathcal{Y}_n), \quad (2.6)$$

where $p(\theta)$ is known as the prior distribution, and $\mathcal{L}(\theta;\mathcal{Y}_n)$ is the likelihood function (see Fig 2.1B2 for a visual representation of Bayes' rule).

Intuitively, Bayes' rule updates our *best guess* about the distribution of the model parameters as new data is being incorporated. The prior distribution $p(\theta)$ represents the *best guess* before any data

are collected and encodes any assumptions on the parameters. For instance, the prior may convey the physiological ranges for parameter values or may weigh known values more heavily (see Section 2.2.8). The likelihood function $\mathcal{L}(\theta; \mathcal{Y}_n) = p(\mathcal{Y}_n | \theta)$ updates our *belief state* by measuring the misfit between the data and model predictions with a specific parameter set. Parameter sets that are more likely to occur will produce model predictions that better match the data and thus have higher likelihood probabilities. For example, although the prior in Fig 2.1B2 places more probability on smaller values of θ , the likelihood in Fig 2.1B2 places more probability mass towards larger values. It is important to note that evaluation of the likelihood function requires model simulations. For example, assuming Gaussian measurement noise with zero mean, a possible likelihood function is

$$\mathcal{L}(\theta; \mathcal{Y}_n) = \frac{1}{(2\pi)^{\frac{m}{2}} |\Gamma|^{\frac{1}{2}}} \prod_{k=1}^n \exp\left(-\frac{1}{2} \|\mathbf{y}_k - \mathbf{H}\hat{\mathbf{x}}_k\|_{\Gamma}^2\right), \quad (2.7)$$

where $\hat{\mathbf{x}}_k$ is the predicted state at time t_k with the parameters θ , $|\cdot|$ denotes the matrix determinant, and the $\mathbf{C}$ -weighted norm is defined as $\|\mathbf{a}\|_{\mathbf{C}}^2 := \mathbf{a}^\top \mathbf{C}^{-1} \mathbf{a}$, where $\mathbf{C}$ is a symmetric positive definite matrix. Lastly, the posterior distribution conveys the *best guess* after collecting and incorporating the data $\mathcal{Y}_n$ into the statistical model and can be further refined if more data are included. In Fig 2.1B2, the posterior illustrates how the likelihood function re-weights the prior to update our *belief state*.

A fundamental difficulty in Bayesian parameter estimation is that Bayes' rule only enables evaluating the posterior distribution at specific points in parameter space. This is in contrast to, for example, the formula for a Gaussian distribution (with mean μ and standard deviation σ)

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

that can be analytically evaluated at all values of x . Therefore, parameter samples are drawn from

the posterior help to characterize the distribution over the entire parameter space. Markov chain Monte Carlo (MCMC) algorithms enable sampling from arbitrary distributions, such as the posterior distribution (see Section 2.2.9 for details). Before performing Bayesian estimation the next section introduces the constrained interval unscented Kalman filter Markov chain Monte Carlo (CIUKF-MCMC) algorithm that accounts for uncertainty in the model and the data.

2.2.6 Constrained interval unscented Kalman filter Markov chain Monte Carlo (CIUKF-MCMC)

Complete uncertainty quantification of dynamical models in systems biology must account for uncertainty in the model form, parameters and noisy data. In [82], Galioto and Gorodetsky suggest adding a process noise term to Eq (2.1) to account for model form uncertainty in the system. Following this suggestion, the model in Eqs (2.1-2.2) is recast as a discrete time stochastic process

$$\mathbf{x}_k = \boldsymbol{\psi}(\mathbf{x}_{k-1}; \boldsymbol{\theta}_f) + \boldsymbol{\xi}_k, \quad \boldsymbol{\xi}_k \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}(\boldsymbol{\theta}_\Sigma)) \quad (2.8)$$

$$\mathbf{y}_k = \mathbf{H}\mathbf{x}_k + \boldsymbol{\eta}_k, \quad \boldsymbol{\eta}_k \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Gamma}(\boldsymbol{\theta}_\Gamma)), \quad (2.9)$$

where k is the discrete time index for t_k , and $\boldsymbol{\psi}(\cdot; \cdot)$ is the discrete state propagator that evolves the state from time t_{k-1} to time t_k . Additionally, $\boldsymbol{\xi}_k$ and $\boldsymbol{\eta}_k$ are Gaussian process and measurement noise (stochastic noise processes) with covariances $\boldsymbol{\Sigma}(\boldsymbol{\theta}_\Sigma)$ and $\boldsymbol{\Gamma}(\boldsymbol{\theta}_\Gamma)$, respectively. The discrete state propagator $\boldsymbol{\psi}(\cdot; \cdot)$ in Eq (2.8) is the discrete operator that evolves the state according to the differential equation in Eq (2.1), and, for example, could be a forward-Euler finite-difference approximation. This work deploys the Matlab function `ode15s()` to construct a state propagator that guarantees the necessary stability to handle systems biology models.

Bayesian estimation of the model parameters, $\theta = [\theta_f, \theta_\Sigma, \theta_\Gamma]^\top$, of the extended system in Eqs (2.8)–(2.9) accounts for uncertainty in the data, model, and parameters. The introduction of process noise increases the dimension of the parameters to estimate by requiring estimates for θ_Σ . Further, the addition of stochastic process noise means the state variables are random variables; Bayes' rule for this system becomes

$$p(\theta, \mathcal{X}_n | \mathcal{Y}_n) \propto p(\theta) \mathcal{L}(\theta; \mathcal{Y}_n, \mathcal{X}_n), \quad (2.10)$$

to account for the additional uncertainty in the states. The key step of the UKF-MCMC algorithm is the marginalization of uncertainty in the states out of Eq (2.10) to enable estimation of the parameter posterior distribution [82].

The UKF-MCMC algorithm begins by constructing an expression for the joint likelihood of the states and the parameters, $\mathcal{L}(\theta; \mathcal{Y}_n, \mathcal{X}_n)$. Two probability distributions implied by the stochastic system in Eqs (2.8-2.9) are needed to define an expression for the joint likelihood. First, the probability of the current state $\mathbf{x}_k$ given the past state $\mathbf{x}_{k-1}$ is

$$p(\mathbf{x}_k | \mathbf{x}_{k-1}, \theta_f, \Sigma(\theta_\Sigma)) = \frac{\exp\left(-\frac{1}{2} \|\mathbf{x}_k - \psi(\mathbf{x}_{k-1}, \theta_f)\|_{\Sigma}^2\right)}{(2\pi)^{\frac{d}{2}} |\Sigma(\theta_\Sigma)|^{\frac{1}{2}}}, \quad (2.11)$$

where the norm $\|\mathbf{x}_k - \psi(\mathbf{x}_{k-1}, \theta_f)\|_{\Sigma}^2$ quantifies the misfit between the past state and the predicted current state. Next, the probability of a measurement $\mathbf{y}_k$ given $\mathbf{x}_k$ is

$$p(\mathbf{y}_k | \mathbf{x}_k, \Gamma(\theta_\Gamma)) = \frac{\exp\left(-\frac{1}{2} \|\mathbf{y}_k - \mathbf{H}\mathbf{x}_k\|_{\Gamma}^2\right)}{(2\pi)^{\frac{m}{2}} |\Gamma(\theta_\Gamma)|^{\frac{1}{2}}}, \quad (2.12)$$

where the norm $\|\mathbf{y}_k - \mathbf{H}\mathbf{x}_k\|_{\Gamma}^2$ quantifies the residual between the measurement and the true states. By

combining Eq (2.11) and Eq (2.12) the joint likelihood is

$$\mathcal{L}(\boldsymbol{\theta}; \mathcal{Y}_n, \mathcal{X}_n) = \prod_{k=1}^n \left[\frac{\exp\left(-\frac{1}{2} \|\mathbf{x}_k - \boldsymbol{\psi}(\mathbf{x}_{k-1}, \boldsymbol{\theta}_f)\|_{\Sigma}^2\right)}{(2\pi)^{\frac{d}{2}} |\Sigma(\boldsymbol{\theta}_{\Sigma})|^{\frac{1}{2}}} \times \frac{\exp\left(-\frac{1}{2} \|\mathbf{y}_k - \mathbf{H}\mathbf{x}_k\|_{\Gamma}^2\right)}{(2\pi)^{\frac{m}{2}} |\Gamma(\boldsymbol{\theta}_{\Gamma})|^{\frac{1}{2}}} \right]. \quad (2.13)$$

Marginalizing out the uncertain states by integration yields the likelihood for the uncertain parameters

$$\mathcal{L}(\boldsymbol{\theta}; \mathcal{Y}_n) = \int_{\mathbb{R}_{\geq 0}^d} \cdots \int_{\mathbb{R}_{\geq 0}^d} \mathcal{L}(\boldsymbol{\theta}; \mathcal{Y}_n, \mathcal{X}_n) d\mathbf{x}_1 \dots d\mathbf{x}_n. \quad (2.14)$$

However, there is no obvious computationally tractable approach to integrate over a set of uncertain states directly. Theorem 1, stated below, provides a recursive algorithm to marginalize the states out of the likelihood, e.g., to perform the integration in Eq (2.14). Although Theorem 1 assumes that the initial condition is uncertain (and it is therefore estimated), we do not use that estimate in this work, as we start with a known initial condition $\mathbf{x}_0$.

Theorem 1 (Marginal likelihood (Theorem 1 of [82] and 12.1 of [161])). *Let $\mathcal{Y}_k$ denote the set of all observations up to time t_k as defined in Section 2.2.2. Let the initial condition be uncertain with distribution $p(\mathbf{x}_0|\boldsymbol{\theta})$. Then the marginal likelihood is defined recursively in three stages:*

for $k = 1, 2, \dots$

1. *Predict the new state from previous data*

$$p(\mathbf{x}_{k+1}|\boldsymbol{\theta}, \mathcal{Y}_k) = \int_{\mathbb{R}_{\geq 0}^d} p(\mathbf{x}_k|\boldsymbol{\theta}, \mathcal{Y}_k) \frac{\exp\left(-\frac{1}{2} \|\mathbf{x}_{k+1} - \boldsymbol{\psi}(\mathbf{x}_k, \boldsymbol{\theta}_f)\|_{\Sigma}^2\right)}{(2\pi)^{\frac{d}{2}} |\Sigma(\boldsymbol{\theta}_{\Sigma})|^{\frac{1}{2}}} d\mathbf{x}_k,$$

2. *update the prediction with the current data*

$$p(\mathbf{x}_{k+1}|\boldsymbol{\theta}, \mathcal{Y}_{k+1}) = p(\mathbf{x}_{k+1}|\boldsymbol{\theta}, \mathcal{Y}_k) \frac{\exp\left(-\frac{1}{2} \|\mathbf{y}_{k+1} - \mathbf{H}\mathbf{x}_{k+1}\|_{\Gamma}^2\right)}{(2\pi)^{\frac{m}{2}} |\Gamma(\boldsymbol{\theta}_{\Gamma})|^{\frac{1}{2}}},$$

3. and marginalize out uncertainty in the states

$$\mathcal{L}_{k+1}(\boldsymbol{\theta}|\mathcal{Y}_{k+1}) = \int_{\mathbb{R}_{\geq 0}^d} p(\mathbf{x}_{k+1}|\boldsymbol{\theta}, \mathcal{Y}_k) \frac{\exp\left(-\frac{1}{2} \|\mathbf{y}_{k+1} - \mathbf{H}\mathbf{x}_{k+1}\|_{\Gamma}^2\right)}{(2\pi)^{\frac{m}{2}} |\Gamma(\boldsymbol{\theta}_{\Gamma})|^{\frac{1}{2}}} d\mathbf{x}_{k+1}.$$

The recursion defined in Theorem 1 closely resembles a Bayesian filter [161]; thus, it is evaluated with Kalman filtering algorithms [82]. For linear models, the standard linear Gaussian Kalman filter can be used to evaluate the recursion (see Algorithm 2 in [82]). However, exact solutions to the recursion are not possible if the model or measurement processes are nonlinear. Therefore, approximations such as extended Kalman filters (EKF), unscented Kalman filters (UKF), or ensemble Kalman filters (EnKF) [82], [161] must be employed. The original implementations of the UKF-MCMC algorithm use the UKF [162] for this approximation because the UKF is generally stable and can handle nonlinear models and measurement processes [82], [145]. However, the UKF is not suitable for systems biology models because it ignores constraints on the state variables, such as the nonnegativity of chemical concentrations. Section 2.2.7 describes the constrained interval unscented Kalman filter (CIUKF) [148] implemented in this work to enforce state constraints during filtering. Thus, we refer to the UKF-MCMC from [82] that uses the constrained interval unscented Kalman filter [148] as CIUKF-MCMC.

2.2.7 Constrained interval unscented Kalman filter (CIUKF)

We implement the constrained interval Unscented Kalman filter (CIUKF) [148], [163] algorithm to enforce all state constraints in CIUKF-MCMC. This algorithm assumes that the state is subject to an interval constraint $\mathbf{x}_{\text{LB}} \leq \mathbf{x} \leq \mathbf{x}_{\text{UB}}$. Since we only seek to enforce nonnegativity in systems biology,

the interval constraint is $\mathbf{0} \leq \mathbf{x} \leq \infty$. We choose the CIUKF over alternative state constrained Kalman filters [148], [164] because it enforces constraints in both the predict and update steps of the algorithm and retains the same structure as the standard UKF [161]. We outline the steps of the CIUKF below.

The CIUKF algorithm predicts the state $\mathbf{x}_{k+1}$ from all preceding measurement data $\mathcal{Y}_n$. Following the structure of the linear Kalman filter, CIUKF is a recursive algorithm that iterates over all data and performs prediction and update steps at each time point [161]. For simplicity, we outline a single iteration of the CIUKF that moves the state forward in time from $\mathbf{x}_k$ to $\mathbf{x}_{k+1}$. Let $\mathbf{C}_k^{xx}$ be the state covariance matrix at time t_k , $\Sigma(\theta_\Sigma)$ be the process noise covariance matrix, $\Gamma(\theta_\Gamma)$ be the measurement noise covariance matrix, and θ_f be the model parameters. There are two steps to the CIUKF, prediction, and update.

First, the prediction step uses the the interval constrained unscented transform [148], [163] (the state-constrained equivalent to the unscented transform [162], [165]) to predict the state and its covariance matrix at the next time point after propagation by the nonlinear model, e.g., Eq (2.8). The interval constrained unscented transform constructs a set of sigma points that capture the covariance $\mathbf{C}_k^{xx}$ at time t_k . Each sigma point is propagated in time by the nonlinear model to approximate the new state and its covariance at the next time t_{k+1} . The set of $2d + 1$ sigma points, $\mathcal{X}$, is given by

$$\mathcal{X}_k^{(0)} = \mathbf{x}_k \quad (2.15a)$$

$$\mathcal{X}_k^{(i)} = \mathbf{x}_k + \xi_i [\sqrt{\mathbf{C}_k^{xx}}]_i \quad (2.15b)$$

$$\mathcal{X}_k^{(i+d)} = \mathbf{x}_k - \xi_{i+d} [\sqrt{\mathbf{C}_k^{xx}}]_i, \quad (2.15c)$$

where ξ_i is the i th sigma point coefficient, $[\mathbf{A}]_i$ is the i th column of $\mathbf{A}$, $\sqrt{\mathbf{A}}$ is the matrix square root of $\mathbf{A}$, and $i = 1, \dots, d$. The sigma point coefficients ξ_i control the distances of the sigma points around the initial

state $\mathbf{x}_k$ and are chosen to ensure that no sigma points violate the state constraints. They are

$$\xi_i = \min([\Xi]_i) \quad (2.16)$$

$$\Xi(i, j) := \begin{cases} \sqrt{d + \lambda} & \text{if } \mathbf{S}(i, j) = 0 \\ \min\left(\sqrt{d + \lambda}, \frac{\mathbf{x}_{\text{LB}} - \mathbf{x}_k}{\mathbf{S}(i, j)}\right) & \text{if } \mathbf{S}(i, j) < 0 \\ \min\left(\sqrt{d + \lambda}, \frac{\mathbf{x}_{\text{UB}} - \mathbf{x}_k}{\mathbf{S}(i, j)}\right) & \text{if } \mathbf{S}(i, j) > 0 \end{cases} \quad (2.17)$$

$$\mathbf{S} := \begin{bmatrix} \sqrt{\mathbf{C}_k^{xx}} & -\sqrt{\mathbf{C}_k^{xx}} \end{bmatrix}, \quad (2.18)$$

where λ is a parameter of the algorithm. Alternatively, in the standard unscented transform, the coefficients are all equal to $\sqrt{d + \lambda}$ [162], [165]. Next, a set of weights, w_i , are assigned to each sigma point as

$$w_0 = b \quad (2.19a)$$

$$w_i = a\xi_i + b \quad (2.19b)$$

$$a := \frac{2\lambda - 1}{2(d + \lambda) (\sum_{i=1}^d \xi_i - (2d + 1)\sqrt{d + \lambda})} \quad (2.19c)$$

$$b := \frac{1}{2(d + \lambda)} - \frac{2\lambda - 1}{2\sqrt{d + \lambda} (\sum_{i=1}^d \xi_i - (2d + 1)\sqrt{d + \lambda})}, \quad (2.19d)$$

where ξ_i are as defined in Eqs (2.16-2.18). Importantly the sum of the weights equals one, $\sum_{i=0}^{2d} w_i = 1$.

The prediction step then uses the nonlinear model, Eq (2.8), to propagate each sigma point forward in time

$$\hat{\mathcal{X}}_k^{(i)} = \psi\left(\mathcal{X}_k^{(i)}, \theta_f\right) \quad (2.20)$$

for $i = 0, \dots, 2d$. The prediction mean $\mathbf{x}_k^*$ and covariance $\mathbf{C}_k^{xx*}$ are then respectively computed as

$$\mathbf{x}_k^* = \sum_{i=0}^{2n} w_i \hat{\mathcal{X}}_k^{(i)} \quad (2.21)$$

$$\mathbf{C}_k^{xx*} = \sum_{i=0}^{2n} w_i \left[\hat{\mathcal{X}}_k^{(i)} - \mathbf{x}_k^* \right] \left[\hat{\mathcal{X}}_k^{(i)} - \mathbf{x}_k^* \right]^\top + \Sigma(\theta_\Sigma). \quad (2.22)$$

Equations (2.15)-(2.22) describe the constrained interval unscented transform that approximates the mean and covariance of the state after propagation by the nonlinear model. The prediction mean, Eq (2.21), and covariance, Eq (2.22), provide the predicted state and its covariance, respectively, that are then updated using the available data $\mathbf{y}_k$.

The update step begins by constructing a new set of sigma points centered around $\mathbf{x}_k^*$, where

$$\mathcal{X}_k^{*(0)} = \mathbf{x}_k^* \quad (2.23a)$$

$$\mathcal{X}_k^{*(i)} = \mathbf{x}_k^* + \sqrt{n + \lambda} \left[\sqrt{\mathbf{C}_k^{xx*}} \right]_i \quad (2.23b)$$

$$\mathcal{X}_k^{*(i+d)} = \mathbf{x}_k^* - \sqrt{n + \lambda} \left[\sqrt{\mathbf{C}_k^{xx*}} \right]_i. \quad (2.23c)$$

Additionally, a new set of weights are

$$\begin{aligned} w_0^{(m)} &= \frac{\lambda}{d + \lambda}, & w_i^{(m)} &= \frac{1}{2(d + \lambda)}, \\ w_0^{(c)} &= \frac{\lambda}{d + \lambda} + (1 - \alpha^2 + \beta), & w_i^{(c)} &= \frac{1}{2(d + \lambda)}, \end{aligned}$$

where $\{w_i^{(m)}\}$ are used to compute the mean and $\{w_i^{(c)}\}$ are used to compute the covariance matrix. These weights are indeed equal for all sigma points and are equivalent to those used in the standard UKF. Next,

the measurement function is applied to each sigma point, yielding a set of predicted measurements

$$\mathcal{Y}_k^{*(i)} = h\left(\mathcal{X}_k^{*(i)}, \theta_h\right),$$

where the measurement function $h(\cdot; \cdot)$ is possibly nonlinear with parameters θ_h . Then the mean and covariance matrices of the predicted measurements are computed with the weighted sums,

$$\begin{aligned}\mathbf{y}_k^* &= \sum_{i=0}^{2n} w_i^{(m)} \mathcal{Y}_k^{*(i)} \\ \mathbf{C}_k^{yy*} &= \sum_{i=0}^{2n} w_i^{(c)} \left[\mathcal{Y}_k^{*(i)} - \mathbf{y}_k^* \right] \left[\mathcal{Y}_k^{*(i)} - \mathbf{y}_k^* \right]^\top + \Gamma \\ \mathbf{C}_k^{xy*} &= \sum_{i=0}^{2n} w_i^{(c)} \left[\mathcal{X}_k^{*(i)} - \mathbf{x}_k^* \right] \left[\mathcal{Y}_k^{*(i)} - \mathbf{y}_k^* \right]^\top,\end{aligned}$$

and the Kalman gain is

$$\mathbf{K}_k = \mathbf{C}_k^{xy*} \left(\mathbf{C}_k^{yy*} \right)^{-1}.$$

Lastly, the updated state $\mathbf{x}_{k+1}$ is found by solving the following constrained nonlinear optimization problem,

$$\mathbf{x}_{k+1} = \arg \min_{\mathbf{x}} f(\mathbf{x}) \tag{2.24a}$$

$$\text{subject to } \mathbf{x}_{\text{LB}} \leq \mathbf{x} \leq \mathbf{x}_{\text{UB}}, \tag{2.24b}$$

where the objective function is

$$f(\mathbf{x}) = [\mathbf{y}_k - h(\mathbf{x}, \theta_h)] \Gamma^{-1} [\mathbf{y}_k - h(\mathbf{x}, \theta_h)]^\top + [\mathbf{x}_k - \mathbf{x}_k^*] (\mathbf{C}_k^{xx*})^{-1} [\mathbf{x}_k - \mathbf{x}_k^*]^\top. \tag{2.25}$$

This optimization problem can be solved in Matlab using the `fmincon()` optimizer. Additionally, the updated covariance matrix is given by

$$\mathbf{C}_{k+1}^{xx} = \mathbf{C}_k^{xx*} - \mathbf{K}_k \mathbf{C}_k^{yy*} \mathbf{K}_k^\top.$$

In offline state estimation problems, such as CIUKF-MCMC, this filter is iterated over all available data, e.g. from time t_0 to time t_n if n data points are available [82], [145].

In practice, the CIUKF algorithm is substantially more compute-intensive than the standard UKF [148] because the CIUKF update step involves solving a constrained nonlinear optimization problem, e.g., Eq (2.24). However, the objective function in Eq (2.25) can be simplified given the linear measurement assumptions made in Section 2.2.2. The simplified objective function becomes

$$f(\mathbf{x}) = [\mathbf{y}_k - \mathbf{H}\mathbf{x}] \Gamma^{-1} [\mathbf{y}_k - \mathbf{H}\mathbf{x}]^\top + [\mathbf{x}_k - \mathbf{x}_k^*] (\mathbf{C}_k^{xx*})^{-1} [\mathbf{x}_k - \mathbf{x}_k^*]^\top. \quad (2.26)$$

Expanding this and recognizing that minimizing $f(x) = y(x) + b$ is equivalent to minimizing $f(x) = y(x)$, gives

$$f^*(\mathbf{x}) = \mathbf{x}^\top \left[\mathbf{H}^\top \Gamma^{-1} \mathbf{H} + (\mathbf{C}_k^{xx*})^{-1} \right] \mathbf{x} - 2 \left[\mathbf{y}_k^\top \Gamma^{-1} \mathbf{H} + \mathbf{x}_k^* \mathbf{C}_k^{xx*} \right] \mathbf{x}, \quad (2.27)$$

which is equivalent objective function to Eq (2.26) and is a quadratic form. Thus, the constrained optimization problem in Eq (2.24) becomes a quadratic program when using Eq (2.27) as the objective function. We use the `quadprog()` function in Matlab to solve the quadratic program with the 'Algorithm' option set to 'trust-region-reflective'. We keep all other settings for the solver as defaults. As expected, solving the quadratic program is substantially more efficient than solving the general nonlinear problem.

Throughout this work, we set $\lambda = 1$, $\alpha = 1 \times 10^{-3}$, and $\beta = 1$ for the CIUKF. Furthermore the Cholesky decomposition, $\mathbf{A} = \mathbf{L}\mathbf{L}^\top$, is used to compute the matrix square roots in Eq (2.15) and Eq (2.23), because covariance matrices are always positive definite. Additionally, we compute all covariance matrices as $\mathbf{P}^* = \frac{1}{2}(\mathbf{P} + \mathbf{P}^\top) + \varepsilon\mathbf{I}$, where $\varepsilon = 1 \times 10^{-10}$, to ensure all they remain symmetric positive definite. The next section discusses how to choose prior distributions for Bayesian estimation.

2.2.8 Prior distribution

The prior distribution, $p(\theta)$ encodes our *belief state* about parameters before collecting data and performing Bayesian estimation [5], [84]. The form of the prior distribution allows us to incorporate varying levels of prior knowledge into our models. If the values of a parameter are known, informative priors can be used to shift to the possible values for that parameter towards the known values [84]; for example, a log-normal prior distribution can be used to center the prior around experimental measurements of a parameter [166]. However, the Bayesian inference literature often warns that informative priors should only be used in combination with good information on the parameter values [84] as it can take a large amount of data to overcome a “bad” prior. Alternatively, noninformative or weakly-informative priors reflect a lack of good prior knowledge about the parameters. For example, if we only know a parameter’s physiological ranges, we could construct a uniform prior that states there is an equal probability of any parameter value within this range. Noninformative or weakly-informative priors rely on the data to provide information on the parameters, so the choice of such priors is often safer when prior knowledge of the parameters is limited [84].

In applying the CIUKF-MCMC algorithm, this work constructs prior distributions for two sets of parameters, the biological model parameters, and the noise covariance parameters. We choose to use uniform priors for the biological model parameters, θ_f , to replicate the typical modeling setting where

only the possible ranges for model parameters are known. Supplemental Tables in A.2 list the upper and lower bounds of all biological model parameters. Furthermore, we follow the choices in [82] and use right-half-normal priors for the measurement and process noise covariance parameters, θ_Σ and θ_Γ , respectively (this is further motivated in section 7.1 of [167]). The choice of covariance and upper bound of these priors was found to significantly affect the convergence of MCMC sampling. Thus, the prior distributions for the noise covariance parameters were scaled to match the respective state variable. For example, the prior for a measurement noise covariance θ_{Σ_i} that corresponds to measurements $\{y_1^i, y_2^i, \dots, y_n^i\}$ would be a right-half-normal distribution with mean zero and standard deviation equal to a fraction of the standard deviation σ_{y^i} of the data. The prior is then $\theta_{\Sigma_i} \sim \text{Right-Half-Normal}(0, b\sigma_{x_i})$ truncated to $[0, \sigma_{x_i}]^\top$, where $b < 1$ and σ_{x_i} is the standard deviation of the state x_i . We chose $b = \frac{1}{3}$ unless otherwise specified. These choices are motivated by the observation that measurement noise and process noise covariances will always be smaller than or equal to the covariance of the available data if there is any meaningful information in the data. The next section introduces MCMC sampling to characterize posterior distributions that use CIUKF to approximate the likelihood function.

2.2.9 Markov chain Monte Carlo sampling

Markov chain Monte Carlo (MCMC) algorithms enable sampling from arbitrary probability distributions [5], [110], [168], [169]. The key idea of MCMC is to construct a Markov chain of samples $\theta^1, \theta^2, \dots, \theta^N$ whose distribution converges to the target distribution, $\pi(\theta)$, that we wish to sample [84], [110]. We apply MCMC sampling to Bayesian parameter estimation by constructing a Markov chain where the target distribution is the posterior distribution, that is $\pi(\theta) = p(\theta | \mathcal{Z}_n)$. In this work, we use two MCMC sampling algorithms, delayed rejection adaptive Metropolis (DRAM) [152] and affine invariant ensemble sampler (AIES) [153]. These samplers build upon the classical Metropolis-Hastings algorithm [170], [171]

that we introduce in Section 2.2.9. We outline DRAM in Section 2.2.9 and AIES Section 2.2.9. Lastly, we discuss convergence assessment with the integrated autocorrelation time [153] in Section 2.2.9. We focus these discussions on the practical aspects of MCMC and refer the reader to [5], [84], [110] for additional theoretical details.

Metropolis-Hastings

The Metropolis-Hastings (MH) algorithm [170], [171] constructs a Markov chain whose probability distribution is guaranteed to converge to the target distribution and forms the foundation for a large family of MCMC samplers [84], [110]. The MH algorithm consists of two steps, a proposal and an accept-reject step, repeated to draw the set of samples $\theta^1, \theta^2, \dots, \theta^N$. The Markov chain starts with an initial sample θ^0 that the user chooses. We outline one iteration of the MH algorithm moving from step i to step $i + 1$.

The first step in Metropolis-Hastings, called the proposal step, is to propose a new sample θ^* . We draw θ^* from the proposal distribution $q(\theta^*|\theta^i)$. The proposal distribution is specific to each MCMC algorithm, but, for example, could be a normal distribution centered around the previous sample such that $\theta^* \sim \mathcal{N}(\theta^i, \sigma^2 \mathbf{I})$ where σ is specified (this is random walk Metropolis (RWM) [110]).

Next, the accept-reject step decides if the proposal is accepted, set $\theta^{i+1} = \theta^*$, or rejected, set $\theta^{i+1} = \theta^i$. In the MH algorithm, this decision is probabilistic, e.g., the proposal is accepted or rejected with acceptance probability $\alpha(\theta^*|\theta^i)$. The acceptance probability is

$$\alpha(\theta^*|\theta^i) = \min\left(1, \frac{\pi(\theta^*) q(\theta^i|\theta^*)}{\pi(\theta^i) q(\theta^*|\theta^i)}\right), \quad (2.28)$$

which guarantees that the stationary distribution of the samples (the distribution that the samples converge to in the infinite sample limit) equals the target distribution [84], [110]. These steps, propose and accept-reject, are repeated until the distribution of the set of samples has *converged* to its stationary distribution.

Although convergence is guaranteed in the infinite sample limit [84], [110], assessing convergence in practice is nontrivial. We chose to use an approach from [153] that uses the integrated autocorrelation time to assess convergence as outlined in Section 2.2.9 (see [5], [84], [110] for additional approaches).

Delayed rejection adaptive Metropolis (DRAM)

The delayed rejection adaptive Metropolis (DRAM) algorithm is based on the random walk Metropolis algorithm and combines delayed rejection with adaptive Metropolis [152]. DRAM closely follows the Metropolis-Hastings algorithm but specifically uses a Gaussian proposal distribution such that $\theta^* \sim \mathcal{N}(\theta^i, \mathbf{C})$, where $\mathbf{C}$ is the covariance matrix. Delayed rejection (DR) [152], [172] and adaptive Metropolis (AM) [152], [173] are two modifications to the random walk Metropolis algorithm that help to improve the convergence rate of the Markov chain.

First, delayed rejection adds an additional round of Metropolis-Hastings (propose and accept-reject) if the first proposal is rejected. That is, if θ^* is rejected a new proposal, $\theta^{*2} \sim q_2(\theta^{*2}|\theta^*, \theta^i)$ is drawn, where $q_2(\cdot)$ is the new proposal distribution. The new proposal distribution is also Gaussian, however the covariance is scaled to a fraction of the original proposal covariance, e.g. $q_2(\theta^{*2}|\theta^*, \theta^i) = \mathcal{N}(\theta^i, \gamma\mathbf{C})$, where $\gamma < 1$ and is a tuning parameter of the algorithm [152]. Thus the second proposal is closer to the previous point and is more likely to be accepted. DRAM evaluates the new proposal with a Metropolis-Hastings accept-reject step with the acceptance probability

$$\alpha_2(\theta^{*2}|\theta^*, \theta^i) = \min \left(1, \frac{\pi(\theta^{*2})}{\pi(\theta^i)} \frac{q(\theta^*|\theta^{*2})}{q(\theta^*|\theta^i)} \frac{(1 - \alpha(\theta^*|\theta^{*2}))}{(1 - \alpha(\theta^*|\theta^i))} \right),$$

where $\alpha(\cdot|\cdot)$ is as defined in Eq (2.28). Although most implementations impose a single DR step [5], delayed rejection can be repeated more than once, where the proposal distribution and acceptance

probability are modified accordingly at additional each round.

Adaptive Metropolis acts separately from delayed rejection and aims to move the proposal distribution closer to the target distribution [152], by replacing the covariance matrix of the proposal, $\mathbf{C}$, with the covariance matrix of the samples. In practice, adaptation begins after a set number of i_0 samples have been drawn and updates the covariance matrix at every step as

$$\mathbf{C}_i = \begin{cases} \mathbf{C}_0 & i \leq i_0 \\ s_p \text{Cov}[\theta_1, \dots, \theta_i] + s_p \varepsilon \mathbf{I} & i > i_0, \end{cases}$$

where $\mathbf{C}_0$ is the initial covariance matrix, s_p is an algorithm tuning parameter that is often set to $s_p = 2.38^2/p$ [152], [168] where p is the dimension of θ , and ε is a small positive number.

We implement the DRAM algorithm with a single round of delayed rejection following [82] with $\gamma = 0.01$, $i_0 = 200$, and $\varepsilon = 1 \times 10^{-10}$. Furthermore, the initial covariance matrix $\mathbf{C}_0$ is tuned to accept between 20–40% of proposals. We refer the reader to [5], [152] for further details on DRAM and tuning the initial covariance matrix. Additionally, each Markov chain is initialized to the MAP point by using `fmincon()` (default options except: 'UseParallel' set to `true` and 'MaxFunctionEvaluations' set to 10,000) in Matlab to minimize the negative log-posterior (equivalent to maximizing the posterior) as in [82].

Affine invariant ensemble sampler (AIES)

While random walk Metropolis-based algorithms such as DRAM can adequately sample complex posterior distributions, these methods will show very slow convergence when the target distribution is highly anisotropic [153]. The posterior distribution with CIUKF-MCMC in systems biology are anisotropic because the scales of model parameters can vary by several orders of magnitude, and the noise covariance

parameters often have different scaling than the model parameters. Fortunately, AIES provides an algorithm to sample such anisotropic distributions [153] effectively. The motivation for affine invariance is that anisotropic distributions can be transformed to isotropic distributions with an affine transformation. Thus an algorithm that is invariant to such transformations will effectively sample an isotropic distribution when sampling an anisotropic distribution [153].

The AIES algorithm differs from DRAM and random walk Metropolis because it leverages an ensemble of Markov chains rather than a single chain. Each chain in the ensemble of N_e Markov chains is called a walker, and we denote the set of walkers at step i with $\{\theta_1^i, \dots, \theta_{N_e}^i\}$ where the subscript is the walker index and the superscript is the step-index. Note that the number of walkers must be larger than the dimension of θ that is $N_e > p$, for $\theta \in \mathbb{R}^p$. At the end of N steps, a set of N_e Markov chains with N samples is obtained. The first chain in the ensemble is, for example, $\theta_1^1, \theta_1^2, \dots, \theta_1^N$. Note that the total ensemble will have $N \cdot N_e$ samples. Each step of the AIES algorithm involves a proposal and a Metropolis-Hastings update for each chain in the ensemble. One iteration of these steps to move from i to $i + 1$ for a single walker, θ_n^i , is outlined below and these steps are repeated to update the entire ensemble.

First, a proposal for the current walker θ_n^i is chosen using the stretch move [153] that ensures affine invariance of the sampler. The stretch move proposes a new point that lies along the line

$$\theta_n^* = \theta_k^i + Z \cdot (\theta_n^i - \theta_k^i)$$

that connects the current walker θ_n^i and another walker in θ_k^i randomly chosen from the ensemble.

Here, Z is a random variable that is sampled following

$$g(z) = \begin{cases} \frac{1}{\sqrt{z}} & \text{if } \frac{1}{a} \leq z \leq a \\ 0 & \text{else,} \end{cases}$$

where $a > 1$ is an algorithm tuning parameter that the user must specify.

Second, the proposal is accepted or rejected using a Metropolis-Hastings-like accept-reject step.

The acceptance probability is

$$\alpha(\theta_n^* | \theta_n^i) = \min \left(1, Z^{p-1} \frac{\pi(\theta_n^*)}{\pi(\theta_n^i)} \right),$$

where Z is as defined above, and p is the dimension of θ . This formulation of the acceptance probability guarantees convergence to the target distribution [153].

We use the Matlab implementation of the AIES algorithm [174] from the UQLab toolbox [158]. Unless otherwise specified, the ensemble size is $N_e = 150$ because we observed improved sampling with a large ensemble. To accelerate sampling, the likelihood is evaluated for each ensemble member in parallel using a parallel for loop (e.g., `parfor` in Matlab) with at most 24 parallel threads. Additionally, default value of $a = 2$ is used for the stretch move tuning parameter. Lastly, each Markov chain in the ensemble is initialized to a random point drawn uniformly over the support of the prior.

Markov chain burn-in

In MCMC, Markov chains (or ensembles of chains) often display an initial transient, called *burn-in*, before converging to their stationary distributions [5], [84], [169]. Importantly, these samples during burn-in are not distributed according to the stationary distribution and should therefore be excluded from the final set of samples. Common practice in the MCMC literature is to simply discard these initial samples to remove the effects of burn-in [5], [84]. The choice of the burn-in length is often nontrivial

and is best informed by an analysis of the Markov chain [169]. Unless otherwise specified the integrated autocorrelation time (described below) dictates the number of samples to discard as burn-in in this work. Specifically, we compute the integrated autocorrelation time after collecting many samples and set the burn-in length to 5–10 times the computed value.

Convergence assessment with the integrated autocorrelation time

A key challenge in Markov chain Monte Carlo sampling is determining the appropriate number of samples N to collect. In this work, we use the integrated autocorrelation time [153], [169] to determine when the Markov chain has approximately converged to its stationary distribution. We outline the theory and motivation behind the integrated autocorrelation time for a single Markov chain and refer the reader to [153] for a discussion of ensemble methods. The use of the integrated autocorrelation time is motivated by the typical use of MCMC sampling to compute an expectation

$$\mathbb{E}[\mathbf{x}] = \int \mathbf{x} \pi(\mathbf{x}) d\mathbf{x}.$$

Given a Markov chain of length N , we can estimate the expectation with the Monte Carlo estimator

$$\hat{\mu} = \frac{1}{N} \sum_{i=1}^N \mathbf{x}^i.$$

In general it is common to consider the variance of $\hat{\mu}$, $\text{var}(\hat{\mu})$, as the estimation error [153]. This variance is given by

$$\text{var}(\hat{\mu}) = \frac{\text{var}(\mathbf{x})}{N/\tau_s},$$

where the integrated autocorrelation time τ_s is given by

$$\tau_s = \sum_{T=-\infty}^{\infty} \frac{C_s(T)}{C_s(0)}.$$

Here, the autocovariance function $C_s(T)$ with lag $T \in \mathbb{N}$ is given by

$$C_s(T) = \lim_{t^* \rightarrow \infty} \text{cov}[\mathbf{x}^{t^*+T}, \mathbf{x}^T],$$

where $t^* \in \mathbb{N}$. Thus, the Monte Carlo estimation error is proportional to integrated autocorrelation time for a fixed chain length. The integrated autocorrelation time can be interpreted as the time it takes for the samples in a Markov chain to become uncorrelated [169]. Additionally, the effective number of samples, N_{eff} , can be defined using the integrated autocorrelation time as, $N_{\text{eff}} = N/\tau_s$.

In this work, we use the integrated autocorrelation time for two purposes. First, the computation of the integrated autocorrelation time is used to choose the correct burn-in length. We compute the integrated autocorrelation time after collecting many samples and then discard between 5–10 times τ_s as burn-in. Second, after discarding the initial samples, the integrated autocorrelation time helps to determine if enough samples have been collected, e.g., N_{eff} is large. Should the effective sample size be small, the MCMC sampler is run longer to collect more samples. We compute the integrated autocorrelation time using a Matlab function associated with [175] (available at https://www.physik.hu-berlin.de/de/com/UWerr_fft.m) with default algorithm parameters. We compute τ_s for each parameter and take the maximum of these values. For an ensemble, we compute the mean τ_s for the walkers of each parameter and take the maximum value of the means. After completing MCMC and evaluating the chains, we leverage the posterior samples to quantify uncertainty in model predictions.

2.2.10 Ensemble simulation and output uncertainty analysis

Markov chain Monte Carlo sampling provides a set of samples, $\{\hat{\theta}_1, \dots, \hat{\theta}_n\}$, that converge in distribution to the posterior distribution (e.g. Fig 2.1B3). One is often interested in how uncertainty in the parameter estimates, which is conveyed by the posterior distribution, propagates to uncertainty in the model predictions. Fortunately, an ensemble of simulations (see Fig 2.1B1) distributed according to the posterior can be run using the posterior samples. This approach to uncertainty propagation is known as sampling-based uncertainty propagation [5] and is feasible because the simulation of dynamical systems biology models is computationally efficient. We refer the reader to [134], [176] for examples of sampling-based uncertainty propagation in systems biology.

Each simulation is run by solving the differential equation with the `ode15s()` integrator in Matlab with a unique sample from $\{\hat{\theta}_1, \dots, \hat{\theta}_n\}$ for the parameters. We use default tolerances for the integrator and supply the Jacobian matrix (specify the `'Jacobian'` option) to improve computations. We compute the Jacobian by hand and evaluate it with the same parameter as the differential equation.

Following ensemble simulation, statistical analyses of the ensemble of predicted trajectories and any relevant quantities of interest (QoI) can be performed. For example, Fig 2.1B2 highlights the uncertainty in the trajectory with 95% credible intervals to show the region where 95% of the trajectories fall. Additionally, we compute the statistics of relevant QoIs, such as the steady state values or limit cycle period, from the ensemble. The next section discusses how to choose single point estimates that best represents the estimated parameters.

2.2.11 Parameter point estimates from posterior samples

One often wants to compute a point estimate for each of the parameters in addition to characterizing the entire posterior distribution (black line in Fig 2.2). Common choices for point estimates in Bayesian

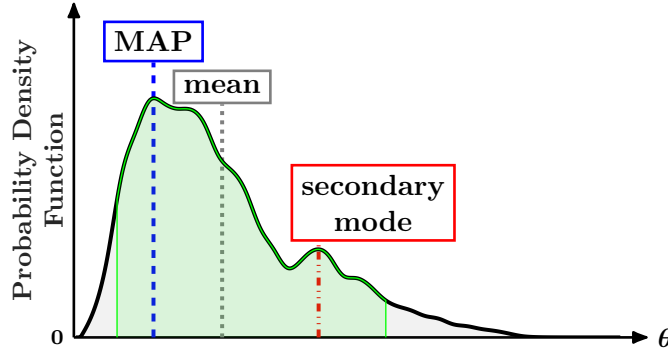


Figure 2.2: The MAP (maximum *a posteriori*) point is located at the most probable point (blue dashed line). The long tail of the distribution shifts the mean (gray dotted line) away from the MAP point. Secondary modes (red dashed-dotted line) can effect the quality of a point estimate. Additionally, the green shaded region highlights the 95% credible interval, the region between the 2.5th and 97.5th percentiles, that is used to capture the uncertainty in an estimate.

statistics include the mean, median, or mode of the posterior distribution [84]. Fig 2.2 highlights the mean and mode (denoted MAP), along with a secondary mode that may confound choosing a point estimator. The mode of the posterior distribution is a strong choice because it provides the most probable set of parameters and is often called the maximum *a posteriori* (MAP) point. However, the mean or median may provide better point estimates when the posterior is multimodal (there are multiple modes).

These point estimates are computed from posterior samples acquired via MCMC sampling. While sample statistics, such as the mean and median, are computed directly from the samples, computing the MAP point requires estimating the posterior of the probability density function and finding the maximal point. Direct computation of the sample mode will not yield the MAP point because the parameters are continuous random variables, so no two-parameter samples are expected to be identical. One approach to estimate the MAP point involves computing a histogram of samples and using the center of the bin with the most associated probability as the MAP. However, we found that the histogram of posterior samples is often noisy and can be sensitive to the bin size, so that the MAP estimate may be erroneous. We chose to take an alternative approach that fits a kernel density estimator [177] to approximate the posterior distribution and subsequently computes that MAP point. The kernel density estimator provides a

non-parametric approximation of the posterior distribution and, intuitively, smooths the posterior histogram. We use the `ksdensity()` function for kernel density estimation in Matlab. The `'support'` option is set to be the region bounded by the prior distribution and the `'BoundaryCorrection'` option is set to use the `'reflection'` method to account for these bounds. Lastly, we use the default values for the `'Bandwidth'` unless otherwise specified. All other options are kept to the defaults as defined in the Matlab documentation. The MAP point is the point with maximum probability in the `ksdensity()` output. The next section moves from the details of CIUKF-MCMC and Bayesian estimation to outline how synthetic data is generated for the examples in this paper.

2.2.12 Synthetic data generation for numerical experiments

The synthetic data in this work aims to replicate noisy data found in biological experiments. We generate noisy synthetic data by drawing samples from deterministic model simulations and simulate measurement noise by adding independently and identically distributed (iid) perturbations to each sample. First, a nominal set of biological model parameters and an initial condition are chosen for data generation. These values become the *ground truth* for the estimation problem, and are informed by the available literature when possible. Next, a numerical solution to the system provides *true* trajectories of the state variables. Unless otherwise specified the `ode15s()` integrator in Matlab is used with default tolerances, and the Jacobian matrix is supplied (specify the `'Jacobian'` option with the analytical Jacobian Matrix) to improve computations. We then apply the measurement function and sub-sample the *true* trajectory to simulate sparse sampling. Lastly, we corrupt the data by adding a realization of an iid noise stochastic process to each sample. To meet the assumptions of the CIUKF-MCMC algorithm (see Section 2.2.6) we use mean-zero, normally distributed perturbations with the diagonal covariance matrix, Γ . We chose the entries of Γ to be proportional to the variances of the respective state variables, e.g.

$$\Gamma = b \begin{bmatrix} \text{var}(x_0^1, \dots, x_n^1) & & 0 \\ & \ddots & \\ 0 & & \text{var}(x_0^d, \dots, x_n^d) \end{bmatrix},$$

where b is a positive constant typically chosen to be less than one that controls the noise level. The last section discusses how to compute the amplitude and period of a limit cycle oscillation.

2.2.13 Limit cycle analysis

Limit cycle amplitude and period are used to characterize limit cycle oscillations for global sensitivity and output uncertainty analysis. These quantities are relevant in intracellular signaling because the strength and timing of signals, amplitude and period, respectively, are thought to encode different inputs [178]. The limit cycle amplitude, y_{lca} , quantifies the difference between the maximum and minimum values of the oscillations and is defined as

$$y_{\text{lca}} := x_{\text{max}} - x_{\text{min}}, \quad (2.29)$$

where x_{min} and x_{max} denote the minimum and maximum values of the state x over a single complete oscillation. Further, the limit cycle period, y_{period} , is the time to complete an oscillation and is defined as

$$y_{\text{period}} := t^* \iff x(t) = x(t + t^*) \quad \text{for all } t. \quad (2.30)$$

To compute these quantities we find trajectories that show limit cycle oscillations and then extract the two quantities of interest. This approach leverages the `findpeaks()` function that returns the locations of the local maxima (peaks) in a trajectory for these computations in Matlab. The `findpeaks()` function can

also find the local minima by applying it to the negative of the trajectory.

The first task in computing the limit cycle features is to detect actual limit cycle oscillations. A trajectory is discarded if it reaches a fixed point (steady state) when no peaks are detected (`findpeaks()` returns an empty set). Next, the difference between the heights of the identified peaks is used to discard trajectories that show decaying oscillations to a fixed point. A threshold on one half of this difference, called for the `helpPeakThreshold`, is set to 17.0 unless otherwise specified; a trajectory is discarded if its difference in peak values exceeds this threshold. Any remaining trajectories will show limit cycles or will contain numerical artifacts which are falsely detected as limit cycles.

Lastly, the limit cycle amplitude and period are computed. The limit cycle amplitude is the mean difference between each pair of detected peaks and minima that correspond to one oscillation of the limit cycle. The limit cycle period is then computed as the mean time between two peaks and the time between two minima. Trajectories with numerical artifacts are eliminated by discarding those with a limit cycle amplitude that is smaller than the `LCAminThresh`, with a range of limit cycle values greater than the `Decaythresh` or those that return no limit cycle period (e.g., the empty matrix on Matlab). Unless otherwise specified, we set the `LCAminThresh` to 1.0 and the `Decaythresh` to 5.0. These computations assume that the period of the limit cycle is stable and that no frequency modulation occurs. We note that a small number of oscillating trajectories can be misclassified as fixed points due to numerical errors in the simulated trajectory. We specifically observed this when there was large peak-to-peak variability. However, misclassification is rare, so we do not attempt manual relabelling of these trajectories.

2.3 Results

We applied the Bayesian parameter estimation framework to three different models that represent signal transduction cascades of increasing biological and mathematical complexity. Section 2.3.1 uses

the first model, a simple kinetic scheme, to describe a series of computational experiments that illustrate the effects of measurement noise and data sparsity on estimation uncertainty (Fig 2.3). Next, we tested our framework on two models that are more representative of the nonlinearities and overparameterization observed in systems biology models. Section 2.3.2 analyzes a representative model of the mitogen-activated protein kinase (MAPK) pathway [150] that exhibits multistability depending on the choice of model parameters, and Supplemental Section 2.3.3 uses this model to illustrate the necessity of structural identifiability and global sensitivity analyses. Finally, Section 2.3.4 analyzes a simplified model of synaptic plasticity [151] that illustrates the effects of higher parameter uncertainty even in phenomenological biological representations.

2.3.1 Measurement noise and sparsity increase estimation uncertainty in a simple model of signal transduction

The first model we consider is a relatively simple two-state model from [149] shown in Fig 2.3A.

The governing equations for this model are

$$\frac{dx_1(t)}{dt} = -(k_{1e} + k_{12}) \cdot x_1(t) + k_{21} \cdot x_2(t) + b \cdot u(t) \quad (2.31a)$$

$$\frac{dx_2(t)}{dt} = k_{12} \cdot x_1(t) - k_{21} \cdot x_2(t), \quad (2.31b)$$

where the states are $\mathbf{x}(t) = [x_1(t), x_2(t)]^\top$, the biological model parameters are $\theta_f = [k_{1e}, k_{12}, k_{21}, b]^\top$, and $u(t)$ is the input function. The input function (illustrated in Fig 2.3A)

$$u(t) = \begin{cases} t + 0.5 & \text{if } 0 \leq t \leq 1 \\ 1.5e^{1-t} & \text{if } t > 1 \end{cases},$$

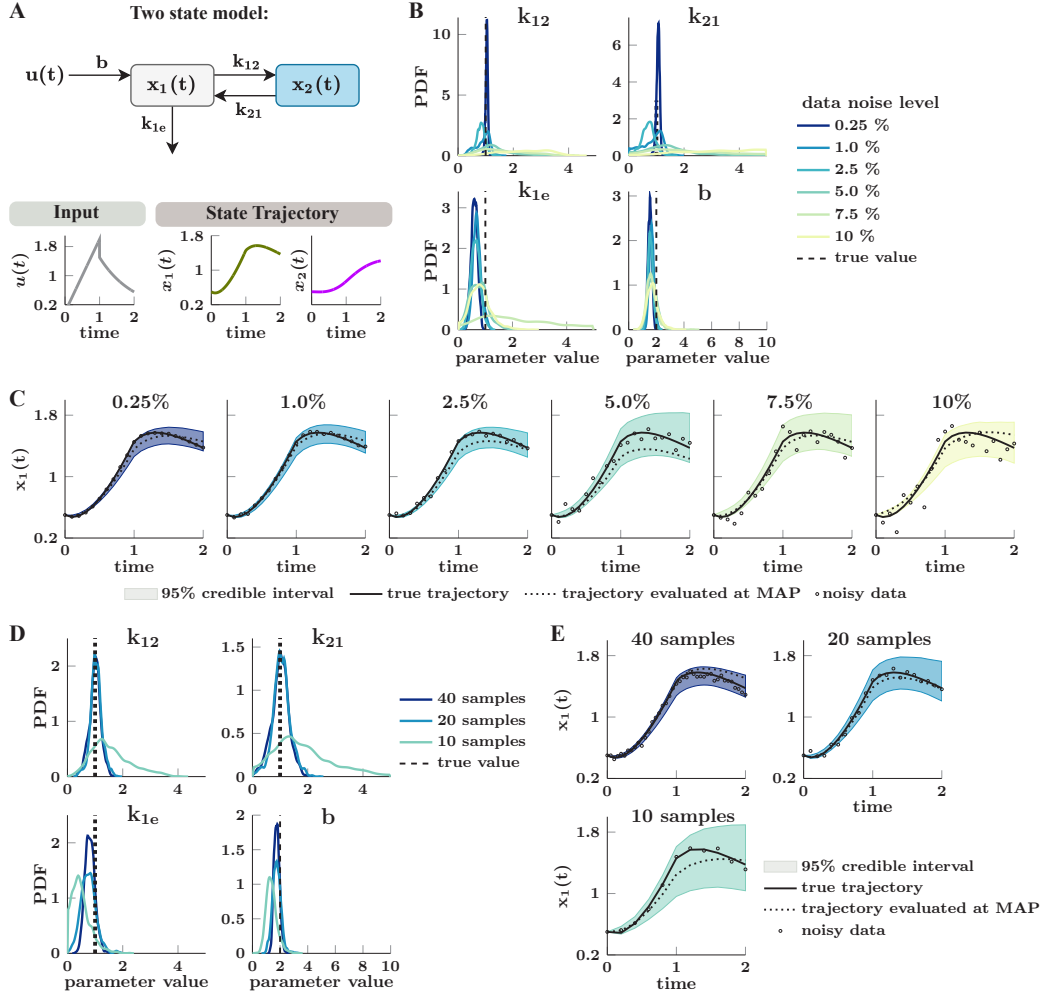


Figure 2.3: (A) *Top row:* Network diagram of the two-state model with states, $x_1(t)$, $x_2(t)$, input function $u(t)$, and four unknown parameters, $\theta_f = [k_{1e}, k_{12}, k_{21}, b]^T$. *Bottom row:* Trajectories of the input function $u(t)$ and corresponding state trajectories. The input has at least one non-zero derivative to ensure that all model parameters are globally structurally identifiable following [149]. (B) Marginal posterior distributions of the model parameters show increasing uncertainty in the parameter estimates (e.g. widening and flattening) with increasing levels of additive normally distributed measurement noise with mean zero. We control the noise level by setting the noise covariances to the specified percentage of the standard deviation of each state variable. The dashed black vertical lines indicate each parameter's nominal (true) value. Marginal posteriors are visualized by fitting a kernel density estimator to 20,000 MCMC samples obtained using CIUKF-MCMC with the delayed rejection adaptive Metropolis (DRAM) MCMC algorithm after discarding the first 10,000 samples as burn-in. (C) Posterior distributions of the trajectory of $x_1(t)$ reflect increasing parameter estimation uncertainty in panel B. The true trajectory (solid black line) shows the dynamics with the nominal parameters, dashed black lines show that trajectory with the most probable set of parameters (MAP point), and the empty circles show the noisy data at the specified noise level. The 95% credible interval shows the region between the 2.5th and 97.5th percentiles that contains 95% of the 5,000 trajectories. (D) Marginal posterior distributions of the model parameters show increasing uncertainty (widening and flattening) with increasing data sparsity (fewer samples). We simulate data sparsity by sampling the simulation from $0 \leq t \leq 2$ with three time steps, $\Delta t = 0.05$ (40 experimental samples), $\Delta t = 0.1$ (20 experimental samples) and $\Delta t = 0.2$ (10 experimental samples). Marginal posteriors are fit to 20,000 MCMC samples obtained as in panel B. (E) Posterior distributions of the trajectory of $x_1(t)$ reflect increasing parameter estimation uncertainty seen in panel (D).

ensures that all four biological model parameters are structurally identifiable because the input function has at least one nonzero derivative [149]. Supplemental Table A.1 lists the nominal parameter values and the initial condition was $\mathbf{x}_0 = [0.5, 0.5]^\top$. Sensitivity analysis was not performed on this model because the state variables are linearly dependent on the parameters.

Synthetic data with full state measurements (see Section 2.2.12) was used to perform two parameter estimation experiments—one with increasingly noisy data and another with increasingly sparse data—to investigate how measurement noise and data sparsity affect parameter and output uncertainty. The noise levels of each dataset were controlled by taking a fraction of the maximum values of the true trajectory for the corresponding state variables. In each estimation experiment, we chose the prior distributions as outlined in Section 2.2.8 and used CIUKF-MCMC (Section 2.2.6) with DRAM (Section 2.2.9) to draw 30,000 posterior samples conditioned on this noisy data. To ensure sample counts were constant across noise and sparsity levels, we chose a constant burn-in length of 10,000 samples to discard from every Markov chain (see Section 2.2.9 for details). Supplemental Fig A.2 shows the Markov chains for the measurement noise experiments, and Supplemental Fig A.3 shows those for data sparsity experiments. The following remark highlights an important distinction between data points and MCMC samples.

Remark 1. *Data points are different than MCMC samples. The experiments in this work produce at most 40 (simulated) data points, i.e., noisy measurements of the states, for parameter estimation. However, MCMC algorithms draw 10,000s-1,000,000s of sample parameter sets to characterize the posterior distribution, which requires evaluating the likelihood and, therefore, simulating the model.*

A first hypothesis we tested was that noisy data increases uncertainty because measurement noise limits the ability to constrain the dynamics of the state variables. To test this hypothesis, we performed Bayesian parameter estimation from synthetic data with increasing measurement noise (circle marks in Fig 2.3C). We observed that the marginal posterior distributions in Fig 2.3B (kernel density estimator fit

to the posterior samples as in Section 2.2.11) widen and flatten with increasing noise levels, indicating increased uncertainty. However, the most probable value for each parameter, the MAP point, lies close to the nominal parameter values (dashed lines in Fig 2.3B) for every noise level, suggesting that the data provide information about the parameter irrespective of the noise level. Additionally, Fig 2.3C shows that the width of the 95% credible interval for the dynamics of $x_1(t)$ grew as the noise increased from the lowest level (2.5%) to 5.0% and then remained similarly wide at the highest values (see Supplemental Fig A.1A for the respective trajectories of $x_2(t)$). While the uncertainty bound did not widen above the 5.0% noise level, the shape of the trajectories began to shift further from the truth (dotted line in Fig 2.3C), indicating the estimates began to take on a bias. These experiments validated our hypothesis that even in a simple dynamical system measurement noise increases estimation uncertainty of kinetic parameters.

Next, we hypothesized that data sparsity (fewer data points) would increase the uncertainty in parameter estimates. To test this, we fixed the measurement noise to the 2.5% level (see Fig 2.3) and varied the number of measurements (e.g., the sampling rate) that were included in the data used for estimation. We tested three sparsity levels with 40 experimental samples ($\Delta t = 0.05$), 20 experimental samples ($\Delta t = 0.1$) and 10 experimental samples ($\Delta t = 0.2$) over the simulation time $0 \leq t \leq 2$. Fig 2.3D highlights the widening of the estimated marginal posterior distributions for each model parameter as we decreased the number of data points (increased sparsity). Additionally, Fig 2.3E shows that the increased parameter estimation uncertainty translates to increased uncertainty (wider 95% credible interval) in the trajectory of $x_1(t)$ (see Supplemental Fig A.1B for the trajectories of $x_2(t)$). In both of these experiments, the proposed uncertainty quantification framework qualitatively and quantitatively confirmed that increasing the noise or sparsity level increases estimation uncertainty.

2.3.2 Parameter estimation for a model of the MAPK cascade

We chose a simplified model of the highly conserved mitogen-activated protein kinase (MAPK; also known as the MEK/ERK cascade) signaling pathway [179] as a second test case for our parameter estimation framework. This pathway is known to exhibit bifurcations in its dynamical behavior [178], [180]; the system can reach a stable steady state or exhibit limit cycle oscillations. We focused on a phenomenological model of the MAPK pathway from [150] (see diagram in Fig 2.4A) that includes the mixed feedback (negative and positive feedback) necessary to predict the range of dynamical behavior observed in experiments. This model has three states, $x_1(t)$, $x_2(t)$, $x_3(t)$ that correspond to phosphorylated RAF, MEK, and MAPK/ERK, respectively [35], [181] and 14 model parameters. The differential equations are

$$\frac{dx_1(t)}{dt} = k_1 \cdot (S_{1t} - x_1(t)) \cdot \left[\frac{K_1^{n_1}}{K_1^{n_1} + x_3(t)^{n_1}} \right] - k_2 \cdot x_1(t) \quad (2.32a)$$

$$\frac{dx_2(t)}{dt} = k_3 \cdot (S_{2t} - x_2(t)) \cdot x_1(t) \cdot \left[1 + \frac{\alpha \cdot x_3(t)^{n_2}}{K_2^{n_2} + x_3(t)^{n_2}} \right] - k_4 \cdot x_2(t) \quad (2.32b)$$

$$\frac{dx_3(t)}{dt} = k_5 \cdot (S_{3t} - x_3(t)) \cdot x_2(t) - k_6 \cdot x_3(t), \quad (2.32c)$$

with the biological model parameters $\theta_f = [k_1, k_2, k_3, k_4, k_5, k_6, K_1, K_2, S_{1t}, S_{2t}, S_{3t}, \alpha, n_1, n_2]^\top$. A previous analysis of the model in [150] found that it can predict three regimes of dynamical behavior that depend on the model parameters; these regimes are limit cycle oscillations, bistability and mixed multistability. Here we focus on using CIUKF-MCMC to estimate model parameters that produce two of the three dynamical regimes—bistability and limit cycle oscillations (see Fig 2.4B for example trajectories). Supplemental Tables A.2 and A.3 list the nominal parameter values and initial conditions (as defined in [150]) used to produce each of these dynamics.

First, we performed identifiability and sensitivity analysis to find the subsets of relevant parameters

to estimate each dynamical regime. The SIAN software [146] (see Section 2.2.3) showed that 12 of the 14 biological model parameters are structurally identifiable from measurements of all three state variables; however, K_1 and K_2 are only locally structurally identifiable. SIAN cannot assess parameters that appear in an exponent [130], [146], so we fixed n_1 and n_2 to their nominal values listed in Supplemental Table A.2. To avoid global versus local identifiability complications, we fixed K_1 and K_2 to their nominal values. Additionally, we omitted the total concentration parameters, S_{1t} , S_{2t} , and S_{3t} , from further analysis because we assume they would be specified according to the cell type that corresponds to the available data. As a result, we narrowed the free parameters down to a set of 9 parameters, from a set of 14 originally.

Next, we used global sensitivity analysis as described in Section 2.2.4 to further reduce the number of free parameters. We choose the quantities of interest for sensitivity analysis for the bistable and oscillatory regimes separately. The quantities of interest for the bistable regime are the steady state values of x_2 and x_3 (the values at $t = 30$ min). Those for the oscillatory regime are the limit cycle amplitude and limit cycle period (computed following Section 2.2.13). In the computations, we allowed the biological model parameters to vary uniformly over the ranges listed in Supplemental Table A.2 with 5,000 samples for the bistable cases and 15,000 samples for the limit cycle regime. Figures 2.4C and 2.4D show the computed Sobol sensitivity indices for the parameters ranked by decreasing sensitivity index. We selected the parameters with the *greatest* sensitivity indices, that is, the parameters to which the output quantities of interest are *most* sensitive. For the bistable case, we selected the four most sensitive parameters, which were k_2 , k_4 , k_5 , and k_6 . For the oscillatory case, we selected the parameters with a first-order sensitivity index S_i greater than 10^{-3} , which were k_2 , k_4 , k_6 , and α . All remaining biological model parameters were fixed to the nominal values listed in Supplemental Table A.2. Sensitivity analysis highlighted that k_2 , k_4 , and k_6 are important in predicting both dynamical regimes. Meanwhile, parameters such as k_5 and α are only important for the bistable and oscillatory regimes, respectively.

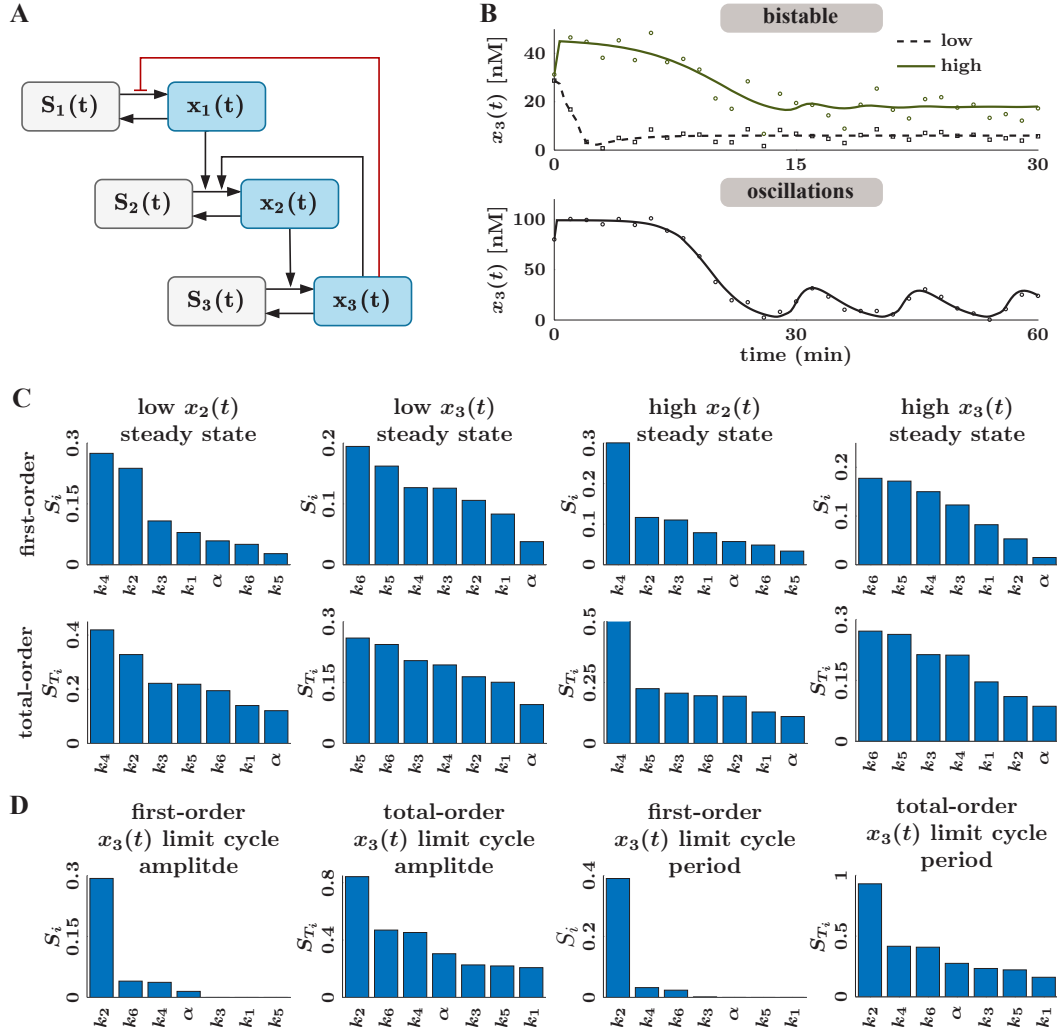


Figure 2.4: (A) Network diagram of the model of the core MAPK signaling cascade. The red line indicates inhibition; the black lines indicate activation. (B) Trajectories of $x_3(t)$ with the sets of nominal parameters that produce bistability (top) and limit cycle oscillations (bottom). The two dynamical regimes correspond to two different sets of nominal parameter values. The low (black dashed line) and high (solid green line) steady states are reached by manipulating the initial condition $\mathbf{x}_0$. The initial condition for the high steady state is $\mathbf{x}_{0,\text{high}} = [0.1245, 2.4870, 31.2623]^\top$ and that for the low steady state is $\mathbf{x}_{0,\text{low}} = [0.0015, 3.6678, 28.7307]^\top$. (C and D) Sobol sensitivity for the MAPK model parameters. All parameters except the total concentrations, S_{1t} , S_{2t} and S_{3t} , exponents, n_1 and n_2 , and locally identifiable K_1 and K_2 , are varied uniformly over the identified ranges (see Supplemental Table A.2). We use 5,000 and 15,000 samples for the bistable and limit cycle regimes, respectively. (C) Sensitivity indices for bistable behavior dynamics. We use the steady state value of $x_2(t)$ and $x_3(t)$ for both the high and low steady states as quantities of interest. By selecting the two most sensitive parameters for the four quantities of interest, we reduce the set of free parameters to $\theta_f = [k_2, k_4, k_5, k_6]^\top$. (D) Sobol sensitivity indices for a set of free parameters that contribute to limit cycle behavior. We show the first-order sensitivity indices S_i and the total-order indices S_{Ti} for the limit cycle amplitude and period of $x_3(t)$. We reduce the number of free parameters by selecting those with $S_i > 10^{-3}$ across both output quantities, that is, $\theta_f = [k_2, k_6, k_4, \alpha]^\top$.

After identifiability and sensitivity analyses, we applied the CIUKF-MCMC method to estimate model parameters that predict the correct steady state in the bistable case. To simulate noisy experimental data, we generated two synthetic datasets (see Section 2.2.12), sampled from the high steady state and the low steady state (circle and square markers in Fig 2.4B). Each dataset had 30 full-state measurements evenly spaced over $0 \leq t \leq 30$ (min) with measurement noise covariances set to 2.5% of the variances of the true trajectories. The prior assumptions were specified according to Section 2.2.8 for all model parameters with the bounds listed in Supplemental Table A.2. Using the CIUKF-MCMC algorithm with AIES (Section 2.2.9), we ran an ensemble of 150 Markov chains with 3,500 samples per chain (525,000 total samples) for each of the datasets (as shown in Supplemental Fig A.10A-B). The maximum integrated autocorrelation times (Section 2.2.9) were 120.06 for the low steady state and 169.03 for the high steady state, leading us to discard 840 samples for the low steady state and 1,183 samples for the high steady state as burn-in.

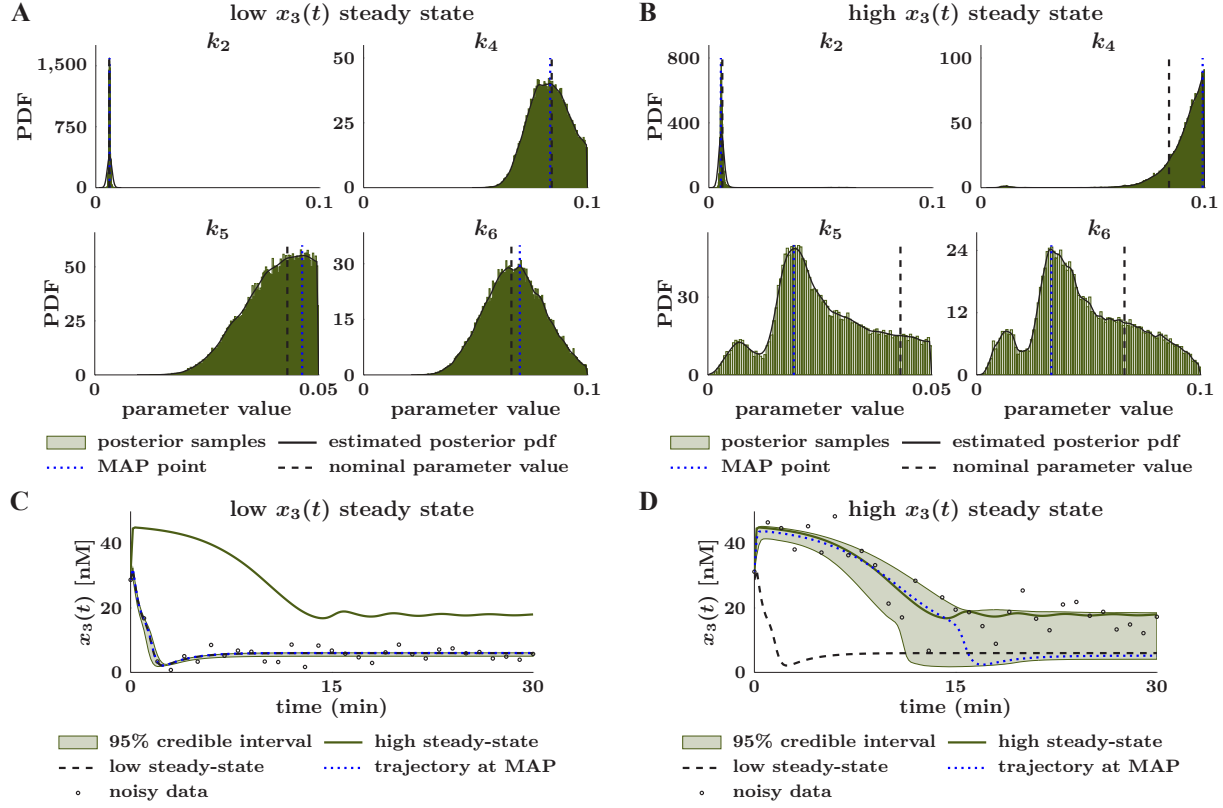


Figure 2.5: (A) Marginal posterior distributions of the model parameters for parameter estimation from noisy data of the low steady state. Posterior distributions are visualized by fitting a kernel density estimator to 325,200 (150 walkers with 2,660 steps each) MCMC samples obtained using CIUKF-MCMC with the affine invariant ensemble sampler (AIES) for MCMC after discarding the first 840 samples per walker as burn-in. (B) Marginal posterior distributions of the model parameters for parameter estimation from noisy data of the high steady state reveal larger uncertainty in the model parameters when compared to the low steady state. We visualize distributions by fitting a kernel density estimator to 347,700 (150 walkers with 2,644 steps each) MCMC samples obtained using CIUKF-MCMC with the affine invariant ensemble sampler (AIES) for MCMC after discarding the first 856 samples per walker as burn-in. (C) Posterior distribution of the trajectory of $x_3(t)$ with initial conditions that yield the low steady state highlights low uncertainty in the predicted dynamics. The true trajectory (dashed black line) shows the dynamics with the nominal parameters, the dotted blue line shows the trajectory evaluated at the MAP point, and the empty circles show the noisy data (covariance is 50% of the standard deviation of the true trajectory). The 95% credible interval shows the region between the 2.5th and 97.5th percentiles that contains 95% of 30,000 posterior trajectories. (D) Posterior distribution of the trajectory of $x_3(t)$ with initial conditions that yield the high steady state highlights the ambiguity between which steady state is reached. All lines and computations are the same as in panel (A), except simulations were run using an initial condition that results in the high steady state.

The estimated marginal posterior distributions (solid black line; Fig 2.5A and Fig 2.5B for the low and high steady states, respectively) indicated varying levels of uncertainty between the model parameters and across the two steady states. For example, in both the low and high steady states, the marginal posterior for k_2 has most of its probability mass centered around the nominal value (dashed black lines), while that for k_6 has probability mass spread over a broader range of the prior support (range of the prior bounds). Additionally, the MAP points (dotted blue line) for the low steady state closely correspond to the nominal values for all model parameters, whereas there is a significant discrepancy between the MAP and the nominal values for k_4 , k_5 and k_6 in the high steady state case.

An ensemble of 30,000 simulations with randomly selected posterior samples (see Section 2.2.10) represented the posterior distributions of the dynamics for both steady states. For the low steady state, the trajectory evaluated at the MAP point (dotted blue line in Fig 2.5C) closely matches the true trajectory (dashed black line) and the 95% credible interval (green shaded region) tightly constrains these trajectories. However, the trajectory at the MAP point for the high steady state (Fig 2.5D) reaches the low steady state rather than the high steady state (solid green line). Furthermore, the 95% credible interval for the high steady state closely follows the initial transient ($0 \leq t \leq 10$ (min)), but it covers both steady states by the end of the simulation, e.g., for $10 \leq t$ (min). The considerable uncertainty and bias in the estimated dynamics of the high steady state are unsurprising, given the uncertainty observed in the marginal posterior distributions. This comprehensive uncertainty analysis of the bistable MAPK dynamics showed that the presence of multiple steady states makes parameter estimation harder for the same set of parameters and governing equations. In particular, the estimation uncertainty is much lower when data from the low steady state is supplied for estimation than when data from the high steady state is used.

Next, we used CIUKF-MCMC to estimate posterior distributions for the reduced set of model parameters that predict limit cycle oscillations in $x_3(t)$. Synthetic data with 15 samples evenly spaced over

$30 \leq t \leq 60$ (min) simulated noisy measurements from the oscillating trajectory at a noise level of 1% of the variance of the true trajectory (circle markers in Fig 2.7A). We refer to this data as the *oscillations only* data, because it only includes samples from the limit cycle and not from the initial transient. Using CIUKF-MCMC with AIES, we ran 30 Markov chains (shown in Supplemental Fig A.11) with 51,000 steps per chain to sample the parameter posterior distributions. We discarded 7,477 samples per chain, seven times the integrated autocorrelation time of 1,068, to account for burn-in. The marginal posterior distributions estimated from the remaining posterior samples (Fig 2.6A) tightly concentrate the probability around the nominal parameter values. This high level of certainty in the parameter values leads to a posterior distribution of $x_3(t)$ (represented with an ensemble of 30,000 simulations in Figure 2.6B) that closely follows the true limit cycle oscillations (see Supplemental Fig A.4C for those of $x_1(t)$ and $x_2(t)$). Additionally, the trajectory evaluated at the MAP point (dotted blue line) follows the true trajectory (dashed black line) and shows very similar oscillations.

Closer examination of a subset of 50 out of the 30,000 posterior trajectories (green lines in Fig 2.6C) revealed that the 95% credible interval includes many limit cycles that are similar to the true trajectory (dashed black lines), a smaller number of limit cycles that do not match the true oscillations, and yet an even smaller number of trajectories that reach fixed points (for example, the dotted blue line in Fig 2.6C). We further leveraged the posterior samples to quantify the variability in the predicted limit cycles. First, we characterized all of the trajectories in the ensemble (Fig 2.6D) and found that 90.6% (27,182 samples) correctly produce limit cycle oscillations while 9.6% (2,818 samples) reach a fixed point. Quantification of the limit cycle amplitude and period for each of the 27,182 oscillating trajectories showed that despite the small uncertainties in model parameters, we still observe variability in the characteristics of predicted limit cycles. We find that the most probable limit cycle amplitude and period (see histograms in Fig 2.6E) are close to the true values (26.12 for the amplitude and 13.04 for the period; indicated by

vertical dashed black lines). However, we still observe variability in these quantities despite the high level of certainty in the parameter estimates. Specifically, we find that the limit cycle amplitude often deviates

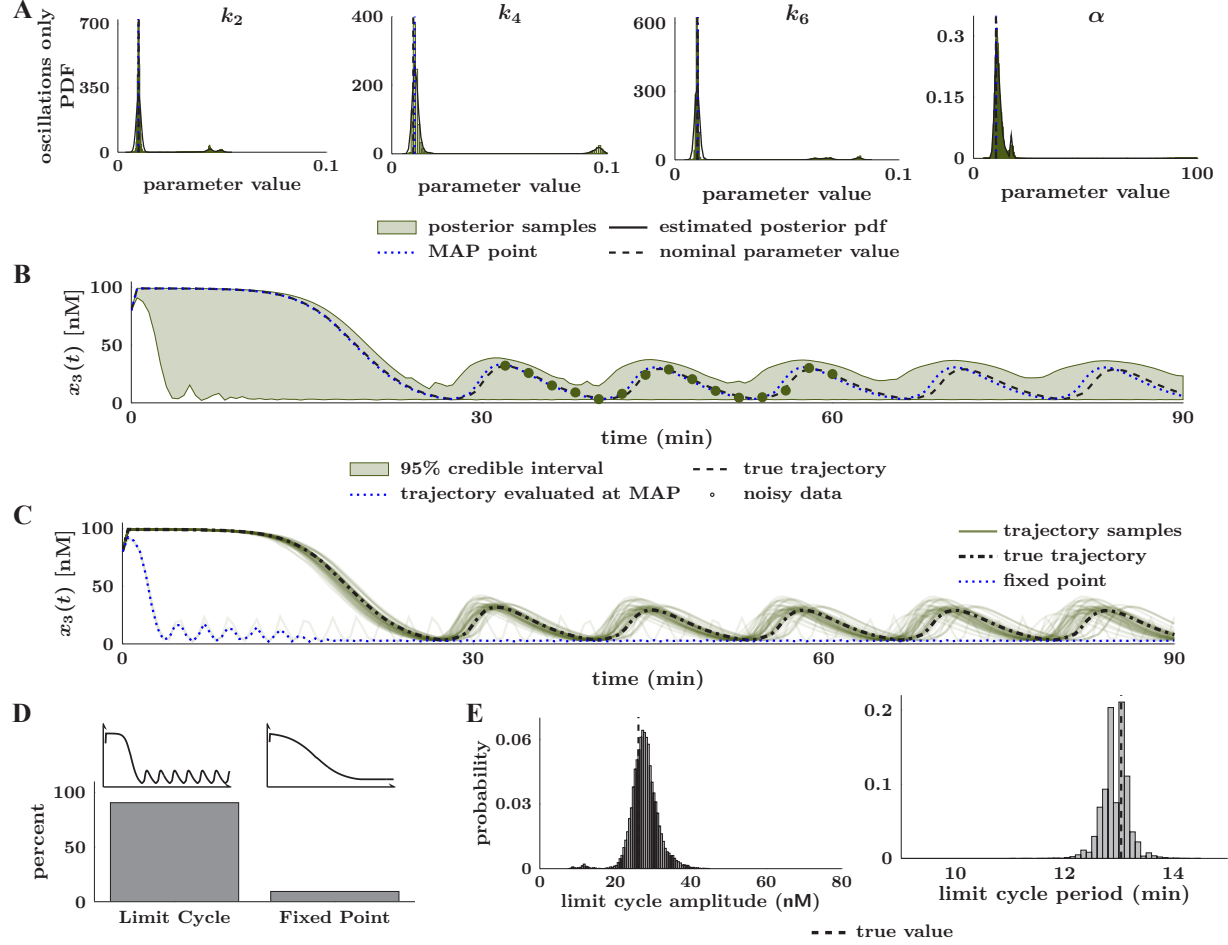


Figure 2.6: (A) Marginal posterior distributions of the model parameters. Distributions are visualized by fitting a kernel density estimator to 1,305,720 (30 walkers with 43,524 steps each) MCMC samples obtained using CIUKF-MCMC with the affine invariant ensemble sampler (AIES) after discarding the first 7,447 samples per walker as burn-in. (B) Posterior distribution of the trajectory of $x_3(t)$ in the limit cycle regime. The true trajectory (dashed black line) shows the dynamics with the nominal parameters, the dotted blue line shows the trajectory evaluated at the MAP point, and the points show the noisy data (covariance is 1% of the variance of the true trajectory). The 95% credible interval shows the region between the 2.5th and 97.5th percentiles that contains 95% of 30,000 posterior trajectories. (C) Sample posterior trajectories (50 out of 30,000 total) reveal that most trajectories closely match the true limit cycles. Additionally, several trajectories that reach a fixed point are shown. (D) Quantification of the fraction of the 30,000 sample trajectories that produce limit cycles oscillations, 90.6% (27,182 samples), or reach a fixed point, 9.4% (2,818 samples). (E) Histograms quantify the variability in limit cycle amplitude and period for the 27,182 trajectories that show limit cycle oscillations. We define the limit cycle amplitude as the peak-to-peak difference for one oscillation, and the period is the time to complete an oscillation. The vertical black lines show these quantities for the true trajectory.

by up to 30% (10 nM) from the true amplitude. While the period deviates by a smaller amount, up to 15% (2 min), we find that its distribution is skewed to periods that are shorter than the true period. Based on these findings, we conclude that small uncertainties in the model parameters result in limit cycles with varied amplitudes and periods, but do not strongly affect our ability to predict oscillatory dynamics.

Next, we investigated whether the nature of the data supplied for estimation affects our ability to predict limit cycle oscillations. To investigate this, we repeated parameter estimation with two additional noisy limit cycle data sets (shown in Fig 2.7A with the noise level set to 1% of the variance of the true trajectory). We refer to these new data as the *equidistant sampling* data and the *non-equidistant sampling* data. The equidistant sampling data includes 20 samples taken every two minutes over $0 < t \leq 60$ minutes and covers the initial transient and several periods of oscillations (see Fig 2.7A). The non-equidistant sampling data also covers the first 60 minutes, however it only has five samples taken every five minutes for the initial transient (the first 30 minutes) and the same 15 samples taken every two minutes for the remaining 30 minutes (20 samples in total). We compared this to the results for the oscillations only data (Fig 2.6) to investigate how samples from the initial transient affects estimation. We performed MCMC with 150 walkers for 22,500 steps for the equidistant sampling data (Supplemental Fig A.12) and 30 AIES walkers for 36,500 steps for the non-equidistant sampling data (Supplemental Fig A.13). To account for the MCMC burn in, we discarded the first 5,669 samples per chain for the equidistant data and 11,065 samples for the non-equidistant data. Additionally, we simulated ensembles of 30,000 trajectories with random posterior samples for the two new data sets; Supplemental Fig A.5 shows the resulting posterior distributions of the dynamics.

Comparison of the fraction of the 30,000 simulated trajectories that show oscillations clearly indicates that the input data affects our ability to predict limit cycles (see Fig 2.7B). As a baseline, we compare the three data sets to simulations with samples from the prior distribution, which we think of

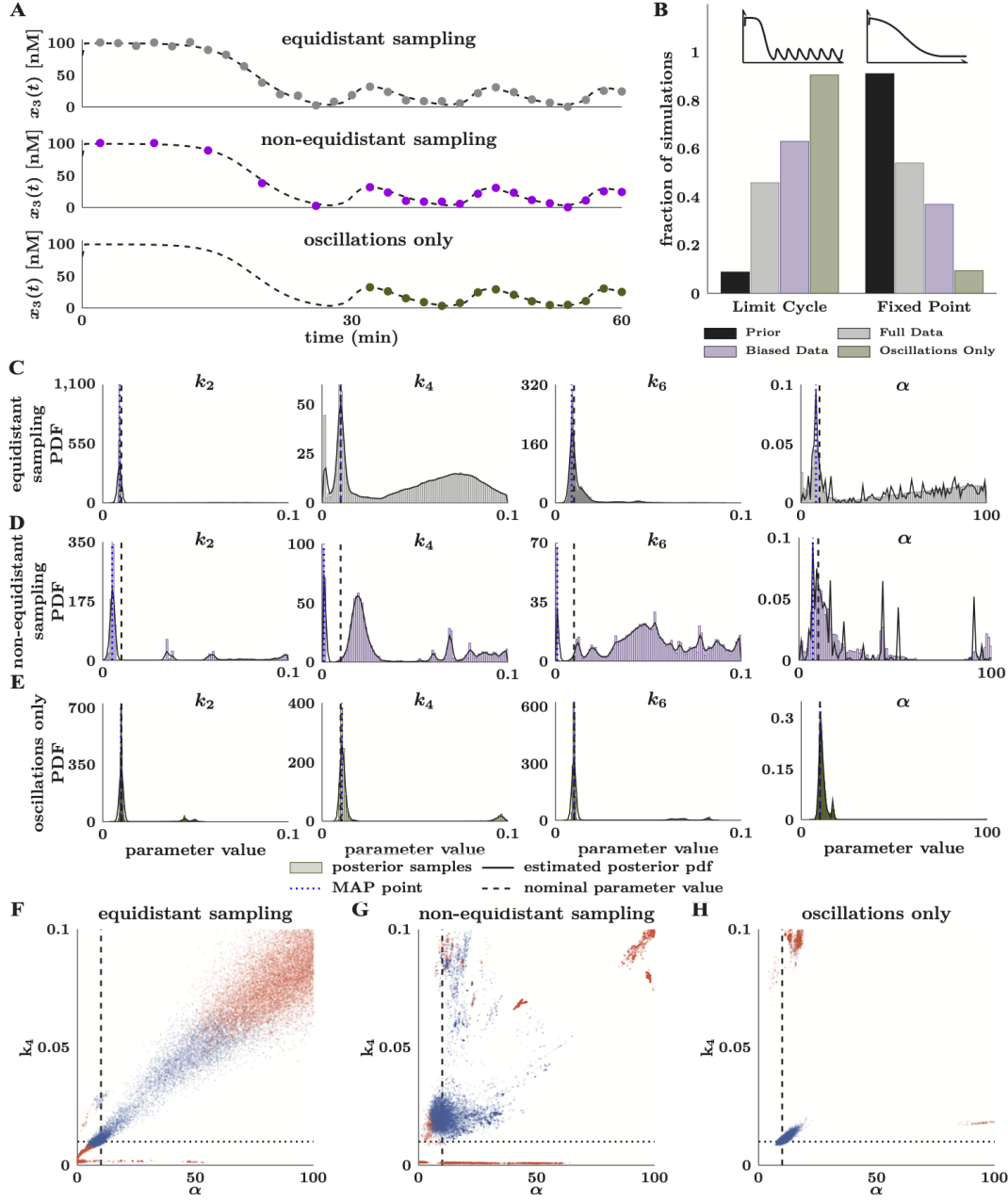


Figure 2.7: (A) The *equidistant sampling* data includes 30 samples taken every 2 minutes over $0 < t \leq 60$ (min). The *non-equidistant sampling* data includes 20 total samples with two sampling rates; there are 5 samples taken every 5 minute for the first 30 minutes and 15 samples taken every 2 minutes for 30 additional minutes. The *oscillations only* data set only includes samples from the oscillations with 15 samples taken every two minutes over the interval $t \in (30, 60]$. (B) The fraction of the 30,000 simulations that yield limit cycle or fixed point trajectories, with parameter samples from the prior, and the posterior distribution associated with each data set. (C–E) Marginal posterior distributions of the model parameters. Distributions are visualized by fitting a kernel density estimator to 2,524,800 samples for the equidistant sampling data, 763,080 for the non-equidistant sampling data, and 1,305,720 for the oscillations only data. (F–H) Two-dimensional scatter plots reveal relationships between k_4 and α that are necessary to produce limit oscillations. Simulations with blue points produce limit cycle oscillations, and those with red points produce fixed points. Darker regions indicate a higher probability of observing the corresponding parameter values.

as the setting where no data are provided. Figure 2.7B shows that all cases with data predict a higher fraction of limit cycle trajectories than the prior samples, where only 17.6% of simulations oscillated. We found that only 45.9% (13,794 samples) of the simulations corresponding to the equidistant sampling data show sustained oscillations. However as the number of samples from the initial transient (data from 0 to 30 minutes) is decreased, the fraction of oscillating simulations increased; 63.1% (18,921 samples) and 90.6% (27,182 samples) of the simulations corresponding to the non-equidistant sampling and the oscillation only data, respectively, showed limit cycle oscillations.

We additionally found that uncertainty in the nature of the dynamics is closely tied to the uncertainty in the model parameters. Figure 2.7C-E shows the marginal posterior distributions corresponding to the equidistant sampling, non-equidistant sampling, and oscillations only data, respectively. Specifically, we observed that we can always predict k_2 with high certainty, but the amount of uncertainty in k_4 , k_6 and α varies between the three data sets. For example, for the equidistant sampling data, the marginal distribution for k_4 (Fig 2.7C) has three distinct peaks of high probability (modes); one concentrated at very small values, another centered near the nominal value and a wider mode at higher values. The k_4 marginal distribution for the non-equidistant sampling data in Fig 2.7D places less probability at the two modes far from the nominal value, but the mode near the nominal value is shifted towards higher values of k_4 . Lastly, for the oscillations only data, we see only one prominent mode for k_4 that is very close to the nominal value (Fig 2.7E). Similar trends are seen for the α marginal distributions. The posterior distribution of dynamics in Supplemental Fig A.5 and the distributions of the limit cycle characteristics in Supplemental Fig A.6 show that the increased uncertainty in these model parameters increases variability in the predicted dynamics. In cases with high uncertainty in the model parameters, the 95% credible intervals for $x_3(t)$, for example in Supplemental Fig A.5A, do not tightly constrain the dynamics, and the limit cycle characteristics vary more (Supplemental Fig A.6A).

Lastly, we used the results of our Bayesian analysis to discover relationships between parameter pairs and the nature of the corresponding predicted dynamics. Based on observations that increased uncertainty in both k_4 and α results in higher fractions of incorrectly predicted fixed point trajectories and prior knowledge of how the model parameters interact to control the interplay of positive and negative feedback [150], we hypothesized that we might observe correlations between pairs of model parameters and the predicted dynamics. To test this, we plotted two-dimensional scatter plots of all pairs of parameters for each data set in Supplemental Fig A.7, Supplemental Fig A.8, and Supplemental Fig A.9, and for k_4 and α in Fig 2.7F–H. The points are colored blue if the corresponding trajectory produced a limit cycle and red if the trajectory reached a fixed point. We observed the strongest correlation between k_4 and α , where, for the equidistant sampling data, (see Fig 2.7F), the values of both parameters needed to be above the smallest values, but not too large, to produce oscillations. The plots for the non-equidistant sampling data (Fig 2.7G) and oscillations only data (Fig 2.7H) further highlight the relationship between k_4 and α . Based on these results, we conclude that k_4 and α are correlated such that their values must be within specific ranges to produce limit cycles. We note that we were able to discover this correlation through a Bayesian analysis; while, it is visually most apparent for the equidistant sampling data due to the high estimation uncertainty, it is a property of the model itself and not of the data.

In summary, comprehensive UQ for the MAPK model highlighted how the existence of multi-stability introduces additional uncertainties into parameter estimation. Specifically, sensitivity analysis identified two parameters of the MAPK model, k_5 and α , that were only important for the bistable dynamics and the oscillatory dynamics, respectively. In both cases, Bayesian parameter estimation was able to predict the correct type of dynamics but showed remaining uncertainty in the specific characteristics of the dynamics. By varying the data supplied for estimation, we found that the nature of the data set provided to CIUKF-MCMC affects our ability to predict limit cycles. Interestingly, we found that more data does

not always improve estimation and that having less data that is more focused on the limit cycles reduced estimation uncertainties and gave better posterior parameter distributions. Additionally, we discovered that the values of both k_4 and α need to be within a certain range to predict oscillatory MAPK dynamics. Overall, the proposed framework for comprehensive UQ found parameters that were important to each dynamical regime and directly quantified how uncertainties in these parameters contributes to uncertainty in the dynamics.

2.3.3 Structural identifiability and global sensitivity analyses are key to successful parameter estimation

In the previous section, we applied structural identifiability and global sensitivity analysis to reduce the number of parameters to be estimated. We found that this preprocessing was necessary, because failing to reduce the parameter set to the identifiable and influential parameters led to a more difficult estimation problem and increased parameter uncertainty. To demonstrate these effects, we varied the preprocessing analysis that we used to reduce the set of free parameters for the MAPK model in the bistable regime. In particular, we tested two additional cases; one where we only used structural identifiability analysis to reduce the parameter space and another where we did not perform any preprocessing. Based on our previous experience, we predicted that parameter uncertainty would increase as we eliminated each parameter reduction preprocessing step.

Based on the marginal parameter distributions in Figure 2.8, we conclude that structural identifiability and global sensitivity analysis are necessary for practical parameter estimation and uncertainty quantification. We estimated all parameters from the same data set generated from the MAPK low steady state in the bistable regime (results shown in Fig 2.5). The posterior distributions of the reduced set of free model parameters $\theta_f = [k_2, k_4, k_5, k_6]^\top$ that we found by applying the complete proposed framework

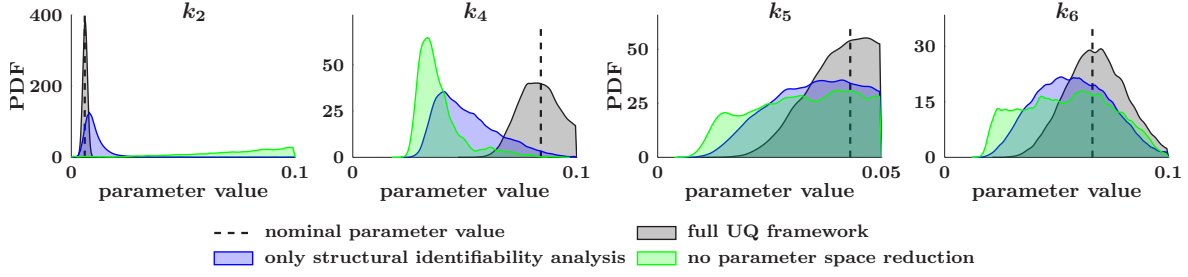


Figure 2.8: We visualize all distributions by fitting a kernel density estimator to MCMC samples. The grey-shaded distributions, labeled “full UQ framework,” are the posterior distributions for the estimated parameters after applying the entire proposed UQ framework to reduce the free parameters to $\theta_f = [k_2, k_4, k_5, k_6]^\top$. The blue-shaded distributions, labeled “only structural identifiability analysis,” are the posterior distributions after using only structural identifiability analysis to reduce the parameter set to $\theta_f^{\text{no GSA}} = [k_1, k_2, k_3, k_4, k_5, k_6, \alpha]^\top$. The green-shaded distributions, labeled “no parameter space reduction,” are the posterior distributions for the full set of free parameters $\theta_f^{\text{no reduction}} = [k_1, k_2, k_3, k_4, k_5, k_6, K_1, K_2, \alpha]^\top$. We only show the distributions for the four parameters that are common to all three analyses. In each case, we used CIUKF-MCMC with AIES to draw 3,500 MCMC samples with 150 walkers. We discarded 840, 1,342, and 1,485 samples per walker as burn-in for the full framework, structural identifiability only, and no reduction cases, respectively.

(black line with grey shading) have peaks well-aligned with the nominal parameter values. Repeating the estimation with the set of free parameters $\theta_f^{\text{no GSA}} = [k_1, k_2, k_3, k_4, k_5, k_6, \alpha]^\top$ that we found by only using structural identifiability analysis for parameter reduction yields estimated posterior distributions (blue line and shading) with much greater uncertainties. Here, the peaks of the distributions for k_2 and k_4 do not align with the nominal values, and all four distributions are wider than in the previous case. Lastly, estimation of the full set of free model parameters (no preprocessing to reduce the parameter set) $\theta_f^{\text{no reduction}} = [k_1, k_2, k_3, k_4, k_5, k_6, K_1, K_2, \alpha]^\top$ resulted in posterior distributions with even greater uncertainty which indicate lack of MCMC convergence. In this case, the posterior distributions place little probability on the nominal values of k_2 and k_4 ; thus, the predictions for these parameters would be incorrect. We used AIES for MCMC with 3,500 samples and 150 walkers for all three cases. These standard MCMC settings highlight how reducing the parameter space with structural identifiability and global sensitivity analyses lead to a better-posed estimation problem. Overall, each preprocessing step improves MCMC convergence and reduces uncertainty in the estimated parameters.

2.3.4 Parameter estimation in a phenomenological model of long-term potentiation/depression

A phenomenological model of coupled kinase and phosphatase switches whose activities affect the level of membrane-bound AMPAR (alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor) as a reporter of synaptic plasticity was proposed in [151] to capture the key events in synaptic plasticity. This kinase-phosphatase model has three states $pK(t)$, $P(t)$, and $A(t)$ that correspond to active forms of kinase (CaMKII in [151]), phosphatase (PP2A in [151]), and membrane-bound AMPAR, respectively. The differential equations are

$$\begin{aligned} \frac{dpK(t)}{dt} = & k_1 \cdot pK(t) \cdot \frac{K_{\text{tot}} - pK(t)}{K_{m_1} + (K_{\text{tot}} - pK(t))} - k_2 \cdot (P(t) + P_0) \cdot \frac{pK(t)}{K_{m_2} + pK(t)} \\ & + k_3 \cdot K_0 + k_4 \cdot (K_{\text{tot}} - pK(t)) \cdot \frac{Ca^{2+}(t)^4}{K_{m_3}^4 + Ca^{2+}(t)^4} \end{aligned} \quad (2.33a)$$

$$\begin{aligned} \frac{dP(t)}{dt} = & k_5 \cdot P(t) \cdot \frac{P_{\text{tot}} - P(t)}{K_{m_4} + (P_{\text{tot}} - P(t))} - k_6 \cdot (pK(t) + K_0) \cdot \frac{P(t)}{K_{m_5} + P(t)} \\ & + k_7 \cdot P_0 - k_8 \cdot (P_{\text{tot}} - P(t)) \cdot \frac{Ca^{2+}(t)^3}{K_{m_3}^3 + Ca^{2+}(t)^3} \end{aligned} \quad (2.33b)$$

$$\frac{dA(t)}{dt} = (c_1 \cdot pK(t) + c_3) \cdot (A_{\text{tot}} - A(t)) - (c_2 \cdot P(t) + c_4) \cdot A(t), \quad (2.33c)$$

where the 24 biological model parameters are $\theta_f = [k_1, k_2, k_3, k_4, k_5, k_6, k_7, k_8, K_{m_1}, K_{m_2}, K_{m_3}, K_{m_4}, K_{m_5}, K_0, P_0, K_{\text{tot}}, P_{\text{tot}}, A_{\text{tot}}, c_1, c_2, c_3, c_4]^\top$. The nominal values and physiological ranges for these parameters are listed in Supplemental Table A.4.

This model predicts tristability (three steady states) in the level of excitatory postsynaptic potential (EPSP) as a function of the calcium $Ca^{2+}(t)$ input. The normalized EPSP is the membrane-bound AMPAR $A(t)$ level, normalized to the initial condition, e.g., normalized EPSP = $A(t)/A(t=0)$, as defined in [151]. Fig 2.9B shows simulations of the three expected responses with the nominal parameter values from [151].

The initial condition $\mathbf{x}_0 = [0.0228, 0.0017, 0.4294]^\top$ used for all simulations was determined by allowing the system to reach steady state with the baseline $Ca^{2+}(t)$ level of $Ca^{2+}(t) \equiv 0.1 [\mu\text{M}]^\top$. The three steady states are the initial baseline, the higher long-term potentiation state (LTP; trajectory depicted in dashed black Fig 2.9B), and the lower long-term depression state (LTD; trajectory depicted in solid green Fig 2.9B). The LTP state is obtained by applying a constant stimulus of $Ca^{2+}(t) \equiv 4.0 [\mu\text{M}]$ from $1 \leq t \leq 3$ (sec), while the LTD state is reached by applying a constant stimulus of $Ca^{2+}(t) \equiv 2.2 [\mu\text{M}]$ in the same time interval; $Ca^{2+}(t)$ is set to the baseline level before ($t < 1$ (sec)) and after ($t > 3$ (sec)) the stimulus is applied. We investigated how well the proposed uncertainty quantification framework could estimate the model parameters for LTP and LTD from synthetic data of an LTP-inducing calcium input.

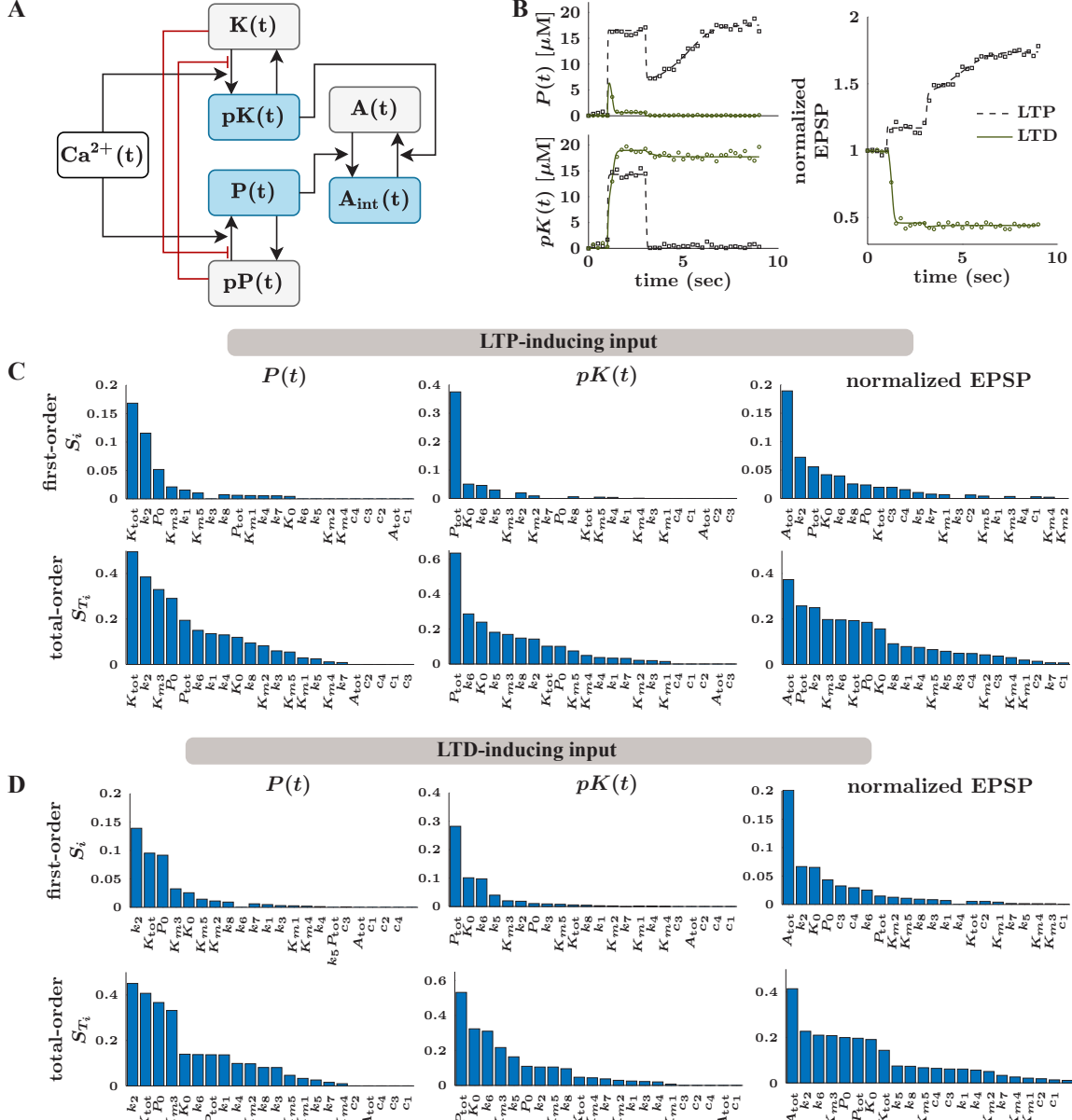


Figure 2.9: (A) Network diagram of the simplified coupled kinase-phosphatase signaling model where calcium $Ca^{2+}(t)$ acts as the input. (B) Trajectories of the three state variables in response to long-term potentiation (LTP; pulse of $Ca^{2+}(t) \equiv 4.0$ [μM] from $2 \leq t \leq 3$ (sec)) and long-term depression (LTD; pulse of $Ca^{2+}(t) \equiv 2.2$ [μM] from $2 \leq t \leq 3$ (sec)) inducing calcium inputs. The calcium level is set to a baseline of $Ca^{2+}(t) \equiv 0.1$ [μM] before and after stimulus. We compute normalized EPSP by normalizing $A(t)$ to its initial condition as described in [151]. The synthetic noisy data for the LTP and LTD cases are indicated by the black square and green circle marks, respectively, with the noise covariance equal to 1% of the variance of the data. (C and D) Sobol sensitivity indices for all free model parameters in response to LTP-inducing and LTD-inducing inputs, respectively. The quantities of interest are the steady state values of each state variable. We show both the first-order sensitivity indices S_i and the total-order indices S_{T_i} . We select a reduced set of free parameters by choosing the parameters whose first-order sensitivity index is greater than 0.05, e.g., $S_i > 0.05$. This gives us the same set of free parameters, $\theta_f = [k_2, k_6, K_0, P_0, K_{tot}, P_{tot}, A_{tot}]^T$, for both the LTP and LTD cases. Remaining model parameters are fixed to the nominal values in Supplemental Table A.4.

Following the proposed framework, the parameter space is reduced by performing identifiability and sensitivity analysis. First, identifiability analysis showed that all model parameters are globally structurally identifiable from full-state measurements. Next, global sensitivity analysis of the steady states in response to an LTP-inducing and an LTD-inducing input ranked the 22 globally identifiable model parameters. All free parameters were uniformly varied over two orders of magnitude centered around the nominal values, θ_i^* , for each parameter, e.g. $\theta_i \sim \mathcal{U}(0.1 \cdot \theta_i^*, 10 \cdot \theta_i^*)$. In order to maintain conservation of mass, the lower bounds of the total concentration parameters, $K_{\text{tot}}, P_{\text{tot}}, A_{\text{tot}}$, were chosen to be greater than or equal to the initial condition. Figure 2.9C shows the computed Sobol sensitivity indices for the LTP-inducing input and Fig 2.9D shows those for the LTD-inducing input. The sensitivity analyses point to the same group of seven model parameters for both the LTP-inducing and LTD-inducing inputs. These are $\theta_f = [k_2, k_6, K_0, P_0, K_{\text{tot}}, P_{\text{tot}}, A_{\text{tot}}]^\top$, whose first-order indices were greater than 0.05, e.g., $S_i > 0.05$. We chose to estimate these seven parameters and fix the remaining model parameters to the nominal values listed in Supplemental Table A.4.

Next, we used the CIUKF-MCMC algorithm with AIES to estimate the posterior distribution for the reduced set of parameters. The parameters were estimated from noisy synthetic data with 36 full-state measurements of the LTP response. The data are spread uniformly over the domain $0 \leq t \leq 9$ (sec) at a noise level of 1% of the variance of the true trajectory for the respective states. The maximum integrated autocorrelation time of the ensemble 150 Markov chains with 8,000 steps per chain was 621.58, leading us to discard 4,351 samples as burn-in. Traces of the ensemble of Markov chains for all parameters are shown in Supplemental Fig A.14. Figure 2.10A shows the estimated marginal posterior distributions for the free model parameters, $k_2, k_6, P_0, K_0, P_{\text{tot}}, K_{\text{tot}}, A_{\text{tot}}$. We observed different levels of uncertainty across the estimated parameters. For instance, the marginal posterior for A_{tot} indicated a very high level of certainty with almost all of the probability mass, and thus the MAP point, aligned to the nominal

value. However, the marginal posterior distributions for P_0 and K_0 show much more significant uncertainty because we observe posterior probability spread over the entire support of the prior. Additionally, the marginal posterior distributions for k_2 , k_6 , P_{tot} , and K_{tot} show a large mode around the MAP point that is shifted from the nominal value and a smaller mode at the nominal value.

Using the posterior samples, an ensemble of 30,000 simulations (see Section 2.2.10) represented the posterior distribution of the predicted dynamics in response to an LTP-inducing input, as shown in Fig 2.10B. We observed that the 95% credible interval for the normalized EPSP covered both LTP (normalized EPSP > 1) and LTD (normalized EPSP < 1) responses even though the input was LTP-inducing. However, the trajectory evaluated at the MAP point (dotted blue line) matched the true trajectory (dashed black line), indicating that most trajectories align with the expected LTP response. Examination of the individual trajectories within the ensemble simulation (Fig 2.10C, top row) and the normalized EPSP steady state values (Fig 2.10D, top row) confirmed that there are both LTP (blue traces) and LTD (black traces) responses to the LTP-inducing input. Specifically, we found that 76.59% (22,972 samples) of the responses correctly predict LTP, and 23.41% (7,024 samples) of response incorrectly predict LTD (Fig 2.10E). Therefore, despite the high-quality data (many measurements and low noise), we still observed substantial uncertainty in the predicted normalized EPSP.

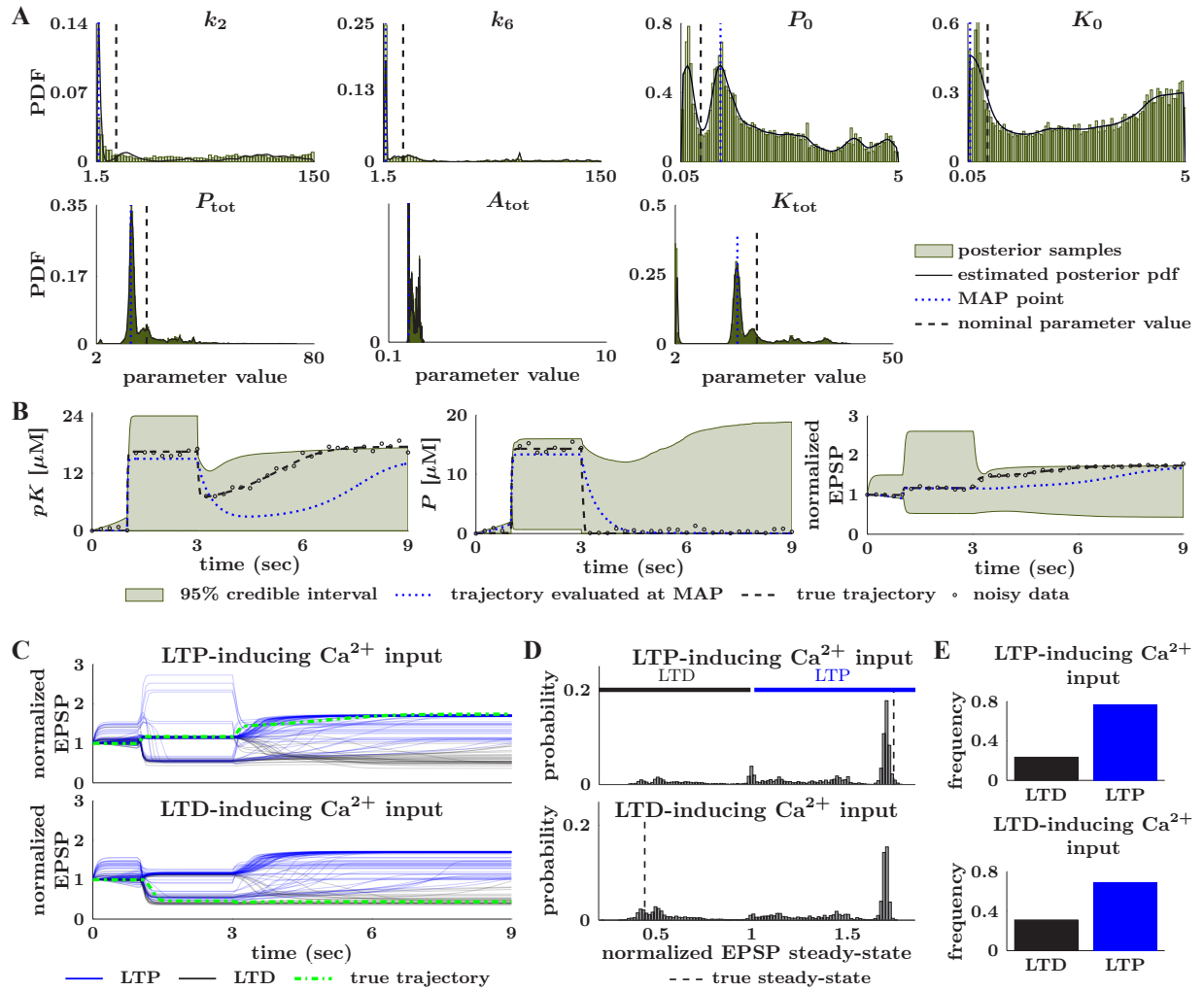


Figure 2.10: (A) We estimated marginal posterior distributions of the model parameters from noisy data with an LTP-inducing calcium input. Distributions are visualized by fitting a kernel density estimator to 547,500 (150 walkers with 3,649 steps each) MCMC samples obtained using CIUKF-MCMC with AIES for MCMC after discarding the first 4,351 samples per walker as burn-in. (B) Posterior distribution of the trajectories of the state variables show LTP (normalized EPSP > 1) and LTD (normalized EPSP < 1) responses for an LTP inducing input. The true trajectory (dashed black line) shows the dynamics with the nominal parameters, the dotted blue line shows the trajectory evaluated at the MAP point, and the points show the noisy data (covariance is 1% of the variance of the true trajectory). The 95% credible interval shows the region between the 2.5th and 97.5th percentiles that contains 95% of 30,000 posterior trajectories. (C) Sample posterior trajectories (100 out of 30,000 total) highlight the LTP (blue lines) and LTD (black lines) in response to the LTP inducing input (top) and the LTD inducing input (bottom). The dashed green lines show the true trajectories for the respective calcium inputs. (D) Histograms reveal the distribution of the long-term responses to the LTP-inducing input (top) and the LTD-inducing input (bottom). The dashed black lines show the true response for the respective calcium inputs. (E) Quantifying the percentage of the 30,000 sample trajectories that produce LTD and LTP responses for each calcium input. The LTP-inducing input yields 76.59% (22,972 samples) of the responses in the LTP state and 23.41% (7,024 samples) in the LTD state. The LTD inducing input yields 68.93% (20,680 samples) of the responses in the LTP state and 31.07% (9,320 samples) in the LTD state.

Lastly, we investigated if the posterior distribution estimated in the LTP regime can predict the response to an LTD-inducing input. An initial hypothesis was that in response to an LTD-inducing input, most trajectories would predict LTD, but a significant subset would predict the incorrect response, LTP. An additional ensemble of 30,000 simulations, with an LTD-inducing input of $Ca^{2+}(t) \equiv 2.2 \text{ } [\mu\text{M}]$ (pulse from $2 \leq t \leq 3$ (sec)) was used to determine the posterior distribution of the dynamics. Visualization of 100 of the 30,000 trajectories in Fig 2.10C (bottom row) again showed both LTP (blue) and LTD (black) responses; however, we unexpectedly observed more LTP than LTD in response to the LTD-inducing input. The distribution of normalized EPSP responses (bottom row of Fig 2.10D) confirms that there are a large number of responses in the LTP regime with a minor mode around the correct response. Quantification of these results highlights that only 31.07% (9,320 samples) of responses are in the LTD state (expected response) while 68.93% (20,680 samples) of the responses are in the LTP state (unexpected response). In summary, this model formulation can correctly capture the same LTP behavior over a range of model parameters while losing the ability to predict LTD behavior.

Despite the LTP and LTD responses being sensitive to the same set of parameters (see Fig 2.9C-D), the posterior distribution estimated from measurements of LTP places more probability on parameter sets that robustly predict LTP over those that correctly predict both responses. From these results, we conclude that for this model, we need to learn the model parameters with a high degree of certainty in order to disambiguate the LTP versus the LTD response because sensitivity analyses revealed that the same set of parameters governs these two different outputs. This finding highlights that sensitivity analyses are not sufficient to distinguish parameter uncertainty for systems with multistability and a comprehensive framework as outlined here is necessary to shed light on such model complexities.

2.4 Discussion

In this work, we developed a framework (see Section 2.2.1) for comprehensive uncertainty quantification of dynamical models in systems biology. The proposed framework leverages identifiability and sensitivity analysis to reduce the parameter space (sections 2.2.3 and 2.2.4) followed by Bayesian parameter estimation with CIUKF-MCMC (Section 2.2.6). We applied this framework to three systems biology models to demonstrate its applicability and highlight how a focus on uncertainty can transform modeling-based studies. First, we performed two computational experiments on a simple two-state model that showed how noise and data sparsity contribute to estimation uncertainty. Next, we applied our framework to two models, the MAPK and the synaptic plasticity models, which better resemble the models used to capture biological readouts. Using these models, we highlight how comprehensive uncertainty quantification enables quantitative analysis of two biologically relevant dynamical behaviors, limit cycles and steady state responses. We also found that good quality data cannot always overcome uncertainty due to the model structure. These examples provide an essential starting point for applying our framework in practice and interpreting systems biology studies under uncertainty.

Our results highlight how a focus on uncertainty quantification can give new insights in modeling-based studies. For example, in Section 2.3.2 we were able to learn a posterior distribution for the parameters that predict limit cycles with high probability. The posterior distribution is our *best guess* for the distribution of the model parameters after incorporating the available data into the statistical model. Therefore, the posterior distribution for the dynamics (approximated via an ensemble simulation as in Section 2.2.10) provides our *best guess* for the dynamics, given everything we know about the model and the data. For the MAPK limit cycle oscillations, this *best guess* includes dynamics with a range of limit cycle properties. Additional analysis of the parameters that produce limit cycle oscillations revealed correlations between k_4 and α and that their values must both be within a certain range to produce the expected dynamics. Therefore,

incorporating uncertainty quantification into modeling provides additional context and understanding of the system of interest.

We also observe that predictions do not always capture the correct bistability in the example of high MAPK steady state (Section 2.3.2) and the LTP/LTD response (Section 2.3.4). In both cases, the 95% credible intervals of the ensemble of predicted dynamics cover both the higher and lower steady states (LTP and LTD in the synaptic plasticity example). These results imply that the estimated posterior distributions for these models include parameter sets that no longer show bistability at the specified input (the initial condition for the MAPK model or the calcium level for the synaptic plasticity model). Further analysis of these systems could test if these parameter sets lead to a complete loss of bistability or merely shift the bifurcation point, the value of the input that changes which steady state is reached. Overall our results point to a complex interplay between model parameters and inputs that potentially confounds parameter estimation of multistable systems.

As highlighted in Sections 2.3.1 and 2.3.2, we showed that estimation uncertainty is a result of the quality, quantity, and nature of the available experimental data. As expected, we demonstrated, for the simple two state model (Fig 2.3,) that data with more noise and fewer samples leads to estimates with wider uncertainty intervals. Furthermore, for the MAPK oscillations (Fig 2.7), we found that the sampling strategy, either including data from the initial transient or focusing on sampling from the oscillations, drastically affects estimation uncertainty. This finding contradicts the common intuition that *more data is always better*, as we had the smallest estimation uncertainty in predictions made using data that had the fewest samples, the oscillations only data. We suspect that these observations can be understood by considering that the likelihood function measures the mismatch between the data and estimated trajectory with an l^2 -like norm, for example see Eq (2.12). When we include data from both the initial transient and oscillations, as in the equidistant and non-equidistant sampling cases in Section 2.3.2, the likelihood

probability is similar for all trajectories that capture the initial transient, regardless of whether they oscillate or reach a fixed point. However, the oscillations only data focuses the likelihood on the long-time dynamics and places lower probability on trajectories that reach a fixed point. While l^2 -norm-based likelihood functions are often used in Bayesian methods, one can also design a likelihood function to explicitly determine if a trajectory displays the same features of the data, such as the feature-based approach in [182]. These findings imply that thoughtful experimental design can reduce unnecessary and potentially costly data collection and can increase certainty in parameter estimates and model predictions.

In the proposed framework, we stress that global identifiability and sensitivity analyses are essential to practical parameter estimation and uncertainty quantification. By running variations of the proposed framework without these analyses in Supplemental Section 2.3.3, we found that each analysis reduces estimation uncertainty and leads to an easier estimation problem. In the complete framework, we isolate globally identifiable parameters and choose a cutoff value of the Sobol sensitivity index (see Section 2.2.4) to select the influential parameter subset. Throughout this work, we choose a logical sensitivity index threshold that corresponds either to a sharp drop in or a logical value of the sensitivity index, such as the $S_i \geq 0.1$ criterion that we apply in several of the example problems. However, we caution that selecting the subset of influential parameters is an intricate problem that requires users to decide the acceptable amount of information to discard from the uncertainty analysis [183], [184]. Alternative methods to the thresholding approach include those discussed in [183], group sensitivity analysis [131] or lasso regression [185]. While these methods aim to select a subset of influential parameters, they are similar to the thresholding approach in that each method hinges upon determining the acceptable number of degrees of freedom to exclude from the analysis. Future research should develop methods for fully quantitative parameter subset selection methods.

While the computations remained tractable for the three example models in this paper, it is essential

to consider how complete uncertainty quantification can increase the computational costs of parameter estimation. In [82], the authors showed that the original UKF-MCMC algorithm required more floating point operations than methods that do not account for all sources of uncertainty. Similarly, we found (see Supplemental Section A.1.1) that a single CIUKF likelihood evaluation takes 81 times longer than an evaluation of a similar likelihood function that does not account for uncertainty in the model form. A deeper look into the execution of the CIUKF likelihood function revealed that approximately 60% of the runtime was spent on calls to `ode15s()`, which we use to discretize the differential equation model. The added computational cost of CIUKF may be reduced in the future by developing optimized codes for discretizing the differential equation model and evaluating the likelihood function.

Bayesian methods, such as CIUKF-MCMC, enable complete uncertainty quantification; however, the need for MCMC sampling and the associated repeated likelihood evaluations can pose an increased computational expense. In particular, as models include greater numbers of state variables and parameters, the added expense of simulation, the increased dimensionality, and the potentially complex geometry of the posterior distributions can challenge standard MCMC approaches. Innovative methods from the broader UQ community, combined with the increasing accessibility of high-performance computing resources, enable Bayesian analysis of models with large numbers of state variables and parameters [5], [186], [187]. Additionally, while we showed that identifiability and sensitivity analyses could significantly reduce the dimensionality of the parameter space, this reduced space may remain relatively high dimensional for specific applications. Many MCMC sampling algorithms that incorporate information about the geometry of the distribution or run parallel Markov chains can sample high-dimensional and possibly complex posterior distributions. These methods include parallel MCMC samplers [188], [189], Hamiltonian Monte Carlo [84] and dimension independent, likelihood-informed methods [190]. Although comprehensive uncertainty quantification comes with added computational costs, recent innovations in UQ methods can

alleviate some of the added computational burdens of comprehensive analysis.

Throughout this work, we make several assumptions regarding the availability of the model equations and the statistical properties of associated errors. First, we assume that the model equations are known prior to parameter estimation. This assumption reflects standard modeling practices where models use biochemical theories that assume equations for the kinetics of biochemical reactions. In using the CIUKF-MCMC we somewhat weaken this assumption because it introduces process noise to account for uncertainty in the model form [82]. In reality, all models have some level of uncertainty because they rely on assumptions about the system. Therefore, accounting for model form uncertainty regularizes the dynamics to account for a mismatch between the predicted dynamics and the data [82]. However, in cases where the model is not known *a priori*, it may be necessary to estimate the model structure and parameters from the data simultaneously. One approach to disambiguate a model structure is to learn the biochemical reaction network [191], [192] or the mathematical model directly from data [193]–[197]. Additionally, it is possible to cast these problems in the Bayesian perspective to learn the model form and the associated uncertainties [82], [198].

Furthermore, we make several assumptions when choosing statistical models for measurement and process noise and simulating biological measurement data. First, we assume Gaussian measurement and process noises in applying the CIUKF-MCMC method. Standard Kalman filtering approaches revolve around assumptions of normally distributed noise processes [161]. However, we may be better able to describe noise in biological systems and experiments with alternate probability distributions (see, e.g., [140]). Assuming a non-Gaussian distribution for the measurement noise would require reformulating the likelihood function (it would alter the distribution in Eq (2.12)); however, incorporating alternative distributions for the process noise would require approximations for the marginal likelihood that go beyond Kalman filtering. Next, this work considers linear measurement functions, e.g., Eq (2.3), but

CIUKF-MCMC is also well-suited to handle nonlinear measurement functions [82]. Lastly, we assume that we only have access to a single time series of measurements, e.g., one trial of an experiment; however, most experiments in biology perform several repeated trials. A straightforward approach to incorporate all available data to inform the statistical model would use the mean of each time point and estimate parameters from the time series of means. Additionally, one could estimate parameters from each time series separately and then analyze several posterior distributions that one could merge via meta-analysis or information fusion principles [199]–[201], or construct a statistical model that accounts for the multiple time series [84] simultaneously.

The framework introduced in this work can be applied and extended to enable comprehensive uncertainty quantification for most dynamical models encountered in systems biology. Future research should focus on rigorous methods for parameter subset selection, applying MCMC samplers that are well suited for complex, high-dimensional distributions to systems biology, and incorporating more specific statistical models of biological data and noise. In summary, research at the intersection of uncertainty quantification and systems biology modeling will strengthen parameter estimation and enable models that more accurately predict the dynamics observed in experimental measurements.

2.5 Code availability statement

The code necessary to reproduce this work is available on GitHub at <https://github.com/RangamaniLabUCSD/CIUKF-MCMC>. We store all models in the systems biology markup language (SBML) for reproducibility and provide a function to translate these to MATLAB. Additionally, we provide a tutorial in the GitHub as a starting point for future applications of the proposed uncertainty quantification framework. In the tutorial, we also list Python and Julia-based software for parameter estimation and sensitivity analysis as alternatives to the MATLAB-based software we use in this study.

2.6 Acknowledgements

Chapter 2, in full, is a modified reprint of the material as it appears in Bayesian parameter estimation for dynamical models in systems biology. Nathaniel Linden, Boris Krämer, and Padmini Rangamani, PLoS Computational Biology, 2022. The dissertation author was the primary researcher and author of this paper.

Chapter 3

Increasing Certainty in Systems Biology

Models using Bayesian Multimodel

Inference

3.1 Introduction

Current innovations in molecular tools [69], [70] and high-resolution microscopy [91], [202] have led to discoveries in intracellular signal transduction, such as the spatial regulation of signaling pathways. These discoveries often require new mathematical models to give rise to mechanistic insights. Furthermore, mathematical models also enable the generation of experimentally testable predictions of intracellular signaling [9], [36], [203]. However, one key challenge in systems biology is formulating a model when there are many unknowns, and the available data do not observe every species in a system. As a result, different mathematical models that vary in their simplifying assumptions and model formulations

can describe the same signaling pathway. For example, searching the popular BioModels database for models of the *extracellular-regulated kinase* (ERK) signaling cascade yields over 125 results for models that use ordinary differential equations [204], [205]. While each model was developed for specific experimental observations, accounting for the range of simplifying assumptions can help improve the certainty and accuracy of resulting predictions. In this work, we explore two important questions: (1) *How can we quantify the effects of uncertainty in the model formulation, called model uncertainty, on model predictions?* and (2) *How can we reduce model choice uncertainty to increase the certainty of intracellular signaling predictions?* Using ERK signaling as our model system, we show that *Bayesian multimodel inference* (MMI) helps to address model uncertainty by leveraging the available data and accounting for all user-specified models.

Uncertainty quantification in systems biology aims to understand how model assumptions, inferred quantities, and data uncertainties impact model predictions [206], [207]. Bayesian parameter estimation quantitatively assesses parametric uncertainty by estimating a probability distribution for unknown parameters, such as reaction rate constants and equilibrium coefficients, from training data [206], [208], [209]. However, at the model level, explicit approaches to handle model uncertainty are yet to be routinely employed in systems biology. Model selection using information criteria, such as the Akaike information criterion (AIC), [210], [211] or Bayes Factors [212] has been the preferred approach in systems biology to select a single “best” model when multiple models are available [210], [213]. Reasoning with a single set of predictions, traditionally from a single model, is much easier than with predictions from multiple models. However, given that the available experimental data are often sparse and noisy, these approaches may limit predictive performance by introducing selection biases and misrepresenting uncertainty [210], [214].

Multimodel inference reduces selection biases and accounts for model uncertainty by including

contributions from every specified model [209], [210], [214]–[217]. Theoretical results have shown that MMI can improve predictive performance by reducing uncertainty and increasing robustness to modeling assumptions [210], [214], [218]–[220]. MMI methods combine model predictions by taking a weighted average over all supplied models [219]–[221]. Methods for choosing the weights range from parametric consensus estimation [219], [220] and frequentist Akaike-weighing [210] to *Bayesian model averaging* (BMA) [215], pseudo-Bayesian model averaging (pseudo-BMA) [214], [218], and stacking of predictive densities (stacking) [218].

Previous applications of MMI to systems biology have focused on a limited subset of the available methods, primarily information-criterion-based and Bayesian model averaging approaches. Stumpf et al. conducted a theoretical analysis of MMI for biological network inference, where the network structure was inferred from data [216], [217]. Based on these analyses and limited applications to protein-protein interaction data, the authors concluded that choosing a good set of models is generally important for constructing accurate MMI estimates and that MMI could improve predictions of the biological network structures even when the true networks were not among the set of candidate models. More recently, Beik et al. utilized MMI with BMA to select candidate tumor growth mechanisms consistent with several experimental datasets [222]. However, MMI has not yet been investigated to increase the certainty of intracellular signaling predictions in systems biology.

In this work, we expand the Bayesian systems biology toolkit to employ MMI as a powerful approach to address model uncertainty. We apply MMI to extracellular-regulated kinase signaling (ERK) as an example problem for deeper analysis. Specifically, we select ten ERK signaling models that emphasize the core of the ERK pathway and estimate the kinetic parameters with Bayesian inference using experimental data from Keyes et al. [223]. First, we show that MMI increases the certainty of model predictions using several different experimental and synthetic datasets. Next, we show that MMI-based

predictions are robust to changes in the composition of the set of supplied ERK models and to increases in data uncertainty. Finally, we apply MMI to study the mechanisms driving ERK activity, which is localized to subcellular locations observed by Keyes et al. [223]. We refer to this ERK activity as subcellular location-specific ERK activity. Subcellular localization of intracellular signaling is hypothesized to play a role in differentiating cellular responses to heterogeneous stimuli and is observable using modern molecular tools [223]. We find that Bayesian MMI enables us to compare hypotheses about what drives location-specific ERK signaling and, in doing so, suggests that location-specific differences in both Rap1 activation and negative feedback strength are necessary to capture the observed dynamics. We conclude that MMI increases predictive certainty when multiple models of the same signaling pathway are available via a structured approach to handle model uncertainty and selection simultaneously.

3.2 Materials and Methods

3.2.1 Standardized EGF-inputs

To standardize the EGF input across models, we modified all model formulations to include EGF as a state variable. Then, we set the time derivative of EGF to zero to simulate a sustained EGF stimulus, which assumes that modeled reactions do not deplete extracellular EGF pools. Additionally, we define all EGF stimuli in nanomolar (nM) concentrations. To convert from concentrations in ng/mL or pg/mL to nM we assume EGF has a molecular weight of 6,048 g/mol [224]. Additionally, to convert from nM to molecules/cell we assume a cell volume of 1 nanoliter (nL) [225]. Lastly, in models that include additional growth factors, we set the concentrations of the corresponding receptors to zero. Specifically, in Birtwistle et al. 2007 [226] we fixed concentrations for H, E3, and E4, and in von Kreigsheim et al. 2009 [227] we fixed the NGF and NGFR concentrations.

Experimental data preprocessing From live-cell fluorescence microscopy experiments, Keyes et al. report all EKAR measurements as the YFP/CFP emission ratio, $R(t)$ [223]. In Figure 3.3 and Supplemental Figure B.3, we normalize each cell-wise trajectory by removing the cell-wise minimum and dividing by the difference in the cell-wise maximum and minimum. Further, in Figure 3.8, we define the normalized EKAR4 data as $\tilde{R}(t) := R(t)/\overline{R_{\max}}$, where $R(t)$ is the EKAR4 emission ratio, and $\overline{R_{\max}}$ is the mean (across all cells with and without Rap1 inhibition) of the maximum (in time) emission ratio. This normalization choice retains location-specific differences in ERK signaling and the effects of Rap1 inhibition. All original data are reported as measurements over individual cells; thus, to yield a single dataset with an uncertainty estimate, we computed the average and standard deviation over all cells (e.g., black and red curves in Figure 3.8A). We found that normalization introduced substantial uncertainty in Rap1 inhibition data, so in constructing a likelihood model, we took the standard deviation of the data to be one-half of that computed directly (shown as gray and red bars in Figure 3.8A). To ensure that the data and model predictions were on the same scale, we similarly normalized model predictions of active ERK to the trajectory-wise maximum.

Synthetic data generation We generated two experimentally realistic synthetic datasets to analyze MMI. First, using the H' 1996 model, we recreated the EGF-ERK dose-response curve shown in Figure 2B of [228] by varying the input EGF concentration from 0.001 nM to 0.106 nM over 10 levels and computing the resulting steady-state ERK activity (Figure B.4A). The range of EGF input concentrations is also similar to experiments in [229]. Specifically, we normalized each response to the maximum across input levels and denoted this by the % maximal ERK activity. Further, we assumed that each measurement is subject to mean-zero independent Gaussian measurement error with a standard deviation of 0.1, similar to measurement uncertainties observed in experiments [228], [229]. To account for differences in the total available ERK concentration across models and ensure that model predictions are on the same scale of the

data, we computed the % maximal ERK activity by normalizing each predicted dose-response curve to the maximum ERK activity in that curve.

3.2.2 Bayesian parameter estimation and sequential Monte Carlo

In this section, we provide a brief overview of Bayesian parameter estimation. For more details in the context of systems biology, see [207], [208], and more theory in general, see [206], [209].

Systems biology models use systems of *ordinary differential equations* (ODEs) to describe the rates of change in the concentration of included biochemical species. For a particular signaling pathway, we often have a set of K models, $\mathfrak{M}_K = \{\mathcal{M}_1, \dots, \mathcal{M}_K\}$, that vary in their representation of the system. Each model connects the dynamics of the state variables $\mathbf{x}_k(t) \in \mathbb{R}_+^{n_k}$, ($\mathbb{R}_+ = [0, \infty)$ are the nonnegative real numbers) which correspond to the concentration of biochemical species, to observations $\hat{\mathbf{y}}_k(t) \in \mathbb{R}_+^{m_k}$ using the system of equations

$$\frac{d\mathbf{x}_k(t)}{dt} = f_k(\mathbf{x}_k(t); \boldsymbol{\theta}_k), \quad (3.1)$$

$$\hat{\mathbf{y}}_k(t) = h_k(\mathbf{x}_k(t); \boldsymbol{\theta}_k) + \boldsymbol{\eta}(t), \quad \boldsymbol{\eta}(t) \sim \mathcal{N}(\mathbf{0}, \Gamma), \quad (3.2)$$

where $f_k(\cdot; \cdot) : \mathbb{R}_+^{n_k} \times \mathbb{R}_+^{p_k} \rightarrow \mathbb{R}_+^{n_k}$, and $h_k(\cdot; \cdot) : \mathbb{R}_+^{n_k} \times \mathbb{R}_+^{p_k} \rightarrow \mathbb{R}_+^{m_k}$. The model parameters $\boldsymbol{\theta}_k \in \boldsymbol{\Theta}_k \subseteq \mathbb{R}_+^{p_k}$ control model predictions and include quantities such as reaction rates and equilibrium coefficients. We additionally assume that a Gaussian measurement noise process $\boldsymbol{\eta}(t) \in \mathbb{R}^{m_k}$ with covariance matrix $\Gamma \in \mathbb{R}^{m_k \times m_k}$ reflects uncertainty in the measurements. Beyond predictions of the observations, we assume that each model predicts a biologically relevant quantity of interest (QoI) $q(t) \in \mathbb{R}$. The QoI is predicted by a function of the internal states and parameters, $\hat{q}_k = g_k(\mathbf{x}_k(t), \boldsymbol{\theta}_k)$, where $\hat{q}_k$ is the QoI prediction with model $\mathcal{M}_k$, and $g_k(\cdot, \cdot) : \mathbb{R}_+^{n_k} \times \mathbb{R}_+^{p_k} \rightarrow \mathbb{R}$.

Bayesian parameter estimation learns a probability density for the parameters of each model conditioned on the training data $p(\theta_k | \mathcal{D}_{\text{train}}, \mathcal{M}_k)$ [206], [209]. The training data $\mathcal{D}_{\text{train}} = \{\mathbf{y}^1, \dots, \mathbf{y}^{N_{\text{train}}}\}$ consists of N_{train} noisy experimental observations, and can correspond to time points t^i in dynamic responses $\mathbf{y}^i = \mathbf{y}(t^i)$ or to input stimuli u^i in dose-response curves $\mathbf{y}^i = \mathbf{y}(u^i)$. Bayesian estimation applies Bayes' rule

$$\underbrace{p(\theta_k | \mathcal{D}_{\text{train}}, \mathcal{M}_k)}_{\text{posterior}} \propto \underbrace{p(\theta_k | \mathcal{M}_k)}_{\text{prior}} \underbrace{p(\mathcal{D}_{\text{train}} | \theta_k)}_{\text{likelihood}}, \quad (3.3)$$

which relates the *posterior probability density* of the model parameters, $p(\theta_k | \mathcal{D}_{\text{train}}, \mathcal{M}_k)$, to the product of the *prior density* $p(\theta_k | \mathcal{M}_k)$ and the *likelihood function* $p(\mathcal{D}_{\text{train}} | \theta_k)$. The posterior density is the probability density of the model parameters given the available training data and the model. The prior density encodes assumptions about the parameters before considering training data. The likelihood function measures the probability that model predictions match the training data and is a function of the model parameters. In most systems biology problems, we do not have a closed-form equation to evaluate the posterior density directly, so we instead must rely on methods such as *Markov chain Monte Carlo* (MCMC) [206], [208], [209] or *sequential Monte Carlo* (SMC) [230], [231] to characterize the posterior through the S samples drawn from it, $\{\theta_k^1, \dots, \theta_k^S\} \sim p(\theta_k | \mathcal{D}_{\text{train}}, \mathcal{M}_k)$.

This work uses log-normal prior densities for all unknown model parameters. To let the data inform the estimation procedure, we choose to center the mean of each prior on the logarithm of the nominal values for the unknown parameters $\theta_k^{\text{nominal}}$ and let the standard deviation be suitably large such that 95% of the probability mass of the prior is in the range $[10^{-2} \times \theta_k^{\text{nominal}}, 10^2 \times \theta_k^{\text{nominal}}]$. One can show that this corresponds to setting the prior standard deviation to 2.350. Substantial empirical evidence suggests that wide, weakly informative priors such as the lognormal prior with wide variance greatly enable sampling of the posterior density compared to completely uninformative priors such as the uniform prior [209].

Next, the likelihood function is a Gaussian density because we assume a Gaussian measurement noise in Eq. (4.2) [208]. We use the PyMC probabilistic programming library in Python to build statistical models efficiently and enable posterior sampling [232].

We use sequential Monte Carlo to sample from the posterior density because it estimates the marginal likelihood without additional computation [230], [231]. SMC is a particle-based sampler that sequentially adapts the particles from prior samples to posterior samples using a tempering scheme. The particles are weighted and mutated at each sampler stage with an importance sampling step. At the final stage, the weights assigned to the particles correspond to the marginal likelihood and can be averaged to provide a marginal likelihood estimate. We use the implementation provided by the PyMC python package with the independent Metropolis-Hastings transition kernel. Additionally, we set the `correlation_threshold` parameter to 0.01 and the `threshold` parameter to 0.85. Unless otherwise noted, we run four independent SMC chains with at least 500 posterior samples per chain.

3.2.3 Forward uncertainty propagation with ensemble simulation

Given S samples from the posterior density, we perform ensemble simulations to propagate uncertainty forward to model predictions. Specifically, we solve the ODE model with each of the S posterior samples to generate a set of ODE solutions for resulting predictive densities. There are three important predictive densities that we use to assess predictive uncertainty. First, the *posterior push-forward density of the QoI* (also called the predictive density of the QoI) is defined as

$$p(\hat{q}_k | \mathcal{D}_{\text{train}}, \mathcal{M}_k) := \int_{\Theta_k} g_k(\mathbf{x}_k(t), \boldsymbol{\theta}_k) p(\boldsymbol{\theta}_k | \mathcal{D}_{\text{train}}, \mathcal{M}_K) d\boldsymbol{\theta}_k \quad (3.4)$$

and directly propagates parametric uncertainty to the QoI. To sample from the posterior push-forward density of the QoI, we evaluate the QoI function $g_k(\cdot, \cdot)$ at each sample, yielding a set of QoI samples

$\{\hat{q}_k^1, \dots, \hat{q}_k^S\} \sim p(\hat{q}_k | \mathcal{D}_{\text{train}}, \mathcal{M}_k)$. Next, the *posterior push-forward density of the measurements*,

$$p(\hat{\mathbf{y}}(t)_k | \mathcal{D}_{\text{train}}, \mathcal{M}_k) := \int_{\Theta_k} h_k(\mathbf{x}_k(t)) p(\boldsymbol{\theta}_k | \mathcal{D}_{\text{train}}, \mathcal{M}_K) d\boldsymbol{\theta}_k, \quad (3.5)$$

similarly propagates parametric uncertainty to the measurements. We evaluate the measurement function at each ODE solution to sample from the posterior push-forward density of the measurements. Lastly, the *posterior predictive density* is defined as

$$p(\tilde{\mathcal{D}}_{\text{train}} | \mathcal{D}_{\text{train}}, \mathcal{M}_k) := \int_{\Theta_k} p(\mathcal{D}_{\text{train}} | \boldsymbol{\theta}_k, \mathcal{M}_k) p(\boldsymbol{\theta}_k | \mathcal{D}_{\text{train}}, \mathcal{M}_K) d\boldsymbol{\theta}_k, \quad (3.6)$$

and accounts for both parametric and data uncertainty. To obtain posterior predictive samples, we sample the posterior push-forward density of the measurements and subsequently add independent samples from the measurement noise process defined in Eq. (4.1). We refer the reader to [233] and [209] for more details on predictive densities and Bayesian model analysis in general.

In this work, we use both the posterior push forward of the QoI and the posterior predictive density where appropriate. In Figures 3.3, 3.4, 3.5 and Supplemental Figures B.3, B.5, and B.8 we use the posterior push forward of the QoI. Alternatively, we use the posterior predictive density in Figure 3.8 and Supplemental Figures B.4.

3.2.4 Bayesian model averaging

Bayesian model averaging weighs each model with the model probability $w_k^{\text{BMA}} = p(\mathcal{M}_k | \mathcal{D}_{\text{train}})$ [215]. Notably, the model probability is the realization of a discrete probability mass function over the set of models. The model probability is computed by applying Bayes' rule a second

time (the first time is for model parameter estimation) at the model level, that is

$$p(\mathcal{M}_k | \mathcal{D}_{\text{train}}) = \frac{p(\mathcal{D}_{\text{train}} | \mathcal{M}_k) p(\mathcal{M}_k)}{\sum_{l=1}^K p(\mathcal{D}_{\text{train}} | \mathcal{M}_l) p(\mathcal{M}_l)}, \quad (3.7)$$

where $p(\mathcal{M}_k)$ is the prior model probability and $p(\mathcal{D}_{\text{train}} | \mathcal{M}_k)$ is the marginal likelihood. Importantly, the marginal likelihood is included in Eq. (3.7) because it depends on the model, whereas we exclude the denominator in parameter estimation and replace equality with proportionality in Eq. (4.3). The *marginal likelihood*,

$$p(\mathcal{D} | \mathcal{M}_k) = \int_{\Theta_k} p(\mathcal{D} | \theta_k, \mathcal{M}_k) p(\theta_k | \mathcal{M}_k) d\theta_k, \quad (3.8)$$

quantifies the probability of observing the data under model $\mathcal{M}_k$. Direct computation of Eq. (3.8) can become intractable because the integral is often over a high-dimension space and thus requires approximations by Monte Carlo integration, Bridge Sampling, or sequential Monte Carlo [215], [234]–[236]. In this work, we use the PyMC implementation of sequential Monte Carlo, which estimates the log marginal likelihood alongside posterior samples [230]–[232]. PyMC returns a log marginal likelihood estimate for each independent chain, so we take the average over all chains to generate an estimate for that model. Additionally, we assume that each model is equally probable *a priori* with prior model probability $p(\mathcal{M}_k) = 1/K$. We employ the log-sum exponential trick to avoid numerical overflow in evaluating Eq. (3.7), which involves computing the sum of the exponent of log marginal likelihoods. That is, given log marginal likelihoods, Eq. (3.7) can be written abstractly as

$$p(\mathcal{M}_k | \mathcal{D}_{\text{train}}) = \frac{\exp(x_k)}{\sum_{l=1}^K \exp(x_l)}, \quad (3.9)$$

where $x_k = \log(p(\mathcal{D}_{\text{train}}|\mathcal{M}_k)p(\mathcal{M}_k))$ can be quite large and thus Eq. (3.9) is susceptible to overflow errors.

We can rearrange Eq. (3.9) to

$$p(\mathcal{M}_i|\mathcal{D}_{\text{train}}) = \exp \left[x_i - \log \sum_{l=1}^K \exp(x_l) \right], \quad (3.10)$$

where we use the log-sum-exponential trick

$$\log \sum_{l=1}^K \exp(x_l) = c + \log \sum_{l=1}^K \exp(x_l - c), \quad (3.11)$$

with $c = \max[x_1, \dots, x_K]$ for numerical stability.

3.2.5 Pseudo-Bayesian model averaging

Pseudo-Bayesian model averaging weighs models based on the expected predictive performance of future data measured with the expected log pointwise predictive density (ELPD). Similar to Akaike-type weighting, which uses the Akaike information criterion instead of the ELPD [210], we define the scalar pseudo-BMA weights as

$$w_k^{\text{pBMA}} := \frac{\exp(\widehat{\text{ELPD}}_{\mathcal{M}_k})}{\sum_{i=1}^K \exp(\widehat{\text{ELPD}}_{\mathcal{M}_i})}, \quad (3.12)$$

given an estimate of the ELPD for each model $\widehat{\text{ELPD}}_{\mathcal{M}_k}$. Further, we stabilize Eq. (3.12) with the log-sum-exponential trick as in the previous section. To account for uncertainty in the ELPD estimates, we use the Bayesian bootstrap as was done in [218], [237].

We adapt the definitions from [214] to introduce approximations to the ELPD that can be used for pseudo-BMA. First, we let the training data $\mathcal{D}_{\text{train}} = \{\mathbf{y}^1, \dots, \mathbf{y}^{N_{\text{train}}}\}$ consist of statistically independent data points. It follows that $p(\mathcal{D}_{\text{train}}|\theta_k, \mathcal{M}_k) = \prod_{i=1}^{N_{\text{train}}} p(\mathbf{y}^i|\theta_k, \mathcal{M}_k)$ due to the independence of the data.

The *expected log pointwise predictive density* of model $\mathcal{M}_k$ is defined as

$$\text{ELPD}_{\mathcal{M}_k} := \sum_{i=1}^{N_{\text{train}}} \int \log p(\tilde{\mathbf{y}}_i | \mathcal{D}_{\text{train}}, \mathcal{M}_k) p_{\text{true}}(\tilde{\mathbf{y}}_i) d\tilde{\mathbf{y}}_i, \quad (3.13)$$

and quantifies the expected predictive performance of the model compared to the true data generating distribution $p_{\text{true}}(\tilde{\mathbf{y}}_i)$. The *log posterior predictive density* for model $\mathcal{M}_k$ and new data-point $\tilde{\mathbf{y}}_i$ is defined as

$$\log p(\tilde{\mathbf{y}}_i | \mathcal{D}_{\text{train}}, \mathcal{M}_k) := \log \int_{\Theta_k} p(\tilde{\mathbf{y}}_i | \theta_k, \mathcal{M}_k) p(\theta_k | \mathcal{D}_{\text{train}}, \mathcal{M}_k) d\theta_k \quad (3.14)$$

and measures the log probability of observing the new data point. Given S posterior samples $\hat{\theta}_k^s$, the log posterior predictive density can be approximated with

$$\log p(\tilde{\mathbf{y}}_i | \mathcal{D}_{\text{train}}, \mathcal{M}_k) \approx \log \frac{1}{S} \sum_{s=1}^S p(\tilde{\mathbf{y}}_i | \hat{\theta}_k^s, \mathcal{M}_k). \quad (3.15)$$

In general, we do not know the true probability of the data $p_{\text{true}}(\tilde{\mathbf{y}}_i)$, so we must rely on approximations to the ELPD such as the leave-one-out cross-validation estimator (LOO-CV) [214], which we use in this work, or the widely applicable information criterion (WAIC) [238].

The simplest approximation to the ELPD, called the *sample-approximated log pointwise predictive density* (LPD), is simply the sum of Eq. (3.15) over all data points,

$$\widehat{\text{LPD}}_{\mathcal{M}_k} = \sum_{i=1}^{N_{\text{train}}} \log \frac{1}{S} \sum_{s=1}^S p(\mathbf{y}_i | \hat{\theta}_k^s, \mathcal{M}_k); \quad (3.16)$$

however, the LPD is known to overestimate the ELPD [214]. Thus, we can use LOO-CV to approximate

the ELPD as

$$\widehat{\text{ELPD}}_k^{\text{LOO}} := \sum_{i=1}^n \log p(\mathbf{y}_i | \mathcal{D}_{\text{train}}^{-i}, \mathcal{M}_k), \quad (3.17)$$

where the *leave-one-out predictive density* is defined as

$$p(\mathbf{y}^i | \mathcal{D}_{\text{train}}^{-i}, \mathcal{M}_k) := \int p(\mathbf{y}^i | \boldsymbol{\theta}_k, \mathcal{M}_k) p(\boldsymbol{\theta}_k | \mathcal{D}_{\text{train}}^{-i}) d\boldsymbol{\theta}_k. \quad (3.18)$$

Here, $\mathcal{D}_{\text{train}}^{-i} := \{\mathbf{y}^1, \dots, \mathbf{y}^{i-1}, \mathbf{y}^{i+1}, \dots, \mathbf{y}^{N_{\text{train}}}\}$ indicates all of the training data $\mathcal{D}_{\text{train}}$ excluding point $\mathbf{y}^i$. We refer the reader to Vehtari et al. [214] for further details and motivation. This work uses the LOO-CV-based approximation to the ELPD because both approximations provided similar weights for each model (results not shown here). However, direct computation of Eq. (3.18) can be prohibitively computationally expensive because it requires repeating parameter estimation for each held-out data point. To enable efficient computation of the LOO-CV ELPD estimator, we use Pareto smoothed importance sampling [214], [239], the details of which we omit for brevity. We denote the Pareto smoothed importance sampling leave-one-out cross-validation estimator of the ELPD as PSIS-LOO-CV.

In this work, we use the ArviZ Python library [240] to compute the PSIS-LOO-CV and the resulting model weights. Given a dictionary of inference data objects with log-likelihood samples for each model, we use the `Arviz.compare()` function with the `ic` parameter set to `loo` and the `method` parameter set to `BB-pseudo-BMA`. We use default settings for all computations. To ensure that PSIS-LOO-CV accurately estimates the LOO-CV estimator, we compared PSIS-LOO-CV to direct leave-one-out cross-validation (LOO-CV) for the S' 2014 model using synthetic dose-response data (Figure B.4A). Notably, the ELPD predictions for PSIS-LOO-CV and LOO-CV are within five percent, predicting -5.69 and -5.43 , respectively.

3.2.6 Stacking of predictive densities

Stacking of predictive densities [218] aims to assign scalar weights w_k^{stack} to each model such that the new estimator is optimal according to a specified optimality criterion. The stacking optimization problem

$$\max_{\{w_k^{\text{stack}}\}} S\left(\sum_{k=1}^K w_k^{\text{stack}} p(\tilde{\mathcal{D}}|\mathcal{D}_{\text{train}}, \mathcal{M}_k), p_{\text{true}}(\tilde{\mathcal{D}})\right) \quad (3.19)$$

aims to maximize a score, $S(\cdot, \cdot) : \mathcal{P} \times \Omega \rightarrow \mathbb{R}$, where $\mathcal{P}$ is a probabilistic forecast and Ω is the sample space on which the true predictive density is defined (see [218] for a more detailed definition). The score is computed between the consensus predictive density evaluated at new data $\tilde{\mathcal{D}}$ and the true predictive density. However, the true predictive density is unknown, so Yao et al. [218] suggest using a logarithmic scoring rule and replacing the predictive density evaluated at new data with the LOO estimator. The approximate stacking optimization problem becomes

$$\max_{\{w_k^{\text{stack}}\}} \frac{1}{n} \sum_{i=1}^n \log \sum_{k=1}^K w_k^{\text{stack}} p(\mathbf{y}_i | \mathcal{D}_{\text{train}}^{-i}, \mathcal{M}_k), \quad (3.20)$$

where the LOO-CV predictive density $p(\mathbf{y}_i | \mathcal{D}_{\text{train}}^{-i}, \mathcal{M}_k)$ can again be estimated with PSIS-LOO-CV as in the previous section.

We again use the ArviZ Python package [240] to compute the stacking weights in this work. Specifically, we use the `Arviz.compare()` function with the `ic` parameter set to `loo` and the `method` parameter set to `stacking`. All other parameters are set to the defaults.

3.2.7 Numerical solution of ODEs and steady-state simulation

As systems biology models are often stiff ODE systems, we solve all systems of ordinary differential equations with the Kvaerno 4/5 implicit Runge-Kutta method that is implemented in the DiffraX

Python package and is well-suited for integrating stiff systems of ODEs [241], [242]. Unless otherwise stated, we set the maximum number of solver steps, `max_steps`, to 6×10^6 . To account for potential stiffness, in conjunction with the implicit scheme, we use a PID controller-based method for adaptive time stepping [243]–[245] with tolerances `atol=1e-6` and `rtol=1e-6` for all problems unless otherwise noted in Supplemental Table B.2. Such tolerances are suggested to provide sufficient accuracy when solving most systems. We run the solver from the initial condition to the desired time point to obtain the entire solution over time. However, we choose the following methods to balance accuracy and computational efficiency to obtain each model’s steady-state solution. The first approach runs the ODE solver until the solution and time derivative satisfy $\|\mathbf{dx}(t)/dt\|_2 \leq \text{atol} + \text{rtol}\|\mathbf{x}(t)\|_2$, where $\|\cdot\|_2$ is the standard vector two-norm. To evaluate this criterion, we use a `DiffraX SteadyStateEvent` with tolerances `atol=1e-5` and `rtol=1e-6` for all problems unless stated otherwise in Supplemental Table B.2. The second approach uses Newton’s method to directly solve $\mathbf{dx}(t)/dt = 0$. We use the standard Newton method implemented in the `Optimistix` Python library [246] with `max_steps = 100` and tolerances `atol=1e-10` and `rtol=1e-10`. To mitigate convergence issues, we initialize the Newton solver at a point along the solution of the ODE by running the previous ODE-based steady-state method with crude tolerances `atol=1e-5` and `rtol=1e-5` for all problems unless stated otherwise in Supplemental Table B.2.

3.2.8 Structural Identifiability Analysis

We perform an *a priori* local structural identifiability analysis to reduce the unknown model parameters with unique values given the observables. We refer the reader to [247], [248] for background and mathematical details on structural identifiability analysis. We perform a local analysis instead of a global analysis as suggested in our previous work [208], because computing global identifiability proved computationally intractable for larger models such as H’ 2005 and B’ 2007. Specifically, we used the structural

identifiability analysis method from [248] that is implemented in the `StructuralIdentifiability.jl` package in the Julia programming language. We use default settings for the `assess_local_identifiability()` function, and set the probability of correctness to $p = 0.99$ for all models except Birtwistle et al. 2007 [226], for which we take $p = 0.95$ to improve computational efficiency. Additionally, due to further software limitations, we exclude any parameters in exponents from the identifiability analysis. If these parameters are integers, we fix them to their nominal value, otherwise we set them to 1.0. Specifically, for Hornberg et al. 2005 [249] we fix n to 1.0, and for von Kreigsheim et al. 2009 [227] we fix k_{57} , k_{61} , k_{64} , k_{66} , k_{70} , and k_{72} to 1.0.

3.2.9 Global Sensitivity Analysis

We use the Morris screening method for global sensitivity analysis [250], [251]. First, we assume that all parameters vary independently and can take on values in the range $[10^{-2} \times \theta^{\text{nominal}}, 10^2 \times \theta^{\text{nominal}}]$ (nominal values are listed in Supplemental Materials). We then use the steady-state activated ERK concentration as the quantity of interest for models that predict sustained ERK activation at the nominal parameter values, i.e., H' 1996, O' 2009, S' 2014, R' 2015, and K' 2017, or the maximal activated ERK concentration as the QoI for models that predict a transient activation, i.e., K' 2000, L' 2000, H' 2005, B' 2007, and vK' 2009. Additionally, we assume that the EGF concentration is fixed at 0.1 nM for all models. To perform the Morris screening, we use the SALib Python package [252], [253] with the method of Morris sampler (`SALib.sample.morris.sample()`) and analysis (`SALib.sample.morris.analyze()`) functions with default settings. We draw 256 samples per parameter direction for all models except H' 2005 and B' 2007, for which we use 30 and 10 samples, respectively. Morris screening provides two measures of sensitivity, the mean of the distribution of the absolute value of elementary effects μ^* and the standard deviation of the distribution of elementary effects σ [251]. We refer the reader to [100], [251] for

more details. We assume that parameters are influential to ERK activation when $\mu_i^* / \max\{\mu_i^*\} > 0.1$ or $\sigma_i / \max\{\sigma_i^*\} > 0.1$.

3.2.10 Error and uncertainty metrics

We measure predictive error with either the root mean square error (RMSE) or the relative error.

Given a prediction $\hat{\mathbf{y}}$ and a reference (ground truth) $\mathbf{y}$, the *RMSE* is defined as,

$$\text{RMSE} := \frac{\sqrt{\sum_{i=1}^N (\hat{y}_i - y_i)^2}}{N}, \quad (3.21)$$

where $\mathbf{y} \in \mathbb{R}^N$, and y_i is the i th element of $\mathbf{y}$. We additionally define the *relative error* as

$$\text{Relative Error} := \frac{\|\hat{\mathbf{y}} - \mathbf{y}\|_2}{\|\mathbf{y}\|_2}, \quad (3.22)$$

where the norm is the standard vector two-norm, $\|\mathbf{a}\|_2 = \sqrt{\sum_{i=1}^N a_i^2}$.

We measure predictive uncertainty using the average width of the 95% credible interval. Given a set of predictive samples prediction, $\hat{\mathbf{y}}^1, \dots, \hat{\mathbf{y}}^S$, we define the 95% *credible interval* as the interval between the 2.5th and 97.5th percentiles,

$$\text{95\% credible interval} := [P_{2.5}, P_{97.5}], \quad (3.23)$$

where P_i denotes the element-wise i th percentile of the set of samples. Each element of P_i is the percentile of the corresponding set of elements from the set of vector-valued predictive samples. We define the average width of the 95% credible interval as the mean over all elements of the difference between $P_{97.5}$ and $P_{2.5}$.

3.3 Results

3.3.1 Bayesian multimodel inference

Bayesian multimodel inference systematically constructs a new consensus estimator of important systems biology *quantities of interest* (QoIs) that accounts for model uncertainty. Here, we consider ODE-based intracellular signaling models, $\mathcal{M}_k$, with fixed model structure and unknown model parameters. We leverage the Bayesian framework to estimate unknown parameters and characterize predictive uncertainty through predictive probability densities [206], [209] (see Methods and Materials for more details). The training data, $\mathcal{D}_{\text{train}} = \{\mathbf{y}^1, \dots, \mathbf{y}^{N_{\text{train}}}\}$, consists of N_{train} noisy experimental observations (or synthetic data), and can correspond to time points t^i in dynamic responses $\mathbf{y}^i = \mathbf{y}(t^i)$ or to input stimuli u^i in dose-response curves $\mathbf{y}^i = \mathbf{y}(u^i)$. Given that we use Bayesian methods to estimate unknown model parameters from training data, each model predicts a probability density for the QoI, $p(\hat{q}_k)$ [108], [208]. The goal of Bayesian MMI is to build a multimodel estimate of the QoI, q , defined as $p(q|\mathcal{D}_{\text{train}}, \mathfrak{M}_K)$, that leverages the entire set of specified models, $\mathfrak{M}_K = \{\mathcal{M}_1, \dots, \mathcal{M}_K\}$ [210], [215], [218], [254]. For intracellular signaling, the QoIs can be either the time-varying trajectories of activities or concentrations of key biochemical species ($q(t)$ at time t) or the steady-state dose-response curve ($q(u_i)$ where u_i is a stimulus concentration, and $u_i < u_{i+1}$). Time-varying trajectories characterize the dynamic response of a signaling network to prescribed stimuli. On the other hand, dose-response curves capture the input-output response of the entire signaling network over a range of stimuli concentrations. Measurements of both quantities are fundamental to experimentally characterizing intracellular signaling; for example, see [228] for EGF-ERK dose-response data, and see [223] for time-varying ERK activity data.

The Bayesian multimodel inference workflow used in this work is shown in Figure 3.1, where we first calibrate available models to training data by estimating unknown model parameters with Bayesian

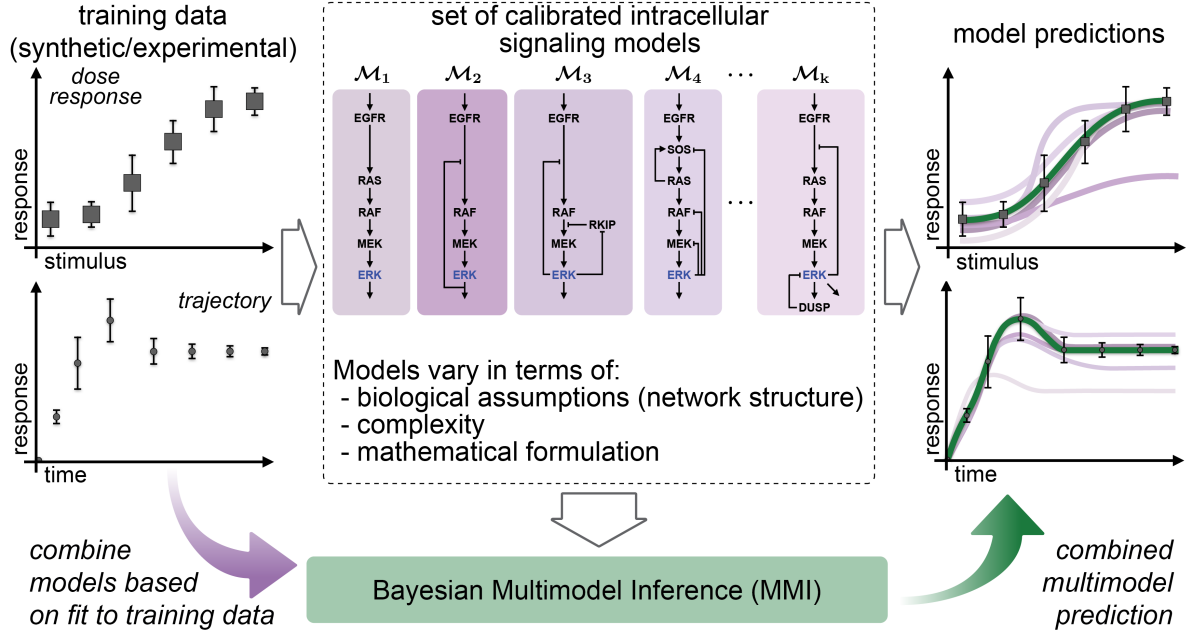


Figure 3.1: First, we use Bayesian parameter estimation to calibrate the parameters in each model to the training data. Each calibrated model can then be used to predict biological quantities; however, the accuracy and uncertainty of each model’s predictions may vary due to differences in the model structure and mathematical formulation. Next, Bayesian MMI constructs a new estimator by combining specified models based on how well each predicts the training data or based on the expected predictive performance of future data. Finally, the multimodel inference-based estimator improves predictive accuracy and robustness compared to individual models. In this work, we use models of extracellular-regulated kinase signaling (ERK) to apply Bayesian MMI for systems biology. The models vary in mathematical and biological complexity. The Bayesian MMI workflow is broadly applicable whenever multiple models are available.

inference, then we combine the resulting predictive probability densities using MMI, and finally provide improved multimodel predictions of important quantities in systems biology studies. Bayesian multimodel inference constructs predictors by taking a linear combination of predictive densities from each model,

$$p(q|\mathcal{D}_{\text{train}}, \mathfrak{M}_K) := \sum_{k=1}^K w_k p(q_k|\mathcal{M}_k, \mathcal{D}_{\text{train}}), \quad (3.24)$$

with weights $w_k \geq 0$ and $\sum_k w_k = 1$ [210], [215], [221], [233], [254]. We note that the weights can either be scalars or realizations of a probability mass function defined over the set of models. The key challenge is estimating the weight to assign each predictive density. Here, we compare three methods for

choosing the weights, w_k : *Bayesian model averaging* (BMA) [215], *pseudo-Bayesian model averaging* (pseudo-BMA) [218], [254], and *stacking of predictive densities* (stacking) [218]. Notably, the potential for MMI to bring new insights into intracellular signaling has not previously been explored, and because each method for MMI has distinct advantages and disadvantages, we chose to compare all three. We briefly summarize the methods below. Importantly, the MMI workflow and methods we investigate in this work broadly apply beyond systems biology whenever a set of candidate models are available.

Bayesian model averaging uses the probability of each model conditioned on the training data to assign weights $w_k^{\text{BMA}} = p(\mathcal{M}_k | \mathcal{D}_{\text{train}})$ [215] to each model. The model probability quantifies the probability of model $\mathcal{M}_k$ correctly predicting the training data relative to the other models in the set. While BMA is the natural Bayesian approach to MMI, the method suffers from several key challenges, including the necessary computation of the marginal likelihood, strong dependence on prior information, and reliance on data-fit alone instead of on predictive performance [209], [215], [218], [255]. Furthermore, as with all fully Bayesian methods, the resulting BMA weights vary with the amount of training data provided, and in the limit of infinite data, the weights converge to a single model [218]. Due to these potential challenges of BMA, we also investigate pseudo-BMA and stacking.

Pseudo-Bayesian model averaging assigns model weights based on the expected predictive performance of unseen data measured with the *expected log pointwise predictive density* (ELPD) [214], [218]. The ELPD quantifies expected predictive performance on new data by computing the distance between the predictive and true data-generating densities and is defined as

$$\text{ELPD}_{\mathcal{M}_k} := \sum_{i=1}^{N_{\text{train}}} \int \log p(\tilde{\mathbf{y}}_i | \mathcal{D}_{\text{train}}, \mathcal{M}_k) p_{\text{true}}(\tilde{\mathbf{y}}_i) d\tilde{\mathbf{y}}_i, \quad (3.25)$$

where $\mathcal{M}_k$ is the k th model and $p_{\text{true}}(\tilde{\mathbf{y}}_i)$ true data generating distribution $p_{\text{true}}(\tilde{\mathbf{y}}_i)$. Higher ELPD values in-

indicate a greater probability of correctly predicting unseen data and, therefore, better predictive performance. Notably, the ELPD is a direct estimate of the predictive performance of a model and does not suffer from the same challenges that affect the marginal likelihood in BMA [214], [218], [256]. However, the ELPD is intractable directly because we do not know the true data-generating density. Instead, we estimate the ELPD with Pareto smoothed importance sampling leave-one-out cross-validation (PSIS-LOO-CV) [214]. Pseudo-BMA constructs model weights, w_k^{pBMA} , by normalizing the estimated ELPD of each model, $\widehat{\text{ELPD}}_k^{\text{LOO}}$, to the sum of that quantity across all models. The two key challenges of pseudo-BMA are that it relies on potentially erroneous estimates of the ELPD, such as PSIS-LOO-CV, and that pseudo-BMA weights do not account for correlations between individual model predictions [218].

Stacking of predictive densities selects optimal model weights, w_k^{stack} , to maximize the ELPD of the consensus density. We follow the approach introduced in [218], which maximizes the log score between the consensus predictive density and the true data-generating distribution estimated by the ELPD. In contrast to BMA and pseudo-BMA, which weigh models independently, stacking weighs all models simultaneously [233]. Thus, stacking finds the best-estimating density that is closest (in terms of the log scoring rule used to define the optimality criterion) to the data-generating process [218]. However, similar to pseudo-BMA, stacking relies on ELPD estimates for MMI.

3.3.2 Variations in ERK signaling models show model form uncertainty

Extracellular-regulated kinase signaling plays a key role in controlling many cellular processes, including proliferation, growth, metabolism, and differentiation [32], [257]. Due to this widespread importance, ERK is one of the most extensively modeled intracellular signaling pathways [258]. As a test problem for MMI, we specified a set of ten models of the core ERK signaling network [226]–[228], [249], [259]–[264]. Throughout this work, we refer to each model by the first letter of the first author’s last

name and the publication year, e.g., Huang and Ferrell 1996 [228] is H' 1996. Even though each model was initially used to answer different research questions, we chose these ten because each one represents the ERK signaling pathway from input to ERK activation. Figure 3.2 depicts the epidermal growth factor (EGF)-to-ERK signaling network spanned by each model and denotes which model includes specific structural components. We note that in addition to each model containing a subset of the complete ERK signaling network, the models also vary in mechanistic complexity and the mathematical formulation used to represent the reaction kinetics. Examples of these variations include using only mass action kinetics in H' 1996 while K' 2017 combines Michaelis-Menten and Hill-equations. Additionally, all models use single equation representations of EGF-EGFR interactions, except for H' 2005 and B' 2007, which use detailed multi-step ligand-binding and receptor internalization mechanisms. Supplemental Table B.1 highlights how extensive variations in the number of model parameters and state variables reflect the variations in the model formulation. Furthermore, each model's equations were readily available from original or secondary sources. Therefore, these models exemplify real-world uncertainties while remaining comparable in their representation of the ERK signaling pathway. Thus, we reasoned that this set of ERK models serves as a good test case for MMI.

We previously showed that *a priori* structural identifiability and global sensitivity analyses are critical to successful Bayesian parameter estimation for intracellular signaling models [208]. Therefore, we performed local structural identifiability analysis on each model to determine which parameters can be uniquely identified locally in parameter space. Next, we performed global sensitivity analysis on each model using Morris screening [100] to determine which identifiable parameters significantly influence predicted ERK activity (Supplemental Figure B.1). Based on the results of both analyses, we reduced the number of free parameters by fixing all nonidentifiable and noninfluential parameters to nominal values from the literature. Supplemental Table B.1 shows the number of parameters remaining after each of these

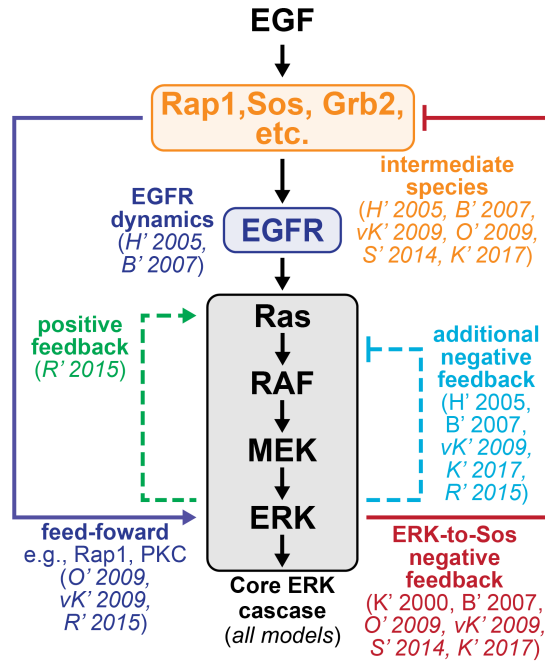


Figure 3.2: Parentheses list the models that include a representation of each component. Abbreviations for models show the first letter of the first author’s last name and the publication year, e.g., H’ 1996 corresponds to Huang and Ferrell 1996 [228]. Supplemental Table B.1 lists the abbreviations and references for all models.

a priori analyses. We only estimated the remaining free parameters.

3.3.3 Bayesian MMI increases certainty in ERK activity predictions

As a first test, we applied MMI to predict EGF-induced ERK activity from experimental observations of cytoplasmic ERK kinase activity from Keyes et al. [223]. Here, we sought to understand better how model averaging with MMI impacts predictions of time-varying EGF-induced ERK activity. Using an improved ERK kinase activity reporter called EKAR4, Keyes et al. measured ERK activity at subcellular locations, including the cytoplasm (Figure 3.3A) and plasma membrane (Supplemental Figure B.3A). Using the cytoplasmic ERK activity, we separately estimated the parameters of the ten ERK signaling models. Figure 3.3B shows the resulting posterior densities for the ERK activity predictions with four of the ten models (Supplemental Figure B.2A shows the densities for all ten models). All ten models predict

the cytoplasmic ERK activity data qualitatively well. We attribute minor variations in the predictions to differences in the formulation of each model.

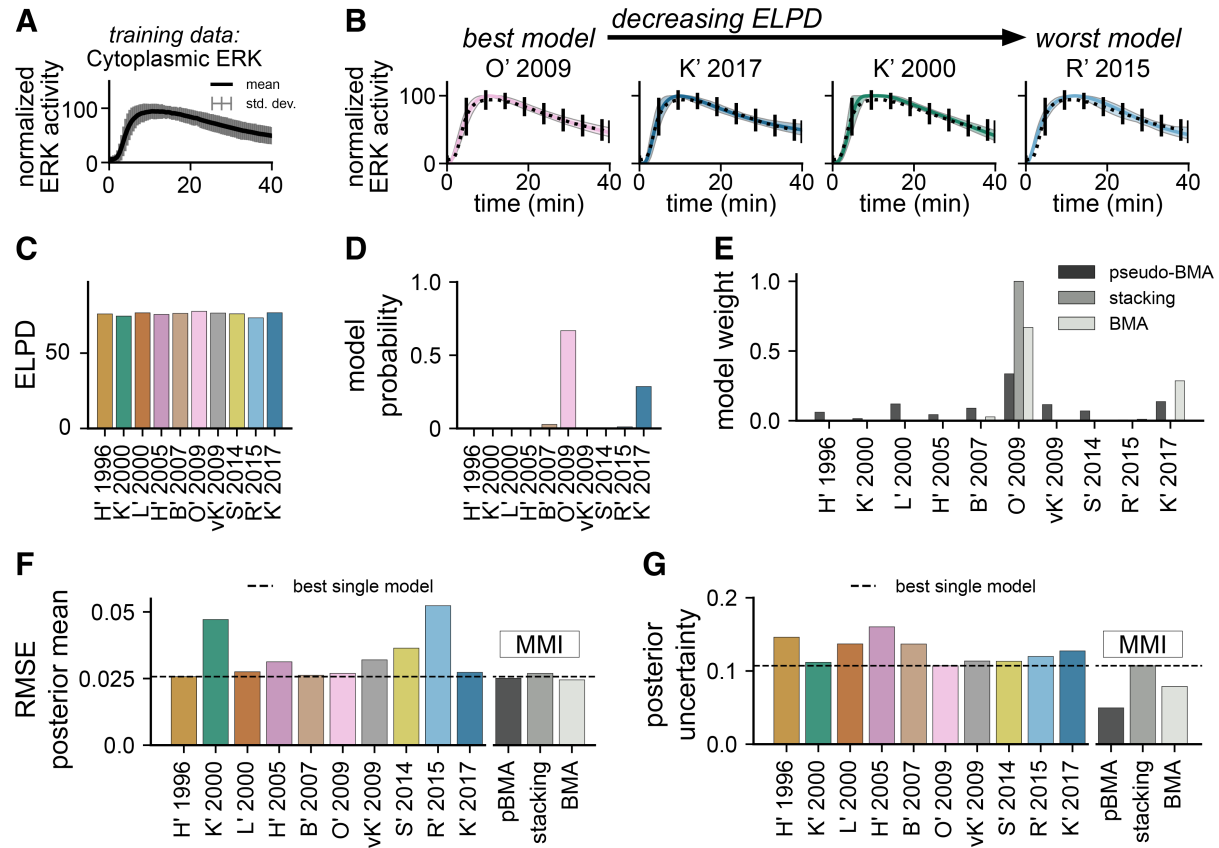


Figure 3.3: (A) Normalized experimental measurements of cytoplasmic ERK activity. Mean (black trace) and standard deviation (grey bars) show statistics computed across all single-cell trajectories. Data originally presented in figure 1 of [223]. (B) Posterior densities of cytoplasmic ERK activity trajectories for four out of nine models ordered by decreasing expected log pointwise predictive density (ELPD). Additional densities are shown in Supplemental Figure B.2A. Shaded regions show the 95% credible interval (the interval between 2.5th and 97.5th percentiles) of the posterior, while the solid line shows the posterior mean. Densities are estimated from 2,000 posterior samples. (C) ELPD values for all models computed using Pareto smoothed importance sampling leave-one-out cross-validation (PSIS-LOO-CV). Models with higher relative ELPD are more likely to predict future data correctly. Estimated ELPD values are used to construct MMI estimates with pseudo-Bayesian model averaging (pseudo-BMA) and stacking of predictive densities (stacking). (D) Model probabilities are computed using sequential Monte Carlo estimates of the marginal likelihood. The model probability is the probability of each model conditioned on the training data and is used for MMI with Bayesian model averaging (BMA). (E) MMI model weights for all models using pseudo-BMA, BMA, and stacking. (F) Root mean square error (RMSE) of the posterior mean cytoplasmic ERK activity prediction. The dashed black line shows the lowest RMSE of any single model. (G) Posterior uncertainty measured by the mean 95% credible interval width in time. The dashed black line shows the lowest uncertainty of any single model.

Each of the three MMI methods distributed MMI weights differently between the models, but the

MMI predictions are similar. The ranking of the models using the ELPD values (Figure 3.3C) and the model probabilities (Figure 3.3D) varies between the two methods, despite the overall trends remaining similar. We observe that the models that predict cytoplasmic ERK activity qualitatively better had higher ELPD values and higher model probability than those that did poorly. Both pseudo-BMA and BMA assigned non-zero weights to more than one model, but they weighed different subsets of the model set (Figure 3.3E). Meanwhile, stacking placed all weight on the O' 2009 model. Despite the different weight distributions, each of the three MMI methods predicted trajectories of cytoplasmic ERK activity that all appear qualitatively similar to the data (Supplemental Figure B.2B). Interestingly, the models with higher ELPD values were not necessarily the most complex or the models with the most free parameters. The best models, O' 2009 and K' 2017, had a moderate number of states and included components such as intermediate species and negative feedback (Figure 3.2). This implies that there is minimal mechanistic complexity which is required to capture time-varying cytoplasmic ERK activity.

Quantitative differences in the predictions for each method show that only pseudo-BMA and BMA increased predictive certainty. The *root mean square error* (RMSE) of the posterior mean and the predictive uncertainty measured by the average width of the posterior 95% credible interval are slightly lower than that of the best model for pseudo-BMA and BMA (Figure 3.3F–3.3G). However, these quantities are the same as the best model for stacking, which only weighs a single model. These results show that MMI predictions of cytoplasmic ERK activity retain the predictive accuracy of the best models and can also yield reductions in uncertainty when a weighted average is used.

To test if MMI improves predictions for additional ERK signaling data, we repeated MMI for the same set of models using experimental measurements of plasma membrane ERK activity (Figure B.3A) and synthetic EGF-ERK dose-response curve data (Supplemental Figure B.4A). For both of these new datasets, predictions varied between individual models and thus, the ELPD estimates and model probabilities

varied as well (Supplemental Figure B.3B–B.3D and B.4B–B.4D). Accordingly, each MMI method yields different weight assignments (Supplemental Figure B.3E and B.4E). We note that several models in the dose-response example, e.g., O’ 2009, failed to capture the trends in the dose-response data (Supplemental Figure B.4B). Importantly, these models received nearly no weight in the MMI averages and do not contribute to MMI predictions. For both datasets, MMI predictions had predictive errors on par with the best individual models. Additionally, MMI predictions had uncertainty that was lower than that of individual models when multiple models were combined via weighted averaging (Supplemental Figure B.3F–B.3G and B.4F–B.4G). Therefore, we concluded that MMI handles model uncertainty in the set of specified models for predictions of plasma membrane ERK activity and the EGF-ERK dose-response curve in addition to cytoplasmic ERK activity. Importantly, from all three datasets, MMI predictions had lower errors than choosing a model at random and showed reduced uncertainty when using pseudo-BMA and BMA. Thus, MMI can increase predictive certainty for this set of models across multiple types of data.

3.3.4 MMI increases robustness to model form uncertainty

To understand how increasing model uncertainty affects model-averaged MMI predictions, we systematically varied the composition of the model set. Specifically, we asked how (i) adding an ill-fitting model, (ii) removing the best model, and (iii) changing the size of the model set affects MMI predictive performance.

First, we found that single-model perturbations to the model set tended to have minor effects on the performance of MMI predictions. To test this for predictions of cytoplasmic ERK activity, we first excluded the worst-fit model, R’ 2015 (lowest ELPD in Figure 3.3C) and computed the MMI weights (Figure 3.4A). Adding the *bad* model back to the set resulted in little change to the MMI weights (Figure 3.4B), and thus, had minimal impact on the predictive error (Figure 3.4D) and the posterior uncertainty (Figure 3.4E)

because the additional *bad* model received nearly no weight. Next, we removed the best-fit model, O' 2009, (highest ELPD in Figure 3.3C) and recomputed the MMI weights (Figure 3.4C). Without the *best* model, stacking reassigned all weight to K' 2017, the next best model, while BMA redistributed the weight across K' 2017 and B' 2007. Pseudo-BMA redistributed weight away from O' 2009 across several other models. Accordingly, the predictive error increased only a small amount because all models predict cytoplasmic ERK activity similarly well. Despite the small increases in error, removing the model with the

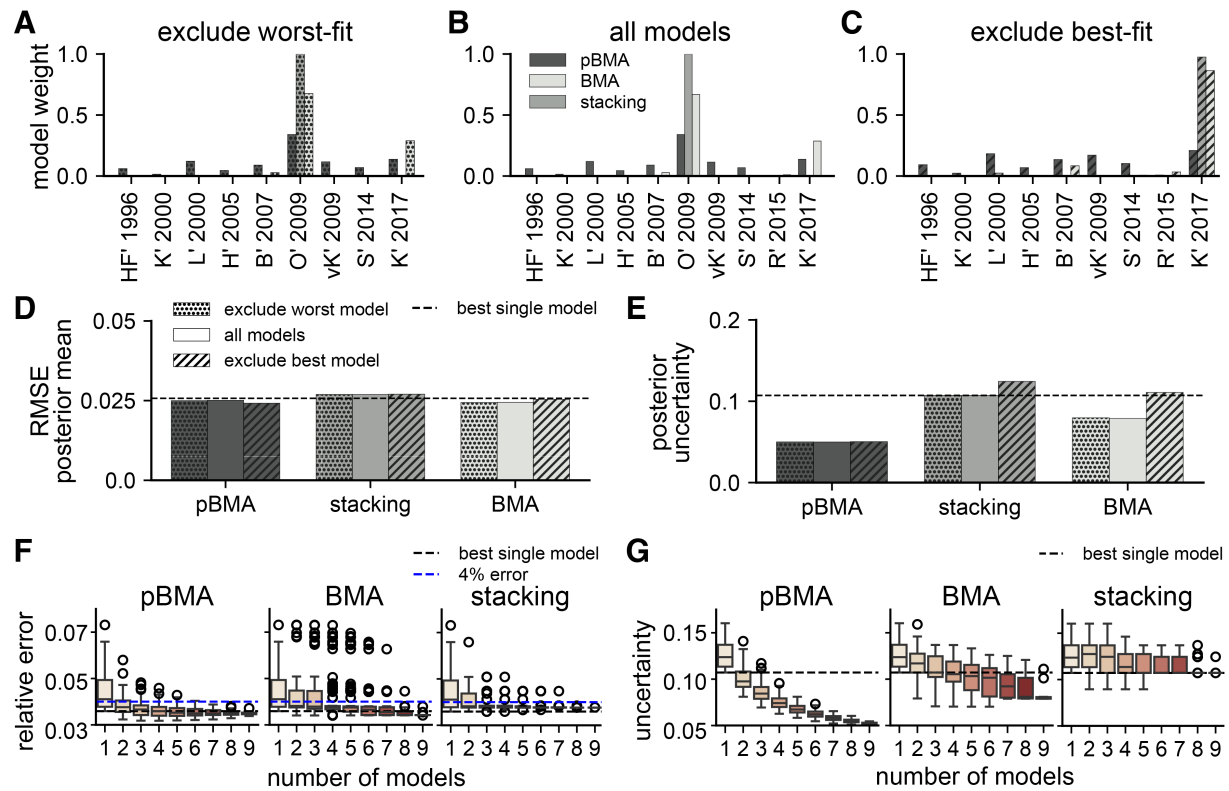


Figure 3.4: (A)–(C) Weights assigned to models in three model sets: (A) excluding the worst-fit model, R' 2015, (B) all models, (C) excluding the best-fit, O' 2009. (D) RMSE of the posterior mean cytoplasmic ERK activity prediction for each MMI method and model set. Dotted patterning corresponds to (A), no patterning to (B), and dashed patterning to (C). The dashed horizontal line is the RMSE of the model with the lowest average error, HF' 1996. (E) Posterior uncertainty is measured by the mean of the 95% credible interval taken over all EGF levels. (F) Relative error of the posterior mean for MMI predictions with increasing model set size. All possible combinations of models were tested at each size. The dashed blue line shows 5% relative error (0.05), and the dashed black line shows the lowest error of any single model, HF' 1996. (G) Average posterior uncertainty of ERK response for MMI predictions with increasing numbers of models. The dashed black line shows the uncertainty of the best model, R' 2015. (F),(G) Boxplots show the mean and interquartile range, and whiskers show the 95% credible interval. Open circles show outliers.

highest ELPD model had little effect on MMI predictive uncertainty (Figure 3.4E). We repeated MMI with these single-model perturbations to the model set using both the plasma membrane data (Supplemental Figure B.5A–B.5E), and the synthetic EGF-ERK dose-response data (Supplemental Figure B.6A–B.6E). We observed similar results for both additional datasets; however, the predictive error increased when we removed the best model for the dose-response example because the difference in the performance of the best and next-best models was bigger for that example. Therefore, these results show that MMI predictions of ERK activity are robust to changes in this model set because including an additional *bad* model had little effect, and excluding the *best* model led to relatively small changes in uncertainty and accuracy.

Next, we investigated how many models are necessary for effective MMI and how the number of models in the model set affects MMI predictive performance. We constructed model sets with all possible combinations of the nine ERK models ranging from sets of two to sets of eight models for all three datasets, cytoplasmic ERK activity (Figure 3.4F), plasma membrane ERK activity (Supplemental Figure B.5F) and EGF-ERK dose-response data (Supplemental Figure B.6F). First, we found that the error of MMI predictions on average decreases and approaches the lowest error of any individual model as the number of models increases for all three datasets. In particular, the errors of pseudo-BMA and stacking predictions tended to decrease faster, while the errors of BMA predictions remained elevated even with larger model set sizes. In particular, for BMA predictions of the EGF-ERK dose-response curve, the relative error remains centered around 5%, which is well above the error of the best individual model (Supplemental Figure B.6F). Based on these findings, we conclude that MMI can increase the probability of correct predictions beyond single models, even with just two specified models. However, the improvements were much greater with pseudo-BMA and stacking than with BMA.

How does the number of models affect MMI predictive uncertainty? For all three datasets, the predictive uncertainty decreased as model set size was increased for pseudo-BMA and BMA, but not

for stacking (cytoplasmic ERK activity, Figure 3.4G; plasma membrane ERK activity, Supplemental Figure B.5G; EGF-ERK dose-response Supplemental Figure B.6G). For pseudo-BMA and BMA, the uncertainty approached a level that is below that of the best model (i.e., the dashed black line in Figure 3.4G) as the number of models increased. However, for stacking, the uncertainty is similar to the prediction made with the model with the greatest ELPD, even after more than three models are used. Therefore, we conclude that MMI reduces uncertainty even with a moderate number of models and that more models lead to greater uncertainty reductions for pseudo-BMA and BMA in these three examples. Additionally, these results show that even with the increased model uncertainty from having more models, MMI can reliably increase predictive certainty and generate predictions as accurately as the best single models.

3.3.5 MMI is robust to increased training data uncertainty

We have previously observed that data uncertainty can greatly impact predictive uncertainty in intracellular signaling models [208]. Furthermore, the quantity and quality of training data affect how different MMI methods weigh models [218]. To explore how MMI performs in the presence of data uncertainties, we simulated increasing data uncertainty by decreasing the quantity and quality of the training data supplied for MMI.

First, we found that MMI predictions of future cytoplasmic ERK activity have lower uncertainty than single-model predictions. We varied the length of the training data by truncating the original 40-minute long recordings at the 10-, 20-, and 30-minute time points (Figure 3.5A), and used MMI to predict the remainder of the trajectory up to 40 minutes. We evaluated the predictions by assessing the testing error and uncertainty on the final 10 minutes of the trajectory, which was not used for parameter estimation. Similar to all individual models, the posterior uncertainty for the best single model, O' 2009, increased with decreasing training data length (top row in Figure 3.5B; additional individual model predictions are

shown in Supplemental Figure B.7). However, the pseudo-BMA predictions showed substantially lower predictive uncertainty at all data lengths (bottom row in Figure 3.5B). Quantitatively, the testing error of MMI predictions is on par with that of the lowest-error models at all training data lengths (Figure 3.5C), and

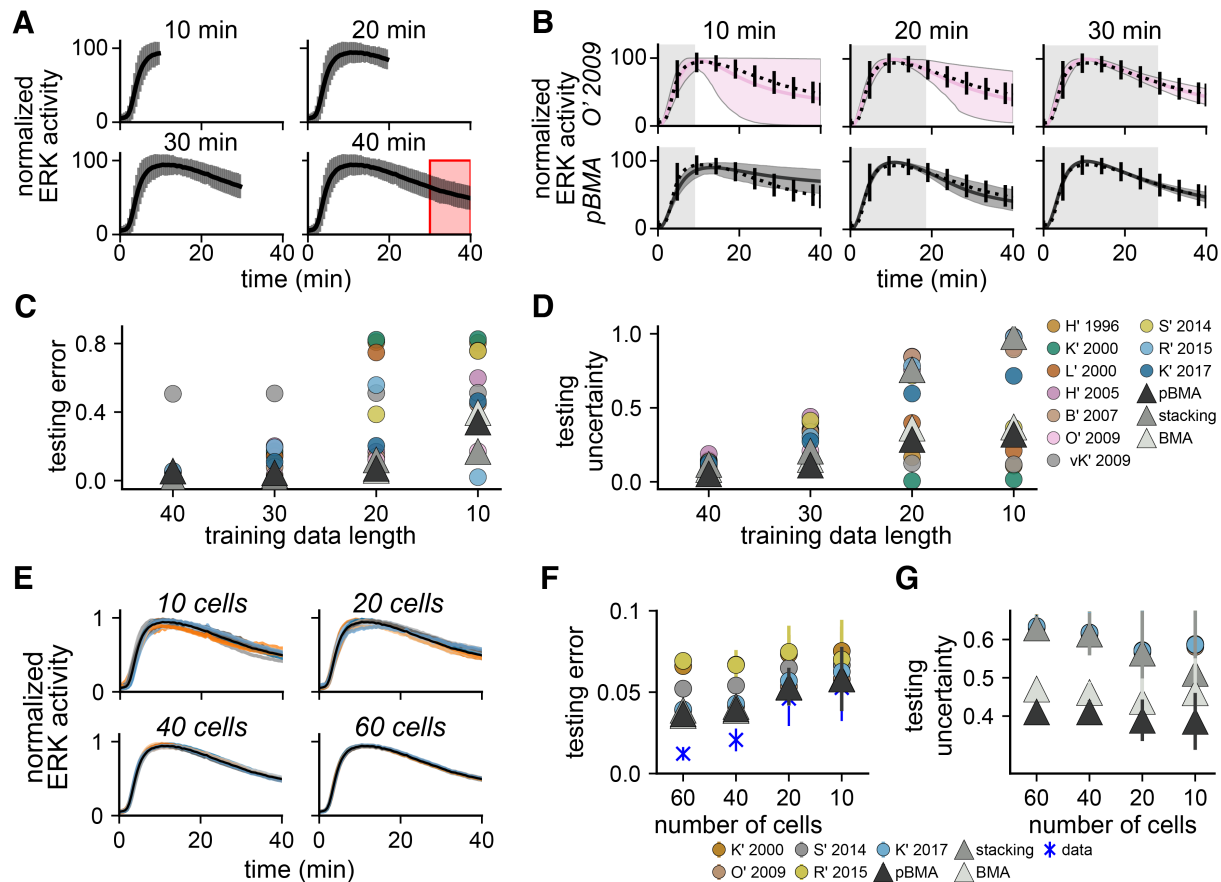


Figure 3.5: (A) Shorter training data was constructed by truncating the original 40-minute cytoplasmic ERK trajectories at the 10-, 20-, and 30-minute time points (mean, black trace; and standard deviation, grey bars). Predictive performance was assessed by computing errors and uncertainties in the final 10 minutes (red box). (B) Posterior predictions from decreased training data for the best model, O' 2009 (highest ELPD across all training datasets), and MMI with pseudo-BMA. Predictions for additional models and MMI methods are shown in Supplemental Figure B.7. (C) Predictive error (relative error) for the final 10 minutes ($t = 30 \rightarrow t = 40$ min) of cytoplasmic ERK activity. (D) Predictive uncertainty (average width of the 95% credible interval) for the final 10 minutes of cytoplasmic ERK activity. (E) Lower-quality training data was generated by averaging random subsets of 10, 20, 40, and 60 imaged cells. The black trace shows an original average of 76 cells. Colored traces show averages of 40 replicate random subsets. (F) The predictive error (relative error) of cytoplasmic ERK activity with lower quality data compared to average activity trajectory using all cells. (G) Predictive uncertainty of cytoplasmic ERK activity with lower quality data. (F),(G) Filled circles indicate an average error of 40 replicates for individual models and triangles for MMI predictions. Error bars show the standard deviation over replicates. Blue markers show the error of the raw training data at each subset size compared to the original full-data mean.

the testing uncertainty of MMI predictions is substantially lower than the best single models (Figure 3.5D). We note that at the 10- and 20-minute data lengths, the L' 2000, K' 2000, and H' 2005 predictions have lower uncertainty than those for MMI. However, they are highly erroneous, with almost all simulations predicting maximal activation after 40 minutes (Supplemental Figure B.7). We additionally repeated the same simulations with data from the plasma membrane. We observed similar trends as with cytoplasm data (Supplemental Figures B.8A–B.8E). Together, these results show that MMI increases the certainty and accuracy of out-of-sample predictions of future ERK activity from shorter training data than single models alone.

Second, MMI predictions showed lower predictive error and uncertainty than single-model predictions from lower-quality data. We generated lower-quality training data by decreasing the number of individual cells we averaged over to create the normalized data. Specifically, we drew 40 random subsets of 10, 20, 40, and 60 of the 76 original single-cell recordings of cytoplasmic ERK activity and computed the cell-wise average and standard deviation (Figure 3.5F). Data created with lower cell counts (10 or 20 cells) was of lower quality than data with higher counts (40 or 60 cells) because the lower-count data had greater error compared to the original 76-cell dataset and had more variation in the standard deviation (not shown). To reduce the computational burden of repeated parameter estimation, we constructed MMI estimates using five of the ten models, K' 2000, O' 2009, S' 2014, R' 2015, and K' 2017, because these models had the shortest inference times (Supplemental Table B.3). The predictive testing error, computed as the relative error of model predictions compared to the original 76-cell dataset, shows that MMI predictions are more accurate than most of the individual models (Figure 3.5G). Additionally, the BMA and pseudo-BMA MMI predictions had lower predictive uncertainty across all subset sizes (Figure 3.5H). We repeated the same simulations with data from the plasma membrane. Similarly, we found that MMI is robust to lower-quality data (Supplemental Figures B.8F–B.8G). Based on these results, we conclude that MMI predictions are

more robust to uncertainties due to reduced data quality than single-model predictions.

3.3.6 MMI improves out-of-sample predictions of ERK activity

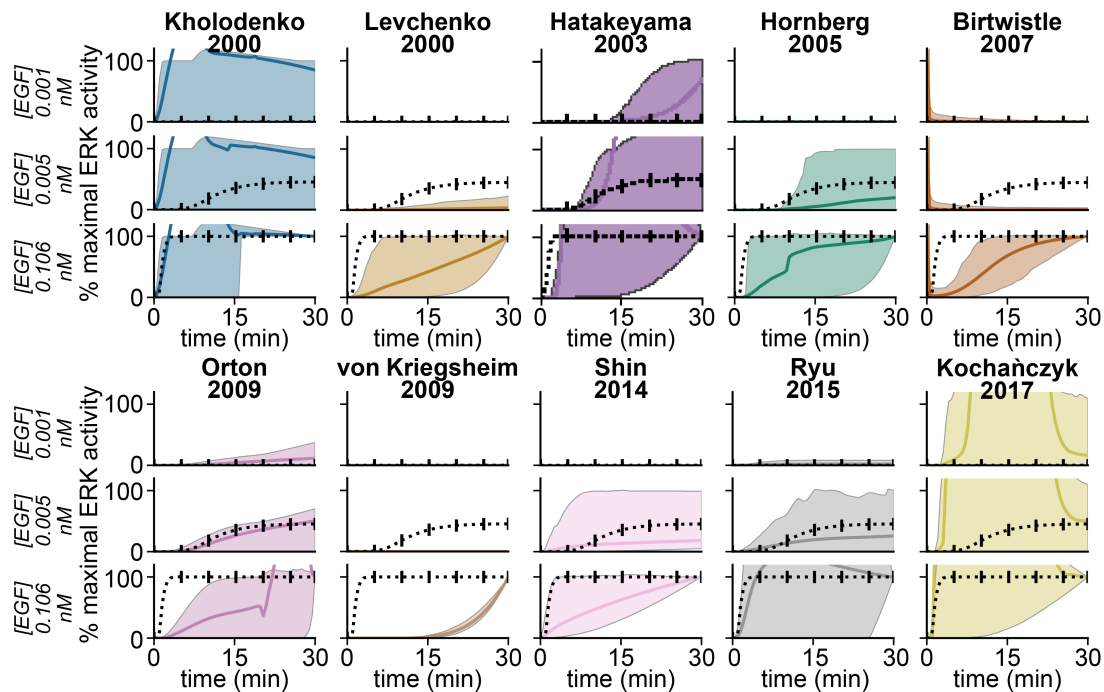


Figure 3.6

In the previous results, we focused on fitting models to data on one type (either trajectory or dose-response) and generating predictions of the same type of behavior. Here, we asked how MMI can improve out-of-sample predictions. Specifically, we fit the models to synthetic data of one type, either dose-response data or trajectory data (Figure 3.7A), and predicted the other response type. Interestingly, we found that none of the models trained on dose-response data could predict the dynamic responses with any reasonable efficacy (Figure 3.6). This is not surprising because dose-response data do not inform the kinetics of the response, only the steady-state levels. We did not apply MMI for this case because the predictions were so poor for all models.

On the other hand, we found that a subset of the models trained on synthetic trajectory data were

able to predict the dose-response curve (Figure 3.7C). This is further captured by the range of ELPDs between the models (Figure 3.7C). Interestingly, in this case the model probability is placed entirely on one model (Figure 3.7E). This is likely an effect of the fact that model probabilities are known to converge to the best model with increasing data [265], and this example has the mode data points. The resulting model weights reflect the ELPD and model probabilities, with pBMA weighing two models and Stacking and BMA only weighing one (Figure 3.7F). This highlights how these methods can select the best models when there are many poor predictors. Finally, we see that the accuracy (Figure 3.7D) and uncertainty of the MMI predictions are similar to those for the models selected by those methods, because MMI is not taking weighted averages here. This shows that MMI can enable out-of-sample predictions and is especially robust when many input models are very poor predictors. However, this example also highlights that MMI predictions are only as good as the candidate models.

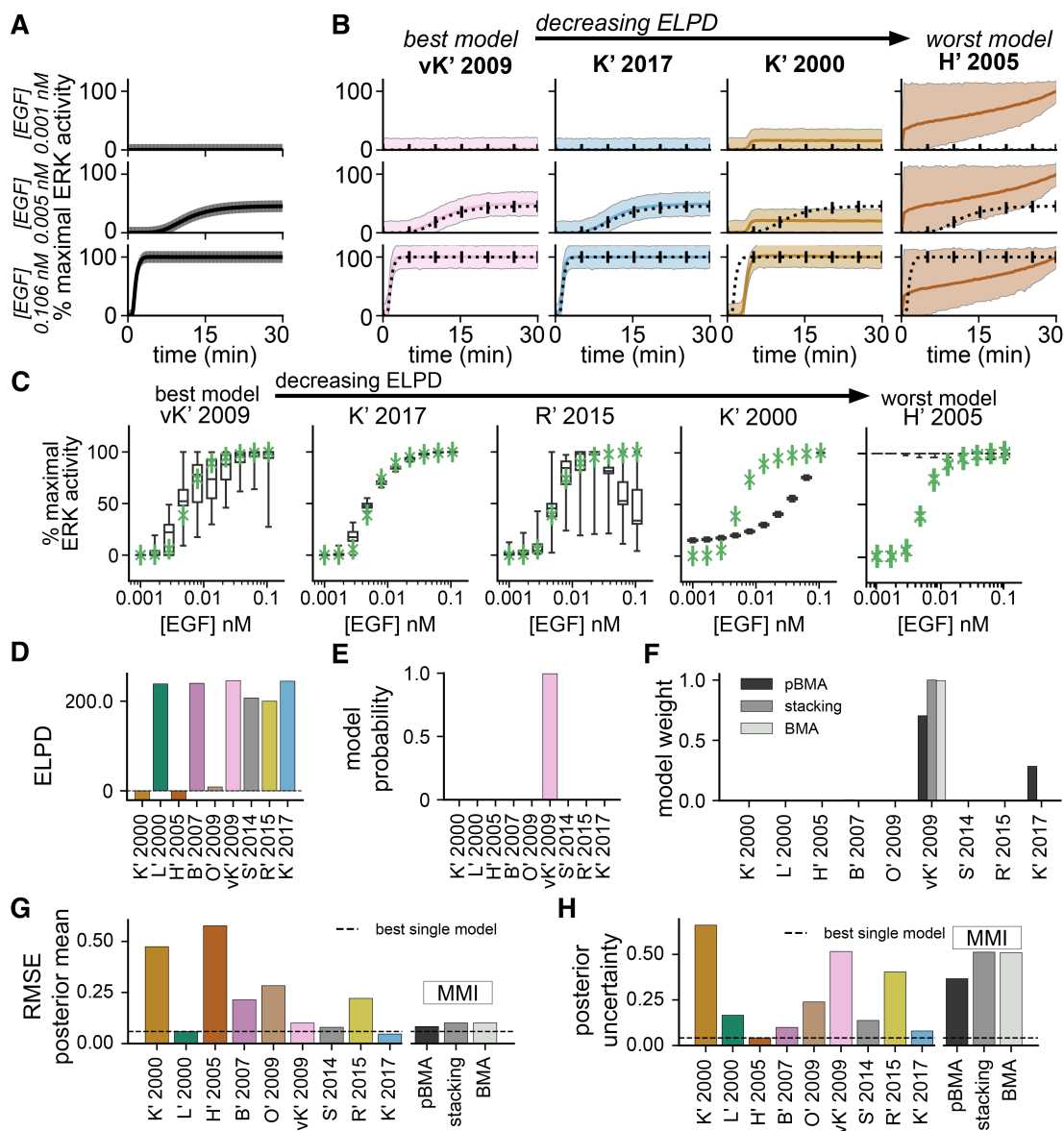


Figure 3.7: (A) Synthetic trajectory data was generated by simulating the H' 1996 mode for 30 minutes at three EGF input levels (0.001, 0.005, and 0.106 mM). (B) Posterior densities of trajectories for a subset of models ordered by ELPD. (C) Posterior densities of predicted dose-response curves using models trained on trajectory data. (D) ELPD of models. (E) Model probabilities. (F) Model weights. *Note that D-F were calculated based on fits to trajectory training data.* (G) RMSE of posterior dose-response predictions. (H) Uncertainty of posterior dose-response predictions.

3.3.7 Subcellular ERK activity depends on Rap1 and negative feedback

Keyes et al. previously revealed spatiotemporal differences between cytoplasmic and plasma membrane ERK signaling [223]. Specifically, the authors observed that ERK activity is more sustained at the plasma membrane compared to the cytoplasm (Figure 3.8A, black traces; data reproduced from figures 1C and 1D of [223]). Furthermore, Keyes et al. showed that plasma membrane ERK activity depended strongly on the non-canonical ERK activator Rap1, but cytoplasmic activity did not (figures 3C and 3D from [223]). Here, we used MMI to investigate whether subcellular differences in the model parameters could predict subcellular variations in ERK activity by selecting models that accurately explain the data.

Initially, we hypothesized that location-specific probability densities for model parameters could explain location-specific ERK activity observed by Keyes et al. [223]. In the previous sections, we observed that the same models could predict both cytoplasmic and plasma membrane ERK activity well when all parameters can vary independently between the two compartments (Figure 3.3B and Supplemental Figure B.3B). Supplemental Figure B.9 highlights each model's most significant subcellular specific parameter estimates. For example, there were subcellular differences in Raf deactivation, EGFR dynamics, ERK activation, and total kinase concentrations. However, since we allowed all model parameters to vary independently between compartments, these findings show which subcellular parameter differences can lead to more accurate predictions of subcellular ERK activity but do not show which specific combinations of independent parameters are required to capture the data.

To explore which combinations of location-specific parameters could capture location-specific ERK activity, we used Bayesian MMI to evaluate all possible combinations of parameters varying between compartments. Specifically, we focused on the O' 2009, S' 2014, and R' 2015 models because these three models comprehensively cover the core ERK network. Furthermore, limiting the investigation to three models ensured the computations remained tractable. In these three models, we observed that the parameters

that control Raf deactivation, ERK phosphorylation, ERK dephosphorylation, ERK negative feedback, and Rap1 activity varied significantly between compartments for these three models (Supplemental Figure B.9 and Supplemental Materials). Therefore, we allowed combinations of the related parameters to vary between compartments. Since the Rap1-dependence of plasma membrane ERK activity was a key finding of Keyes et al. [223], we introduced a Rap1-like activator into the S' 2014 and R' 2015 models, such that all three models could be used to predict ERK activity after Rap1 inhibition. To do so, we assumed that active EGFR activates Rap1, which in turn activates downstream Raf kinase, and that Rap1 activation is not modulated by negative feedback from active ERK [266]. We then performed 94 sets of parameter estimation, allowing every combination of the relevant parameters to either vary independently between compartments or remain the same globally. We fit the models under each hypothesis to EKAR4-only and EKAR4 with Rap1 inhibitor (Rap1GAP) data (Figure 3.8A). To simulate Rap1 inhibition, we set the total Rap1 concentration to zero.

Next, we used MMI to compare which of the combinations of location-specific parameter estimation best captured the subcellular ERK activity data. MMI weighted various parameter combinations (Supplemental Table B.4 shows the best and worst ten). Eight parameter combinations received weights greater than 0.01. These all similarly allowed ERK phosphorylation and/or ERK dephosphorylation parameters to vary between compartments. Additional combinations of location-specific parameters included Rap1 activation, ERK (-) feedback, and Raf deactivation (Supplemental Figure B.4). The identified location-specific parameters controlling ERK phosphorylation can indicate species not included in the models are responsible for location-specific dynamics, such as potential cross-talk or other feedback mechanisms. However, it is also possible that these parameters are simply the *easiest* to estimate. Thus, the location-specific differences may not represent biologically realistic mechanisms because of the known ability of systems biology models to fit a wide range of data [267].

To further refine our understanding of location-specific parameters that can capture the experimental data, we used MMI to compare a smaller set of hypotheses about which parameters impact location-specific ERK activity based on what is known to be biologically realistic [223], [266], [268]. Specifically, we focused on the noncanonical ERK activator Rap1 and ERK negative feedback. We hypothesized that Rap1 signaling would vary between compartments because Keyes et al. [223] found that plasma membrane ERK activity depends on Rap1 while cytoplasmic activity does not (Figure 3 from [223]). Additionally, we hypothesized that ERK negative feedback could vary between compartments because previous findings suggest that the strength of ERK negative feedback can drive differences in sustained versus transient ERK activity [266], [268]. To test these hypotheses about location-specific signaling, we performed nine sets of parameter estimation in which we allowed only the Rap1 parameters (Rap1 diff.), only the ERK negative feedback parameters ((-)FB diff.), or both Rap1 and ERK negative feedback parameters (Rap1 & (-)FB diff.) to vary independently between locations. Parameters not associated with Rap1 or ERK negative feedback were assumed to not vary between locations, and a single probability density was estimated for those parameters.

Location-specific Rap1 signaling and ERK negative feedback were both necessary to predict ERK activity's subcellular variability. Qualitatively, the posterior predictive distributions of the models that allowed only ERK negative feedback parameters to vary between locations ((-)FB diff.) predict the data worse, while those that allowed Rap1 parameters to vary (Rap1 diff.) predict the data more accurately. The models that allow both Rap1 and ERK negative feedback parameters to vary, Rap1 & (-)FB diff., appeared to predict the data best (Figure 3.8C and Supplemental Figure B.10). The predicted ELPDs show that Rap1 & (-)FB diff. models do indeed predict the data best, followed by Rap1 diff. and (-)FB diff., respectively (Figure 3.8C). Furthermore, all three MMI methods only assign weight to hypotheses that allow both Rap1 and negative feedback differences simultaneously (Figure 3.8D). Interestingly, pseudo-BMA and stacking

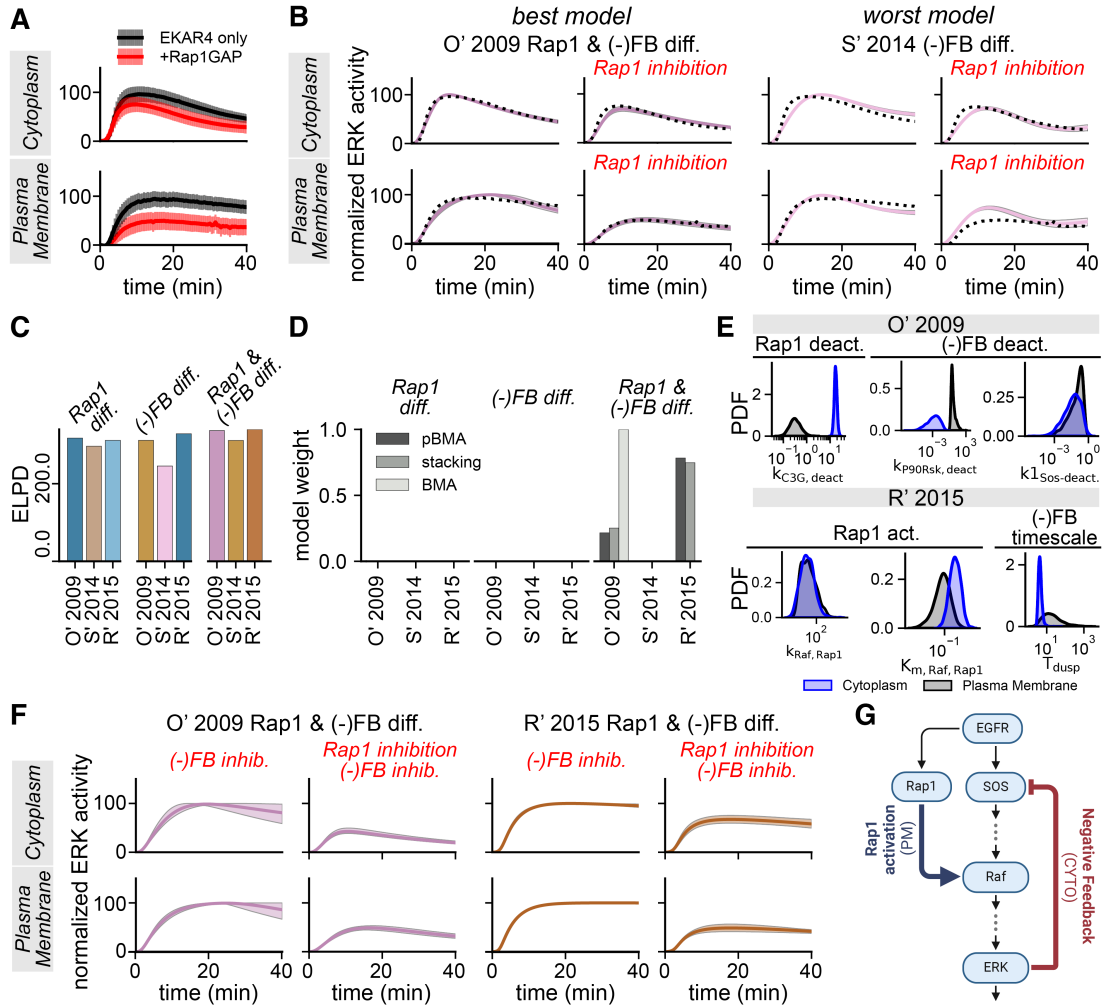


Figure 3.8: (A) EKAR4 measurements from [223] Figure 3C–3D normalized to the average maximum EKAR4 emission ratio in each location separately. Mean and standard deviation for EKAR4-only (black) and EKAR4 with Rap1 inhibition measurements (+Rap1GAP; red). (B) Posterior predictions for the best and worst models indicated by ELPD, i.e., O' 2009 with location differences in Rap1 and negative feedback (O' 2009 Rap1 & (-)FB diff.) and S' 2014 with location differences in negative feedback (S' 2014 (-)FB diff.), respectively. Predictions from all models are shown in Supplemental Figure B.10. (C) ELPD values for all models. (D) MMI weights assigned to all models using pseudo-BMA, stacking, and BMA. (E) Marginal posterior densities for the Rap1 activation/deactivation and ERK negative feedback parameters in each location for the O' 2009 and R' 2015 models with Rap1 and negative feedback location differences selected by MMI. Cytoplasm densities (CYTO) are blue, and those for the plasma membrane (PM) are black. All densities are statistically independent between locations ($p < 0.005$ by the Kolmogorov-Smirnov test with a two-sided hypothesis). (F) Posterior predictions for selected models with negative feedback inhibition alone ((-)FB inhib.) and simultaneous Rap1 and negative feedback inhibition (Rap1 inhibition + (-)FB inhib.). (G) The proposed mechanism of subcellular differences in ERK activity discovered using MMI. Rap1 activation is stronger at the plasma membrane, while negative feedback is stronger in the cytoplasm.

select both the O' 2009- and R' 2015-based models, while BMA selects the O' 2009-based model. This suggests that location-specific ERK negative feedback plays a role in location-specific ERK activity in addition to the Rap1 dependence at the plasma membrane.

Interestingly, in the hypotheses selected by MMI, Rap1 activation is stronger at the plasma membrane, and ERK negative feedback is stronger in the cytoplasm as indicated by the marginal posterior densities of the associated parameters (Figure 3.8E). Specifically, the deactivation rate of the Rap1 activator C3G ($k_{1C3G-deact.}$) is lower in the plasma membrane for the O' 2009 model, and the Rap1–Raf equilibrium constant activation rate ($K_{m,Raf,Rap1}$) is lower at the plasma membrane for the R' 2015 model. Additionally, the parameter estimates suggest stronger ERK negative feedback in the cytoplasm than in the plasma membrane. These results led us to conclude that stronger Rap1 activity drives sustained plasma membrane ERK activity, while stronger negative feedback is necessary for transient cytoplasmic ERK activity. Indeed, inhibition of ERK negative feedback nearly eliminates location-specific differences in ERK activity (Figure 3.8F). Inhibiting ERK negative feedback at the cytoplasm leads to ERK activation for both models and, therefore, to the loss of location-specific signaling. However, at the plasma membrane, negative feedback inhibition had little effect on ERK activity. Thus, these results show that ERK negative feedback and Rap1 activity are necessary for location-specific ERK activity.

Here, MMI allowed us to select hypotheses that *best* agreed with the data from a set of preselected, biologically relevant hypotheses. These findings suggest that cytoplasmic ERK activity depends on ERK negative feedback, while plasma membrane ERK activity depends on Rap1. Based on the MMI-selected hypothesis, we propose a new ERK negative feedback- and Rap1-dependent model for subcellular variability in ERK activity (Figure 3.8F).

3.4 Discussion

In this study, we demonstrate a novel application of Bayesian multimodel inference to the ERK signaling pathway. MMI enabled us to account for model form uncertainty in systems biology, specifically when multiple models of the same system were available. We compared three methods for MMI—Bayesian model averaging, pseudo-Bayesian model averaging, and stacking of predictive densities. Using ten ERK signaling models, we applied Bayesian MMI for two fundamental tasks: model averaging and model selection. We found that MMI for model averaging reduced predictive uncertainty by combining multiple models (Figure 3.3 and Supplemental Figures B.3, B.4), retained the accuracy of the best individual models and was robust to perturbations in the set of models (Figure 3.4 and Supplemental Figure B.5). We also found that MMI predictions were more robust to additional data uncertainty than individual models (Figure 3.5) and that MMI enabled better out-of-sample forecasts (Figure 3.5). Finally, when used for model selection, MMI allowed us to systematically compare many hypotheses and select the most consistent with the data (Figure 3.8). While our findings are based on a set of analyses that use ERK signaling as an example problem, the implications of this work—that leveraging multiple models can improve predictions and increase robustness—are likely to have impacts across systems biology and the broader modeling community.

It is standard to choose *a good model* based on the biological assumptions and scope of the model without assessing predictive performance. In contrast to using a single model for predictions, we found that Bayesian multimodel inference has several advantages. MMI increases the chances of making accurate predictions by weighting models based on predictive performance. Further, we found that MMI predictions show reduced uncertainty when averaged over multiple models. Therefore, the contribution to the overall uncertainty due to parameter estimation and model structure will be reduced. Posterior predictive simulations (Eq. (3.6) in Methods and Materials) show the total uncertainty, which includes

posterior uncertainty and data uncertainty [209]. We note that we only show posterior densities in this work to better highlight differences between models. This was especially prevalent when the models were used to predict future ERK activity (Figure 3.5 and Supplemental Figure B.8). Finally, we found that MMI is robust to model misspecifications and erroneous assumptions because it discards predictions from the *worst* models. For example, our predictions of the EGF-ERK dose response would have been very poor if we had chosen only to use the O' 2009 model—a reasonable and well-justified choice—(Supplemental Figure B.4); however, MMI assigned nearly no weight to that model. These findings show how MMI accounts for model uncertainty by leveraging all user-specified models and can increase predictive certainty.

Is there a best Bayesian multimodel inference method for systems biology? Here, we found that pseudo-BMA, which is widely applicable whenever any Bayesian inference method is used [214], yielded the most accurate and certain predictions. Stacking tended to only select the model with the highest ELPD in our examples except when multiple data were included, such as in Figure 3.8). It is important to note that stacking may only weight a single model when all of the models make similar predictions [269] (see, e.g., Figure 3.3). Finally, we found that BMA, although effective at model averaging, could substantially reduce predictive uncertainty while increasing predictive error beyond the best models (see, e.g., Figure B.4). Additionally, BMA requires using specific samplers such as sequential Monte Carlo or additional computations to estimate the marginal likelihood [215], [218]. While pseudo-BMA tended to perform well for model averaging and is more widely applicable than BMA, both methods demonstrated benefits over utilizing single models. Furthermore, for model selection, we found that including multiple MMI methods for model selection applications increased confidence in the selected models or hypotheses (see, e.g., Figure 3.8). Therefore, based on our findings, we conclude that pseudo-BMA is a good general approach for Bayesian MMI and recommend using more than one method for model selection. We also note that directly comparing Bayesian and frequentist approaches can improve our understanding of the

available MMI toolkit.

Notably, there are alternatives to the three MMI methods analyzed in this work. First, we chose to focus on Bayesian methods and exclude frequentist approaches that utilize metrics such as the Akaike or Bayesian information criteria [210], [214] because the uncertainty quantification provided by Bayesian methods is most appropriate for the limited-data regimes in systems biology. In the Bayesian literature, methods including stochastic search variable selection [270], probability density fusion [220], and reversible jump MCMC [271] have also been proposed for MMI (see references within [215] and [214] for additional methods). Methods such as ensemble estimation [272] can yield predictions similar to MMI. Causal inference [273], and sparsity-promoting inference [97], [274]–[276] can more-directly select models or mechanisms from data. A comparison between MMI-based model selection and alternative data-driven methods could further expand the tools available in systems biology.

How much data are needed for effective multimodel inference? While we found that MMI estimates are *robust* to the amount of training data provided, the minimum amount of required data will likely vary between applications and model sets. In this work, we used data that directly measured the QoI; however, MMI with data that do not measure the QoI directly might prove more challenging or result in predictions with higher uncertainty. In general, we note that the certainty and accuracy of MMI predictions are directly tied to the quality of the individual model predictions, which is well-known to improve with increasing data, e.g., [208]. One possible approach to assessing how the amount of available data impacts MMI predictions is to include uncertainties in the model weights. For example, Yao et al. [218] utilize the Bayesian bootstrap to assess model weight uncertainty by sampling the approximate distribution of possible pseudo-BMA model weights. Furthermore, in this work, MMI is based on metrics that assess predictive performance using a single dataset, and incorporating multiple datasets can potentially improve the confidence in and accuracy of MMI predictions. Such efforts may require multi-data multimodel inference, which could take

inspiration from meta-modeling to weigh datasets and models.

Future development of MMI for systems biology should focus on utilizing computationally efficient methods for parameter estimation and applying methods to learn time-varying MMI weights. MMI requires repeating Bayesian parameter estimation for every specified model, which can introduce a substantial computational burden to a modeling study (e.g., inference run times in Supplemental Table B.3). Approaches to accelerate Bayesian estimation such as variational inference [277] or Laplace approximation [108] can enable MMI with expensive-to-estimate models while providing approximate uncertainty quantification. Additionally, we presented MMI in a framework where the weights on each model do not vary over time. However, in settings with significant time-dependent changes, such as cell-fate transitions or development, one might expect that the underlying mechanism driving the biological process. Thus, the most applicable model could change over time. In those cases, extensions to the MMI framework, such as dynamic BMA [212] or sequential data assimilation with multiple models [278], can potentially learn model weights that are time-varying to allow for the MMI estimator to change over time.

As mathematical models continue to make critical contributions to studying biological systems, it is essential to account for model assumptions rigorously. While one of the gold standards in modeling is to develop a model that perfectly captures biological reality, limitations in our understanding of biological processes require us to rely on simplifying assumptions to build more tractable, simplified models. Furthermore, developing and working with complex, hyper-realistic models requires the ability to collect large quantities of very high-quality data to enable parameter estimation and make accurate predictions. Given that we often work with incomplete knowledge and limited data, we need to be able to make predictions that account for and are robust to uncertainties related to modeling assumptions. Thus, in these regimes, we propose that Bayesian multimodel inference provides a rigorous approach to handling model uncertainty that can potentially improve systems biology predictions and drive future biological

discoveries whenever multiple models of the same system are available. Additionally, when we have access to large datasets, MMI could be the starting point for developing detailed, biologically realistic models by determining the components of simpler models most consistent with the data. In summary, we conclude that MMI is a powerful approach to account for modeling assumptions in systems biology and can improve predictive certainty and inform future model development.

3.5 Data availability

All data supporting the findings of this study are available in the paper and the supplement. Original subcellular ERK activity data are available from Keyes et al. [223] and the associated supplementary materials.

3.6 Code availability

All code required to reproduce this work is available on Zenodo at .

3.7 Acknowledgments

Chapter 3, in full, is a modified reprint of the material as it appears in Increasing certainty in systems biology models using Bayesian multimodel inference. Nathaniel Linden-Santangeli, Jin Zhang, Boris Krämer, and Padmini Rangamani, *In review*, 2025.

Chapter 4

Systems modeling of AMP-activated protein kinase signaling using uncertainty quantification reveals exercise-type and duration-dependent input integration

Abstract

AMP-activated protein kinase (AMPK) plays a key role in restoring cellular metabolic homeostasis after energy stress. Importantly, AMPK acts as a hub of metabolic signaling, integrating multiple inputs and acting on numerous downstream targets to activate catabolic processes and inhibit anabolic ones. Despite the importance of AMPK signaling, unlike other well-studied pathways, such as MAPK/ERK or NF- κ B, only a handful of mechanistic models of AMPK signaling have been developed. Epistemic

uncertainty in the biochemical mechanism of AMPK activity, combined with the complexity of the AMPK pathway, makes model development particularly challenging. Here, we leveraged uncertainty quantification methods and recently developed AMPK biosensors to construct a new, data-informed model of AMPK signaling. Specifically, we applied Bayesian parameter estimation and model selection to ensure that model predictions and assumptions are well-constrained to measurements of AMPK activity using recently developed AMPK biosensors. As an application of the new model, we predicted AMPK activity in response to exercise-like stimuli. We found that AMPK acts as a time- and exercise-dependent integrator of its input. Our results highlight how uncertainty quantification can facilitate model development and address epistemic uncertainty in a complex signaling pathway, such as AMPK. This work shows the potential for future applications of uncertainty quantification in systems biology to drive new biological insights by incorporating state-of-the-art experimental data at all stages of model development.

4.1 Introduction

AMP-activated protein kinase (AMPK) is a master regulator of cellular metabolism that maintains energy homeostasis [279]–[284]. AMPK senses and responds to decreasing cellular energy availability [281], [282]. As a hub of metabolic signaling, AMPK integrates numerous inputs and acts on a large number of downstream targets [51], [285]–[289]. Inputs to AMPK include cellular energy status [282], [286], intracellular calcium [290]–[292], and glucose levels [280]. AMPK is a heterotrimeric protein comprising an α -catalytic subunit and two regulatory subunits, the β - and γ -subunits [293], [294]. Phosphorylation of a key threonine residue, Thr172, on the γ -subunit leads to a 100-fold increase in the AMPK kinase activity and is catalyzed by many upstream kinases, such as LKB1 and CaMKK2 [283], [293]. Phosphorylated AMPK (pAMPK) is additionally dephosphorylated by phosphatases that include PP2A. Upon activation, AMPK phosphorylates a wide range of targets to promote catabolic processes and restore

falling ATP levels [279], [281], [282], including acetyl-CoA carboxylase (ACC) [295], ULK1 [296], and RAPTOR [297]. The centrality of AMPK makes it a critical drug target [298], with well-known drugs including metformin [299], [300] and salicylate (Aspirin) [301] that act as AMPK activators.

A wide range of experimental approaches that include traditional molecular biology, modern omics, and structural biology have been used extensively to study the AMPK signaling network at the systems level [283], [285], [286], [302]. These studies have led to significant findings, such as the specificity of AMPK targets [286], [303], the mechanism of adenine nucleotide regulation of AMPK activity [293], and the *in vivo* response of AMPK to physiological stimuli such as exercise [302] (Figure 4.1). However, unlike other well-studied pathways such as MAPK/ERK [44], [77] or NF- κ B [40], which have 10s–100s of published mathematical models, fewer than five robust mechanistic models of AMPK signaling have been published [49], [50]. Several previous intracellular signaling models of metabolic pathways have embedded AMPK in a larger signaling network **sonntagModelling2012**, [51], [287], [288]; however, these models often leverage phenomenological representations of AMPK activation as a function of select inputs and do not yet capture the regulation of AMPK activity in mechanistic detail. Alternatively, mechanistic models, such as those developed by Connolly et al. [50] and Coccimiglio et al. [49], aim to capture the details of AMPK activation and its interactions with adenine nucleotides in the contexts of neurotoxicity [50] and exercise [49]. While these previous models of AMPK signaling make well-justified assumptions, they were not rigorously constrained to high-quality AMPK signaling data from experiments. Here, we leveraged AMPK activity data measured using a recently developed fluorescent AMPK biosensor, ExRai-AMPKAR [304], to inform model development. These fluorescent biosensors are AMPK substrates and report relative phosphorylation by AMPK [7], [304], [305].

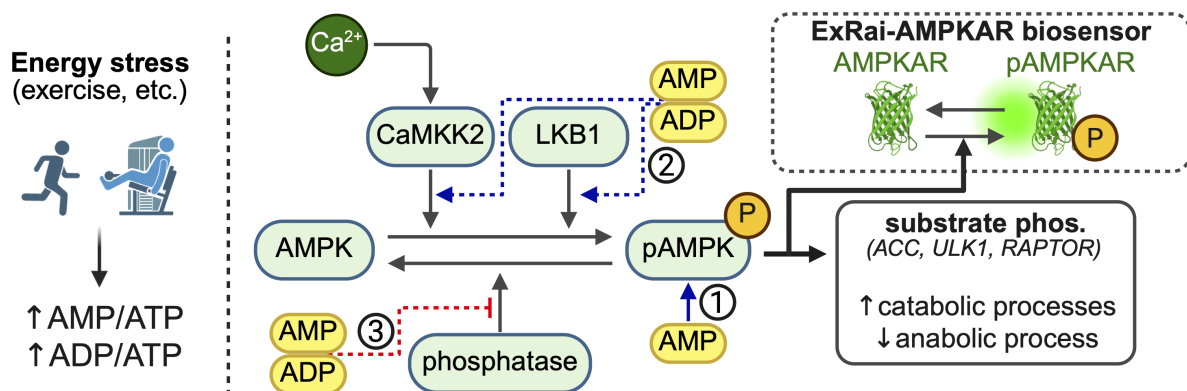


Figure 4.1: Changes in intracellular adenine nucleotide concentrations, AMP/ATP and ADP/ATP, activate AMPK in three ways: (i) AMP acts as a direct allosteric activator of AMPK. (ii) AMP and ADP promote phosphorylation (activation) of AMPK by inducing conformational changes to the AMPK complex. (iii) AMP and ADP protect phosphorylated pAMPK from dephosphorylation. AMPK phosphorylates several downstream targets to promote catabolic and inhibit anabolic processes. AMPK activity can be measured using fluorescent kinase activity reporters such as ExRai-AMPKAR. These activity reporters are substrates of AMPK and report relative levels of phosphorylated by AMPK. Created in <https://BioRender.com>.

Model development for signaling networks typically leverages literature and domain expertise to identify relevant biochemical reactions and formulate a series of modeling assumptions [14], [77], [306]. These assumptions are then encoded into mathematical equations by prescribing kinetic terms, such as mass action or Michaelis-Menten kinetics, to represent reaction fluxes [64]. While such a workflow has led to numerous successful models, applying this approach to systems like AMPK, where new biology is still being discovered, can lead to a strong reliance on model assumptions. Recent research has introduced approaches for data-informed model development that leverage uncertainty quantification (UQ) [307] to ensure that models are well constrained to experimental measurements [308]–[310]. Here, we applied uncertainty quantification methods, including identifiability analysis [311], sensitivity analysis [100], Bayesian parameter estimation [308]–[310], and Bayesian model selection [77], [312], [313] to leverage AMPK activity data throughout the model development process.

In this work, we developed a set of related AMPK signaling models that vary in their mechanistic assumptions. Using Bayesian UQ, we estimated the associated kinetic parameters of these models

using experimental data from Schmitt et al. [72], and selected the model that best captured the data. We then applied the new model to investigate AMPK signaling in the context of exercise. AMPK, which is thought to be activated during exercise, plays a multifaceted and complex role in regulating metabolism **mcconellSkeletalMuscleAMPK2020**, **mcconellItsWellTruly2020** and modulating post-exercise adaptation **hoffmanGlobalPhosphoproteomicAnalysis2015**, **kjobstedAMPKTBC1D1Regulate2019**, **lantierAMPKControlsExercise2014**, [302]. We therefore simulated different exercise regimes to generate physiologically relevant inputs to the AMPK signaling network. We found that AMPK is a robust integrator of high-frequency inputs, but shows exercise-dependent responses to longer timescale, low-frequency inputs. Our findings highlight how uncertainty quantification enables rigorous, data-informed predictive modeling for a system like AMPK, where our understanding of the biology is still evolving.

4.2 Methods and Materials

4.2.1 Workflow for model development with rigorous uncertainty quantification

Uncertainty quantification (UQ) aims to understand how errors in the data and models impact model predictions [307]. Recently, UQ methods for identifiability analysis [311], [314], sensitivity analysis [100], parameter estimation [308], [315], and multimodel inference [77], [313], [316] have been applied in systems biology. Here, we leverage an end-to-end model development workflow to ensure that both model assumptions and predictions are constrained by available data (Figure 4.2). First, we developed a set of models that vary in the assumed mechanism and in the kinetic formulation used to represent the AMPK signaling system (Figure 4.2A). Next, we leveraged our previously developed approach for Bayesian parameter estimation that begins by reducing the dimensionality of the free parameter space using

structural identifiability and global sensitivity analyses [308] (Figure 4.2B). In this work, we added the step of prior elicitation and refinement to ensure that parameters can only vary over regions of parameter space that lead to reasonable model predictions. Finally, we propagated uncertainty in the parameters forward to predictions by running ensembles of simulations, and used Bayesian model selection to choose the model that is most consistent with the data (Figure 4.2B). We summarize the details of each of these methods below. Supplemental Section C.1.3 also highlights prior elicitation and refinement in more detail.

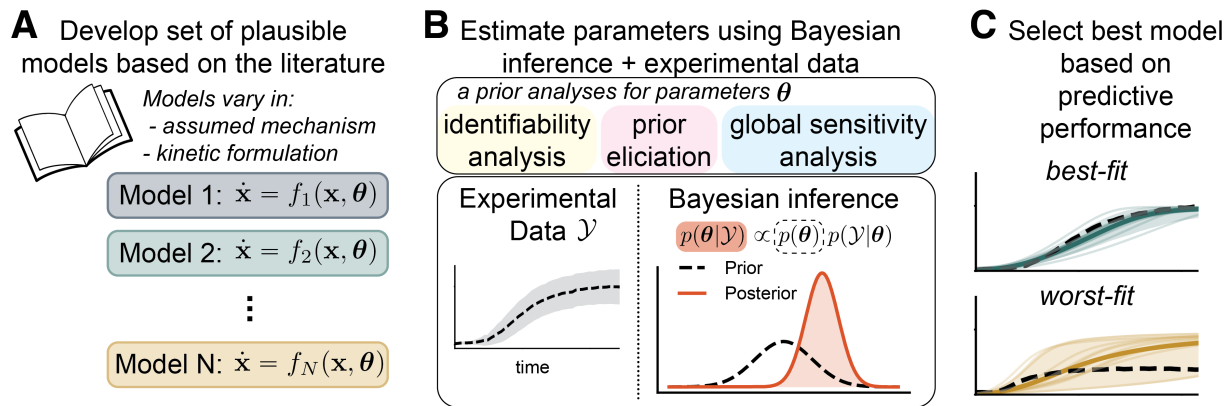


Figure 4.2: (A) First, we developed a set of AMPK signaling models based on available literature. The models vary in the assumed mechanism and the kinetic formulation. (B) Next, we estimated model parameters from experimental AMPK signaling data using Bayesian inference. We performed *a priori* identifiability and sensitivity analyses to reduce the dimension of the parameter space. Prior elicitation was used to ensure that priors lead to reasonable model predictions. (C) Finally, we used Bayesian model selection to select the best model based on predictive performance.

Local structural identifiability analysis

Local structural identifiability implies that a parameter can be uniquely identified within a local region of parameter space [311], [317]. In this work, we used the StructuralIdentifiability.jl package in the Julia programming language for local structural identifiability analysis [318]. Specifically, we used default settings for the `assess_local_identifiability()` function, and set the probability of correctness to $p = 0.99$ for all models. We fixed the parameters for the metabolism module and calcium-calmodulin-CaMKK2 reactions to nominal values. Supplemental Table C.6 lists the identifiability of each

model. We fix all nonidentifiable parameters to nominal values, which are potentially refined following the iterative process in Supplemental Section C.1.3.

Prior elicitation

We used direct prior elicitation [319] to find prior densities for free parameters such that 95% of the probability density falls between a specified range. Specifically, we solved an optimization problem for the prior hyperparameters that seeks to find the prior with the highest entropy that meets the 95% probability bounds using the Preliz Python library **Icazatti_2023**. We specified the bounds to vary by several orders of magnitude around nominal values (Supplemental Table 4.1). We then ran prior predictive samples and performed prior predictive elicitation to refine the priors if the initial densities did not yield reasonable predictions **Icazatti_2023**, [319]. We detail this iterative refinement procedure in Supplemental Section C.1.3. We used lognormal prior densities for all parameters.

Global sensitivity analysis

Sobol global sensitivity analysis (GSA) partitions the total variance of an output quantity into the contributions of varying the uncertain model parameters over specified ranges [100]. Here we used the SALib Python library for GSA [252], [253]. Specifically, we allowed identifiable parameters to vary over ranges defined by the 90% highest-density interval of the prior density. Next, we generated a set of N random parameter samples from those ranges, where N is given by $N = 256(p + 1)$ and p is the number of free parameters. We then simulated the models with these free parameters and calculated the maximum AMPKAR activation and time to half-maximum activation for the wild-type, LKB1 knockout, and CaMKK2 knockout conditions. Finally, we computed the Sobol total sensitivity index for each of the six quantities of interest (Supplemental Tables C.7–C.12). We considered a parameter to be sensitive if it

had a total sensitivity greater than 0.01 for any of the quantities. After this, the non-sensitive parameters were fixed to nominal values.

Bayesian parameter estimation

In this section, we provide a brief overview of Bayesian parameter estimation. For more details in the context of systems biology, see [308], [315], and for general theory, see [307], [309]. We model the AMPK signaling network with a system of parametric ordinary differential equations (ODEs) that describe the rates of change of signaling species. Specifically, the models are defined as

$$\frac{d\mathbf{x}(t)}{dt} = f(\mathbf{x}(t); \boldsymbol{\theta}), \quad (4.1)$$

$$\hat{y}(t) = h(\mathbf{x}(t); \boldsymbol{\theta}) + \boldsymbol{\eta}(t), \quad \boldsymbol{\eta}(t) \sim \mathcal{N}(\mathbf{0}, \gamma). \quad (4.2)$$

Here, the state variables, $\mathbf{x}(t) \in \mathbb{R}_+^n$, correspond to the concentration of biochemical species ($\mathbb{R}_+ = [0, \infty)$ are the nonnegative real numbers). The observations, $\hat{y}(t) \in [0, 1]$, are the fraction of activated AMPKAR biosensor. We assume that an independent and identically distributed Gaussian noise process, $\boldsymbol{\eta}(t) \in \mathbb{R}$ with variance $\gamma \in \mathbb{R}$. The function $f(\cdot; \cdot) : \mathbb{R}_+^n \times \mathbb{R}_+^p \rightarrow \mathbb{R}_+^n$ describes the system dynamics, and $h(\cdot; \cdot) : \mathbb{R}_+^n \times \mathbb{R}_+^p \rightarrow \mathbb{R}_+^m$ maps from states to observables. The model parameters $\boldsymbol{\theta} \in \boldsymbol{\Theta} \subseteq \mathbb{R}_+^p$ include reaction rates, equilibrium coefficients and other quantities that control model behavior.

Bayesian inference characterizes a probability density for the unknown model parameters conditioned on available data, $p(\boldsymbol{\theta}|\mathcal{Y})$ [307], [309]. In this work, the training data, $\mathcal{Y} = \{y_1, \dots, y_{N_{\text{train}}}\}$ consists of N_{train} noisy experimental ExRai-AMPKAR observations. The probability density of the model parameters conditioned on data, called the posterior density, is learned using Bayes' rule, which is defined

as,

$$\underbrace{p(\theta|\mathcal{Y})}_{\text{posterior}} \propto \underbrace{p(\theta)}_{\text{prior}} \underbrace{p(\mathcal{Y}|\theta)}_{\text{likelihood}}. \quad (4.3)$$

Here, $p(\theta)$ is the prior density which encodes assumptions about parameters before data is considered, and $p(\mathcal{Y}|\theta)$ is the likelihood function that measures the probability that the model correctly predicts the training data. Because the models are nonlinear, we cannot evaluate the posterior density directly, so we instead must rely on methods such as *Markov chain Monte Carlo* (MCMC) [307]–[309] or *variational inference* (VI) [277] to characterize the posterior through the S samples drawn from it, $\{\theta_1, \dots, \theta_S\} \sim p(\theta|\mathcal{Y})$. In this work, we use the Pathfinder algorithm for VI [320], which approximates the posterior density with an easier-to-sample approximation by solving an optimization problem, because variational inference can be more computationally efficient than MCMC [277], [320]. We then draw samples from this approximating density. For a review of variational inference, see [277]. We implement all probabilistic models in the PyMC probabilistic programming language [232] and use default settings for the Pathfinder algorithm.

Bayesian model selection

We use the expected log pointwise predictive density (ELPD) to assess predictive performance and direct model selection [309], [321]. Larger ELPD values indicate better predictive performance because they indicate a higher probability of correctly predicting out-of-sample data. We assume that the training data $\mathcal{Y}$ consist of statistically independent data points, such that $p(\mathcal{Y}|\theta) = \prod_{i=1}^{N_{\text{train}}} p(y_i|\theta)$. The ELPD is defined as

$$\text{ELPD} := \sum_{i=1}^{N_{\text{train}}} \int \log p(\tilde{y}_i|\mathcal{Y}) p_{\text{true}}(\tilde{y}_i) d\tilde{y}_i, \quad (4.4)$$

and quantifies the expected predictive performance of the model compared to the true data-generating distribution $p_{\text{true}}(\tilde{y}_i)$, where $p(\tilde{y}_i|\mathcal{Y})$ is the posterior predictive density of out-of-sample data point $\tilde{y}_i$. In

general, the ELPD is computationally intractable because we do not know the true data-generating density, $p_{\text{true}}(\tilde{y}_i)$, so instead we approximate the ELPD using the leave-one-out cross-validation estimator (LOO-CV) [321]. Specifically, we use the ArviZ Python library [240], which uses Pareto-smoothed importance sampling to compute the LOO-CV estimator [322]. We use default settings for all computations.

4.2.2 ODE simulation

We solved all differential equations numerically using the Kvaerno 4/5-order implicit Runge Kutta method, which is implemented in the Diffrax Python library [241], [323]. Additionally, we used a PID controller-based adaptive time-stepping algorithm with tolerances `atol=1e-6` and `rtol=1e-6` [243], [244]. We chose the Diffrax library because it enables us to compute gradients of the ODE solution with respect to the model parameters using autodifferentiation, which is necessary for employing Pathfinder VI for parameter estimation.

4.2.3 Experimental data pre-processing

The ExRai-AMPKAR AMPK kinase activity reporter, developed in [72], is a ratiometric biosensor in which the ratio of emission intensities at two excitation wavelengths increases as the sensor is phosphorylated by active AMPK. Therefore, the emission intensity ratio is considered to vary with the degree of AMPK kinase activity. We aimed to calibrate the ratiometric readout to the fraction of phosphorylated sensor, which can be computed from model outputs. Similar normalizations are reviewed in [70], [324]. To normalize the data, we assumed that (i) the concentration of phosphorylated ExRai-AMPKAR is negligible before 2-DG stimulation, (ii) nearly all expressed ExRai-AMPKAR is phosphorylated when the ratiometric signal is at its maximum, and (iii) the fraction of phosphorylated sensor varies proportionally to the ratiometric signal. Based on these assumptions, we renormalized the average normalized ratiometric

AMPKAR data, defined as $R(t)$, using the normalization

$$R_{\text{norm}}(t) := \frac{R(t) - R_{\min}}{R_{\max} - R_{\min}},$$

where $R_{\min}$ and $R_{\max}$ are the minima and maxima of the signal in time. We additionally normalized the variance of the signaling by the same normalization factor. We used the minimum and maximum signals from the wild-type condition to normalize the kinase knockout conditions. To predict the renormalized signals from our models, we assume that

$$R_{\text{norm}}(t) \approx \frac{[\text{pAMPKAR}](t)}{[\text{AMPKAR}]_{\text{total}}}, \quad (4.5)$$

where $[\text{pAMPKAR}](t)$ is the time-varying concentration of the phosphorylated biosensor, and $[\text{AMPKAR}]_{\text{total}}$ is the total concentration of both active and inactive ExRai-AMPKAR. We drop “ExRai-” for brevity.

4.2.4 Statistical comparison

We use the pairwise Student’s t-tests with the Bonferroni correction to adjust for multiple comparisons to test for statistical significance of the difference between population means. We assess significance at a level of $\alpha = 0.05$. To effectively summarize comparisons between multiple groups, we employ a compact letter display, where groups with the same letter do not exhibit statistically significant differences in their means.

4.3 Results

4.3.1 Model development

To capture AMPK activation in response to energy stresses, we developed a set of six mechanistic models of the AMPK signaling pathway (Figure 4.1) using the UQ workflow outlined in Figure 4.2. We included a minimal representation of metabolic processes, including glycolysis, oxidative phosphorylation, and ATP hydrolysis, and captured key mechanisms of AMPK activation and regulation, including AMPK phosphorylation and interactions with adenine nucleotides [279], [283]. Although we used previous models, including [49], [51], during model development, our new models differ in several ways. First, we included the specific mechanisms by which adenine nucleotides regulate AMPK activity in our models. Second, we explicitly included the upstream kinases, LKB1 and CaMKK2, to capture AMPK activation due to energy stress and calcium signaling. These kinases are known to activate AMPK via phosphorylation [290]–[292], [325]. Third, we varied the associated mechanisms between the six models to investigate how adenine nucleotide binding to AMPK affects AMPK phosphorylation and activity. Importantly, we investigated models that employ different kinetic formulations for the reaction rates before constructing models of AMPK signaling. The remainder of this section delineates key model assumptions and describes our models.

Kinetic formulations impact parameter identifiability and predictive accuracy

Models of intracellular signaling typically prescribe kinetic equations, including mass action, Michaelis-Menten, and Hill-type, to describe the dynamics of biochemical reactions [64], [306]. If the reaction mechanism is well known, then the best practice is to choose the corresponding kinetic formulation, for example, prescribing Hill-type kinetics when a reaction is cooperative [64]. However, when little is known about a reaction mechanism, kinetic formulations are often prescribed in an *ad hoc* or

phenomenological fashion to ensure that a model retains the flexibility to capture available experimental observations [306]. However, if these modeling choices are made while considering impacts on parameter and predictive uncertainties, then subsequent parameter estimation will be more successful [308].

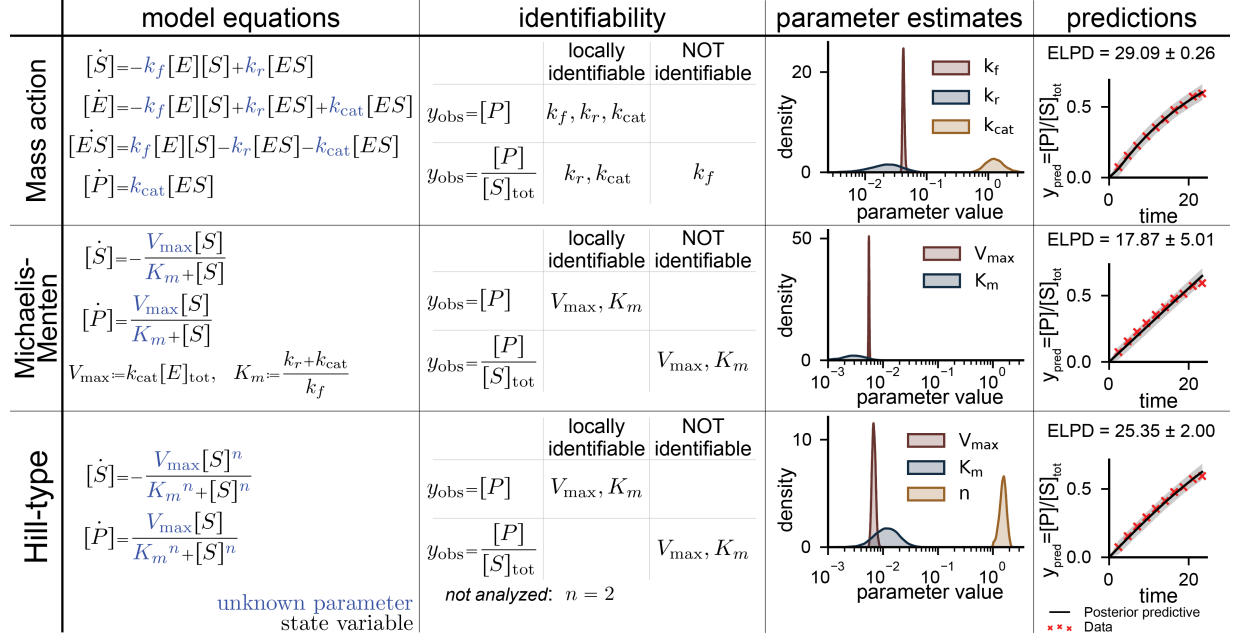


Figure 4.3: Model equations: mass action, Michaelis-Menten, and Hill-type kinetics can be used to formulate differential equation models of a single enzyme-catalyzed reaction. Uncertain free parameters are shown in blue. *Identifiability:* Different kinetic formulations yield different structural local identifiability from direct observations of the concentration of the product ($y_{obs} = [P]$) and more experimentally feasible observations of the ratio of product to total substrate ($y_{obs} = \frac{[P]}{[S]_{tot}}$). *Parameter estimates:* All free model parameters are estimated from data of the ratio of product to total substrate shown in red. Black traces show the posterior mean and gray shading shows the 95% posterior credible interval computed using 8,000 posterior samples.

To understand how different kinetic formulations impact parameter estimation and model predictions, we first modeled a single enzyme-catalyzed reaction using mass action, Michaelis-Menten, and Hill-type kinetics. The reaction



describes the formation of product P from substrate S, catalyzed by conserved enzyme E. The reaction rates k_f , k_r , and k_{cat} describe the rate of substrate-enzyme binding and unbinding, and the rate of formation of

the product, respectively. In Figure 4.3 we show the three models; we note that the number of equations and free parameters varied between the models. Interestingly, the local identifiability of the model parameters was different for these different reaction rate formulations (Figure 4.3). We found that all model parameters for the three models are locally identifiable when the product concentration is observed directly, $y_{\text{obs}} = [P]$. However, the identifiability differed between the model formulations when the ratio of the concentration of product to the total concentration of substrate, $y_{\text{obs}} = [P]/[S]_{\text{tot}}$, was observed. Notably, two of the three parameters of the mass action model were locally identifiable, while none of the Michaelis-Menten or the Hill-type models were identifiable.

Next, to assess how structural nonidentifiabilities impacted parameter estimation, we fit the three models to experimental data of the phosphorylation reaction catalyzed by pyruvate kinase from [326]. In the reaction, phosphoenolpyruvate is converted to pyruvate by the pyruvate kinase enzyme, which, under the experimental conditions, can be considered a single-substrate, single-enzyme reaction. We estimated the distribution of all model parameters using Bayesian inference to gain a better understanding of how nonidentifiabilities affect the estimates. Specifically, we normalized the data to a total substrate concentration of 0.195 mM used in the experiments [326] to mimic the data that observes the product as a fraction of the total substrate. In each of the models, we were able to estimate certain parameters with high certainty (k_f and V_{max}) and others with low certainty (k_r and K_m ; Figure 4.3). Next, we simulated the models using the posterior parameter samples to propagate the parametric uncertainty forward to predictions (Figure 4.3). The mass action model had the highest probability of correctly predicting future data measured by the expected log pointwise predictive density (ELPD; see Eq (4.4) for the definition; ELPD = 29.09). The Hill-type model was next best (ELPD = 25.35), and the Michaelis-Menten model was the worst of the three (ELPD = 17.87).

Since systems biology models typically involve cascades of coupled biochemical reactions, we

constructed nine models of the two-reaction system in which a substrate is phosphorylated by a kinase to form a product, which is then dephosphorylated (Supplemental Section C.1.1). Interestingly, we found that structural identifiability of the model parameters is not impacted by the coupling of two reactions (Supplemental Table C.2). For example, the identifiability of the model parameters for the model that uses mass action kinetics for both the phosphorylation and dephosphorylation reactions (Supplemental Table C.2) was the same as the identifiability for the single reaction mass action model (Figure 4.3).

Overall, for the single-reaction system, the three model formulations were all limited in their ability to capture the pyruvate kinase reaction without introducing structural nonidentifiabilities or increasing the number of equations and parameters in the model. The mass action formulation appeared to be a safe choice because it had the most favorable identifiability, and the additional reaction (and equation) did not introduce a significant computational burden. Contrary to common belief, the Michaelis-Menten formulation did not seem to be the best choice for this example, because it had the lowest predictive accuracy (lowest ELPD). The Hill-type model was almost as good as the mass action model in terms of predictive performance; however, we found that the inclusion of the Hill coefficient, n , made parameter estimation more difficult [308]. Therefore, we concluded that for these simple reactions, the mass action models had more favorable identifiability and yielded higher-quality predictions. However, despite the nonidentifiabilities, we could still estimate the parameters of the Michaelis-Menten and Hill-type models.

Proposed AMPK signaling models describe a range of regulation and signaling mechanisms

Since we did not find a *best* kinetic formulation for describing reaction rates in terms of identifiability in the previous section, we built six related models that use either mass action or Michaelis-Menten kinetics. Additionally, the models differed in their AMPK regulatory mechanisms. Table 4.1 highlights the differences between the models. Each model consisted of three core modules: (i) cellular energetics

and metabolism, (ii) AMPK regulation, and (iii) AMPK kinase activity and biosensor. In this section, we outline our assumptions and summarize the three core modules. We provide the model equations and initial conditions in the Systems Biology Markup Language (SBML) format, as well as Python functions, in the available code. We list nominal parameters in Supplemental Tables C.4–C.5.

Module 1: Cellular metabolism and energy stress AMPK senses changes in the cellular energy state through changes in adenine nucleotide levels [279], [283]. Cellular metabolism is a highly regulated process that tightly controls the production of ATP and the conversion between ADP and AMP [281]. To ensure that our models maintain physiologically consistent adenine nucleotide levels, we adapted the simplified metabolism models from Coccimiglio et al. [49] and Lueng et al. [51] to model cellular ATP production, consumption, and conversion between AMP, ADP, and ATP. We modeled ATP generation by both oxidative phosphorylation and glycolysis, and ATP consumption via a global hydrolysis reaction. Additionally, we included the adenylate cyclase and the creatine kinase reactions to ensure that the appropriate stoichiometry is maintained between adenine nucleotides. Throughout this work, we assumed that the initial concentrations of AMP, ADP, and ATP are 7.03 mM, 1.11 mM, and 7.9×10^{-2} mM, respectively, which we found to be a steady state of the metabolism model (Supplemental Figure C.1A). We modeled energy stresses due to the 2-deoxyglucose (2-DG) stimulus applied in [72] by decreasing the rate of glycolytic ATP production 100-fold (Supplemental Figure C.1B–D). Additionally, we modeled energy stress due to increased ATP consumption, such as during exercise, by increasing the ATP hydrolysis rate k_{hydro} or by prescribing an additional negative ATP flux. We used the same metabolism module across all the models that we developed. Supplemental Section C.1.2 describes the metabolism model in more detail, and Supplemental Tables C.3– C.4 provide the flux equations and reaction parameters, respectively.

Module 2: Adenine nucleotide binding and AMPK phosphorylation AMPK activity is regulated by changes in adenine nucleotide concentrations through a tripartite mechanism that modulates net AMPK phosphorylation and directly activates AMPK (Figure 4.1) [279], [281], [283], [293], [294]. Net AMPK phosphorylation is regulated by several effects of AMP and ADP binding that induce conformational changes to the AMPK complex, including promoting phosphorylation at Thr172 and protecting from dephosphorylation (Figure 4.1). Adenine nucleotides can bind to four cystathionine- β -synthase (CBS) repeats on the γ -subunit of AMPK [283], [293]. However, recent evidence suggests that only binding in the third CBS repeat regulates AMPK activity [283]. Thus, we only accounted for single adenine nucleotide-AMPK binding interactions and assumed that adenine nucleotides can bind and unbind from AMPK.

We modeled each of the three parts of the tripartite AMPK regulatory mechanism (Figure 4.1). First, AMP allosterically activates AMPK, in which AMP-AMPK binding has been found to result in a 10-fold activation in AMPK kinase activity [293], [295], [327]–[329]. Second, AMP and possibly ADP binding have been found to promote AMPK phosphorylation by upstream kinases by inducing conformational changes that expose Thr172 to these kinases [293], [330], [331]. Third, AMP and ADP binding protect Thr172 from dephosphorylation [293], [327], [329]. Notably, ATP is thought to antagonize the effects by competing for the third CBS binding site and induce inhibitory effects on AMPK [283], [294], [298]. Thus, AMPK is sensitive to the AMP/ATP and ADP/ATP ratios.

We constructed three sets of models that make distinct assumptions about the mechanism of adenine nucleotide binding interactions, all of which are potentially valid according to recent literature. In our models, LKB1 activity remains constant, and only interactions between adenine nucleotides and AMPK regulate net LKB1-mediated phosphorylation. We assumed that AMP and ADP are essential activators of LKB1 in Models 1 and 2, where AMP or ADP binding is necessary for LKB1 to phosphorylate AMPK [64].

Table 4.1: Set of AMPK models vary in assumed reaction mechanisms.

Model Number	Kinetics	Allosteric activation by AMP	Promotion of phosphorylation	Protection from dephosphorylation	Antagonism by ATP
1	mass action (MA)	AMP-pAMPK is the only active kinase	only AMP/ADP-AMPK are phosphorylated by LKB1	AMP-pAMPK and ADP-pAMPK cannot be dephosphorylated	ATP-AMPK not phosphorylated by LKB1 & ATP-pAMPK not active
2	Michaelis-Menten (MM)	same as 1	same as 1	same as 1	same as 1
3	MA	AMP is a nonessential act. (activity enhanced by $\beta_{\text{AMP}} \geq 1$)	AMP/ADP reduce LKB1-AMPK K_d by $\alpha_{\text{LKB1}} < 1$	AMP/ADP increase PP-pAMPK K_d by $\alpha_{\text{PP}} > 1$	same as 1
4	MM	same as 3	same as 3	same as 3	same as 1
5	MM	AMP is a nonessential act. (activity enhanced by $\beta_{\text{AMP}} \geq 1$)	AMP/ADP increase LKB1 phosphorylation rate by $\beta_{\text{LKB1}} > 1$	AMP/ADP decrease PP phosphorylation rate by $\beta_{\text{PP}} < 1$	n/a
6	MM	same as 5	same as 5	same as 5	same as 5

Alternatively, in Models 3–6, we assumed that AMP and ADP are nonessential activators of LKB1, where AMP and ADP binding either decrease the K_d of LKB1 binding (Models 3 and 4) or increase the rate at which LKB1 phosphorylates the AMPK complex (Models 5 and 5) [64]. In addition to LKB1, we included CaMKK2, which is known to phosphorylate AMPK in a Ca^{2+} -dependent manner [291], [301]. We assumed that CaMKK2 activity depends on the intracellular Ca^{2+} concentration through calcium-calmodulin binding with the kinetics from [58], [332]. In Models 1–4, we assumed that adenine nucleotide binding does not affect CaMKK2-mediated phosphorylation, such that all AMPK complexes are phosphorylated with the same kinetics, because there has been debate as to whether CaMKK2 activity is impacted by adenine nucleotide binding [283], [286], [293], [298], [327]. However, to capture the possibility that AMP and

ADP promote CaMKK2-, in addition to LKB1-, mediated AMPK phosphorylation, in Models 5 and 6, we modeled these reactions with a nonessential activation mechanism. Lastly, we modeled the net dephosphorylation of phosphorylated pAMPK with a single dephosphorylation reaction, assuming that AMP and ADP, but not ATP, protect from dephosphorylation. In Models 1 and 2, we assumed that AMP and ADP binding completely protect pAMPK from dephosphorylation while free pAMPK and ATP-bound pAMPK are susceptible to dephosphorylation. In Models 3–6, we assumed that AMP and ADP binding reduce the effectiveness of pAMPK dephosphorylation via nonessential inhibition [64]. We assumed the total enzyme concentrations and initial conditions are the same across all six models.

Module 3: AMPK kinase activity and fluorescent biosensor Phosphorylated AMPK is an active kinase with many downstream targets [285]. As its name suggests, AMPK is activated nearly 10-fold by AMP binding [293], [295], [327]–[329]. Here, we modeled AMP as either an essential activator of AMPK (Models 1 and 2) or a nonessential activator of AMPK (Models 3–6). To enable direct calibration of our models to fluorescent microscopy data of AMPK activity from [72], we explicitly modeled the AMPK activity biosensor, ExRai-AMPKAR, as an AMPK substrate [70], [324]. We refer to the ExRai-AMPKAR biosensor as AMPKAR for brevity. In all the models, we assumed that active AMPK phosphorylates AMPKAR [7], [72], [305]. Additionally, we modeled net phosphorylated pAMPKAR dephosphorylation with a single dephosphorylation reaction. We assumed that the fraction of pAMPKAR to total AMPKAR is directly equivalent to the fraction of the above baseline fluorescent sensor response (Eq (4.5)).

4.3.2 Application of the UQ framework suggests that AMP and ADP act as nonessential activators of AMPK

Having developed a set of AMPK signaling models that reflect possible AMPK regulatory mechanisms, we applied the UQ-based model development workflow shown in Figure 4.2. First, we estimated

model parameters to constrain the model predictions to experimental data. Although there are some direct experimental observations of the activities of LKB1, CaMKK2, AMPK, and the related phosphatases, and with AMPK adenine nucleotide binding affinities, these data were not collected in the same cellular context as the available experimental data [293], [325], [328], [333]. Therefore, we aimed to estimate unknown model parameters using physiologically relevant data of energy-stress induced AMPK activity [72]. The number of free parameters ranged from 16 to 25 Model 1 and Model 5, respectively. Here, we used Bayesian parameter estimation because it allows us to leverage previous knowledge of parameters to inform the parameter estimates and to quantify uncertainty in estimates [307]–[309]. Bayesian estimation represents unknown parameters as random variables and aims to infer their probability densities conditioned on experimental data and prior knowledge (see Section 4.2.1).

We have previously found that restricting the set of estimated parameters to identifiable and influential parameters is necessary for successful parameter estimation [308]. So, we performed structural identifiability and global sensitivity analyses before estimating any parameter (Supplemental Table C.6). For the structural identifiability analysis, we fixed the metabolism and calcium-calmodulin model parameters to the nominal values listed in Supplemental Table 4.1, because we had well-established models for those components [49], [51], [58], [332]. Interestingly, contrary to the results for the single-enzyme reactions, we found that all of the mass action model parameters (Models 1, 3, and 5) were structurally locally identifiable. Most of the Michael-Menten model parameters (Models 2, 5, and 6) were identifiable, except for those that control AMPK phosphorylation of the sensor and sensor dephosphorylation. For the subsequent analyses and parameter estimation, we fixed nonidentifiable parameters to the nominal values listed in Supplemental Table 4.1.

Next, we performed global sensitivity analysis using Sobol sensitivity analysis to determine which uncertain parameters have the greatest influence on model predictions in response to a severe

energy-stress-like stimulus [100]. We specified the plausible ranges for parameters spanned by several orders of magnitude around the nominal parameters (Supplemental Table 4.1). However, as these ranges

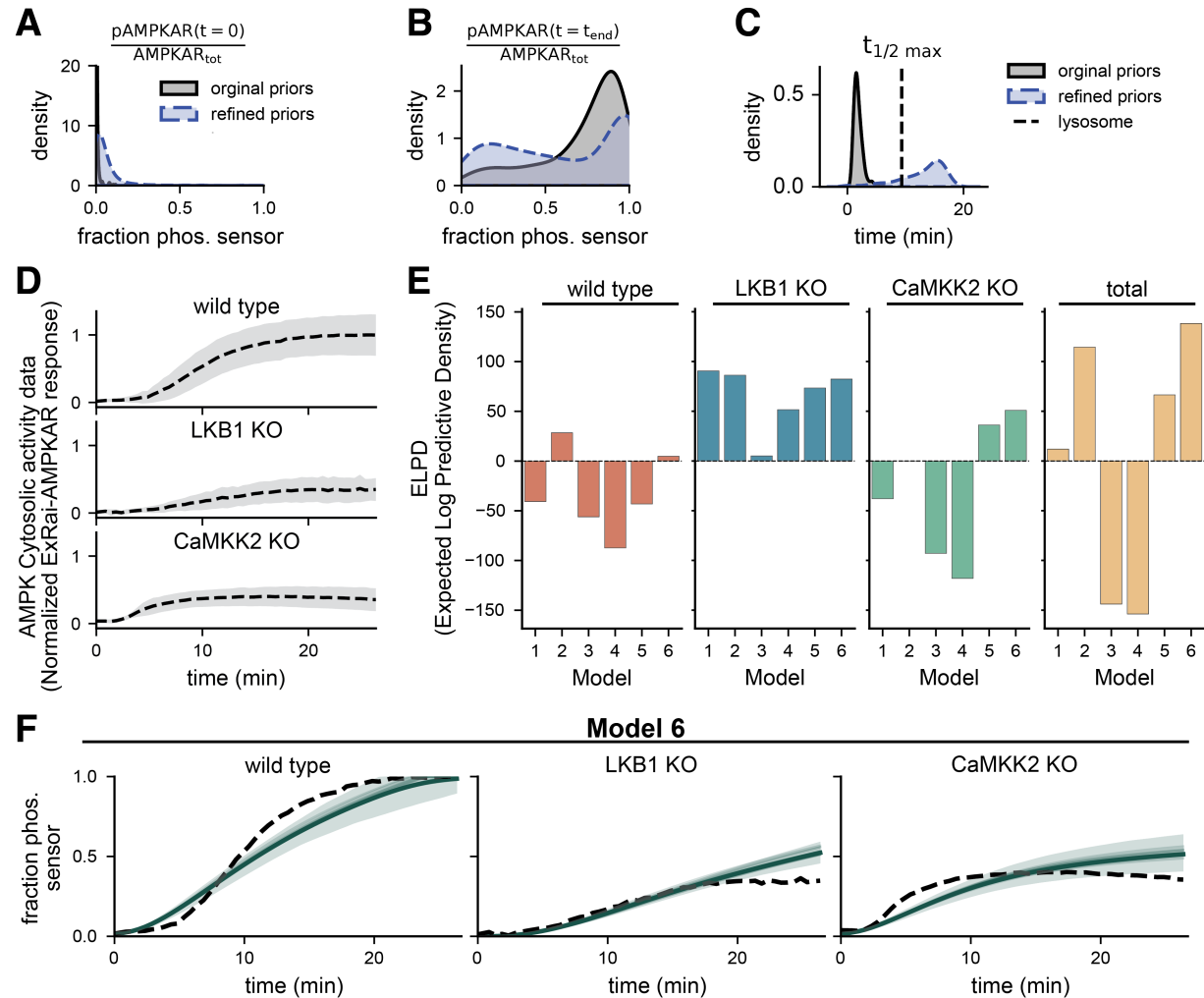


Figure 4.4: (A-C) Estimated probability densities of key quantities-of-interest of the prior predictive density of the mass action model before (black) and after (blue) refining priors with predictive prior elicitation. Priors were refined to ensure the densities placed probability mass on desired values. Supplemental Figure C.3 shows samples from the prior predictive distribution that were used to generate densities of key quantities. (D) Experimental data of AMPK activity in response to 2-deoxyglucose (2-DG) induced energy stress. Data shows the normalized ExRai-AMPKAR signal, which captures AMPK activity by varying proportionally to the fraction of biosensor phosphorylated by upstream AMPK. Experimental conditions include wild type, LKB1 knockout mutants (LKB1 KO), and CaMKK2 knockout mutants (CaMKK2 KO). Data originally from [72]. (E) Predictive performance is quantified by the expected log pointwise predictive density (ELPD) across conditions in the data. Higher ELPD values indicate better predictive performance and lower uncertainty. Models were calibrated to data from all three experimental conditions. (F) Posterior predictions generated using 1000 posterior samples from Model 6, which had the highest total ELPD. The dashed black line shows the data for each condition. The solid green line shows the posterior mean, and the shaded band shows the 95% posterior credible interval.

were loosely based on experimental evidence, we utilized prior predictive simulations to determine if the chosen parameter ranges and nominal values for nonidentifiable parameters were reasonable [309], [319] (see Section 4.2.1). We found that initial parameter ranges and the nominal values for nonidentifiable parameters did not necessarily lead to reasonable predictions (Supplemental Figure C.3). For example, for Model 1, the distribution of initial values was, as expected, mostly centered around zero (Figure 4.4A); however, although the distribution of final values was well spread between zero and one, most of the predictions reached the maximum value too quickly (Figure 4.4B and 4.4C). We hypothesized that a combination of unfeasible parameter ranges for the free parameters and incorrect nominal values for the fixed nonidentifiable parameters was responsible for the poor predictions. Therefore, we took an iterative approach to refine the parameter ranges and nominal values such that the prior predictive density of predictions was reasonable (Supplemental Section C.1.3 and Supplemental Figures C.2 and C.3). After refining the parameter ranges, we analyzed the total sensitivity of the maximum predicted AMPK sensor response and the time to half-maximum for three scenarios: (i) wild-type kinase activity, (ii) LKB1 knockout in which only CaMKK2 is active, and (iii) CaMKK2 knockout (Supplemental Tables C.7–C.12). We found that most free parameters influenced at least one of the quantities since they have a total Sobol sensitivity greater than 0.01. We fixed any parameters not meeting this threshold to their corresponding nominal values and left the remaining parameters free for parameter estimation.

After determining which parameters are identifiable and influential, we used Bayesian inference to estimate probability densities for the uncertain parameters conditioned on fluorescence microscopy AMPK activity data from Schmitt et al. [72]. We used the recordings of cytosolic AMPK activity (Figure 4.4A) for parameter estimation, and we simulated the 2-deoxyglucose (2-DG) stimulus from experiments by decreasing the rate of ATP production via glycolysis 100-fold from 0.5 1/s to 0.005 1/s, because 2-DG blocks glycolysis. Initially, we aimed to estimate model parameters from the wild-type data alone; however,

we found that this data did not inform the individual LKB1 and CaMKK2 activities, so the calibrated models could not accurately predict the kinase knockout conditions (Supplemental Figure C.4). Therefore, we used the data from all three conditions for parameter estimation.

After successfully estimating posterior densities for the parameters (Supplemental Figures C.5 and C.6), we used Bayesian model selection to determine which reaction mechanism and kinetic formulation best capture the data [77], [312], [321]. Specifically, we computed the expected log pointwise predictive density (ELPD; defined in Eq. (4.4)) of each model, which quantifies the probability of correctly predicting unseen data [312], [321]. Higher ELPD values indicate better predictive performance. We found that only Model 6 could predict all three conditions well, so the total ELPD for that model was the greatest (total ELPD=138.25; Figure 4.4E). The next best model, Model 2 (total ELPD=66.48), could not capture the CaMKK2 knockout condition, which suggests that the essential activation mechanism of LKB1-mediated phosphorylation is incorrect in that model. Furthermore, Model 5, which used mass action kinetics and had similar nonessential mechanisms to Model 6, showed poorer predictions than Model 6. Figure 4.4F shows the posterior predictions using Model 6. These results suggest that AMP and ADP act as nonessential activators of AMPK, increasing net AMPK phosphorylation and AMPK kinase activity. Furthermore, in general, the models that prescribed the Michaelis-Menten kinetics performed better than those that used mass action. Based on these results, we used Model 6 for subsequent simulations.

4.3.3 AMPK activity is sensitive to the timing of input stimuli

Next, we sought to understand how stimulus strength and timing affect AMPK activity. We previously predicted the fraction of active AMPKAR for parameter estimation; however, net AMPK activity is more physiologically relevant. To establish AMPK activity as a valid readout, we sought to verify how the AMPKAR sensor response correlates with AMPK kinase activity. For Model 6, we defined

the net AMPK activity as

$$\text{pAMPK activity} = k_{\text{phos,AMPK}} (\beta_{\text{AMPK}} [\text{AMP-pAMPK}] + [\text{ADP-pAMPK}] + [\text{ATP-pAMPK}] + [\text{pAMPK}]),$$

where the parameters $k_{\text{phos,AMPK}}$ and β_{AMPK} were sampled from the posterior distribution. First, we found that the strength of ATP hydrolysis and calcium stimuli led to proportional increases in AMPK activity and decreases in the time-to-half max (Figure 4.5A–D). For a range of energy stress levels and calcium concentrations, the net AMPK kinase activity appeared to be more sensitive to changes in stimulus strength, did not saturate over relevant stimuli, and had faster kinetics than the AMPKAR sensor. Additionally, the response strength and timing were proportional for the AMPKAR and AMPK activity responses over moderate stimuli (Supplement Figure 4.5E–F); however, the sensor response began to saturate at higher input levels, whereas AMPK activity did not.

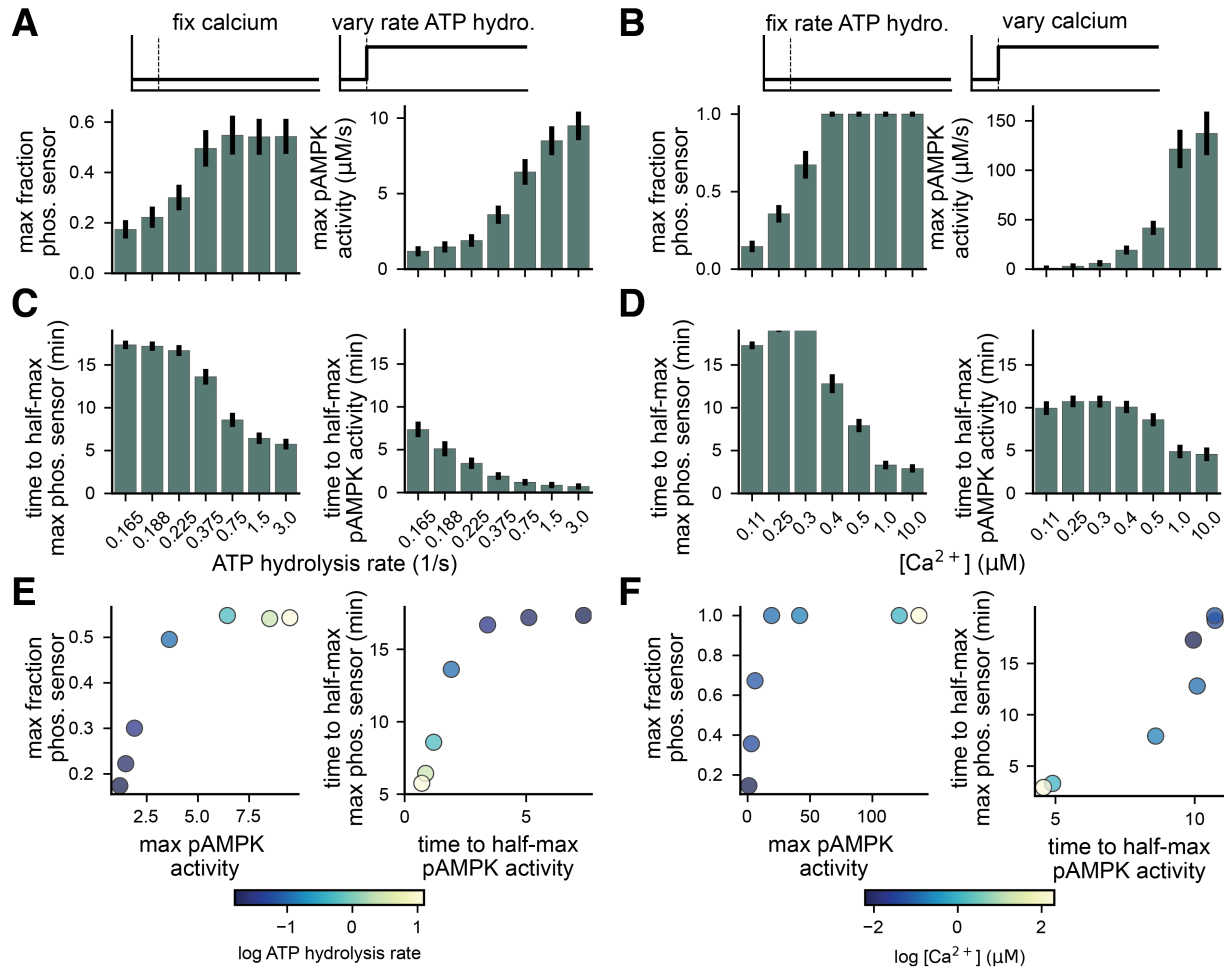


Figure 4.5: (A-B) Maximum fraction of activated AMPKAR sensor (left) and maximum pAMPK activity (right) as a function of ATP hydrolysis rate with fixed calcium (A) and as a function of calcium concentration with fixed ATP hydrolysis rate (B). (C-D) Time to half-maximum fraction of activated AMPKAR sensor (left) and maximum pAMPK activity (right) as a function of ATP hydrolysis rate with fixed calcium (C) and as a function of calcium concentration with fixed ATP hydrolysis rate (D). (E-F) Direct comparison of maximum values (left) and time to half-maximum values (right) as a function of ATP hydrolysis rate (E) and calcium concentration (F).

To understand how AMPK activity responds to more physiologically-relevant stimuli than 2-DG, we simulated transient increases in ATP hydrolysis and intracellular calcium. To do so, we applied a five-second pulse of increased intracellular calcium and ATP hydrolysis and let the system evolve for a total of 40 seconds (Figure 4.6A). We found that the area under the curve of pAMPK activity over the entire 40-second simulation was proportional to the calcium level, but not to the ATP hydrolysis level (Figure 4.6B). Interestingly, there appeared to be a threshold level of increased ATP hydrolysis between 2.0

and 5.0 mM/s that is necessary for the stimulus to begin to increase AMPK activity. We hypothesized that this discrepancy might be caused by slower dynamics in the metabolism module than in calcium-driven AMPK activation.

Next, the discrepancy between the effect of ATP hydrolysis and calcium led us to explore how the duration of a transient input affects AMPK activity. To do so, we fixed the strength of ATP hydrolysis input to 2.0 mM/s and the calcium input to 0.1 μM and varied the input duration from 0.5 to 30 seconds. We found that increasing the duration of the input led to proportional increases in AMPK activity, which were only obtained with calcium inputs at higher concentrations with a 5-second stimulus (Figure 4.6C). Additionally, we repeated these simulations, with either ATP hydrolysis or calcium fixed at baseline, to determine whether energy stress or calcium is more important during longer-duration stimuli. We found that calcium elevations appear to be responsible for most of the AMPK activity at short and moderate durations (0.5–10.0 seconds; Supplemental Figure C.8). However, at longer durations, total AMPK activity with simultaneous stimuli appeared to be greater than the sum of the individual stimuli. This effect is likely because an increase in AMP concentration due to ATP depletion increases the effectiveness of calcium-induced CaMKK2 activity (Supplemental Figure C.6) and thus increases AMPK activity.

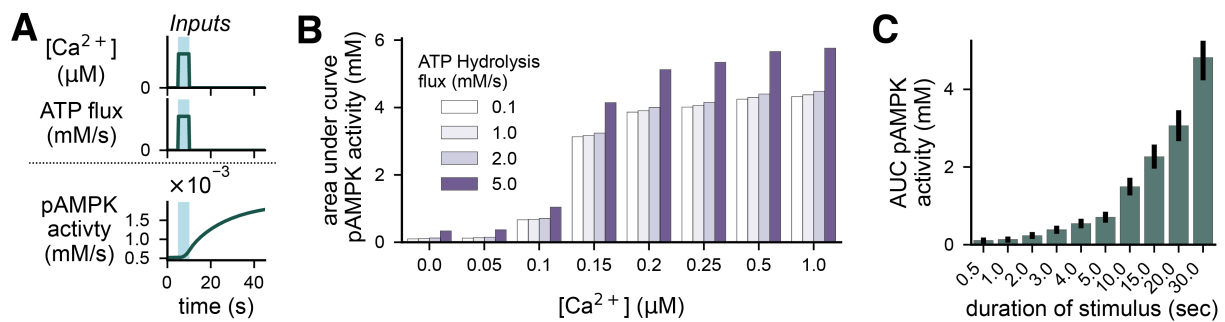
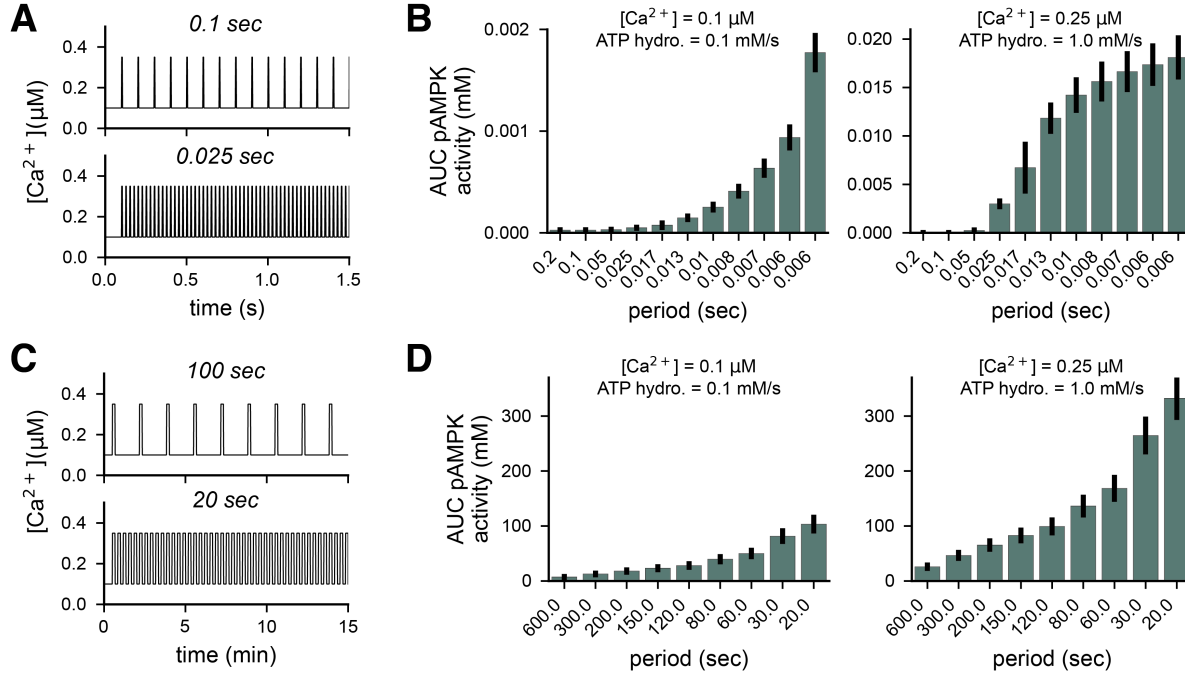


Figure 4.6: (A) Example of 10-second pulses of increased intracellular calcium and ATP flux and the corresponding trajectory of AMPK activity in the following 40 seconds. Blue shading indicates the time during which the stimulus is active. (B) Total AMPK kinase activity over 10 minutes as measured by the area under the curve of active pAMPK and AMP-pAMPK multiplied by the corresponding rate constants as a function of calcium and ATP flux stimuli. (C) Area under the curve of AMPK activity over 10 minutes as a function of stimulation duration for a fixed stimulus strength with $[\text{Ca}^{2+}] = 0.1 \mu\text{M}$ and ATP flux = 0.1 mM/s.

Finally, we investigated how AMPK responds to periodic inputs, which are inspired by physiological signals in the myocyte during exercise **calderon2014excitation**. To do so, we fixed the strength of ATP hydrolysis input to either 0.1 mM/s (low) or 1.0 mM/s (high) and the calcium input to 0.1 μ M (low) or 0.25 μ M (high) and varied the stimulus period over two timescales. On the shorter timescale, with higher frequencies, we applied a five-millisecond pulse and varied the period between the pulses from 0.006–0.2 seconds for a total of 10 seconds (Figure 4.7A). On the longer timescale, with lower frequencies, we applied a 5-second pulse for a total of 10 minutes (Figure 4.7D). For both timescales, we found that total AMPK activity measured by the area under the curve increases proportionally, but nonlinearly, with the input frequency (Figure 4.7B,D). However, at the high stimulus level, the total AMPK activity began to saturate for the short timescale inputs, but not for the long timescale inputs. These results suggest that AMPK responds proportionally to the input frequency.



4.3.4 AMPK acts as an input signal integrator during short-timescale exercise

Finally, to demonstrate the applicability of our calibrated AMPK signaling model, we used the model to predict myocyte AMPK activity during exercise. Although AMPK is activated by a combination of energy stress and intracellular Ca^{2+} , the role of AMPK in regulating metabolism in the muscle remains contested, with recent evidence suggesting that AMPK does not modulate glucose uptake and lipid metabolism during exercise [mcconellSkeletalMuscleAMPK2020](#), [mcconellItsWellTruly2020](#). However, AMPK is responsible for post-exercise metabolic adaptation [hoffmanGlobalPhosphoproteomicAnalysis2015](#), [kjobstedAMPKTBC1D1Regulate2019](#), [lantierAMPKControlsExercise2014](#), [302]. Here, we investigated how AMPK would be activated by physiologically relevant energy stresses and Ca^{2+} inputs during

simulated exercise.

Muscle activation is an inherently multiscale process that involves coordination between muscle fibers, myocytes, and motor units **uchida2021biomechanics**. At the cellular level, muscle activation is driven by rapid, high-frequency neurological inputs that induce rapid changes in intracellular Ca^{2+} to drive force generation and, in turn, rapid changes in ATP consumption **calderon2014excitation**. On a macroscopic scale, muscle contraction requires lower-frequency, slower bouts of increased myocyte activity, such as during a running stride or one repetition of a resistance exercise. Therefore, exercise induces changes in intracellular calcium that evolve over multiple timescales and can include high-frequency information of myocyte activation and lower-frequency information of muscle contraction. Different exercise conditions induce different myocyte activations at varying frequencies. For example, resistance training ranges from 10 to 50 Hz, and high-intensity interval training ranges from 60 to 180 Hz [334].

First, we simulated how AMPK responds to rapid changes in calcium and ATP hydrolysis during short-time-scale muscle activations. To do so, we simulated muscle activation during a resistance-like (Figure 4.8A) and a running-like (Figure 4.8B) stimulus using a detailed model of exercise-driven calcium signaling developed by Francis et al **Francis2025.05.22.655415**. Briefly, this multicompartmental model captures action-potential-driven calcium dynamics and includes store-operated calcium release and cross-bridge cycling. Applying the predicted calcium and ATP hydrolysis transient to our AMPK signaling model allowed us to determine how AMPK responds to physiological inputs. The resistance stimulus consisted of 10 sets of 3 seconds of 40 Hz activation, followed by 3 seconds of rest, for a total of 60 seconds. The running stimulus was 0.1625 seconds of 100 Hz activation for each 0.65-second stride, repeated for a total of 60 seconds. We predicted AMPK activity in response to these exercise stimuli by simulating our model using the predicted calcium and ATP hydrolysis rates as inputs.

Interestingly, the entire AMPK signaling pathway acted as a low-pass filter on the input signals.

Upstream AMP, ATP, and active CaMKK2 only responded to slower variations in calcium and ATP hydrolysis induced by repetitions of contraction (resistance) or stride (running; Figure 4.8B). This aligns with previous findings that ligand-activated kinases, such as CaMKK2, can act as integrators and, thus, low-pass filter input signals [26]. In response, the predicted AMPK activity was nearly identical for the two exercise conditions (Figure 4.8C). Interestingly, the predicted AMPK activity trajectory only showed minor oscillations during the resistance condition. Quantitatively, the maximum AMPK activity was not significantly different in both conditions (WT; Figure 4.8C). Additionally, we found that exercise-induced AMPK activity depended strongly on CaMKK2, with CaMKK2 knockout abolishing nearly all AMPK activity (CaMKK2 KO; Figure 4.8C). These results suggest that they respond to exercise-related inputs in a manner that is robust to the high-frequency information in the signal.

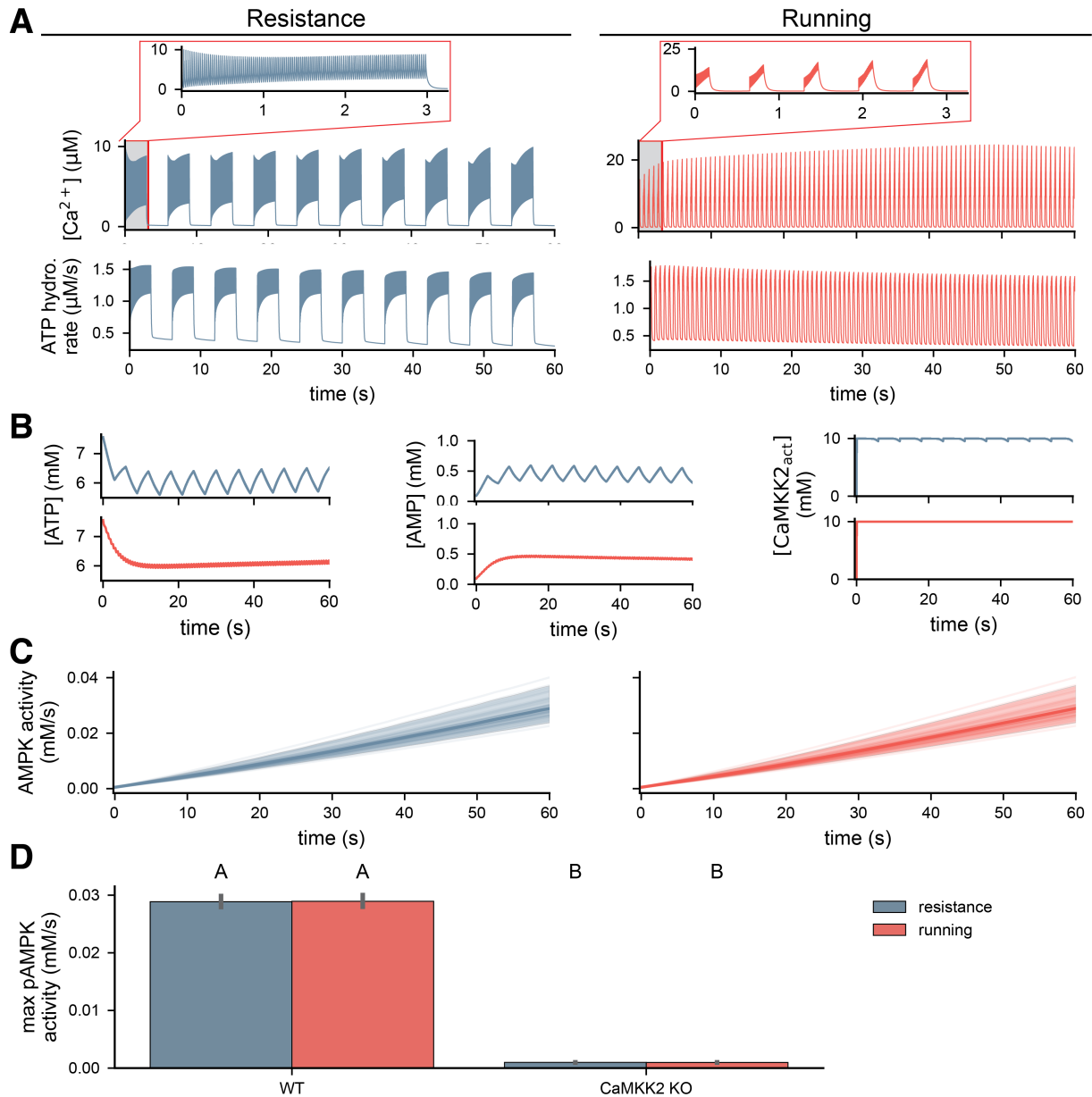


Figure 4.8: (A) Trajectories of $[Ca]^{2+}$ and ATP hydrolysis rate during simulated resistance (left) and running (right) like stimuli. Simulations from myocyte exercise model from [Francis2025.05.22.655415](#). The inset shows the first three seconds of the calcium. (B) Predicted ATP, AMP and activated CaMKK2 concentrations for each exercise type. (C) Posterior densities of predicted pAMPK activity during exercise stimuli. (D) Maximum pAMPK activity during exercise for wild-type and CaMKK2 KO conditions. Letters indicate statistical significance; groups with different letters are significantly different, and groups with like letters are significantly different. Significance was evaluated with a pairwise t-test and assessed at the level of 0.05.

4.3.5 AMPK activity is exercise-type dependent during long-timescale exercise

Finally, we investigated how AMPK responds to exercise-related stimuli at longer timescales. Sprint, endurance, and resistance exercise on the minute-to-hours scale have been shown to induce varied levels of AMPK activation both in vivo [302] and computationally **leung2023computational**. In these studies, endurance training induced greater AMPK activity than resistance training, which was greater than that induced by sprinting. Here, we aimed to investigate AMPK activity in response to similar stimuli using our model. To do so, we simulated the three exercise conditions (Figure 4.9A), which mimic the timing of those prescribed in [302] and have levels informed by the average values of the stimuli in Figure 4.8A. The sprint condition was three 30-second intervals with a Ca^{2+} concentration of 10 μM and a rate of ATP hydrolysis of 3 1/s. The resistance condition was six 1-minute intervals with a Ca^{2+} concentration of 3 μM and a rate of ATP hydrolysis of 1 1/s. The endurance condition was one 90-minute interval with a Ca^{2+} concentration of 0.5 μM and a rate of ATP hydrolysis of 0.25 1/s.

Interestingly, these longer-timescale exercise stimuli induced stimulus-dependent AMPK activity. All three stimuli were slow enough to induce oscillatory (sprint and resistance) or step-wise (endurance) responses in AMP, ATP, and active CaMKK2 (Figure 4.9A). The subsequent trajectories of AMPK activity appeared to vary as a function of stimulus duration. Sprint induced a short and weaker response, resistance induced a stronger response, and endurance induced a moderate response. In all three trajectories, AMPK activity decreased to baseline after the stimulus duration. Resistance led to the highest maximum AMPK activity (WT; Figure 4.9A). Interestingly, endurance had higher maximum activity than sprint, even though the strength of the sprint stimulus was more than 10 times larger than the endurance stimulus. We additionally observed that AMPK activity depended strongly on CaMKK2 (CaMKK2 KO; Figure 4.9A). This result further suggests that AMPK integrates upstream signals and responds proportionally to the duration of the stimulus.

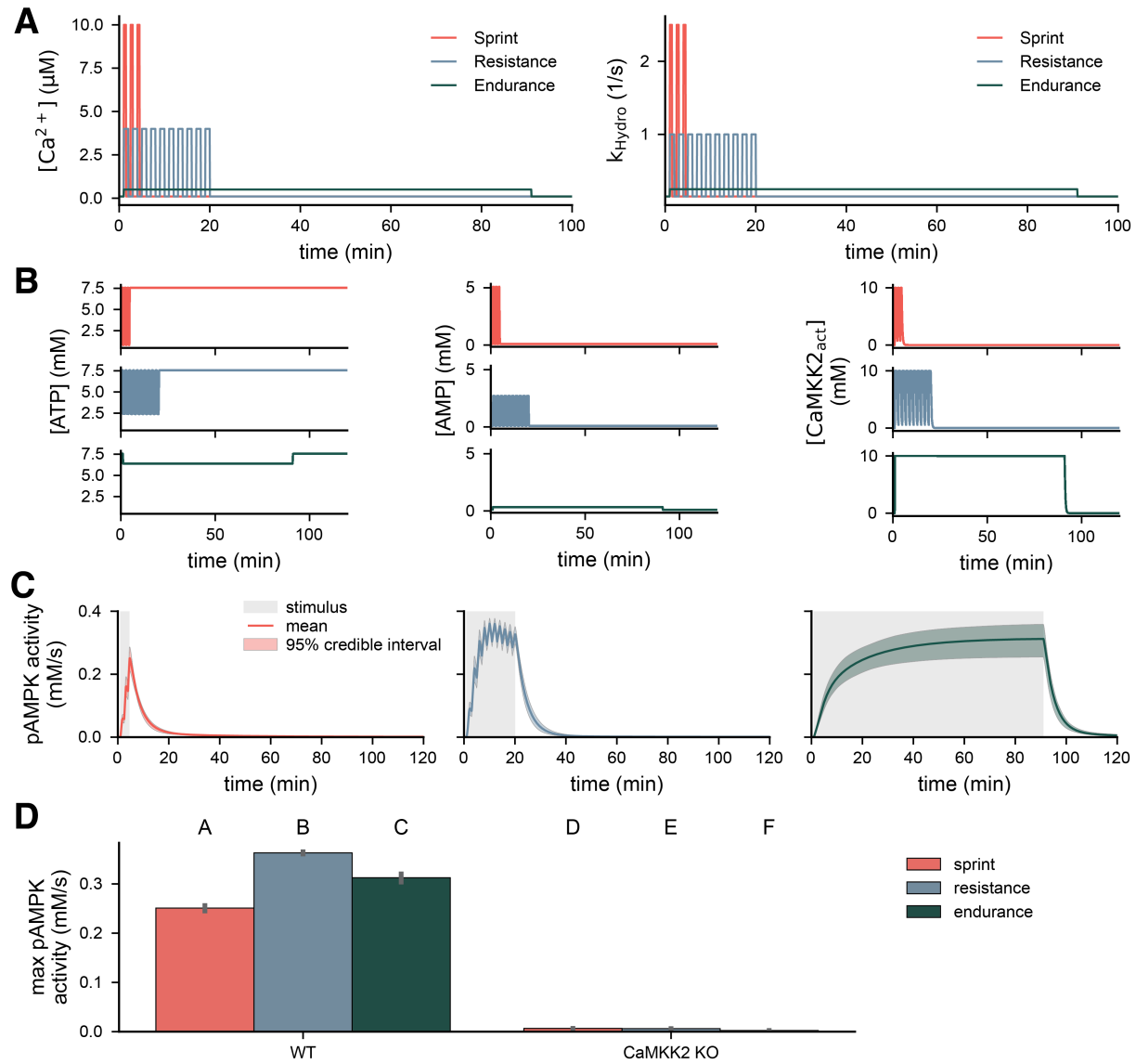


Figure 4.9: (A) Long-time scale exercise stimuli during sprint (3 x 30s), resistance (6 x 1min) and endurance (1 x 90min). (B) Predicted ATP, AMP, and activated CaMKK2 concentrations for each exercise type. (C) Posterior densities of predicted pAMPK activity during exercise stimuli. (D) Maximum pAMPK activity during exercise for wild-type and CaMKK2 KO conditions. Letters indicate statistical significance; groups with different letters are significantly different, and groups with like letters are significantly different. Significance was evaluated with a pairwise t-test and assessed at the level of 0.05.

Discussion

In this work, we developed a new mechanistic model of AMPK activation during energy stress.

We took a data-informed approach to model development and applied Bayesian uncertainty quantification

(Figure 4.2) to leverage recent high-quality experimental AMPK activity data [72]. First, we found that the choice of kinetic formulation used in a model impacted parameter identifiability and that there is no clear *best* formulation from that viewpoint. Next, we developed six related AMPK signaling models and found that the model using Michaelis-Menten kinetics and assuming that AMP and ADP act through nonessential mechanisms to modulate AMPK activity was most consistent with the data. Using this *best fit* model, we found that AMPK was sensitive to the timing of input stimuli, especially for energy-stress inducing inputs. In the context of exercise, we found that AMPK was robust to high-frequency variations in the input signal over short timescales but showed exercise dependence over longer timescales.

Uncertainty quantification, including identifiability analysis, prior elicitation, Bayesian parameter estimation, and Bayesian model selection, played a key role in developing the new model of AMPK signaling. In particular, UQ methods allowed us to utilize available experimental data to select a model and constrain its predictions. While previous applications of UQ in systems biology have enabled parameter estimation [308], [335], [336] and model selection [77], [316], [337], this work is unique in that we apply an end-to-end UQ workflow during the entire model development process. A key advancement here is the use of predictive prior elicitation and prior predictive sampling to iteratively refine the prior densities and values of the nonidentifiable parameters (Supplemental Section C.1.3). This additional step addressed two challenges of our previously developed Bayesian parameter estimation workflow. First, predictive prior elicitation provided a structured approach to construct semi-informative priors when very little information is known about model parameters *a priori*. Second, iterative refinement of fixed nominal values allowed us to update nominal values for parameters that are mathematically nonidentifiable but still influence model predictions. While the UQ workflow presented here provided an end-to-end approach for data-informed model development, we did not *close the loop* and use the findings to update the model structures. To that end, all of the predictions were only as good as the set of input models. Methods for data-driven modeling,

such as model discovery [338], hybrid mechanistic neural network models [339], and continuous model exploration with jump MCMC **andrieu,JointBayesianModel1999**, can enable model development directly from data.

The mechanistic AMPK models capture AMPK activation by upstream kinases (LKB1 and CaMKK2) and activity control by adenine nucleotides. We chose to limit the model's scope to the core of the AMPK signaling pathway to investigate the mechanisms by which adenine nucleotides control AMPK. However, as a hub of metabolic signaling, AMPK acts on several targets that can induce feedback to modulate AMPK activity, especially during longer timescale signaling events. For example, feedback through ULK1 and mTOR signaling could lead to net changes in AMPK activation **garza-lomboMTORAMPKSignaling2018**, [51]. Furthermore, through its action on STIM1, AMPK could modulate Ca^{2+} signaling, which could, in turn, affect calcium-driven AMPK activation **nelsonPhosphoproteomicsRevealsConserved2019**. Future modeling efforts should focus on embedding mechanistic AMPK signaling models into the larger metabolic signaling system, which could be achieved, for example, by replacing the phenomenological representation with our mechanistic model [51], [287], [288]. Additionally, the role of AMPK activation in modulating energy uptake during exercise remains debated **mcconellItsWellTruly2020**, so it is important to consider our predictions of exercise-induced AMPK activity fit into the broader myocyte metabolic signaling network. To better understand this, further modeling efforts should incorporate the relevant mechanisms and data necessary to investigate the roles of AMPK during exercise.

As experimental capabilities continue to improve, there is an ever-growing need to develop well-constrained models to integrate and interpret available data. Here, we showed how uncertainty quantification enables data-informed model development for signaling pathways where few previously developed models had been published. We believe that this work serves as a starting point for future

modeling and development of new UQ methods in systems biology. As we showed here, the application of UQ in systems biology has the potential to improve modeling and better constrain model-based predictions to experimental data and drive new discoveries in intracellular signaling.

Code Availability

We provide all the code and instructions to reproduce these results in the following GitHub repository <https://github.com/natejlinden/AMPK>.

Acknowledgments

Chapter 4, in full, is a modified reprint of the material as it appears in Systems modeling and uncertainty quantification of AMP-activated protein kinase signaling. Nathaniel Linden-Santangeli, Jin Zhang, Boris Krämer, and Padmini Rangamani, *Submitted*, 2025.

Chapter 5

Miscellanea: Quantifying heterogeneity in predictions of beta-cell insulin sensitivity

5.1 Introduction

Early diagnosis and intervention is essential to limiting Type 2 diabetes progression. Type 2 diabetes (T2D) is a growing health concern in the United States that has affected nearly 37.3 million Americans in 2022 [340]; this figure is expected to grow in future years [341]. T2D is characterized by an inability to maintain homeostatic levels of plasma glucose and often begins with changes in pancreatic β -cells that can eventually lead to complications across the body [342], [343]. The hallmarks T2D onset often include impaired fasting glucose (IFG) and impaired glucose tolerance (IGT), both of which are signs of the inability to control plasma glucose with β -cell insulin secretion [341], [344]. A growing body of evidence suggests that early diagnosis and intervention is crucial for limiting the progression of T2D [345], [346] and reducing the risk of diabetes-related complications, including nephropathy, retinopathy, and cardiovascular disease [343], [347], [348]. Despite the importance of early action, 20–50%

of patients with diabetes remain undiagnosed [340], [341], and a significant proportion of those that receive a diagnosis present with substantial disease progression at diagnosis [349]. While increased screening will alleviate some missed diagnoses, predicting disease progression and treatment effectiveness from available clinical data is equally important to improving diabetes care. For instance, many clinical trials and screening protocols rely on glucose challenge tests such as the oral glucose tolerance test (OGTT) and the intravenous glucose tolerance test (IVGTT) to probe pancreatic function and predict diabetes progression. *Here, we address the increasing incidence of T2D and the need for improved disease monitoring tools by substantially improving the quality and trustworthiness of predictions of metabolic function from clinical tests.*

Predictive mathematical modeling is a promising approach for inferring pancreatic function from clinical tests. Mathematical and data-driven models can predict metrics of pancreatic function from clinical data that are not directly evident from the data [350]–[352]. These predictions are vitally important for making clinical decisions, assessing impacts of novel therapeutics on pancreatic function [353], and on classifying diabetes subtypes [354]. A classical example of mathematical modeling in diabetes is the glucose minimal model which has been successfully applied to estimate insulin resistance from IVGTT data since its introduction in 1979 [353], [355]–[357]. Despite the successes of these models, uncertainties due to *ad hoc* modeling assumptions, inpatient and outpatient variability, and limited access to clinical measurements, limit the level of confidence that can be placed on predictions [353]. Specifically, it is unknown whether clinicians should rely on predictions from a single occurrence of a clinical test or if multiple repeated tests are necessary to ensure predictions accurately reflect a patient’s true pancreatic function. Outside of diabetes care, when diagnosing male hypogonadism, known variability in a patient’s testosterone levels often necessitates consecutive measurements of low testosterone to make a diagnosis confidently. In diabetes care, we do not have the same understanding of how variability in predictions

affects the quality of a single prediction.

Uncertainties are introduced into predictions of pancreatic function due to modeling assumptions, variations in clinical tests, and inpatient variabilities in clinical measurements. While we can derive detailed mechanistic models of diabetes physiology, such as β -cell insulin secretion [358], due to their complexity and multi-scale nature, these models do not adequately scale to provide predictions of diabetes status at the patient-level in clinical settings [350], [352]. Thus, we must often rely on phenomenological models that assume a mathematical formulation without adhering to physiological reality. In addition to model uncertainty, significant variations in clinical data have been observed between patient demographics [359] and between test occurrences for the same patient [360], [361]. As stressed by an American diabetes Association consensus panel [351], [362], [363], it is of utmost importance to account for all sources of error and uncertainty when deploying models in clinical settings. Therefore, we hypothesize that these sources of uncertainty—*ad hoc* modeling assumptions, variability in clinical data and non-uniformity in clinical protocols—introduce unknown and unaccounted errors in model-based predictions. *We will provide significant advancements in predictions of pancreatic function in diabetes care by developing new methods to quantify the uncertainties affecting these predictions and by providing a new recommendation as to level of certainty to place in these predictions.*

In this work, we conduct a series of uncertainty quantification studies that applied predictive models to clinical diabetes-related data. Specifically, investigated uncertainties that impact model-based predictions of beta-cell sensitivity to glucose using oral glucose tolerance data. Using Bayesian MMI, we found that different models are better suited for healthy, prediabetic, and diabetic subjects. We additionally found that data uncertainty due to sparse temporal hinders our ability to differentiate between healthy and diabetic subjects.

5.2 Methods

5.3 Oral glucose tolerance test: C-peptide kinetics

5.3.1 The C-peptide model family

In this section we present a family of phenomenological models that estimate of a patient's pancreatic health from the oral glucose tolerance test (OGTT). These models predict glucose-stimulated C-peptide release during the OGTT from which indices of β -cell function can be derived. The oral glucose tolerance test is widely used to aid the clinical diagnosis of glucose metabolic disorders, such as type-II Diabetes Mellitus (T2D) **gavin1997report**, **american2003report**. During an OGTT patient's blood glucose, insulin, and often C-peptide levels are monitored following standardized oral load of 75g glucose [357]. Patients with impaired glucose regulation will show high blood glucose levels during the OGTT **gavin1997report**, **american2003report**.

In addition to determining a patient's glucose tolerance status, mathematical modeling in conjunction with the OGTT can help determine a patient's β -cell function [356], [357]. Here we study three phenomenological models of glucose-stimulated C-peptide release during the OGTT: a Hovorka et al. [364], Cretti et al. [365], Breda et al. [366], which we refer to as the Hovorka 1998, Cretti 2001, and Breda 2001 models, respectively. All three models share the two-compartment structure

$$\frac{dx_1(t)}{dt} = -(k_{01} + k_{21})x_1(t) + k_{12}x_2(t) + u(g(t), t) \quad (5.1a)$$

$$\frac{dx_2(t)}{dt} = k_{21}x_1(t) - k_{12}x_2(t), \quad (5.1b)$$

where $x_1(t)$ is the concentration of C-peptide (nmol/L) in the accessible compartment (or that concentration above baseline) and $x_2(t)$ is the concentration of C-peptide (nmol/L) in an inaccessible peripheral

compartment. The parameter, k_{01} is the loss rate (1/min) of C-peptide from the accessible compartment, and k_{12} and k_{21} are the transport rates (1/min) between the two compartments. The function $u(g(t), t)$ captures the time-varying glucose-stimulated C-peptide secretion rate by the pancreatic β -cells, and is a function of both time and the measured plasma glucose concentration (mmol/L).

The three models differ in their formulation of u . The Hovorka 1998 model assumes that C-peptide secretion varies linearly with $g(t)$ [364], proposing that

$$u(g(t), t) = \begin{cases} M_I(g(t) - g_b) + M_0 g_b & \text{if } M_I(g(t) - g_b) + M_0 g_b > 0 \\ 0 & \text{else.} \end{cases}$$

This formulation introduces three parameters: the baseline plasma glucose concentration $g_b = g(t = 0)$, the post-stimulus sensitivity M_I (1/min), and the basal sensitivity M_0 (1/min). We note that the Hovorka 1998 model was originally proposed for the meal tolerance test, however we adapt it here for the OGTT as these tests are similar. The Cretti 2001 model assumes that the C-peptide secretion rate is the sum of a basal secretion rate B-ISR and a dynamic, glucose-dependent, secretion rate OG-SR(t), such that

$$u(g(t), t) = \text{B-ISR} + \text{OG-SR}(g(t), t).$$

The basal secretion rate B-ISR = [basal C-peptide] k_{01} (pmol/L/min), where [basal C-peptide] is the C-peptide concentration at baseline. The dynamic secretion rate OG-SR($g(t), t$) = $\tau^{-1}X(t)$ (pmol/L/min), where the additional variable $X(t)$ is governed by

$$\frac{dX(t)}{dt} = -\tau^{-1}X(t) + \sigma(g(t) - \theta), \quad X(0) = 0.$$

Here, the constant σ (pmol/L/min per mmol/L), the apparent glucose threshold θ (mmol/L), and the time constant τ , control C-peptide release by pancreatic β -cells. Finally, the Breda 2001 model similarly partitions the C-peptide secretion rate into a static component $SR_s(t)$ and a dynamic component $SR_d(t)$, such that

$$u(g(t), t) = SR_s(t) + SR_d(t). \quad (5.2)$$

The static component $SR_s(t) = X(t)$ (pmol/L/min), where the additional variable $X(t)$ is governed by

$$\frac{dX(t)}{dt} = \alpha [X(t) - \beta (g(t) - h)], \quad X(0) = 0.$$

The dynamic component $SR_d(t)$ is taken to vary with the derivative of the plasma glucose $\dot{g}(t)$ following

$$SR_d(t) = \begin{cases} k(g(t))\dot{g}(t) & \text{if } \dot{g}(t) > 0 \\ 0 & \text{if } \dot{g}(t) \leq 0, \end{cases}$$

where

$$k(g(t)) = \begin{cases} K_d \left(1 - \frac{g(t) - g_b}{g_t - g_b} \right) & \text{if } g_b \leq g(t) < g_t \\ 0 & \text{else.} \end{cases}$$

Here, the constant α (min), the static sensitivity index β (1/min), the static secretion threshold h (mmol/L), the constant K_d (UNITS), the baseline glucose g_b (mmol/L), and dynamic threshold g_t (mmol/L) control the secretion rate.

In the original papers, each model was associated with a model-specific β -cell sensitivity, such as the post-stimulus sensitivity M_I in the Hovorka 1998 model [364], the global β -index in the Cretti 2001 model [365], and the static ϕ_s and the dynamic ϕ_d sensitivities in the Breda 2001 model (see respective

papers for definitions and derivations). However, to enable multimodel inference of a β -cell sensitivity index, we derive β -cell sensitivity to glucose index ϕ for each model following the original derivation for the Breda 2001 model global index. The β -cell sensitivity index ϕ is defined as the ratio of the total C-peptide secretion to the total glucose above threshold

$$\phi = \frac{\int_0^\infty u(t) dt}{\int_0^\infty [g(t) - \theta_{\text{thresh}}] dt}, \quad (5.3)$$

where θ_{thresh} is the threshold parameter for each model, which is g_b , θ , and h for the Hovorka 1998, Cretti 2001, and Breda 2001 models, respectively.

Following the derivations of [366] we can derive the sensitivity index for the Breda 2001 model

$\phi_{\text{Breda2001}}$ as

$$\phi_{\text{Breda2001}} = \frac{\int_0^\infty \text{SR}_s(t) dt}{\int_0^{\text{infy}} [g(t) - h] dt} + \frac{\int_0^\infty \text{SR}_d(t) dt}{\int_0^\infty [g(t) - h] dt}.$$

First, we evaluate $\int_0^\infty \text{SR}_s(t) dt$, by first taking the integral

$$\begin{aligned} \int_0^\infty \dot{X}(t) dt &= \int_0^\infty dX(t) = X(\infty) - X(0) \\ &= -\alpha \int_0^\infty X(t) dt + \alpha\beta \int_0^\infty [g(t) - h] dt. \end{aligned}$$

Noting that $X(0) = X(\infty)$ because we assume that the system returns to basal values as time after the OGTT goes to infinity, we find that

$$\alpha \int_0^\infty X(t) dt = \alpha\beta \int_0^\infty [g(t) - h] dt \implies \int_0^\infty X(t) dt = \beta \int_0^\infty [g(t) - h] dt,$$

and therefore

$$\int_0^\infty \text{SR}_s(t) dt = \beta \int_0^\infty [g(t) - h] dt.$$

Next, we evaluate $\int_0^\infty \text{SR}_d(t) dt$ directly, such that

$$\int_0^\infty \text{SR}_d(t) dt = \int_{G_b}^{G_{\max}} k(g(t)) dG,$$

where $G_{\max}$ is the maximum glucose concentration measured during the OGTT, because $k(G(t))$ is non-zero for $G(t) > G_b$ and only defined up to $G_{\max}$. Thus, we find that

$$\int_{G_b}^{G_{\max}} k(g(t)) dG = \begin{cases} K_d(G_{\max} - G_b) \left[1 - \frac{(G_{\max} - G_b)}{2(G_t - G_b)} \right] & \text{if } G_t > G_{\max} \\ \frac{K_d(G_t - G_b)}{2} & \text{if } G_t \leq G_{\max}, \end{cases}$$

omitting the necessary algebra for brevity. Therefore, we find that

$$\phi_{\text{Breda2001}} = \beta + \frac{1}{\int_0^\infty [g(t) - h] dt} \begin{cases} K_d(G_{\max} - G_b) \left[1 - \frac{(G_{\max} - G_b)}{2(G_t - G_b)} \right] & \text{if } G_t > G_{\max} \\ \frac{K_d(G_t - G_b)}{2} & \text{if } G_t \leq G_{\max}. \end{cases} \quad (5.4)$$

We apply the same formula to find ϕ for the Hovorka 1998 model, and find that

$$\phi_{\text{Hovorka1998}} = \frac{\int_0^\infty M_I [g(t) - g_b] dt}{\int_0^\infty [g(t) - g_b] dt} = M_I. \quad (5.5)$$

Noting that we dropped the $M_0 g_b$ term in $u(t)$ for Hovorka 1998 from the integral, because ϕ characterizes C-peptide secretion above baseline, and that term models baseline secretion. Lastly, for the Cretti 2001

model, we similarly apply the derivation from the Breda 2001 model, taking

$$\phi_{\text{Cretti2001}} = \frac{\int_0^\infty \text{OG-SR}(t) dt}{\int_0^\infty [g(t) - \theta] dt}.$$

Once again assuming that the system return to the basal values after the OGTT, we can find $\int_0^\infty \text{OG-SR}(t) dt$ by letting

$$\begin{aligned} \int_0^\infty \dot{X}(t) dt &= \int_0^\infty dX(t) = X(\infty) - X(0) = 0 \\ &= -\tau^{-1} \int_0^\infty X(t) dt + \sigma \int_0^\infty [g(t) - \theta] dt \\ \Rightarrow \int_0^\infty X(t) dt &= \sigma \tau \int_0^\infty [g(t) - \theta] dt. \end{aligned}$$

Thus,

$$\int_0^\infty \text{OG-SR}(t) dt = \tau^{-1} \int_0^\infty X(t) dt = \sigma \int_0^\infty [g(t) - \theta] dt,$$

by definition, then

$$\phi_{\text{Cretti2001}} = \frac{\sigma \int_0^\infty [g(t) - \theta] dt}{\int_0^\infty [g(t) - \theta] dt} = \sigma. \quad (5.6)$$

We once again drop the basal secretion term in $u(t)$, because the sensitivity index is only concerned with secretion above baseline.

5.4 Results

5.4.1 Model uncertainty impact predictions of β -cell glucose sensitivity from OGTT data

As a preliminary study with the OGTT, we fit the three C-peptide models to population-averaged OGTT data presented in figure 2 of Hovorka et al. [364]. The data, shown in Figure 5.1A–B, contain average OGTT measurements over a population of 16 patients with NGT and 16 patients with T2D. We initially hypothesized that multimodel predictions of β -cell sensitivity would better discriminate between patients with T2D and patients with NGT.

We fit each C-peptide model described in Section 5.3.1 to the data from each patient population using MCMC. After verifying that the MCMC sampling converged, we ran an ensemble of simulations for each model. Figure 5.1C–H, shows the mean and 95% credible intervals for each model and populations. Notably, all models tend to capture equally well the trends in the two sets of data, with some notable exceptions such as the Breda 2001 model performing poorly on the NGT data in Figure 5.1E. Next, we constructed approximated probability densities for β -cell sensitivity index ϕ by fitting kernel density estimators to each model's posterior samples. Figure 5.1I shows the resulting densities for all models and both patient populations. For the NGT population, all models predict similar densities for ϕ , with the Breda 2001 model shifted to larger values than the other models. However, for the T2D population, the predicted densities show greater variation with the Cretti 2001 model placing almost all probability on values less than 10, while the other two models spread probability upwards to 20.

We then used two methods for multimodel inference (see Chapter 3) to combine the predicted densities from all three c-peptide models. Figure 5.1J shows the combined densities using each method. For the NGT population, both multimodel inference methods yield similar predicted densities, because the underlying densities from each model were similar. For the T2D population, we find that model fusion

biases the combined prediction towards the predicted density from Cretti 2001, as that density has smallest variance, while pseudo Bayesian model averaging predicts a combined density that places more probability towards larger values as predicted by the Hovorka 1998 and Breda 2001 models.

Next, we asked if we could similarly discriminate between NGT, IGT and T2D populations in the Hall et al. [367] dataset that only includes a 2-hr OGTT with three timepoints. Figure 5.2A–B shows the averaged OGTT data for each population with the standard deviation shown as shaded regions. We hypothesized that the smaller number of timepoints would reduce our ability to discriminate between populations using individual models, but multimodel inference would bring an improvement.

To test this, we again fit the three c-peptide models to each population’s data using MCMC. After ensuring MCMC convergence, we ran ensemble simulations for each model. Figure 5.2C–E shows the posterior predictions for each model and population, with the lines showing the median and the shaded regions showing the 95% credible intervals. We observe that all three models provide reasonable fits to the data, with minor variations in the predicted C-peptide dynamics and the level of certainty.

5.4.2 OGTT sampling strategy affects predictions

Clinical data in diabetes, such as OGTT and IVGTT measurements, are subject to substantial variability due to heterogeneity between patients, external factors such as exercise and diet, environmental factors, and differences in laboratory procedures [359], [361], [368]. We previously showed that data quality can impact predictive uncertainty [208]. Specifically, we found that increasing measurement error and uncertainty due to the process of data collection reduces predictive certainty. For example, we found that increasing the noise added to measurements drastically increased the uncertainty in predictions made with that noisy data. Furthermore, we found that our ability to make meaningful predictions depended strongly on the quality and information content of the data. In the context of predictive modeling in diabetes, these

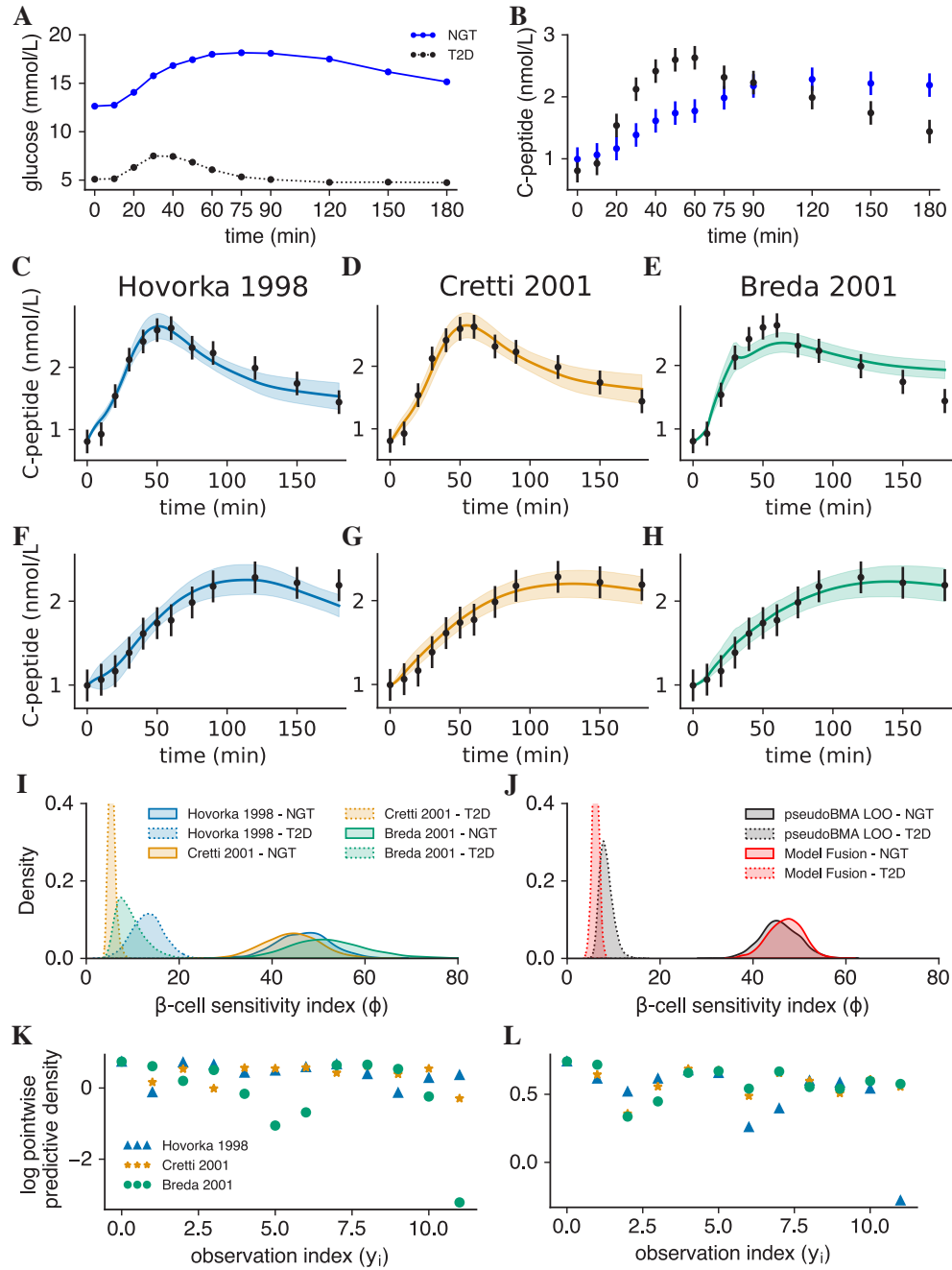


Figure 5.1: (A–B) Population averaged OGTT data from [364]. (C,F) Posterior predictions from the Hovorka 1998 model calibrated to NGT and T2D data, respectively. (D,G) Posterior predictions from the Cretti 2001 model calibrated to NGT and T2D data, respectively. (E,H) Posterior predictions from the Breda 2001 model calibrated to NGT and T2D data, respectively. (I) Approximated probability densities for β -cell sensitivity index ϕ from each model for the NGT (solid line) and T2D (dashed line) populations. (J) Combined probability densities for β -cell sensitivity index ϕ using pseudo Bayesian model averaging (pseudoBMA; black) and model fusion (red). (K–L) Log pointwise predictive densities evaluated at the i th observation y_i show the quality of each model's pointwise c-peptide predictions over the course of the OGTT. Values are averaged over posterior samples.

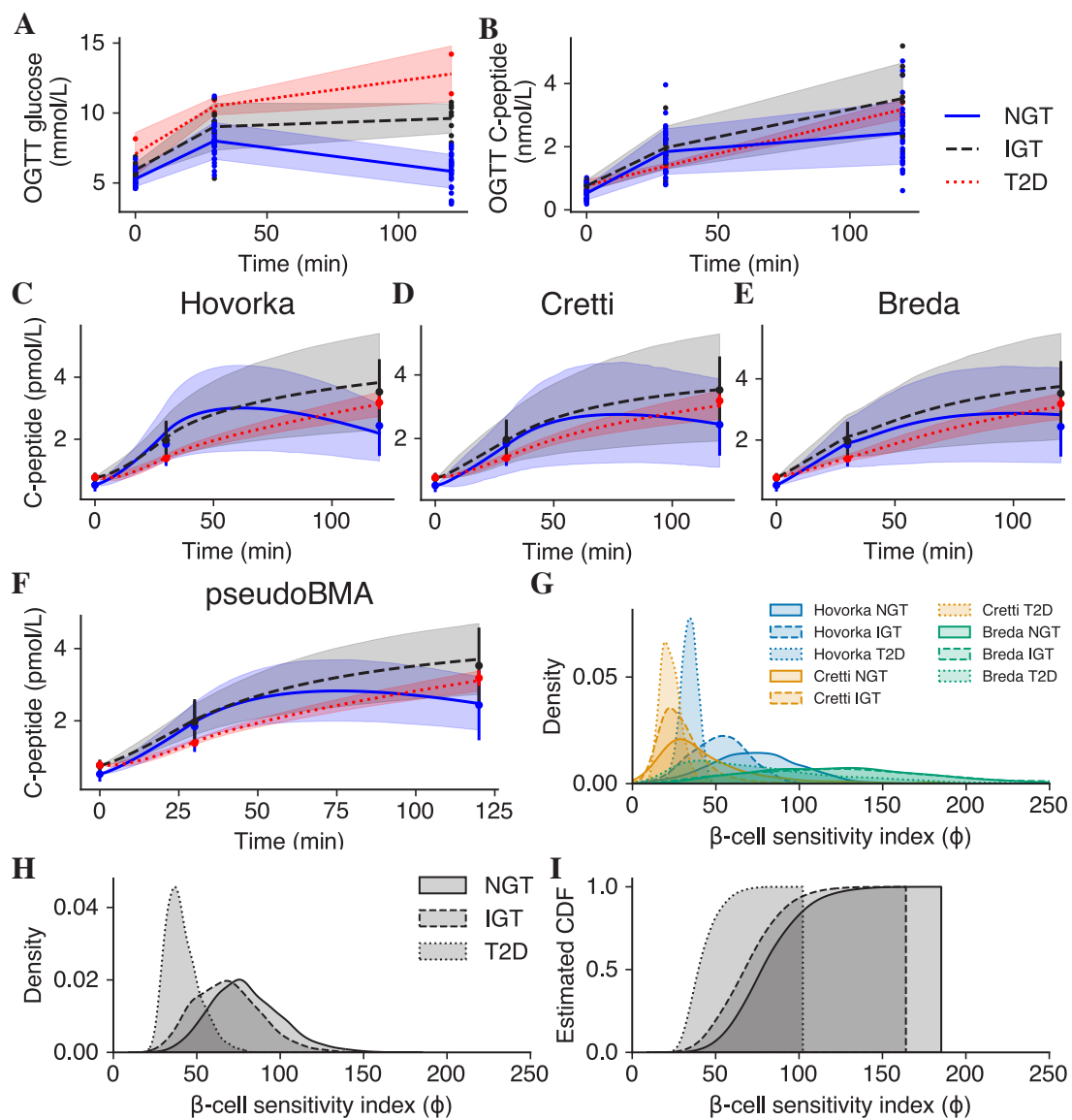


Figure 5.2: (A–B) Population averaged OGTT data from [367].

findings highlight the potential for uncertainties in clinical data to cause substantial predictive uncertainty and error. Clinical protocols for the OGTT often differ in the duration and frequency of sampling. For

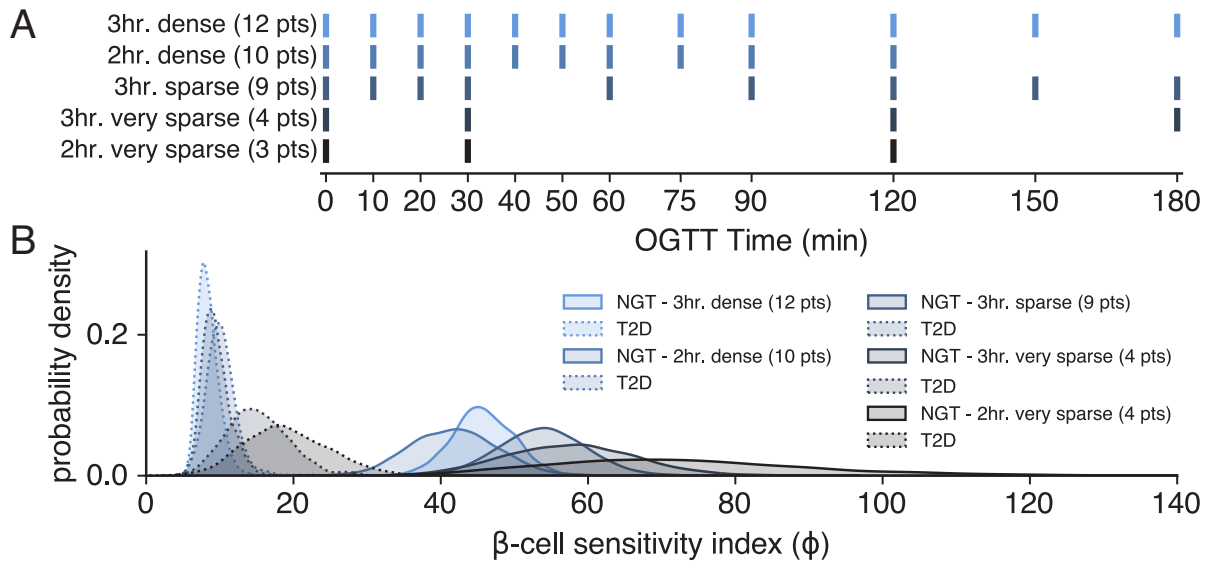


Figure 5.3: (A) Depiction of the five simulated OGTT protocols where we increased sparsity and reduced the total number of samples. (B) Predicted β -cell sensitivity probability densities highlight how protocols with fewer time points lead to greater uncertainty.

example, the common 2-hour OGTT protocol used in the diabetes Prevention Program [345] only collects two or three samples, whereas longer duration protocols, such as the 3-hour OGTT in the RISE studies utilize larger sample counts [369]. To test the effects of OGTT duration and sampling, we simulated five OGTT protocols [364], [367], [369] by resampling the population averaged data in Hovorka et al [364]. Figure 5.3A, shows the different sampling schemes and durations, with each marker indicating the time point when a sample taken. We then applied our previously developed method for Bayesian multimodel inference to predict β -cell sensitivities for NGT and T2D populations. Figure 5.3B, shows the predicted probability densities for each population using the different datasets. Notably, we find that decreasing the OGTT duration and/or the number of samples leads to more spread in the predicted densities, indicating decreased certainty in the predictions. Furthermore, the sparsely and very sparsely sampled datasets tend to introduce systematic errors in the predictions, as indicated by the overall shift of the predicted densities

for these cases. These results support our hypothesis that data variability effects model-based predictions by demonstrating that the information content of the data has significant impacts on predictions.

5.5 Discussion

Our findings demonstrate that model and data uncertainty significantly impact the reliability of β -cell glucose sensitivity predictions from OGTT data, with important implications for clinical diabetes assessment. First, Multimodel inference revealed substantial variation in predicted β -cell sensitivity across the three phenomenological models, particularly for T2D populations, where the Cretti 2001 model predicted markedly lower sensitivity values compared to the Hovorka 1998 and Breda 2001 models. The observation that pseudo-Bayesian model averaging and model fusion yield different combined predictions further underscores the complexity of integrating multiple modeling perspectives. Furthermore, we found that the OGTT sampling strategy dramatically affected prediction quality, with sparse sampling protocols introducing both increased uncertainty and systematic bias in β -cell sensitivity estimates.

These findings have several key clinical implications. First, model-dependent variation highlights a critical limitation in current clinical practice, where reliance on a single modeling approach may lead to inconsistent patient assessments. Second, the finding that reduced sampling frequency and duration lead to broader prediction intervals and shifted probability densities suggests that commonly used abbreviated OGTT protocols may compromise diagnostic accuracy. These results support the hypothesis that unaccounted uncertainties from model assumptions, data variability, and clinical protocols substantially impact the trustworthiness of mathematical predictions in diabetes care, indicating an urgent need for uncertainty quantification methods to guide clinical decision-making and determine when additional testing may be necessary to ensure reliable patient assessment.

Chapter 6

Conclusion—An expanded UQ workflow and the future of UQ in systems biology

6.1 Summary and future directions

Mathematical models of intracellular systems have been integral in driving biological discovery in recent years [14]. In particular, modeling enables researchers to go beyond what is experimentally observable to interpret, explain, and predict biological function. However, uncertainty is inevitable in modeling biology because models are built with incomplete system knowledge and necessarily rely on uncertain data [75]–[77], [308]. As a result, three central questions arise in addressing uncertainty: (i) How can we formulate a model given incomplete knowledge? (ii) Given a model, how can we leverage noisy and sparse experimental data to calibrate its predictions? (iii) How do uncertainties in the data used for model calibration affect downstream predictions? Uncertainty quantification, which was originally developed alongside computational engineering, provides a methodological framework to address such questions [5]. Despite the relative maturity of UQ in other fields, it is only just being applied to systems

biology models [75], [105], [120], [308], [310], [370].

In this thesis, we investigate a series of methods to expand the systems biology UQ toolkit. First, we investigated how identifiability and sensitivity analyses combined with Bayesian parameter estimation enabled parameter estimation from low-dimensional, noisy experimental data (Section 2). Second, we leveraged Bayesian multimodel inference to account for model uncertainty, which arises through multiple models of the same biological system (Section 3). Third, we applied these methods to develop a new data-informed model of AMPK signaling (Chapter 4). Finally, we highlighted how these methods are applicable beyond systems biology by applying them to models of insulin-glucose dynamics (Section 5). In this chapter, I summarize the contributions of this thesis and discuss possible future avenues for research at the interface of computational mathematics and systems biology.

6.1.1 Thesis summary

Bayesian methods enable UQ from noisy and sparse experimental data

Statistical inference is critical to addressing the parameter estimation problem in systems biology. Experiments provide limited information about model parameters, so we often need to estimate parameters from available data. This task is complicated by the nature of systems biology models and the nature of available experimental data. Although traditional frequentist inference methods have been used widely in systems biology [76], [308], [371], [372], these methods do not provide a complete account of uncertainty [5] and can struggle in settings with limited data [84].

We showed how Bayesian parameter estimation can address the challenges of parameter estimation and simultaneously quantify uncertainty in a probabilistic manner. First, we found that *a priori* identifiability and sensitivity analyses enabled successful parameter estimation (Figure 2.8) by reducing the dimension of the parameter space. This *a priori* parameter analysis likely improves the *posedness* of

the inverse problem and therefore leads to more successful inference [93]. Next, we saw how Bayesian posterior densities can highlight uncertainty in training data (Figure 2.3) and that more data is not always better (Figure 2.7). Moving beyond the toy models of Chapter 2, we showed how the Bayesian parameter estimation workflow can be applied to larger models of the ERK signaling pathway with 10s–100s of uncertain parameters. Finally, we addressed a key shortcoming of the original workflow, that is, *what happens if a parameter is nonidentifiable, but its value impacts model prediction?* To do so, we introduced an iterative refinement procedure in Chapter 4 that leverages prior predictive samples to test whether fixed quantities yield reasonable predictions. Taken together, these results show how Bayesian inference enables successful parameter estimation for a wide range of systems biology models and lays the groundwork for future applications.

A multimodel viewpoint accounts for model form uncertainty and allows for data-informed modeling

Model uncertainty is inevitable in systems biology because modeling relies on assumptions and incomplete knowledge for model development [313], [373]. There have been a number of attempts to address model uncertainty in the broader UQ community by extending statistical models to account for the discrepancy between a model and physical reality [79], [374]. However, model uncertainty in systems biology typically arises in having (or being able to construct) multiple models of the same system, each with its own set of assumptions [77], [313], [375]. In this work, we investigated how to take advantage of these multiple models to address model uncertainty. To do so, we leveraged an approach called multimodel inference that constructs new estimates by taking a weighted average of preexisting models (Chapter 3). Importantly, we found that MMI enabled us to account for model form uncertainty in systems biology, specifically when multiple models of the same system were available. Furthermore, we applied Bayesian model comparison (the core tool used to compute the weights on each model) to enable data-informed

selection of a *best* model of AMPK signaling. While MMI can increase the computational demands and complexity of a model discovery, we found that it allowed us to justify modeling decisions with data, which was especially important when few models had been previously published.

A workflow for UQ in systems biology

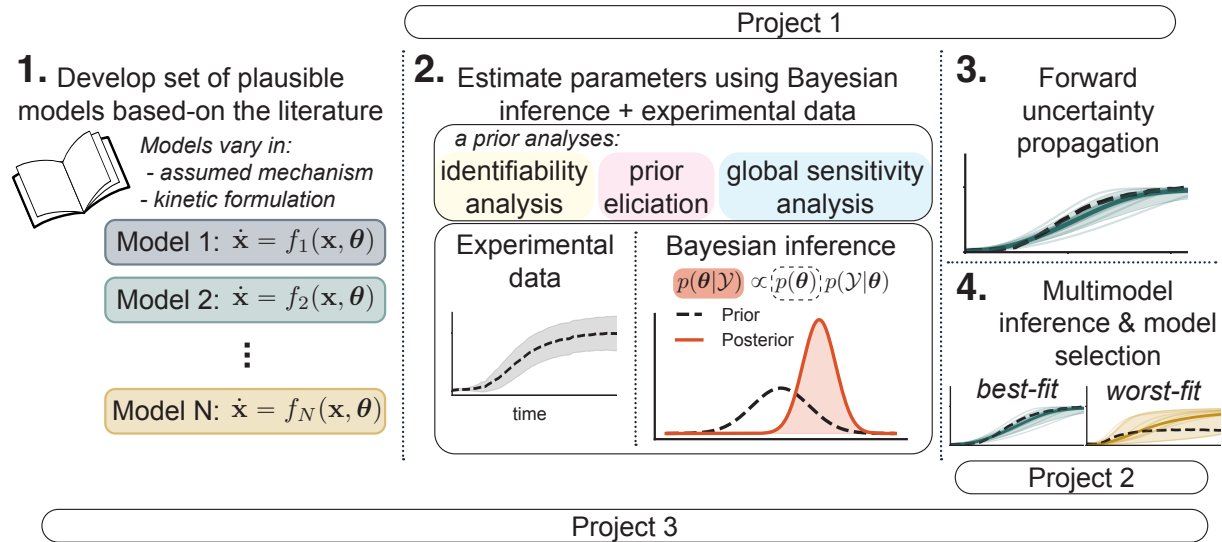


Figure 6.1: (1) Models are either developed or collected using available literature. If multiple models exist, they often vary in model assumptions and kinetic formulation. (2) Next, Bayesian inference combined with *a priori* parameter analysis is used to estimate uncertain model parameters from experimental data. (3) Parametric uncertainty is then propagated forward to model predictions by running ensembles of simulations. (4) Finally, Bayesian multimodel inference can address model uncertainty if multiple models have been calibrated.

In this thesis, we investigated new extensions to the systems biology modeling workflow to enable uncertainty quantification. All-in-all, these extensions combine into a new UQ-based workflow for systems biology (Figure 6.1). In Chapters 2 and 3 we laid the groundwork for the workflow, and in Chapter 4 we applied it to study a new signaling pathway. The workflow begins with model development or curation, where we construct a model (or models) using literature and domain knowledge. Next, we obtain experimental data and apply the Bayesian parameter estimation approach introduced in Chapter 2. We then leverage the resulting posterior density to propagate uncertainty to model predictions. Finally, we

used Bayesian multimodel inference and model selection to address model uncertainty if multiple models were calibrated. This workflow is broadly applicable in systems biology and beyond whenever parametric models and data are available (see Chapter 5 for an application outside of systems biology).

6.1.2 Future directions

Application to study spatiotemporal signaling

Intracellular signaling occurs over several spatial scales in addition to temporal scales. At the sub-cellular level, signaling has been found to be localized to distinct sub-cellular locations, which are hypothesized to play a role in coordinating distinct cellular responses [7], [8], [71], [304]. Furthermore, subcellular spatial features and geometry can introduce spatiotemporal heterogeneity in signaling responses [57]. Modeling such phenomena requires moving beyond the well-mixed ODE framework investigated to incorporate spatial information using compartmental and partial differential equation models [16]. To enable Bayesian parameter estimation for PDE systems biology models, several parameter-related questions need to be addressed, in addition to several technical developments. A typical practice for parameter estimation of spatial models is to first estimate the parameters of the well-mixed ODE model and then use those parameter estimates for the spatial model [47], [376], [377]; however, this approach ignores potential issues with identifiability [378] or changes in parameter behavior that occur due to the introduction of spatial information into the model. A key question is how introducing spatial information into a model affects the parameter identifiability and estimates compared to well-mixed ODE models.

Furthermore, PDE-based reaction-diffusion models are computationally more expensive to solve than ODEs, which introduces new computational challenges when applying Bayesian inference methods. Methods and software frameworks that estimate the parameters of PDE models abound in engineering and computational science [379]–[381], and have recently been developed in the Rangamani lab for systems

biology [17]. For example, ensemble Kalman methods are routinely used to infer parameters of climate and weather models from atmospheric measurements [382] and randomize-then-optimize methods can estimate geophysical properties from measurement data [95]. Furthermore, parametric reduced-order modeling will likely be necessary to accelerate outer-loop tasks such as Markov chain Monte Carlo Sampling [383]–[386]. An additional avenue to address the computational challenges of calibrating PDE models is to develop a randomize-then-optimize approach for approximate posterior sampling [95], [187]. The extension of Bayesian UQ methods to study spatiotemporal signaling will bring many interesting challenges and lead to technological advancements and new biological understanding.

Incorporation of scientific machine learning to enhance systems biology models

There is no escaping *machine learning* or *artificial intelligence* in the science landscape 2025. While these fields make up a vast domain of methodologies, a handful of them have the greatest potential to enhance systems biology. These include data-driven equation discovery [97], [194], [275], universal differential equations [160], neural ODEs [323], and foundation modeling [387]. Data-driven equation discovery enables one to select model terms from a prescribed library that best fit the data using sparse regression [97], [194], [275]. Such models retain the interpretability of a traditional mechanistic model, because the sparse regression framework highlights which terms should be combined into the final model. Universal differential equations retain the structure of a mechanistic model but append a purely data-driven universal approximator (neural network or Gaussian process) to learn interactions and features that may be missing from the original equations [160]. Neural ODEs take this concept one step further and replace the entire right-hand-side of the ODE with a machine-learned function approximator [323]. Finally, foundation models are very large deep learning models that are trained on a vast collection of data [387]. In systems biology, these models have the potential to map from experimental measurements to model

structures or from network diagrams to predictive models. All-in-all, as datasets become larger and more complex, data-driven modeling has the potential to allow systems biology models to keep up and match the complexity of an entire biological system.

Robust software for UQ in systems biology

A number of software tools have been developed to make systems biology more accessible and enable “reproducible” modeling, including VCell ([https://vcell.org/](#)), Copasi (<https://copasi.org/>), SymBiology (<https://www.mathworks.com/products/simbiology.html>), BioNetGen (<https://bionetgen.org/>) and Tellurium (<https://tellurium.analogmachine.org/>). While these are mature software projects that are generally well regarded, they are limited in their ability to interface with modern tools for machine learning, and Bayesian statistics, including Jax (<https://docs.jax.dev/>), PyTorch (<https://pytorch.org/>), Stan (<https://mc-stan.org/>), and PyMC (<https://www.pymc.io/welcome.html>). For that reason, we manually programmed all models in this thesis and did not rely on any of the systems biology software tools. A new generation of systems biology software should leverage these modern software tools to a wide range of users to apply cutting-edge UQ and ML algorithms to their models.

6.2 Closing Remarks

The application of uncertainty quantification to systems biology models of intracellular signaling has led to a rich line of research that has yielded both methodological and biological discoveries. Working at this interface is made possible by interdisciplinary collaboration between experimentalists, modelers, and computational scientists. As we move into an age of big-data biology, such interdisciplinary collaborations will become increasingly important to driving new biological insights from data. I believe that this thesis serves as a starting point for future efforts to apply uncertainty quantification to the biological sciences. The

direct application and extension of the approaches presented here will enable a multitude of data-informed and data-driven studies both within systems biology and in the broader biological and medical sciences.

Appendix A

Supplemental Materials for Chapter 2

A.1 Supplemental Text

A.1.1 Runtime analysis of the CIUKF-based likelihood function

UKF-MCMC, and thus CIUKF-MCMC, poses an increased computational cost compared to estimation methods that do not consider all sources of uncertainty [82]. To better understand how this increased computational burden affects parameter estimation, we used the `timeit()` function in Matlab to time the execution of a CIUKF-based likelihood function and a Gaussian likelihood function Eq (2.7) that ignores model form uncertainty and that is representative of standard approaches [5], [134], [140], [142]. Specifically, we ran `timeit()` 20 times for each likelihood function evaluated for the MAPK model with the data from the low steady state, seen in Fig 2.5A. We found that, on average, the CIUKF-based likelihood function takes 81 times longer (0.278 seconds) than the Gaussian likelihood (0.0034 seconds). All timing was performed on a laptop with a 2.6Ghz 6-core Intel Core i7 (Intel Corporation, Santa Clara, CA) processor with 32 GB of RAM.

To better understand what computations contribute to the observed runtimes of the CIUKF-based

likelihood, we used the `profiler` in Matlab to analyze a single call to the likelihood function. We found that the CIUKF-based likelihood execution spent approximately 60% of the time on calls to `ode15s()`, which we used to discretize the ordinary differential equation model. Additionally, another 30% of the time was spent running `quadprog()` during the update step of the constrained interval unscented Kalman filter. Based on these findings, we conclude that developing better-optimized codes for these steps will likely accelerate the CIUKF-based likelihood function and, thus, shorten MCMC runtimes.

A.2 Supplemental Tables

This section provides supplemental tables.

Table A.1: All listed values have units of one over time.

Parameter	Nominal Value	Range
k_{1e}	1	[0, 5]
k_{12}	1	[0, 5]
k_{21}	1	[0, 5]
b	2	[0, 10]

Table A.2: Note: For the oscillatory dynamics, the range for k_5 is $[1 \times 10^{-5} \ 0.05]$.

Parameter	Nominal Value: Limit-cycle	Nominal Value: Bistable	Range
S_{1t}	100 nM	0.22 nM	[0, 100]
S_{2t}	100 nM	10 nM	[0, 100]
S_{3t}	100 nM	53 nM	[0, 100]
k_1	$0.1 \text{ nM} \cdot \text{s}^{-1}$	$0.0012 \text{ nM} \cdot \text{s}^{-1}$	[0, 0.05]
k_2	$0.01 \text{ nM} \cdot \text{s}^{-1}$	$0.006 \text{ nM} \cdot \text{s}^{-1}$	[0, 0.1]
k_3	$0.01 \text{ nM} \cdot \text{s}^{-1}$	$0.049 \text{ nM} \cdot \text{s}^{-1}$	[0, 0.05]
k_4	$0.01 \text{ nM} \cdot \text{s}^{-1}$	$0.084 \text{ nM} \cdot \text{s}^{-1}$	[0, 0.1]
k_5	$0.01 \text{ nM} \cdot \text{s}^{-1}$	$0.043 \text{ nM} \cdot \text{s}^{-1}$	[0, 0.05]
k_6	$0.01 \text{ nM} \cdot \text{s}^{-1}$	$0.066 \text{ nM} \cdot \text{s}^{-1}$	[0, 0.1]
n_1	10	5	[5, 10]
K_1	1 nM	9.5 nM	[0, 10]
n_2	15	10	[5, 10]
K_2	8 nM	15 nM	[0, 20]
α	10	95	[0, 100]

Table A.3:

State-variable	Limit cycle	Bistable: low steady state	Bistable: high steady state
$x_1(0)$	10 nM	0.0015 nM	0.1245 nM
$x_2(0)$	80 nM	3.6678 nM	2.4870 nM
$x_3(0)$	80 nM	28.7307 nM	31.2623 nM

Table A.4: The ranges are given by $[0.1 \cdot \theta_i^*, 10 \cdot \theta_i^*]$, where θ_i^* is the nominal value. Note: we do not include ranges for n_1 and n_2 because these parameters are always set to the nominal values.

Parameter	Nominal Value	Range
k_1	2 s^{-1}	[0.2, 20]
k_2	15 s^{-1}	[1.5, 150]
k_3	1 s^{-1}	[0.1, 100]
k_4	120 s^{-1}	[12, 1200]
k_5	2 s^{-1}	[0.2, 20]
k_6	15 s^{-1}	[1.5, 150]
k_7	1 s^{-1}	[0.1, 10]
k_8	80 s^{-1}	[8.0, 800]
c_1	1	[0.1, 10]
c_2	1	[0.1, 10]
c_3	6 s^{-1}	[0.6, 60]
c_4	8 s^{-1}	[0.8, 80]
K_{m1}	$10 \mu\text{M}$	[1.0, 100]
K_{m2}	$0.3 \mu\text{M}$	[0.03, 3]
K_{m3}	$4 \mu\text{M}$	[0.4, 40]
K_{m4}	$10 \mu\text{M}$	[1.0, 100]
K_{m5}	$1 \mu\text{M}$	[0.1, 10]
K_0	$0.5 \mu\text{M}$	[0.05, 5]
P_0	$0.5 \mu\text{M}$	[0.05, 5]
K_{tot}	$20 \mu\text{M}$	[2, 200]
P_{tot}	$20 \mu\text{M}$	[2, 200]
A_{tot}	1	[0.1, 10]
n_1	4	—
n_2	3	—

A.3 Supplemental Figures

This section provides supplemental figures.

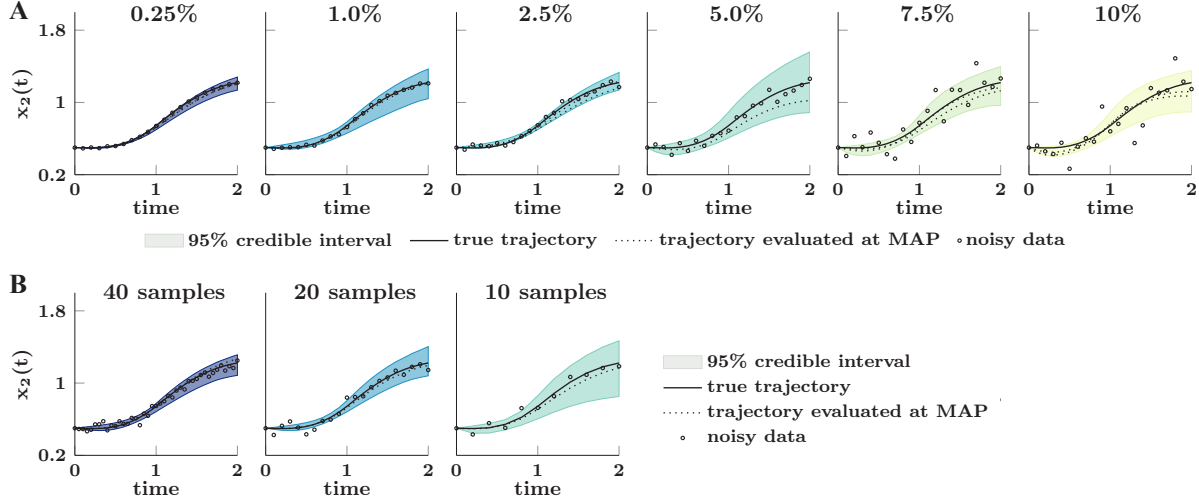


Figure A.1: (A) corresponds to the measurement noise experiment, Fig 2.3C. We observe that increasing the data noise level increased the uncertainty in the predicted dynamics. We control the noise level by setting the noise covariances to the specified percentage of the standard deviation of each state variable. The dashed black vertical lines indicate each parameter's nominal (true) value. (B) corresponds to the data sparsity experiment, Fig 2.3E. We observe that decreasing the number of experimental samples supplied for estimation increased estimation uncertainty. The noise level in the data was fixed to the 2.5% level shown above.

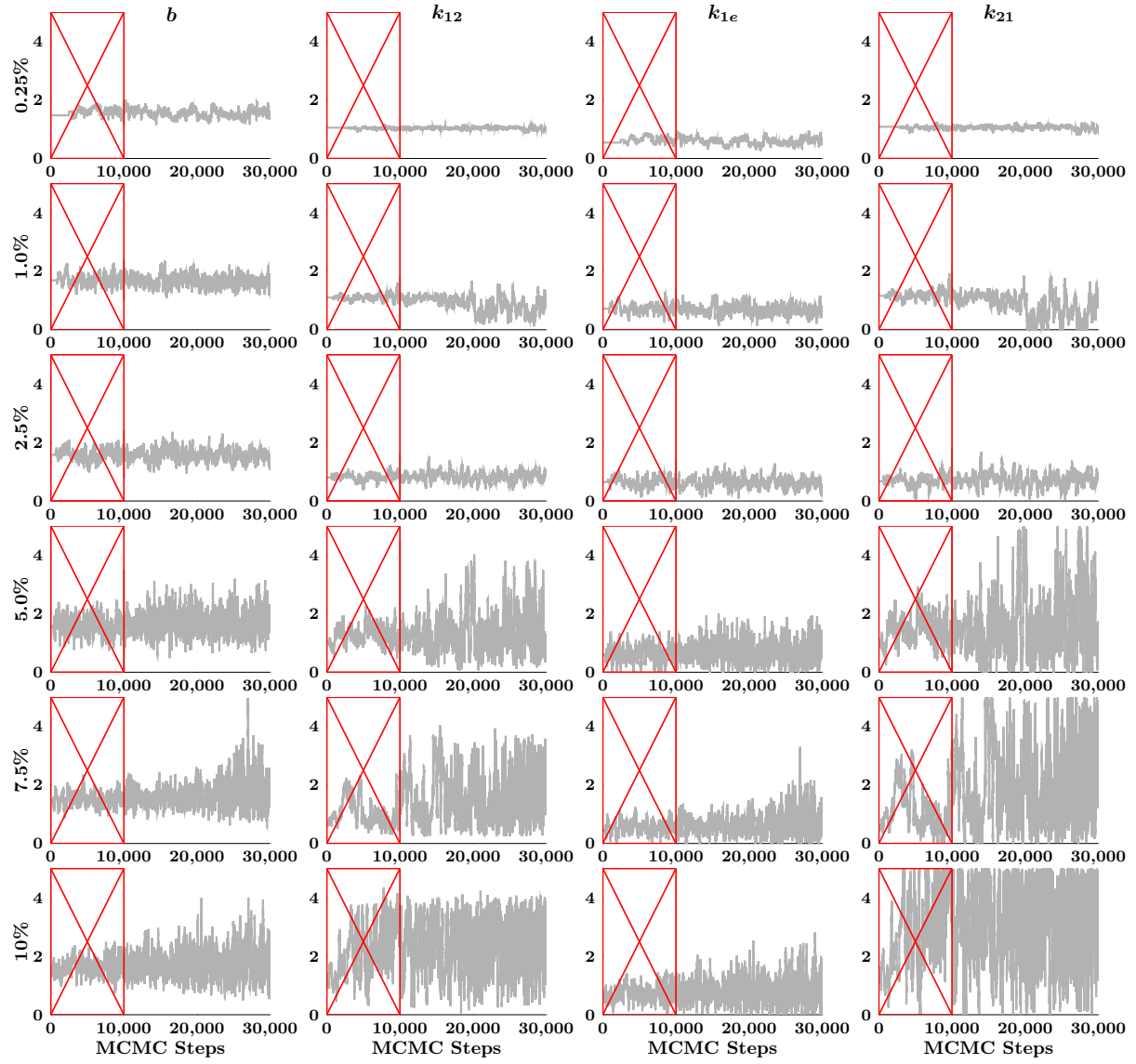


Figure A.2: Each row corresponds to a noise level; noise increases down the figure. The red boxes indicate samples discarded as burn-in.

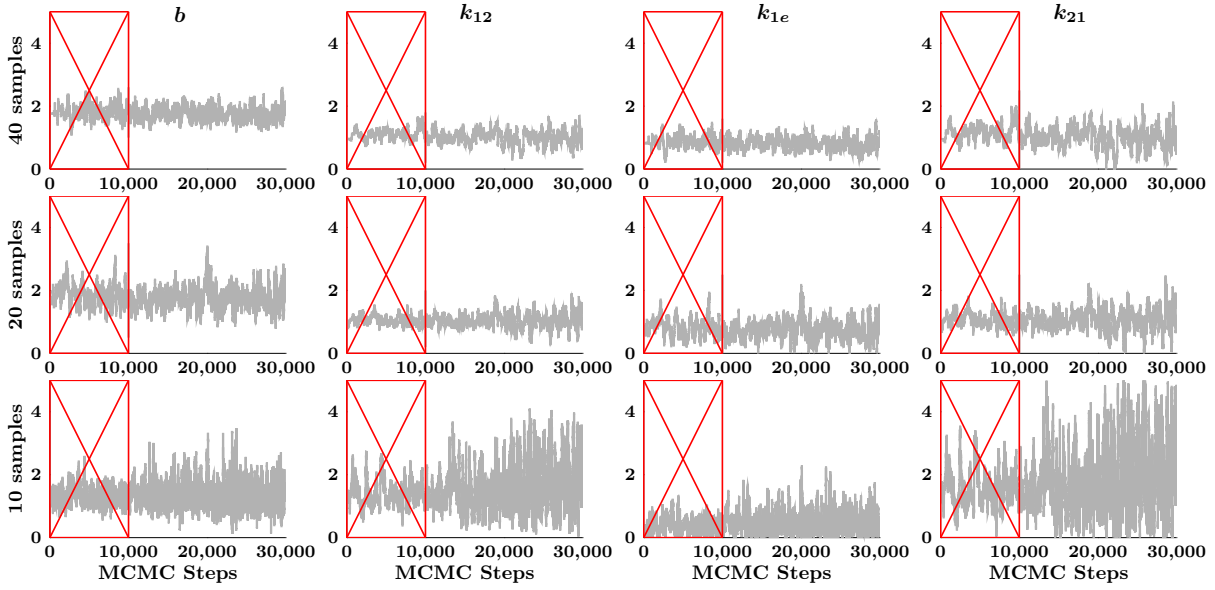


Figure A.3: Each row corresponds to a sparsity level. The red boxes indicate samples discarded as burn-in.

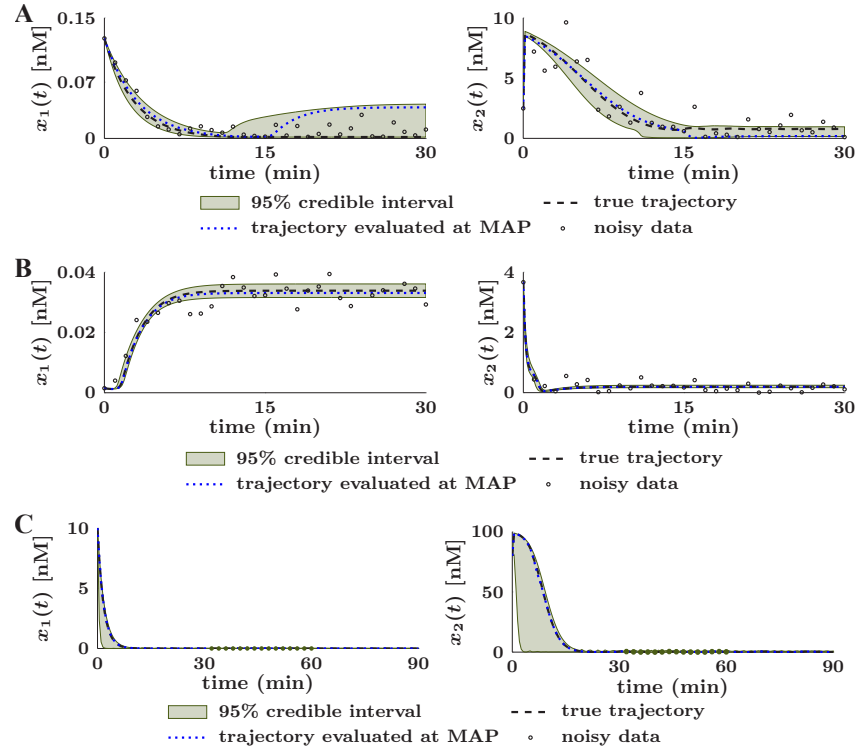


Figure A.4: (A) Trajectories for the high steady-state; corresponds to Fig 2.5D. (B) Trajectories for the low steady-state; corresponds to Fig 2.5C. (C) Trajectories for limit cycle oscillations; corresponds to Fig 2.6B.

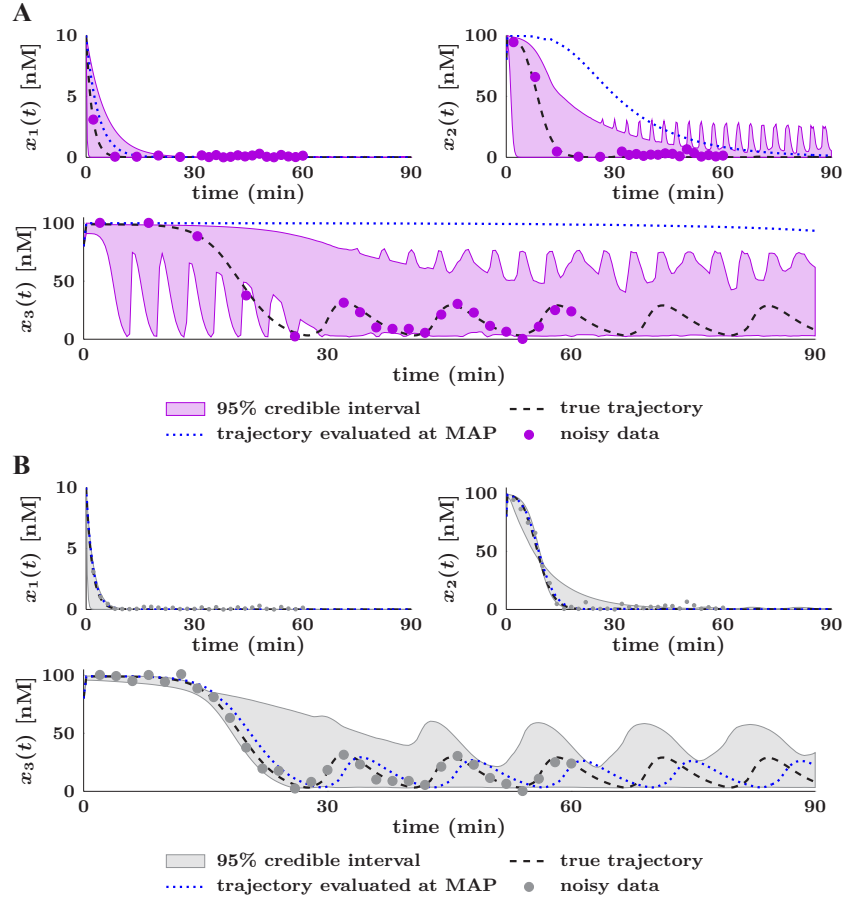


Figure A.5: (A) Distributions for the data with equidistant sampling; corresponds to Fig 2.7C. (B) Distributions for the data with non-equidistant sampling; corresponds to Fig 2.7D.

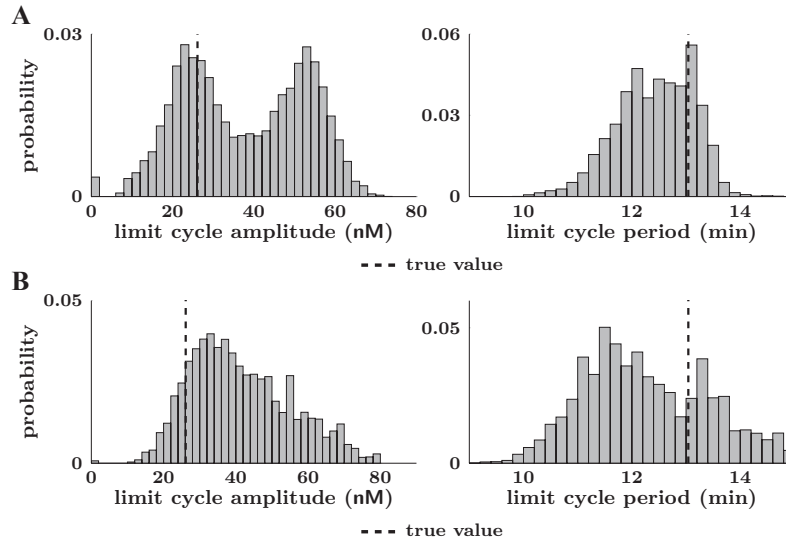


Figure A.6: (A) Corresponds to Fig 2.7C and Supplemental Fig A.5A. (B) Corresponds to Fig 2.7D and Supplemental Fig A.5B.

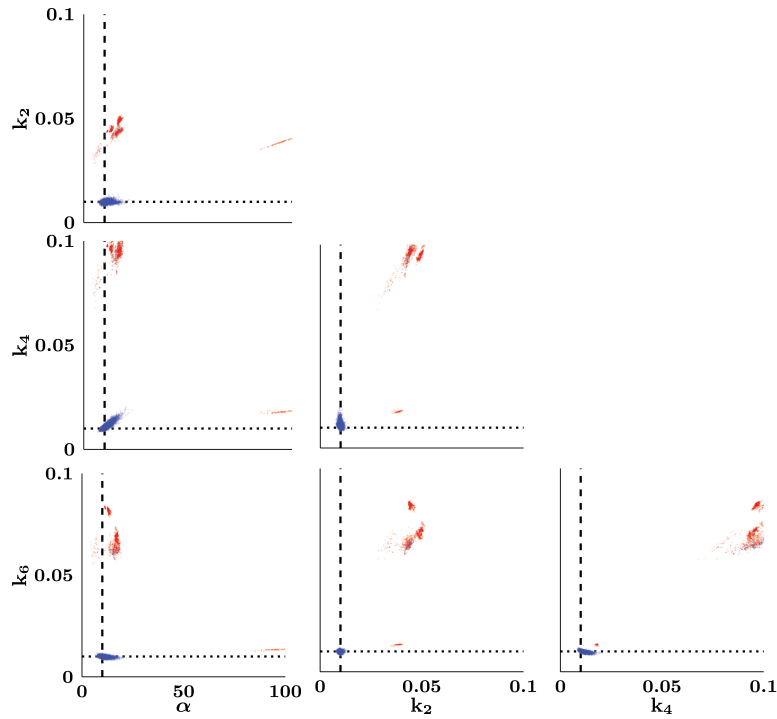


Figure A.7: Each plot has 30,000 points that are colored according to the nature of the simulated dynamics with that parameter set. Simulations with blue points produce limit cycle oscillations, and those with red points produce fixed points. Darker regions indicate a higher probability of observing the corresponding parameter values.

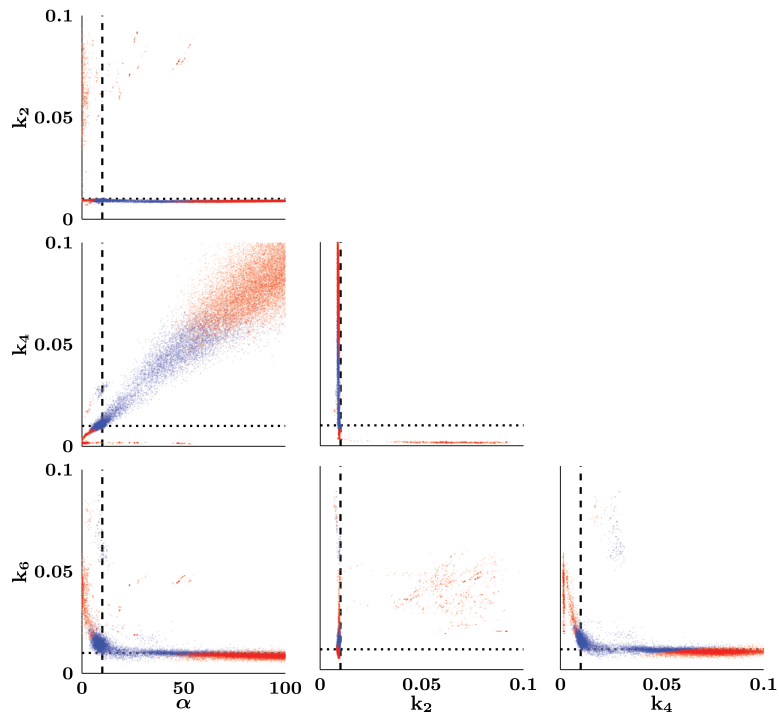


Figure A.8: Each plot has 30,000 points that are colored according to the nature of the simulated dynamics with that parameter set. Simulations with blue points produce limit cycle oscillations, and those with red points produce fixed points. Darker regions indicate a higher probability of observing the corresponding parameter values.

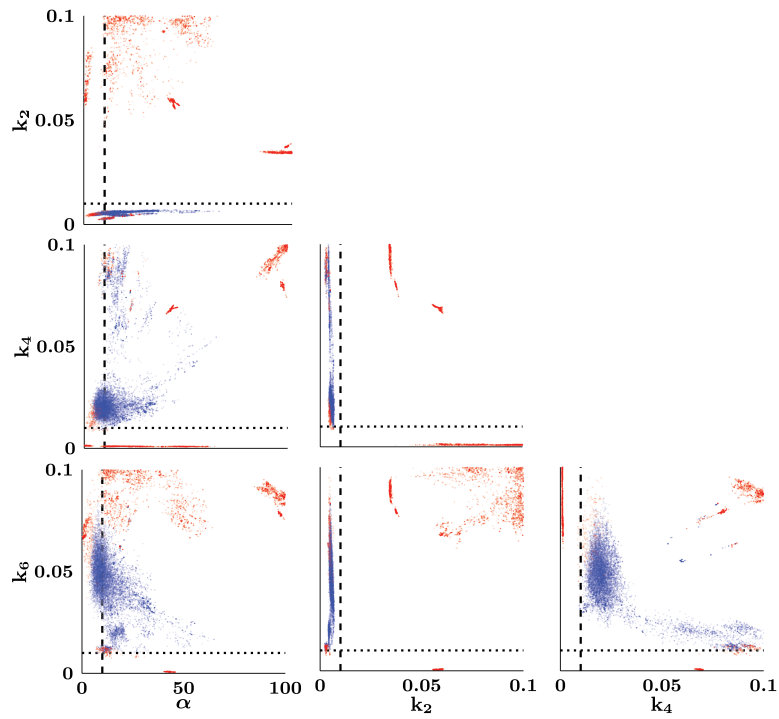


Figure A.9: Each plot has 30,000 points that are colored according to the nature of the simulated dynamics with that parameter set. Simulations with blue points produce limit cycle oscillations, and those with red points produce fixed points. Darker regions indicate a higher probability of observing the corresponding parameter values.

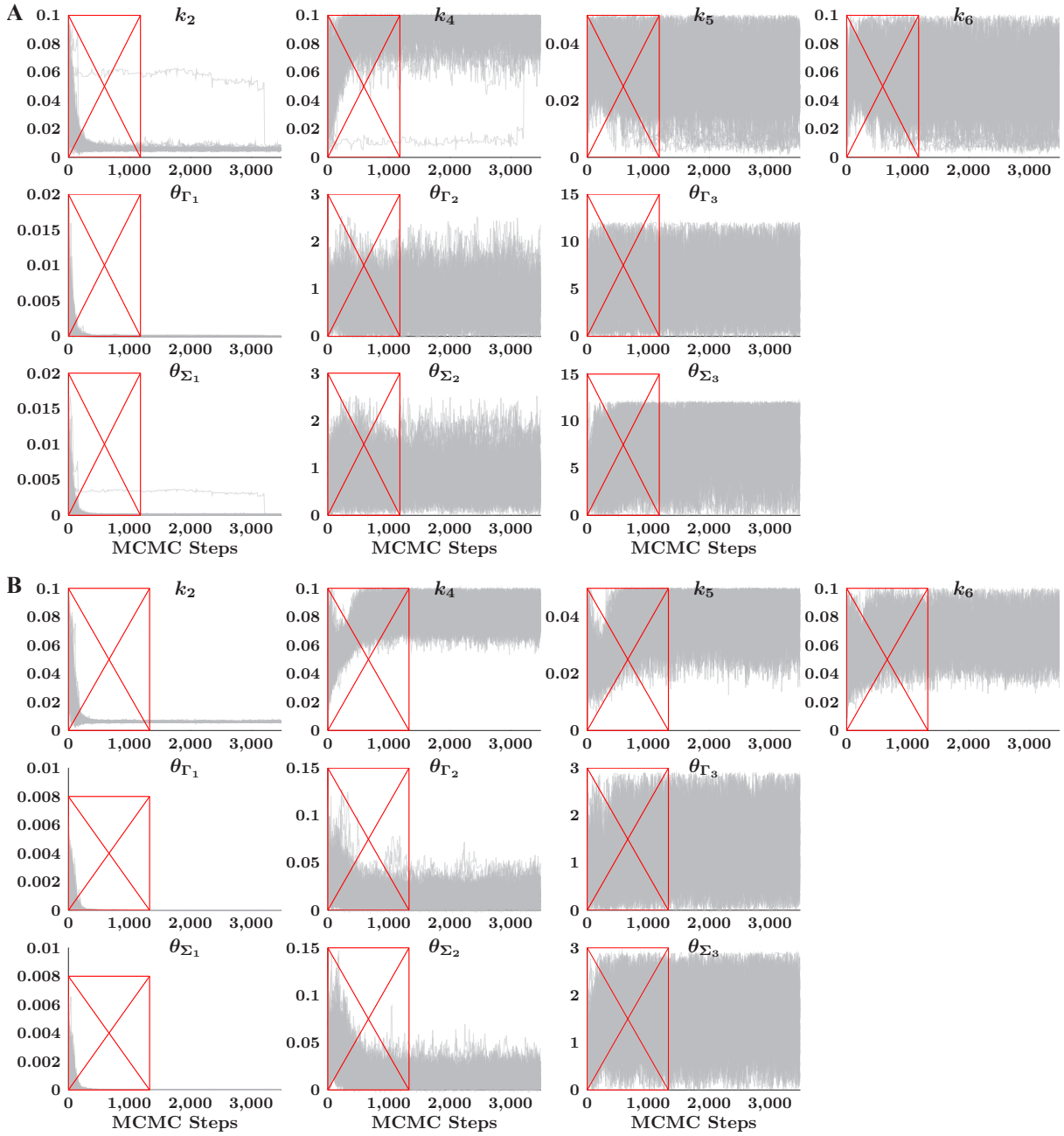


Figure A.10: (A) Low steady-state, Fig 2.5A. (B) High steady-state, Fig 2.5B. The process noise covariances are θ_{Γ_i} and θ_{Σ_i} are the measurement noise covariances. The red boxes indicate samples discarded as burn-in.

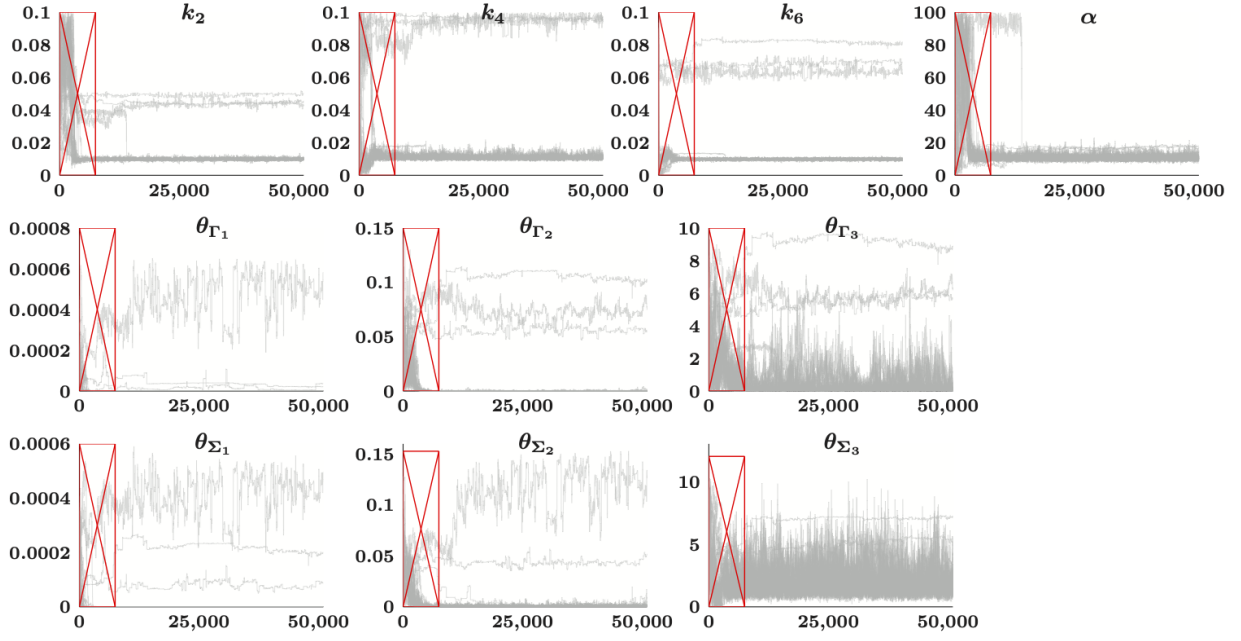


Figure A.11: The process noise covariances are θ_{Γ_i} and θ_{Σ_i} are the measurement noise covariances. The red boxes indicate 7,477 samples per chain discarded as burn-in.

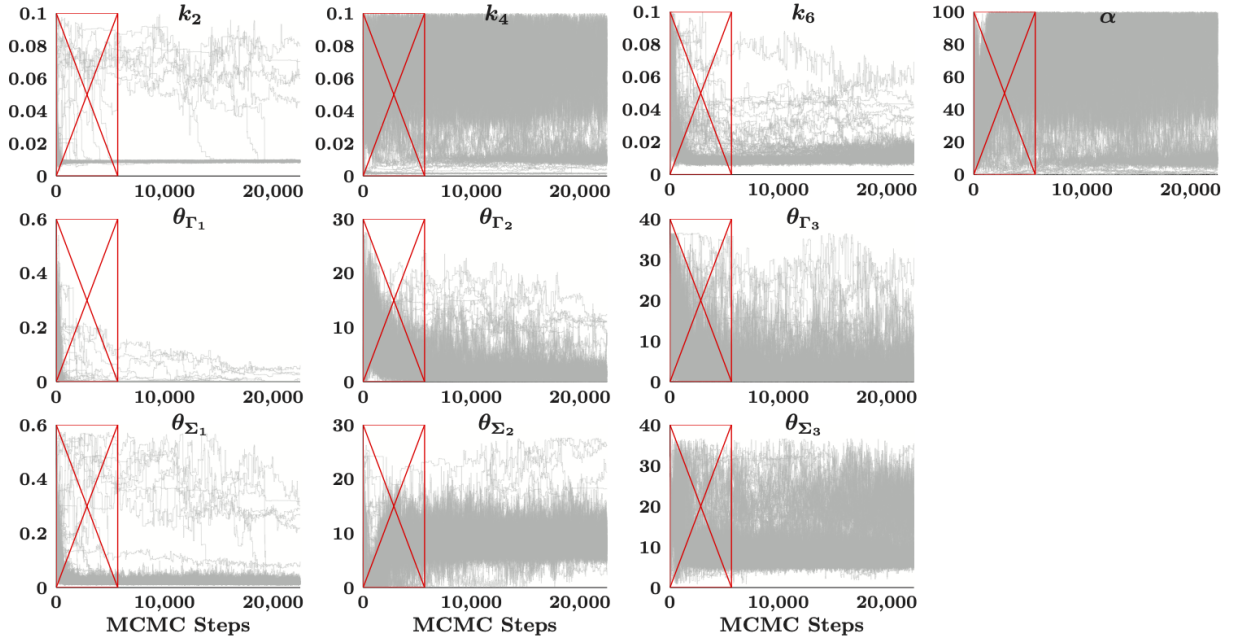


Figure A.12: The process noise covariances are θ_{Γ_i} and θ_{Σ_i} are the measurement noise covariances. The red boxes indicate 5,669 samples per chain discarded as burn-in.

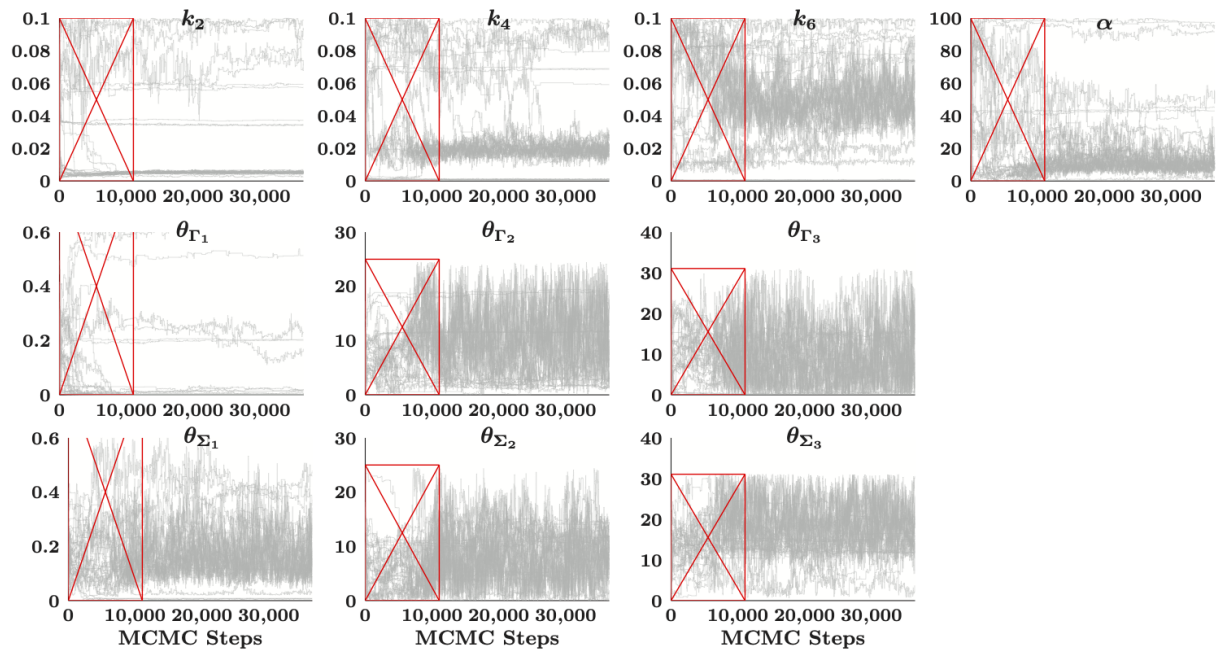


Figure A.13: The process noise covariances are θ_{Γ_i} and θ_{Σ_i} are the measurement noise covariances. The red boxes indicate 11,065 samples per chain discarded as burn-in.

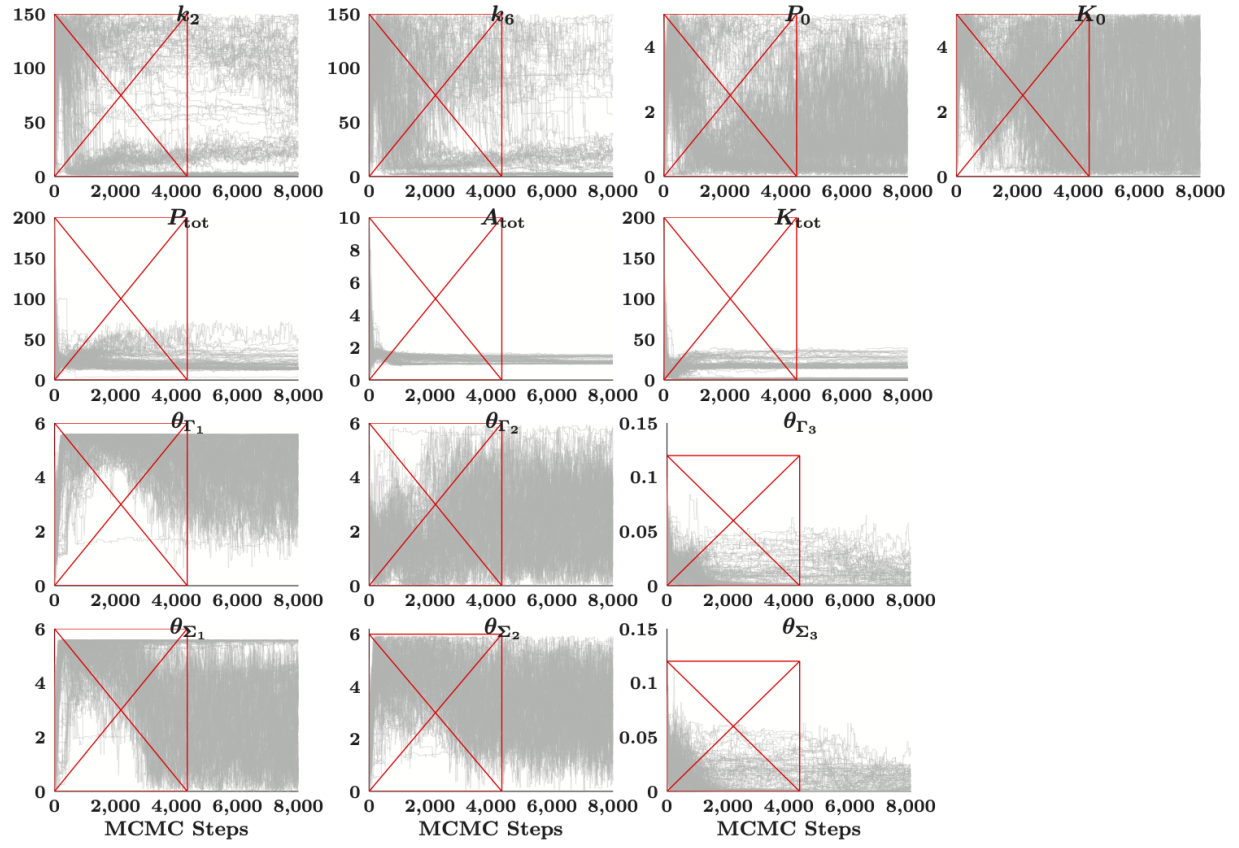


Figure A.14: The red boxes indicate samples discarded as burn-in.

Appendix B

Supplemental Materials for Chapter 3

B.1 Supplemental Tables

Table B.1: The number of state variables (# states (n_k)) represents the number of biochemical species in the model and the number of parameters (# parameters (p_k)) is the total number of free model parameters. The number of locally identifiable (# locally ID) is the number of parameters kept free after local identifiability analysis, and the number of sensitive parameters (# sensitive) is the number of parameters kept free after global sensitivity analysis.

Model Name	# states (n_k)	# parameters (p_k)	# locally ID	# sensitive
H' 1996 [228]	16	36	36	26
K' 2000 [260]	6	25	8	6
L' 2000 [259]	27	40	40	25
H' 2005 [249]	103	96	94	28
B' 2007 [226]	117	216	90	20
O' 2009 [266]	26	59	7	5
vK' 2009 [227]	37	77	31	18
S' 2014 [262]	7	20	7	7
R' 2015 [263]	17	45	5	5
K' 2017 [264]	35	23	14	11

Table B.2: We either use an ODE-solver or Newton iteration to find the steady-state solution. Steady-state ODE tolerances are used to assess the convergence of ODE-based solutions for either approach. We do not compute steady-states for the H' 1996 model, so we omit the corresponding hyperparameters. ODE-based solutions are solved until steady-state convergence or $t = t_{\max}$.

Model	ODE tols (atol; rtol)	steady-state method	steady-state tols (atol; rtol)	time units	$t_{\max}$
H' 1996	1e-6; 1e-6	—	—	sec	—
K' 2000	1e-6; 1e-6	Newton	1e-6; 1e-6	sec	∞
L' 2000	1e-6; 1e-6	ODE	1e-10; 1e-10	sec	100,000 (s)
H' 2005	1e-6; 1e-6	ODE	1e-10; 1e-10	sec	12,000
B' 2007	1e-6; 1e-6	ODE	1e-5; 1e-6	sec	10,000
O' 2009	1e-6; 1e-6	ODE	1e-5; 1e-6	min	1,400
vK' 2009	1e-6; 1e-6	ODE	1e-5; 1e-6	sec	5,400
S' 2014	1e-6; 1e-6	Newton	1e-6; 1e-6	min	∞
R' 2015	1e-6; 1e-6	Newton	1e-5; 1e-5	min	540
K' 2017	1e-6; 1e-6	Newton	1e-6; 1e-6	sec	10,800

Table B.3: Wall-clock times are shown in hours. Parentheses show: (# of chains and # number of samples/chain) if those values differ from: # of chains = 4 and # number of samples/chain = 500, i.e. (4; 500). Sampling for Figure 3.3 and Supplemental Figures B.3–B.4 was performed on a Linux Workstation with dual Intel Xeon 2.60GHz processors and 123Gb of DDR3 RAM. Sampling for Figure 3.8 was performed on a Macbook Pro with the Apple M3 Pro CPU and 18Gb of DDR5 RAM.

Model	Figure 3.3	Supp. Figure B.3	Supp. Figure B.4	Figure 3.8
H' 1996	10.68	0.24	-	-
K' 2000	0.04	0.04	1.72 (4; 1000)	-
L' 2000	0.07	0.08	8.65 (4; 200)	-
H' 2005	1.97 (8; 100)	1.88 (8; 100)	110.66 (4; 100)	-
B' 2007	1.73	1.77	167.87 (4; 100)	-
O' 2009	0.11	0.12	2.10 (4; 250)	-
vK' 2009	0.31	0.40	0.86 (4; 100)	-
S' 2014	0.07	0.07	16.90 (4; 1000)	-
R' 2015	0.04	0.04	18.64 (4; 1000)	-
K' 2017	0.19	0.22	222.76 (4; 1000)	-
O' 2009 Rap1 diff.	-	-	-	0.66
O' 2009 (-)FB diff.	-	-	-	0.58
O' 2009 Rap1 and (-)FB diff.	-	-	-	0.65
S' 2014 Rap1 diff.	-	-	-	0.21
S' 2014 (-)FB diff.	-	-	-	0.16
S' 2014 Rap1 and (-)FB diff.	-	-	-	0.19
R' 2015 Rap1 diff.	-	-	-	0.37
R' 2015 (-)FB diff.	-	-	-	0.33
R' 2015 Rap1 and (-)FB diff.	-	-	-	0.66

Table B.4: Model predictive performance (ELPD) and MMI weights of the top and bottom ten models of the 94 total compartment-specific parameter combinations. Additional models are not shown. Bold entries indicate weight greater than 0.01 by any MMI method.

Model	ELPD	pseudo-BMA weight	stacking weight	model probability
<i>Top ten models:</i>				
O' 2009	343.585	0.404	0.882	0.343
Rap1, ppMEK-ERK & ppERK-pASE diff.				
O' 2009	343.406	0.337	0.0	0.093
all diff.				
O' 2009	341.456	0.072	0.0	0.16
Rap1, (-)FB & ppERK-pASE diff.				
O' 2009	341.371	0.08	0.0	0.402
Rap1 & ppERK-pASE diff.				
R' 2015	339.737	0.065	0.054	0.0
Rap1 & (-)FB diff.				

Continued on next page

Table B.4: Continued: Model predictive performance (ELPD) and MMI weights of the top and bottom ten models of the 94 total compartment-specific parameter combinations. Additional models are not shown. Bold entries indicate weight greater than 0.01 by any MMI method.

Model	ELPD	pseudo-BMA weight	stacking weight	model probability
R' 2015	337.448	0.037	0.063	0.0
Rap1, (-)FB & ppERK-pASE diff.				
O' 2009	337.233	0.002	0.0	0.001
(-)FB, ppMEK-ERK & ppERK-pASE diff.				
O' 2009	336.723	0.001	0.0	0.001
Rap1 & (-)FB diff.				
O' 2009	335.841	0.0	0.0	0.0
Rap1, (-)FB & ppMEK-ERK diff.				
S' 2014	335.068	0.0	0.0	0.0
Rap1, (-)FB, Raf-deact. & ppERK-dephos. diff.				
<i>Bottom ten models:</i>				

Continued on next page

Table B.4: Continued: Model predictive performance (ELPD) and MMI weights of the top and bottom ten models of the 94 total compartment-specific parameter combinations. Additional models are not shown. Bold entries indicate weight greater than 0.01 by any MMI method.

Model	ELPD	pseudo-BMA weight	stacking weight	model probability
R' 2015 all diff.	278.966	0.0	0.0	0.0
R' 2015 (-)FB & ppERK-pASE diff.	276.007	0.0	0.0	0.0
S' 2014 (-)FB & Raf-deact. diff.	262.615	0.0	0.0	0.0
S' 2014 Raf-deact. diff.	260.414	0.0	0.0	0.0
S' 2014 (-)FB diff.	245.325	0.0	0.0	0.0
R' 2015 ppMEK-ERK diff.	208.882	0.0	0.0	0.0
R' 2015 (-)FB & ppMEK-ERK diff.	201.588	0.0	0.0	0.0

Continued on next page

Table B.4: Continued: Model predictive performance (ELPD) and MMI weights of the top and bottom ten models of the 94 total compartment-specific parameter combinations. Additional models are not shown. Bold entries indicate weight greater than 0.01 by any MMI method.

Model	ELPD	pseudo-BMA weight	stacking weight	model probability
R' 2015 (-)FB, ppMEK-ERK & ppERK-pASE diff.	201.08	0.0	0.0	0.0
R' 2015 Rap1, ppMEK-ERK & ppERK-pASE diff.	175.419	0.0	0.0	0.0
R' 2015 ppMEK-ERK & ppERK-pASE diff.	172.626	0.0	0.0	0.0

B.2 Supplemental Figures

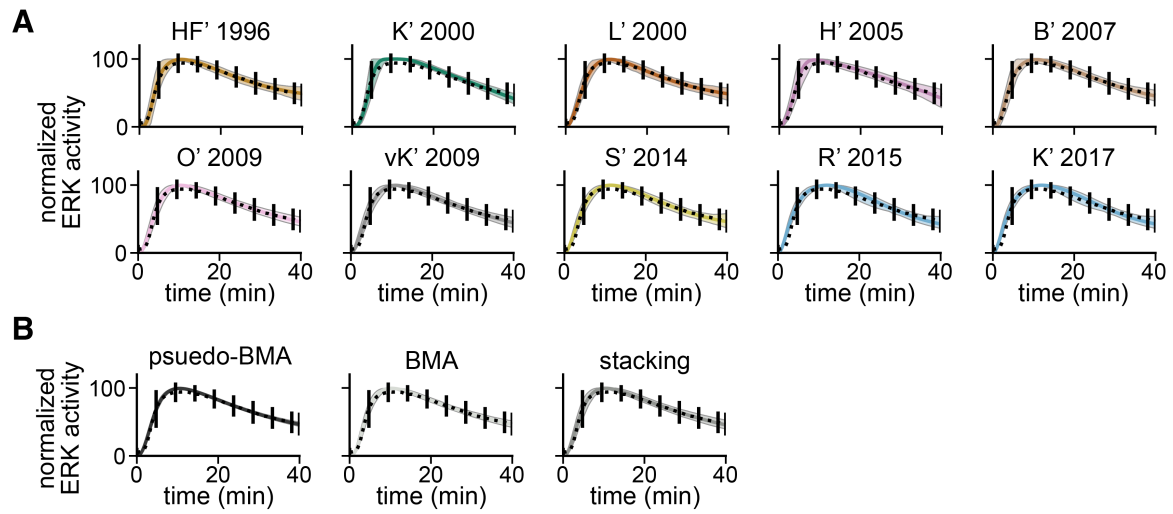


Figure B.2: (A) Posterior densities of cytoplasmic ERK activity trajectories for all models. (B)–(D) MMI predictions of cytoplasmic ERK activity using pseudo-BMA, stacking, and BMA, respectively.

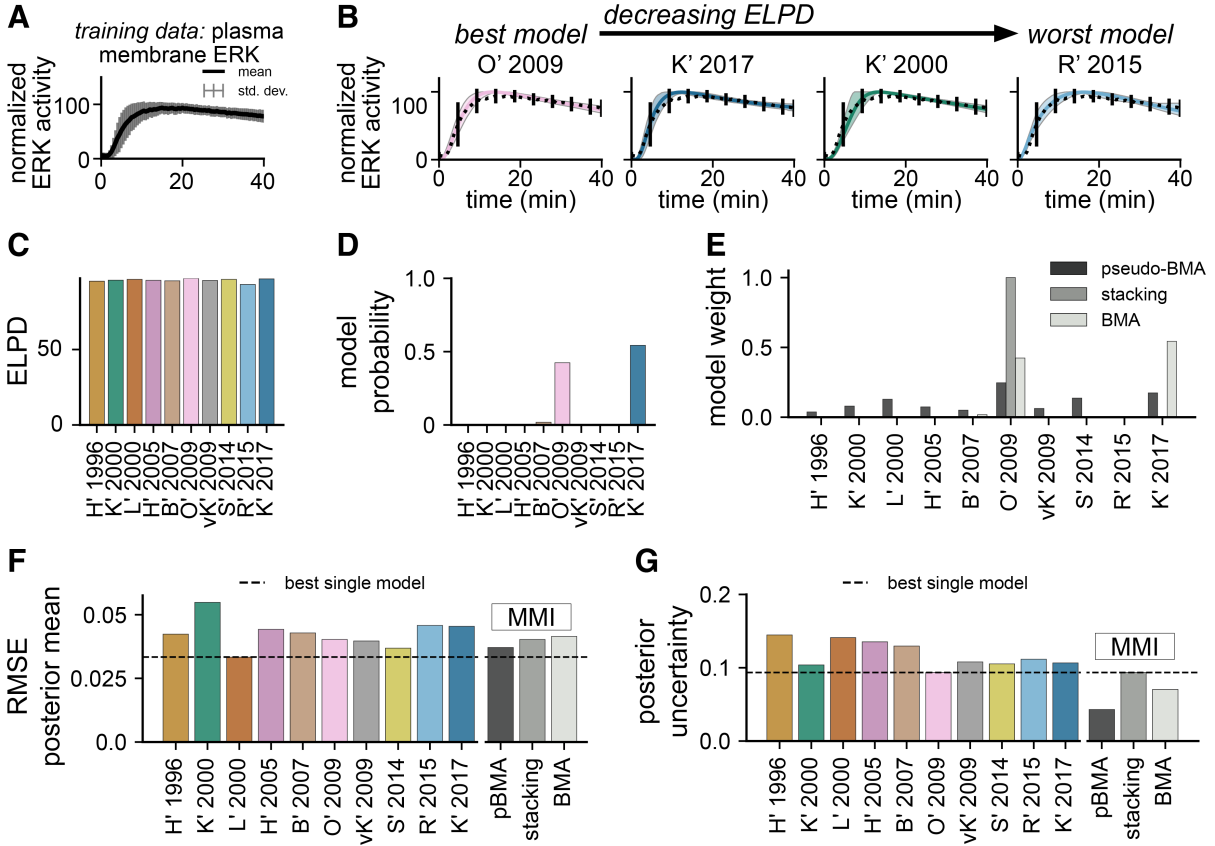


Figure B.3: (A) Normalized experimental measurements of plasma membrane ERK activity. Mean (black trace) and standard deviation (grey bars) show statistics computed across all single-cell trajectories. Data originally presented in figure 1 of [223]. (B) Posterior densities of plasma membrane ERK activity trajectories for four out of nine models were ordered by decreasing ELPD. (C) ELPD values for all models. (D) Model probabilities for all models. (E) MMI model weights for all models using pseudo-BMA, BMA, and stacking. (F) RMSE of the posterior mean plasma membrane ERK activity predictions for each model and the multimodel predictions. The dashed black line shows the lowest RMSE of any single model. (G) Posterior uncertainty measured by the mean 95% credible interval width in time. The dashed black line shows the lowest uncertainty of any single model.

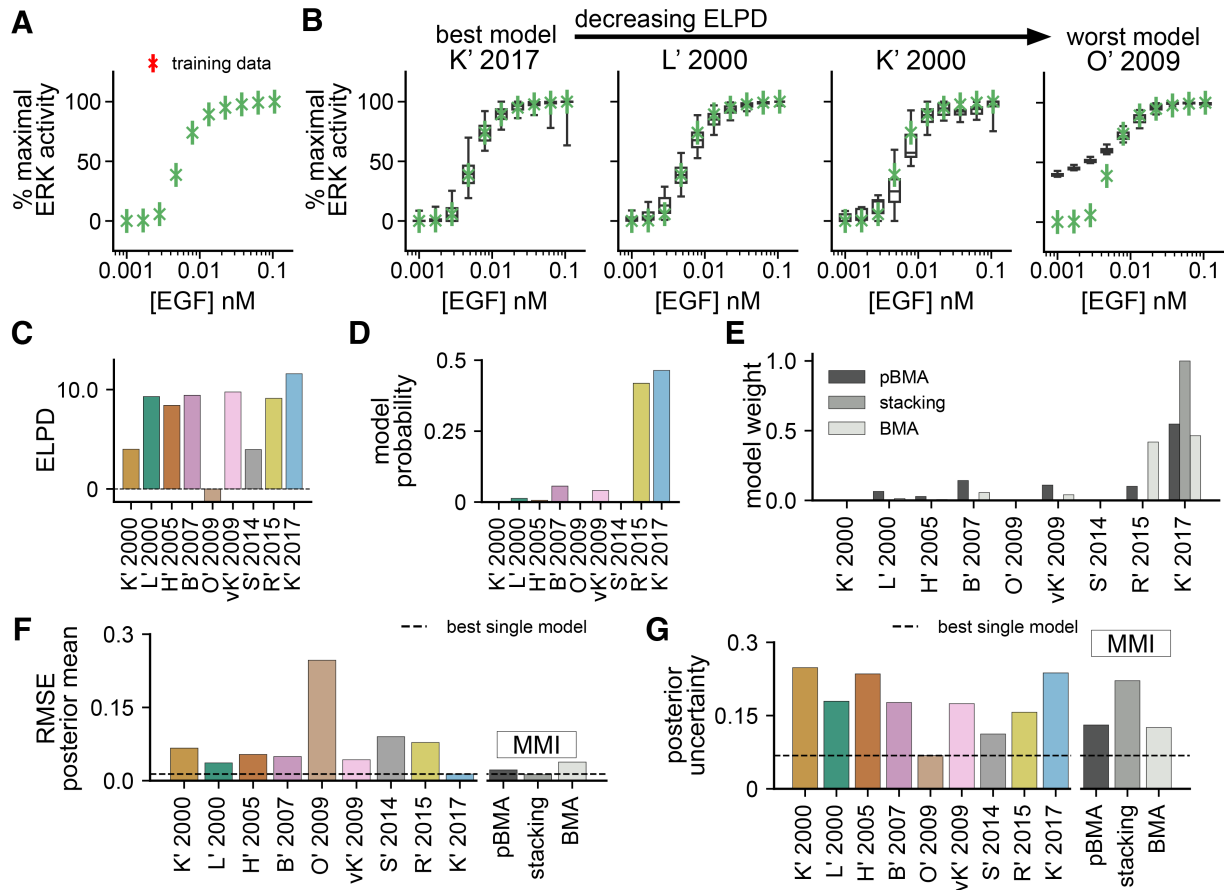


Figure B.4: (A) Synthetic EGF-ERK dose-response curve. Markers show the mean and error bars show plus or minus one standard deviation. (B) Posterior densities of the EGF-ERK dose-response for four out of nine models ordered by decreasing ELPD (additional models not shown). Boxplots show the mean and interquartile range, and whiskers show the 95% credible interval. Data from A are reproduced in each panel. (C) ELPD values for all models. (D) Model probabilities for all models. (E) MMI model weights for all models using pseudo-BMA, BMA, and stacking. (F) RMSE of the posterior-predictive mean dose-response prediction for each model and the multimodel predictions. The dashed black line shows the lowest RMSE of any single model. (G) Posterior uncertainty is measured by the mean 95% credible interval width over all EGF levels. The dashed black line shows the lowest uncertainty of any single model.

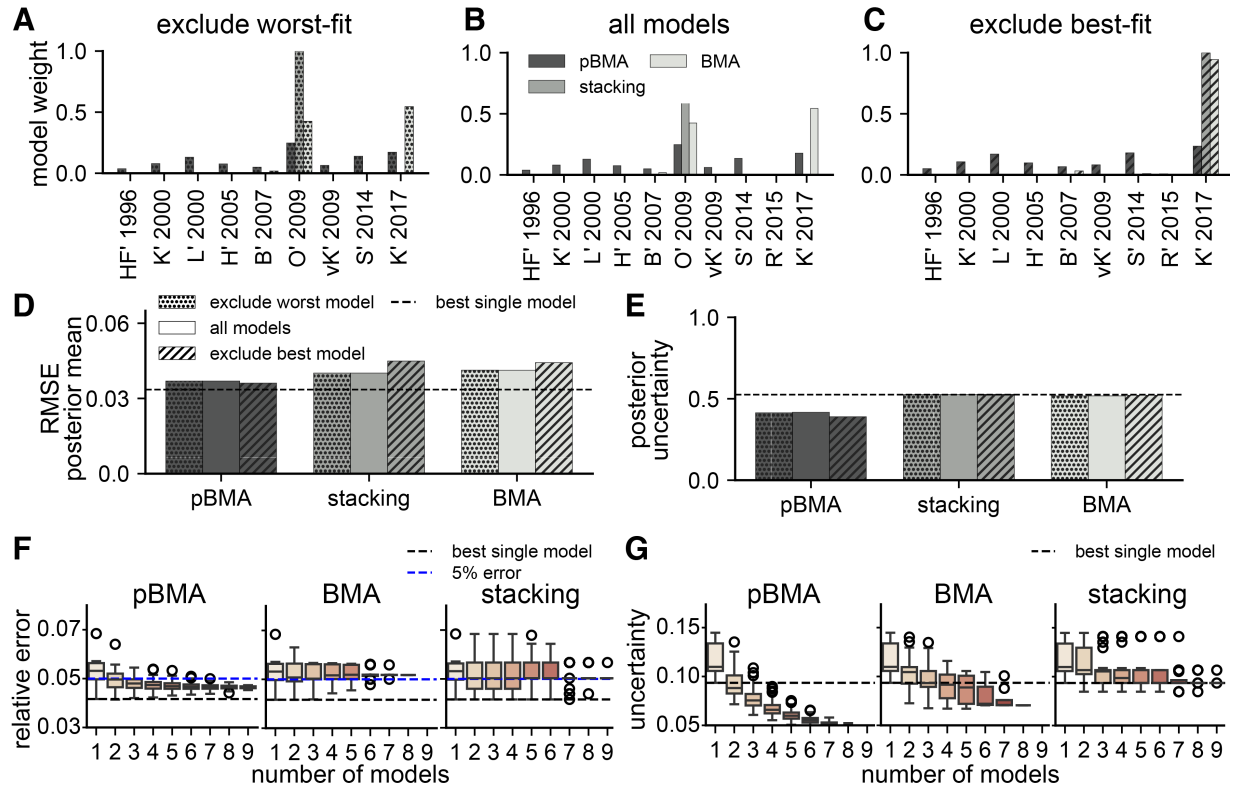


Figure B.5: (A)–(C) Weights assigned to models in three model sets: (A) excluding the worst-fit model, R' 2015, (B) all models, (C) excluding the best-fit, O' 2009. (D) RMSE of the posterior mean cytoplasmic ERK activity prediction for each MMI method and model set. Dotted patterning corresponds to (A), no patterning to (B), and dashed patterning to (C). The dashed horizontal line is the RMSE of the model with the lowest average error, HF' 1996. (E) Posterior uncertainty is measured by the mean of the 95% credible interval taken over all EGF levels. (F) Relative error of the posterior mean for MMI predictions with increasing model set size. All possible combinations of models were tested at each size. The dashed blue line shows 5% relative error (0.05), and the dashed black line shows the lowest error of any single model, HF' 1996. (G) Average posterior uncertainty of ERK response for MMI predictions with increasing numbers of models. The dashed black line shows the uncertainty of the best model. (F),(G) Boxplots show the mean and interquartile range, and whiskers show the 95% credible interval. Open circles show outliers.

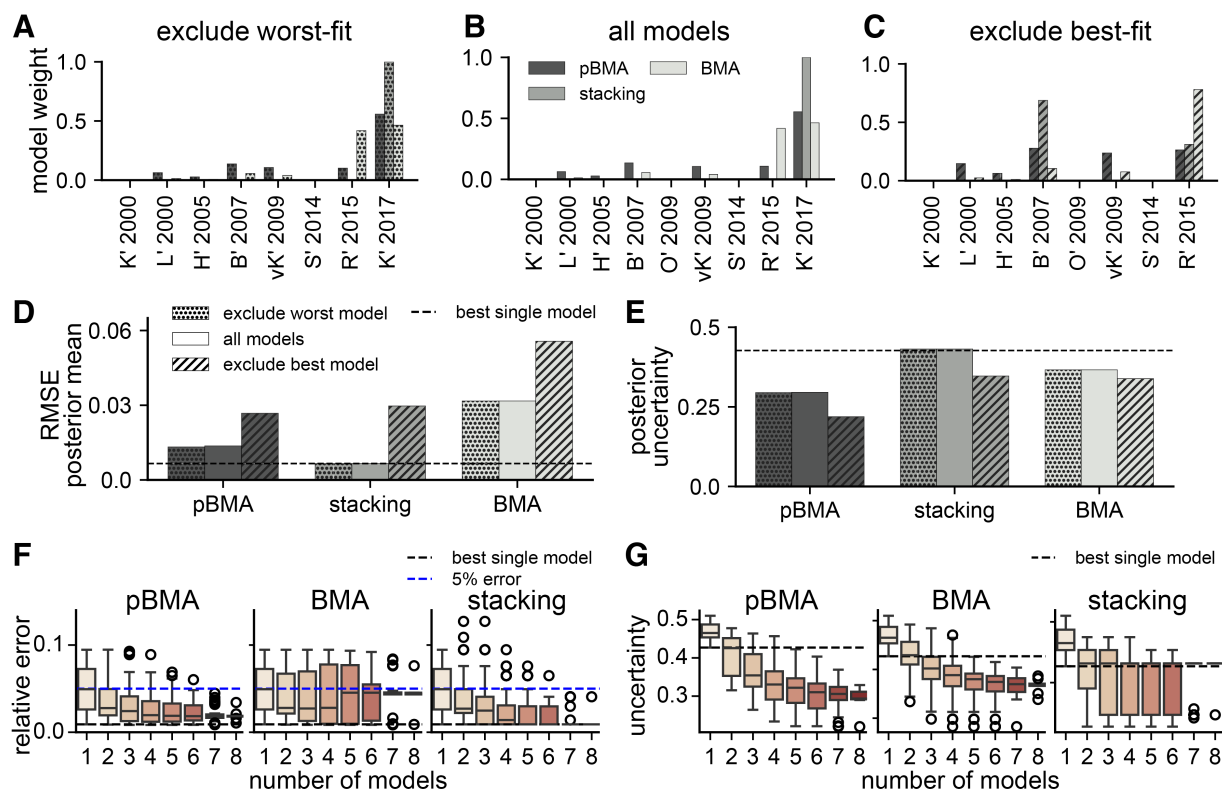


Figure B.6: (A)–(C) Weights assigned to models in three model sets: (A) excluding the worst-fit model, O' 2009, (B) all models, (C) excluding the best-fit, K' 2017. (D) RMSE of the posterior mean dose-response prediction for each MMI method and model set. Dotted patterning corresponds to (A), no patterning to (B), and dashed patterning to (C). The dashed horizontal line is the RMSE of the best-fit model, K' 2017. (E) Posterior uncertainty is measured by the mean of the 95% credible interval taken over all EGF levels. (F) Relative error of the posterior mean for MMI predictions with increasing model set size. All possible combinations of models were tested at each size. The dashed blue line shows 5% relative error (0.05), and the dashed black line shows the lowest error of any single model, K' 2017. (G) Average posterior uncertainty of ERK response for MMI predictions with increasing numbers of models. The dashed black line shows the uncertainty of the best model, R' 2015. (F),(G) Boxplots show the mean and interquartile range, and whiskers show the 95% credible interval. Open circles show outliers.

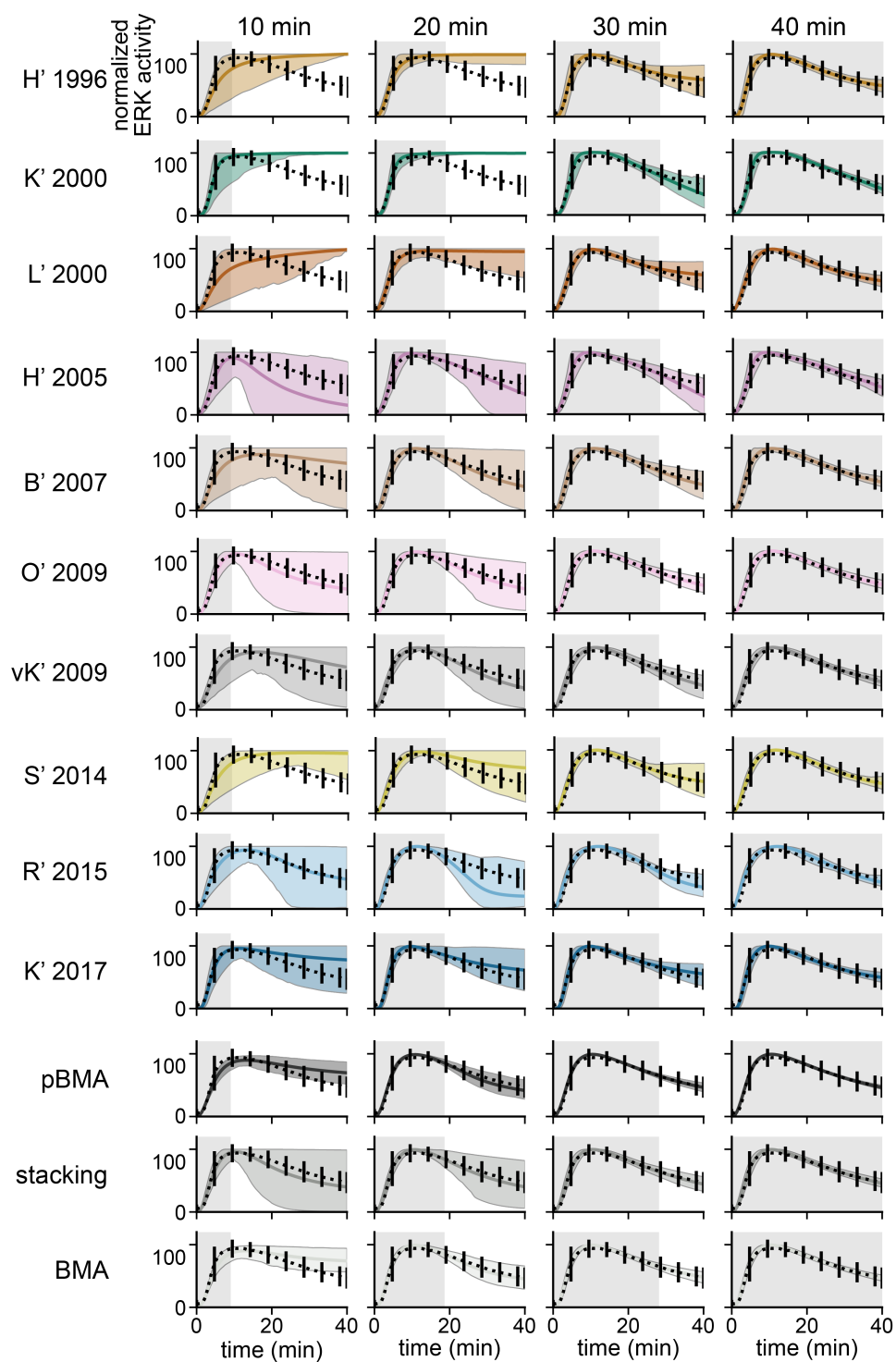


Figure B.7: Black dashed line shows the data with error bars indicating the standard deviation. Solid colored line shows the posterior predictive mean trajectory. Shaded band shows the 95% credible interval of the posterior predictive density.

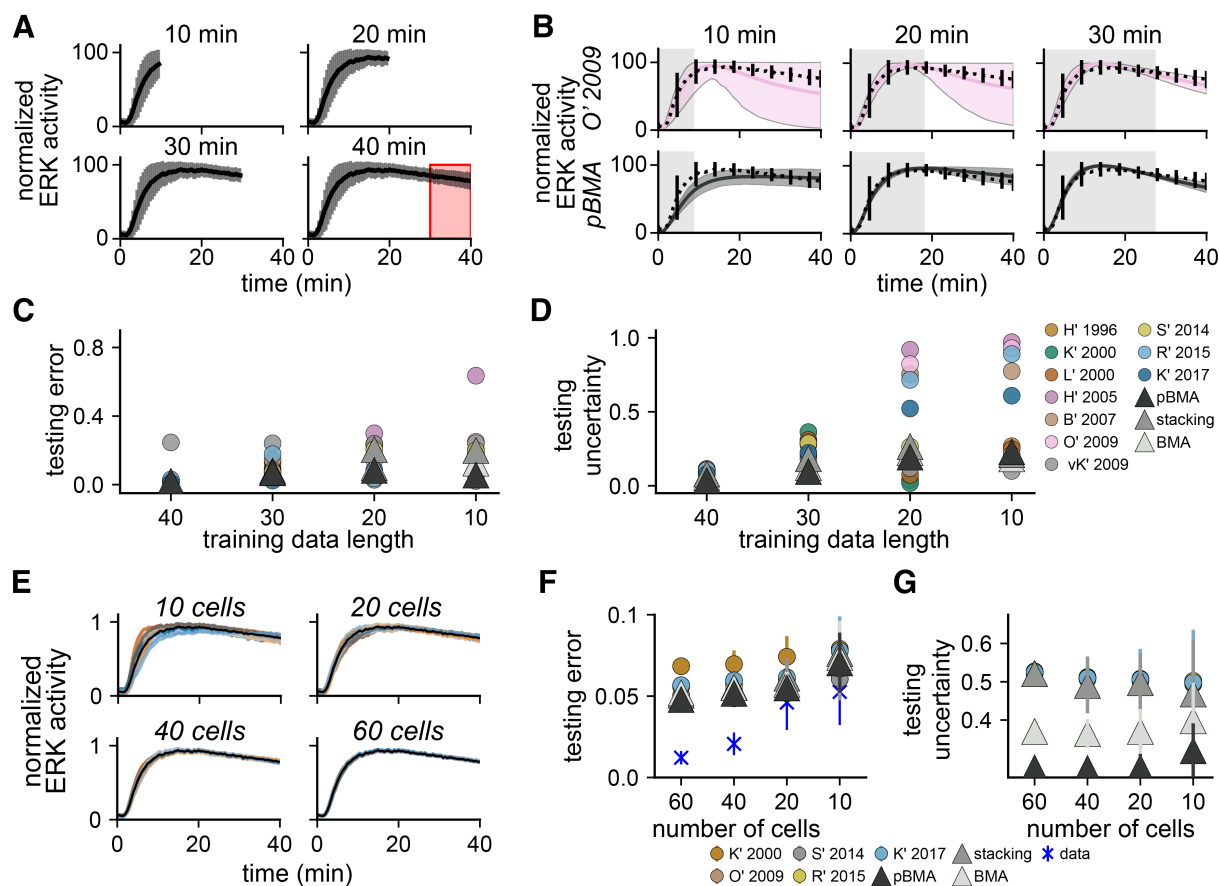


Figure B.8: Effects of decreasing data length. (A) Shorter training data was constructed by truncating the original 40-minute plasma membrane ERK trajectories at the 10-, 20-, and 30-minute time points (mean, black trace; and standard deviation, grey bars). Predictive performance was assessed by computing errors and uncertainties in the final 10 minutes (red box). (B) Posterior predictions from decreased training data for the best model O' 2009 (highest ELPD across all training datasets) and MMI with pseudo-BMA. Predictions for additional models and MMI methods are shown in Supplemental Figure B.7. (C) Predictive error (relative error) for the final 10 minutes ($t = 30 \rightarrow t = 40$ min) of plasma membrane ERK activity. (D) Predictive uncertainty (average width of the 95 % credible interval) for the final 10 minutes of plasma membrane ERK activity. (E) Lower-quality training data was generated by averaging over random subsets of 10, 20, 60, and 60 imaged cells. The black trace shows an original average of 76 cells. Colored traces show averages of 40 replicate random subsets. (F) The predictive error (relative error) of plasma membrane ERK activity with lower quality data compared to average activity trajectory using all cells. (G) Predictive uncertainty of plasma membrane ERK activity with lower quality data. (F),(G) Filled circles indicate average error of 40 replicates for individual models and triangles for MMI predictions. Error bars show the standard deviation over replicates. Blue markers show the error of the raw training data at each subset size compared to the original full-data mean.

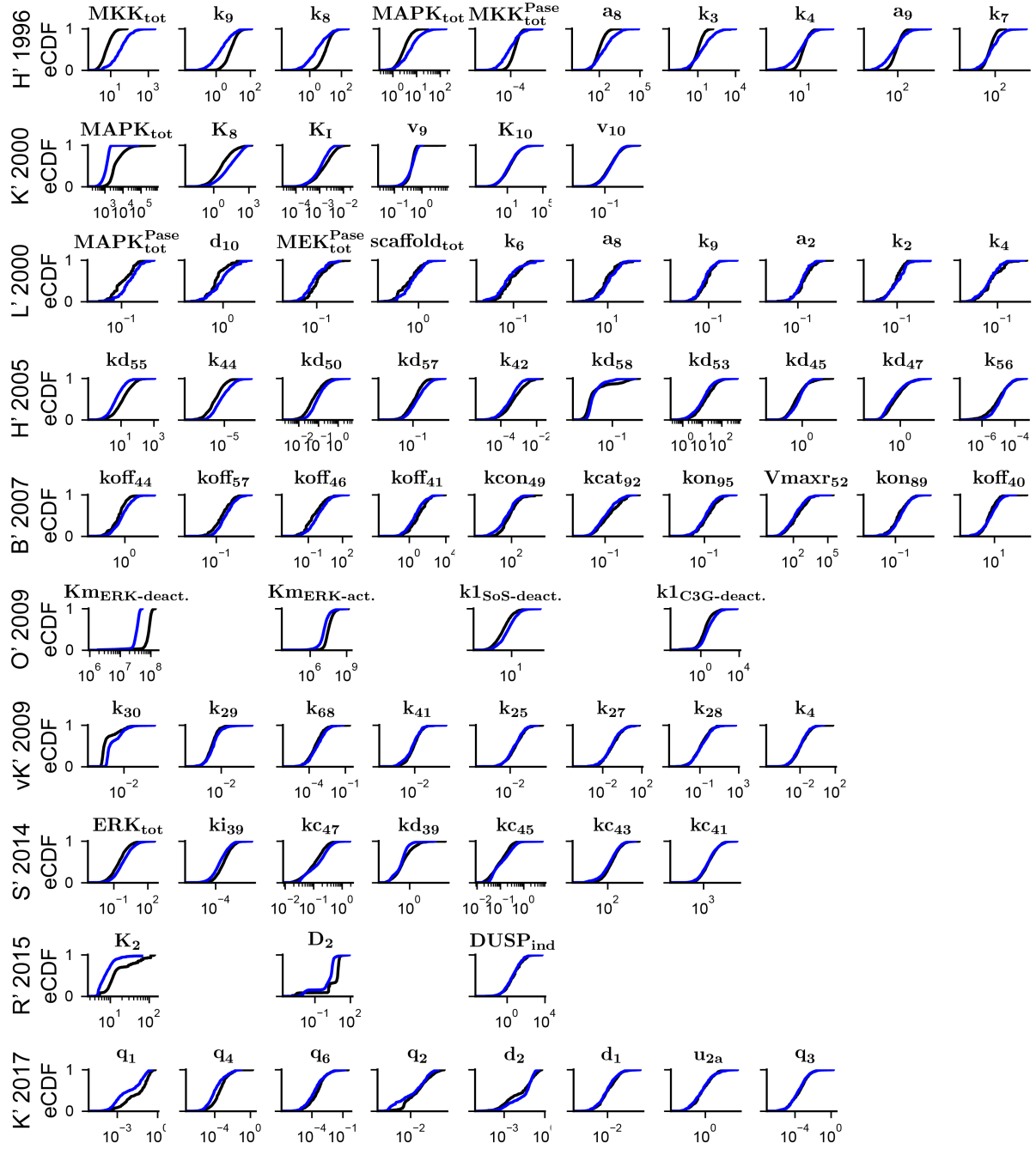


Figure B.9: Figure shows estimated empirical cumulative density functions (eCDF) for the model parameters that showed the greatest variation between locations. All densities are statistically significantly different between locations ($p < 0.05$ by the Mann-Whitney U-test with a two-sided hypothesis). Blue indicates the cytoplasm, and black indicates the plasma membrane. Supplemental Table B.3 lists the sample sizes for each model.

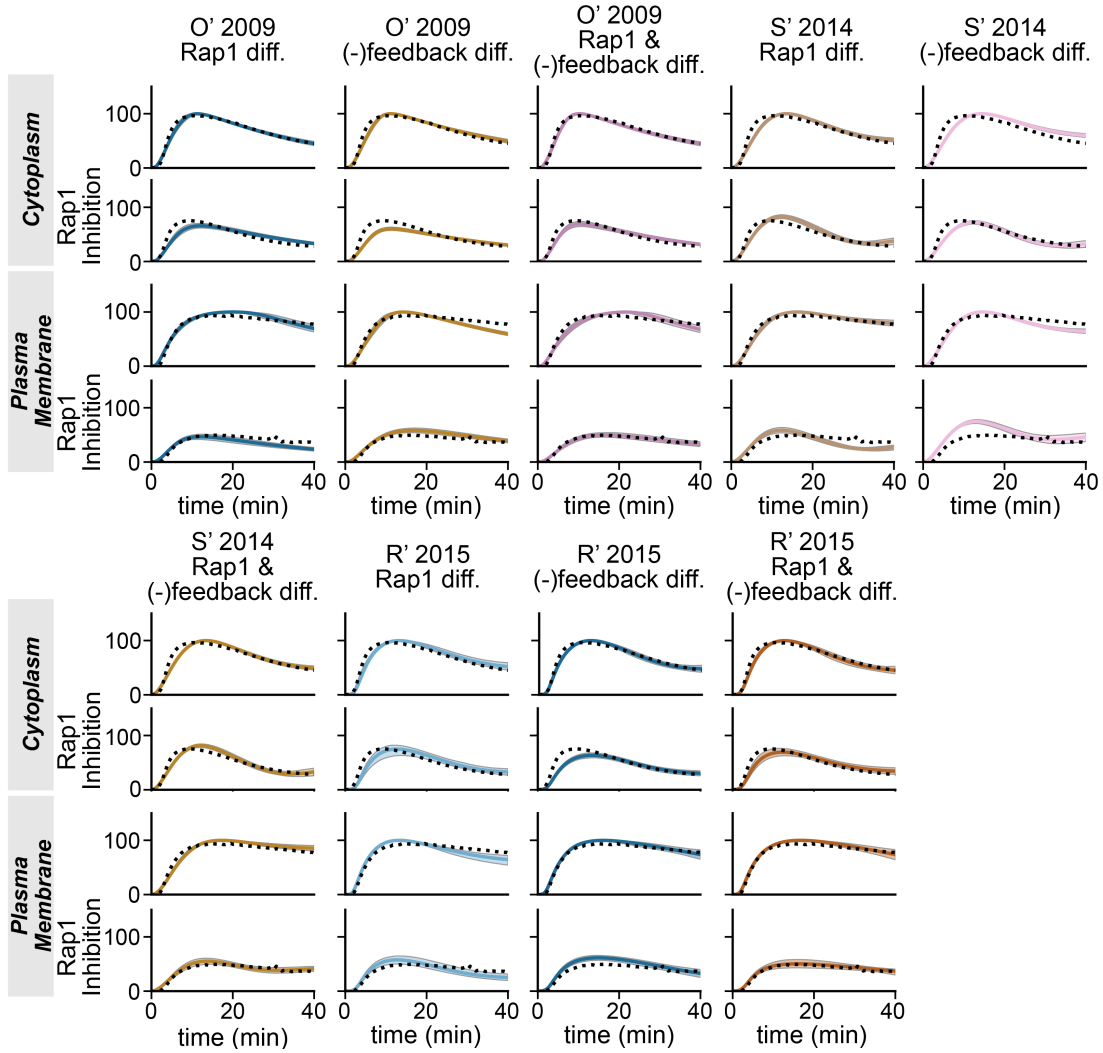


Figure B.10: Black dashed line shows the data with error bars indicating the standard deviation. Solid colored line shows the posterior predictive mean trajectory. Shaded band shows the 95% credible interval of the posterior predictive density. Corresponds to Figure 3.8.

B.3 Supplemental Files

We provide an Excel file that details the nominal parameters and initial conditions for all 10 ERK signaling models used in this work. The file is available for download from our GitHub repository at this [link](#).

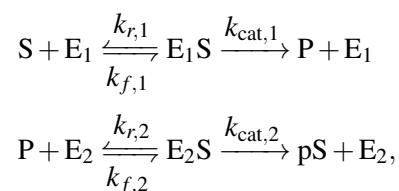
Appendix C

Supplemental Materials for Chapter 4

C.1 Supplemental Text

C.1.1 Two-reaction models

To further investigate how the choice of kinetic formulation affects parameter identifiability, we examined models of a two-reaction loop that was inspired by a phosphorylation-dephosphorylation loop. The models represented the reactions



where S is the substrate, E_1 and E_2 are the enzymes, and P is the product. The product of the first reaction is the substrate of the second P and the product of the second reaction is the original substrate S . To investigate the effects of the kinetic formulation, we constructed nine sets of models that represented the two-reaction systems. The models differed in the kinetic formulation used to represent either reaction, specifically mass

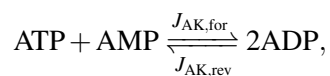
action, Michaelis-Menten, or Hill-type. For example, one of the models employed mass action kinetics for both reactions, while another used mass action for the first reaction and the Michaelis-Menten equation for the second. Supplemental Table C.2 summarizes the models and the identifiability results.

C.1.2 AMPK model details

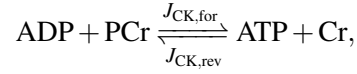
In this section, we provide additional rationale and derivations for the six AMPK signaling models. Unless otherwise noted, all relevant assumptions are outlined in the manuscript.

Module 1: Cellular metabolism and energy stress

We adapted the simplified metabolism models from Coccimiglio et al. [49] and Lueng et al. [51]. Our metabolism model includes phenomenological representations of oxidative phosphorylation, glycolysis, ATP hydrolysis, adenylate kinase, and creatine kinase. First, we represented ATP generation by oxidative phosphorylation with a first-order irreversible reaction from ADP to ATP, which we modeled with the Hill-type equation originally derived in [388]. Second, we represented glycolysis with an irreversible first-order reaction from ADP to ATP, which we modeled with mass action kinetics. Third, we represented net ATP consumption due to hydrolysis with a single hydrolysis reaction from ATP to ADP, which we modeled with mass action kinetics. Fourth, we included both the adenylate and the creatine kinase reactions to ensure the correct buffering of adenine ratios [389]. The adenylate kinase reaction is



which we modeled with a bi-bi reaction mechanism originally derived in [389]. Finally, the reversible creatine kinase reaction is



which we modeled with a Cleland bi-bi reaction that was originally derived in [389]. Table C.3 lists the fluxes, J_{OxPhos} , J_{Gly} , etc., for these reactions.

The associated differential equations for ATP, ADP, AMP, and PCr, are defined as,

$$d[\text{AMP}]/dt = -J_{\text{AK}} \quad (\text{C.1a})$$

$$d[\text{ADP}]/dt = -J_{\text{glycolysis}} - J_{\text{OxPhos}} + J_{\text{hydro}} + J_{\text{CK}} + 2J_{\text{AK}} \quad (\text{C.1b})$$

$$d[\text{ATP}]/dt = J_{\text{glycolysis}} + J_{\text{OxPhos}} - J_{\text{hydro}} - J_{\text{CK}} - J_{\text{AK}} \quad (\text{C.1c})$$

$$d[\text{PCr}]/dt = J_{\text{CK}}. \quad (\text{C.1d})$$

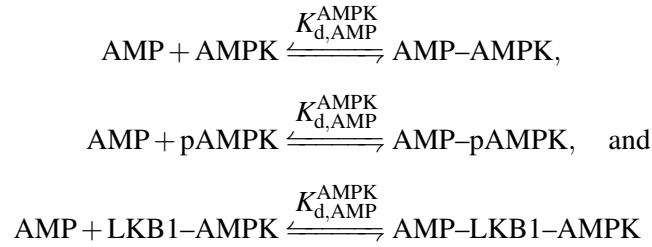
Throughout this work, we assume that the initial concentrations of AMP, ADP, and ATP are 7.03 mM, 1.11 mM, and 7.9×10^{-2} mM, respectively. Supplemental Figure C.1A shows transient AMP, ADP, and ATP concentrations from an arbitrary initial condition to the steady state taken as the initial condition in this work. The cellular metabolism model is the same across all models that we develop. Table C.4 provides values for all of the Module 1 parameters. All parameter values were taken from sources, except for the rate of ATP production by glycolysis. We derived the glycolysis rate $k_{\text{glycolysis}}$ roughly on experimental measurements of glycolytic flux from Salti et al. [390] Figure 2. In MEF cells, those authors measured a flux of 9.16 pmol/ATP/ 10^3 cells, which we converted to mM/s by assuming a cellular volume of 4000 μm^3 , which gives approximately 22 mM/s. We divided by twice initial concentration of ATP (7.03 mM)

and rounded down to yield $k_{\text{gly}} = 1.5 \text{ 1/s}$.

Module 2: Adenine nucleotide binding and AMPK phosphorylation

The models include adenine nucleotide binding to AMPK, phosphorylation of AMPK, and dephosphorylation of AMPK. In this section, we describe additional model assumptions and outline the nonessential and essential activation/inhibition mechanisms used in the models.

Adenine nucleotide binding We assumed that adenine nucleotide binding kinetics do not vary depending on the phosphorylation or binding state of AMPK. For example, in the models, the reactions

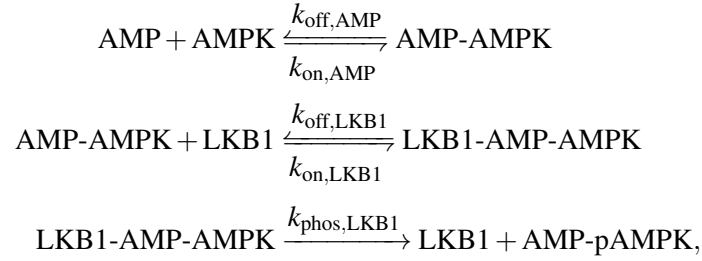


all occur with the same kinetics. We used the binding coefficients from [293] as initial values when constructing prior densities for these parameters.

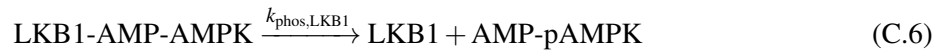
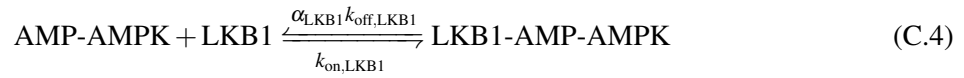
AMPK phosphorylation by upstream kinases LKB1 and CaMKK2 We assumed that AMPK phosphorylation by LKB1 and CaMKK2 occurs as a standard enzyme-mediated reaction [64]. For example, phosphorylation by LKB1 would be described as



In order to capture the substrate-mediated effects of AXP binding, we assumed that AMP and ADP act as either essential activators (Models 1 and 2) or nonessential activators (Models 4–6) of upstream kinases. Essential activation assumes that the activator must be bound to the substrate for the reaction to proceed [64]. For example,



assumes that AMP is the essential activator of LKB1-induced phosphorylation. Alternatively, nonessential activation assumes that the activator increases the reaction rate or decreases the K_d between the enzyme and substrate [64]. For example, in the reaction



the activator AMP decreases the K_d by a factor of $\alpha_{\text{LKB1}} \leq 1$.

AMPK dephosphorylation Similar to phosphorylation, we assumed that dephosphorylation of phosphorylated pAMPK occurs as a standard enzyme-mediated reaction. In Models 1 and 2, we assumed

that AMP and ADP act as essential inhibitors of the dephosphorylation reaction, where AMP-pAMPK and ADP-pAMPK are not dephosphorylated. In Models 3–6, we assumed that AMP and ADP acts as nonessential inhibitors of the dephosphorylation reaction, either increasing the K_d (Models 3 and 4) or decreasing the reaction rate (Models 5 and 6).

C.1.3 Predictive prior elicitation and prior refinement

We encountered two key challenges in constructing prior densities for model parameters: (i) choosing priors that yield reasonable predictions when little prior information is available and (ii) assigning values to nonidentifiable parameters that we fix before estimation. Here, we took an approach that combined predictive prior elicitation with iterative updating of fixed parameters. Predictive prior elicitation is a qualitative approach that allows us to adjust the parameters of the prior densities such that the prior predictive distribution appears reasonable [319]. This allows us to refine our prior beliefs without utilizing any data. Briefly, the prior predictive distribution shows the distribution of data generated by the model given assumed prior densities [309]. In the remainder of this section, we outline the process used to refine prior densities and select values for non-identifiable parameters, using Model 2 as an example.

We began by using direct prior elicitation to choose default priors for uncertain parameters by specifying a range of plausible values (see Section 4.2) [319]. Using these priors for Model 2, we observed that the prior predictive density was unreasonable, because all of the predicted simulations were constantly zero (Supplemental Figure C.2A). We hypothesized that the values of the nonidentifiable parameters, which we fixed to nominal values, could account for the observed behavior. To address this, we varied each of the five nonidentifiable parameters independently over several orders of magnitude and simulated the corresponding prior predictive densities. Supplemental Figure C.2B shows the cumulative density functions of the final value of the predicted sensor response. We found that only k_{AMPK} affected the

predictions, so we updated the nominal value for that parameter to 0.192 because that value yielded more reasonable predictions. We repeated this process for the remaining parameters and found that we only needed one additional iteration of updates, because only k_{CaMKK} affected predictions (Supplemental Figure C.2C). Given that we were satisfied with the updated nominal values for nonidentifiable parameters, we next chose to refine the priors using predictive prior elicitation. Here, we simulated the prior density and iteratively updated the hyperparameters of the priors to yield more reasonable predictions. We refer the reader to the tutorial in the Preliz Python package, at the following link https://preliz.readthedocs.io/en/latest/examples/gallery/predictive_explorer.html for more details. Supplemental Figure C.2D shows the resulting prior predictive distribution and Supplemental Figure C.2E shows the densities of the final value in the simulation and the time to half max. We repeated this process for all six AMPK models (Supplemental Figure C.3).

C.2 Supplemental Figures

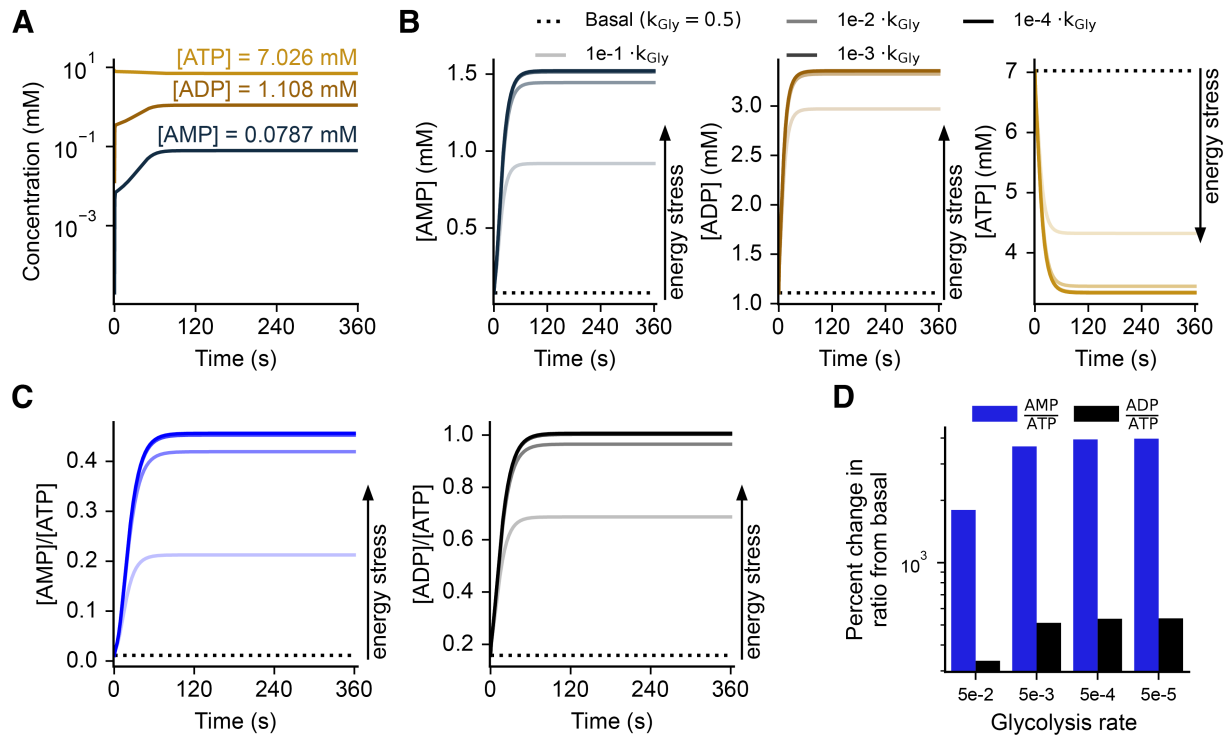


Figure C.1: (A) AMP, ADP, and ATP transients from arbitrary initial conditions to steady state values with basal ATP production. (B) Effect of energy stress due to decreasing the rate of glycolytic ATP production on AMP, ADP, and ATP concentrations over time. The rate of glycolysis is changed from the basal value at $t = 0$. (C) Effect of energy stress on transient AMP/ATP (blue) and ATP/ADP (black) ratios. Stimuli are the same as in B. (D) Percent change in AMP/ATP and ADP/ATP ratios from basal due to increased energy stress.

C.3 Supplemental Tables

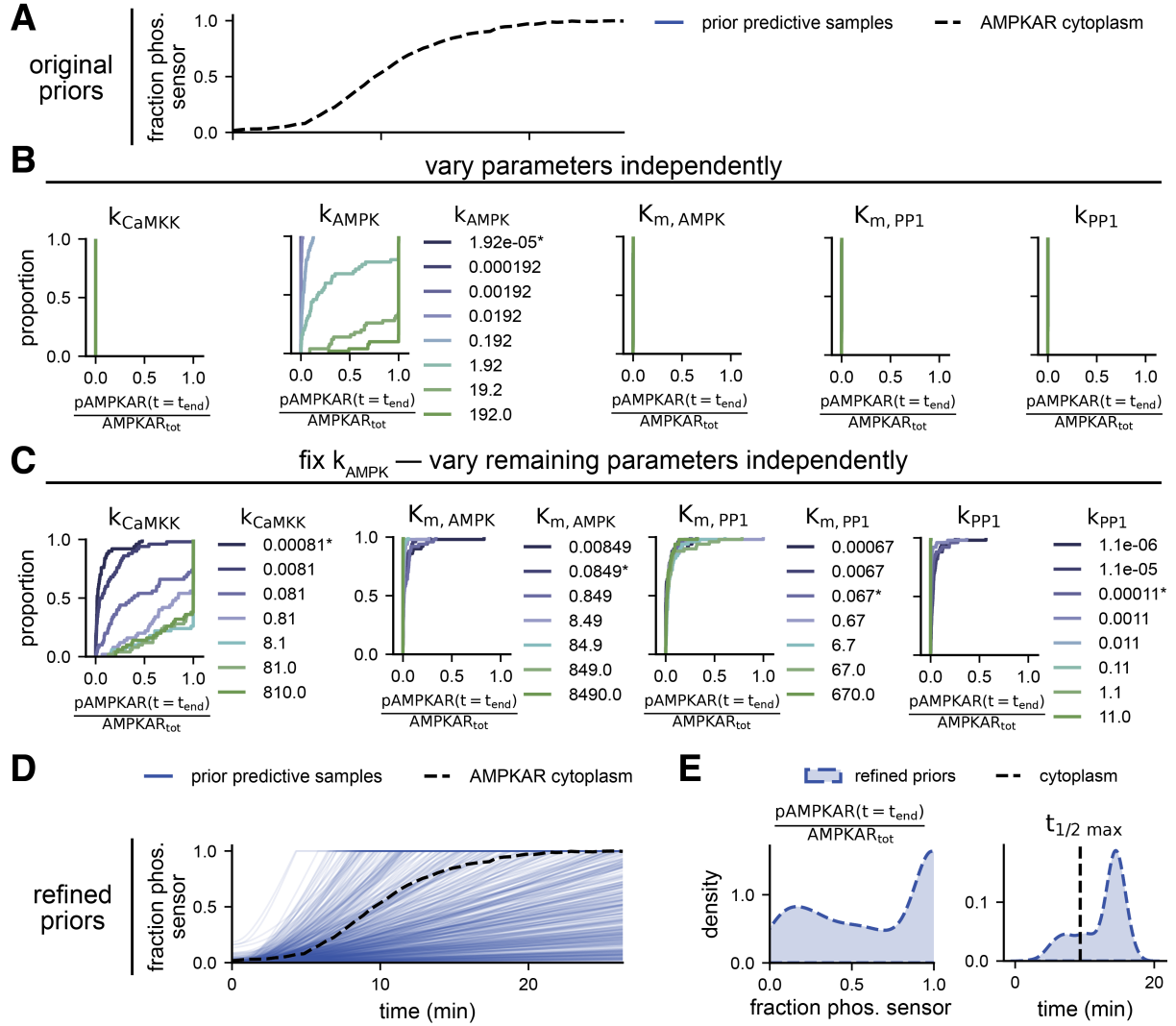


Figure C.2: (A) Prior predictive distribution of predictions for priors based-on parameter ranges before any refinement. All simulations are nearly zero for all time. (B) Empirical cumulative density functions (CDF) for the final fraction of activated AMPKAR from prior predictive samples with single-parameter variations of nonidentifiable parameters. Only k_{AMPK} affects the strength of AMPKAR response. (C) Empirical CDFs for the final fraction of activated AMPKAR from prior predictive samples with single-parameter variations of nonidentifiable parameters with k_{AMPK} fixed to 0.192. Only k_{CaMKK} affects the strength of AMPKAR response. (D) Prior predictive distribution of predictions for refined priors and updated values for k_{AMPK} and k_{CaMKK} . (E) Estimated probability densities of key quantities after prior refinement.

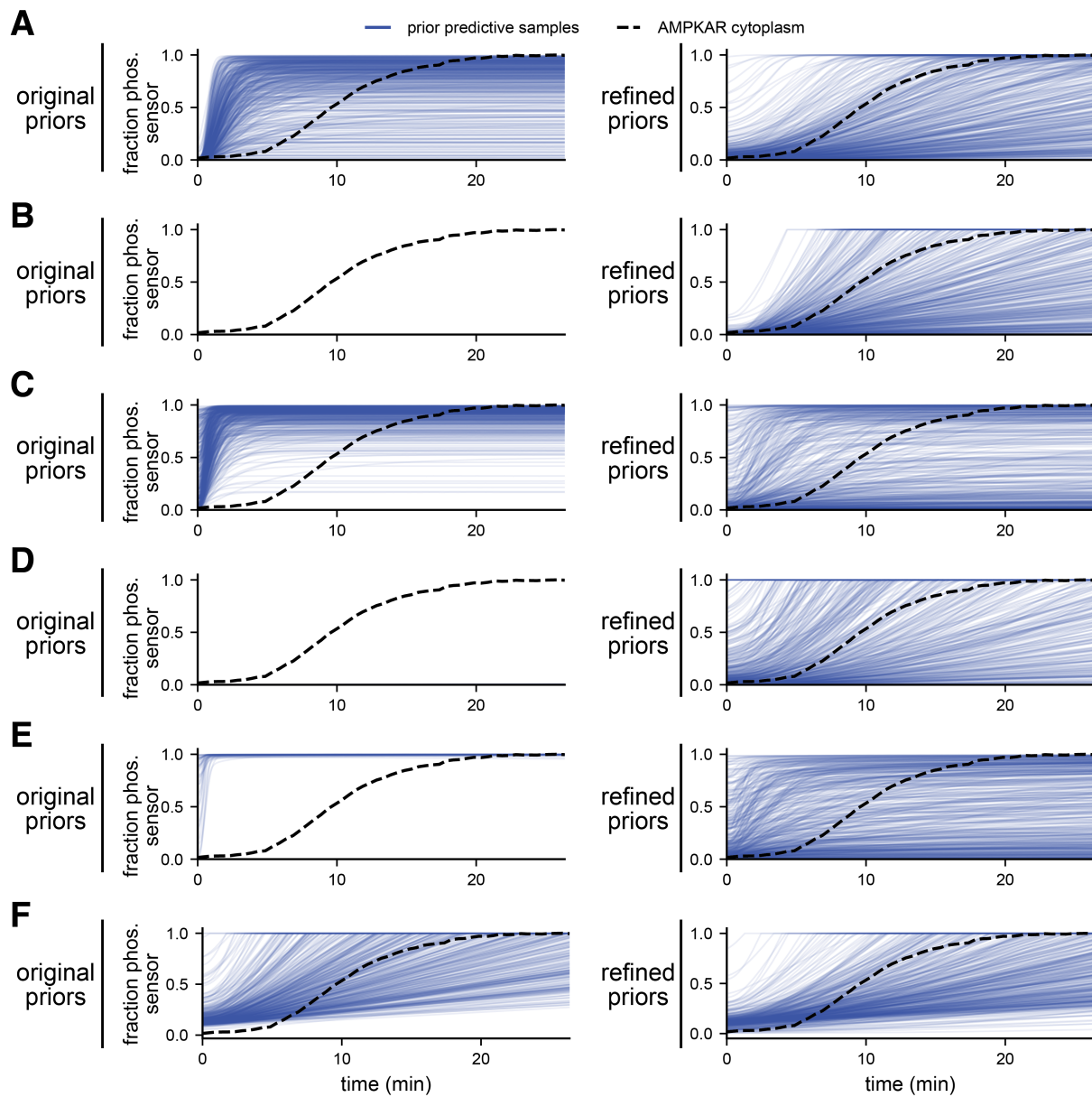


Figure C.3: Dashed black line shows the data mean. Transparent blue lines show simulations with 400 prior samples.

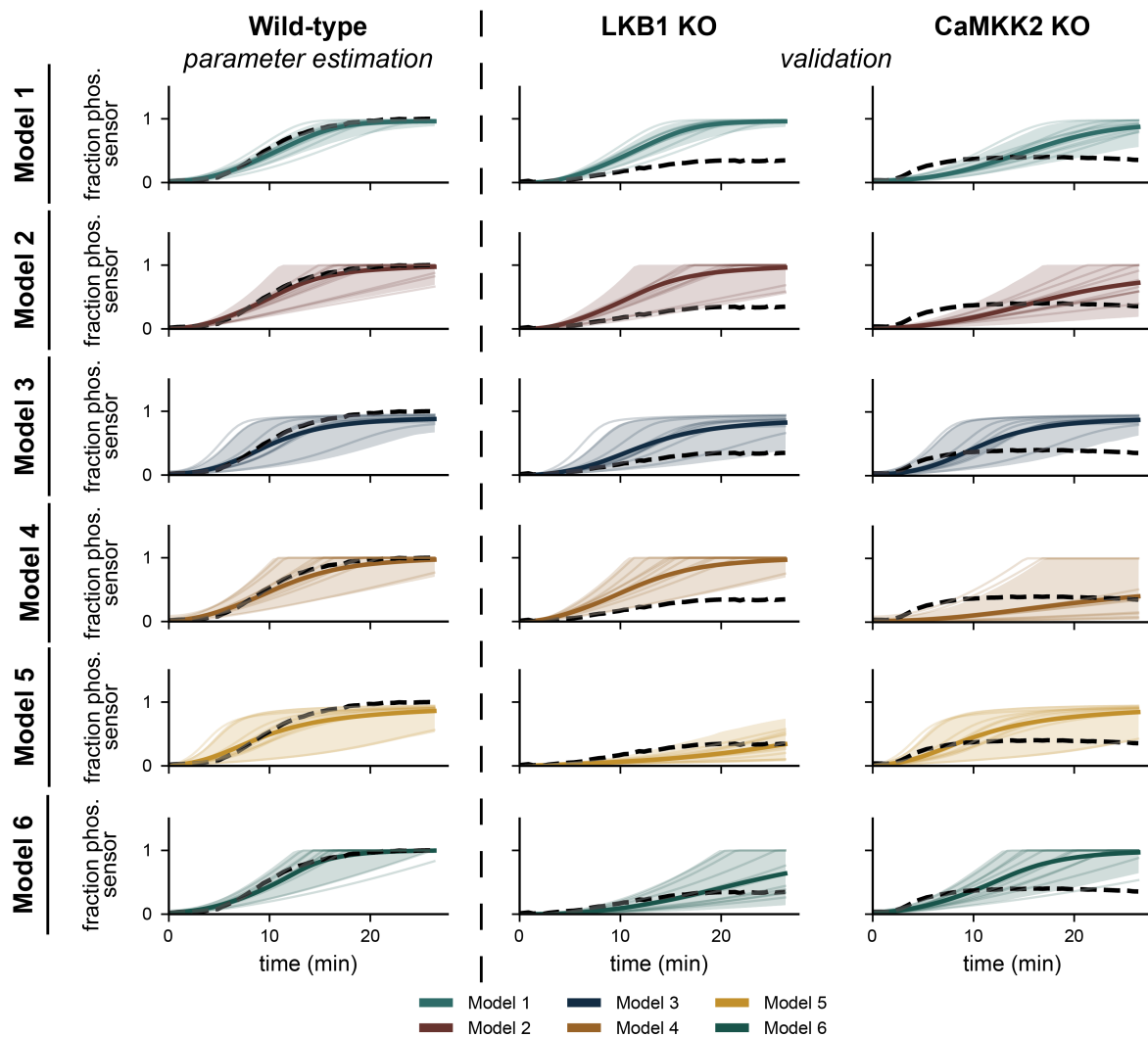


Figure C.4: Dashed black lines show the data mean. Solid colored lines show the posterior mean, the shaded band shows the 95% credible interval, and the transparent lines show 10 samples from the posterior density. LKB1 and CaMKK2 knockout conditions are forward predictions.

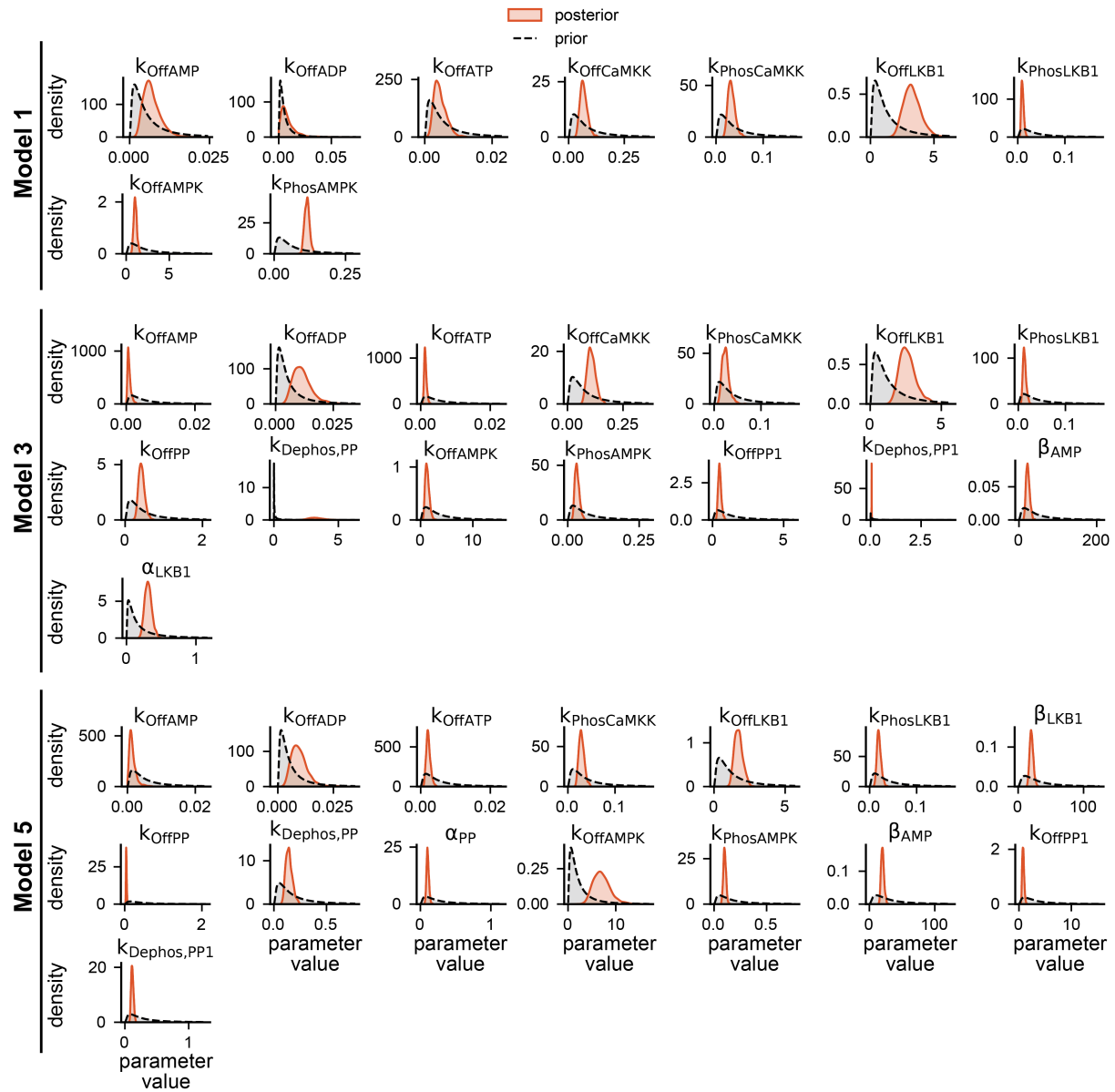


Figure C.5: Densities were estimated by fitting a kernel density estimator to 1000 posterior samples. Posterior was sampled using the Pathfinder variational inference algorithm [320].

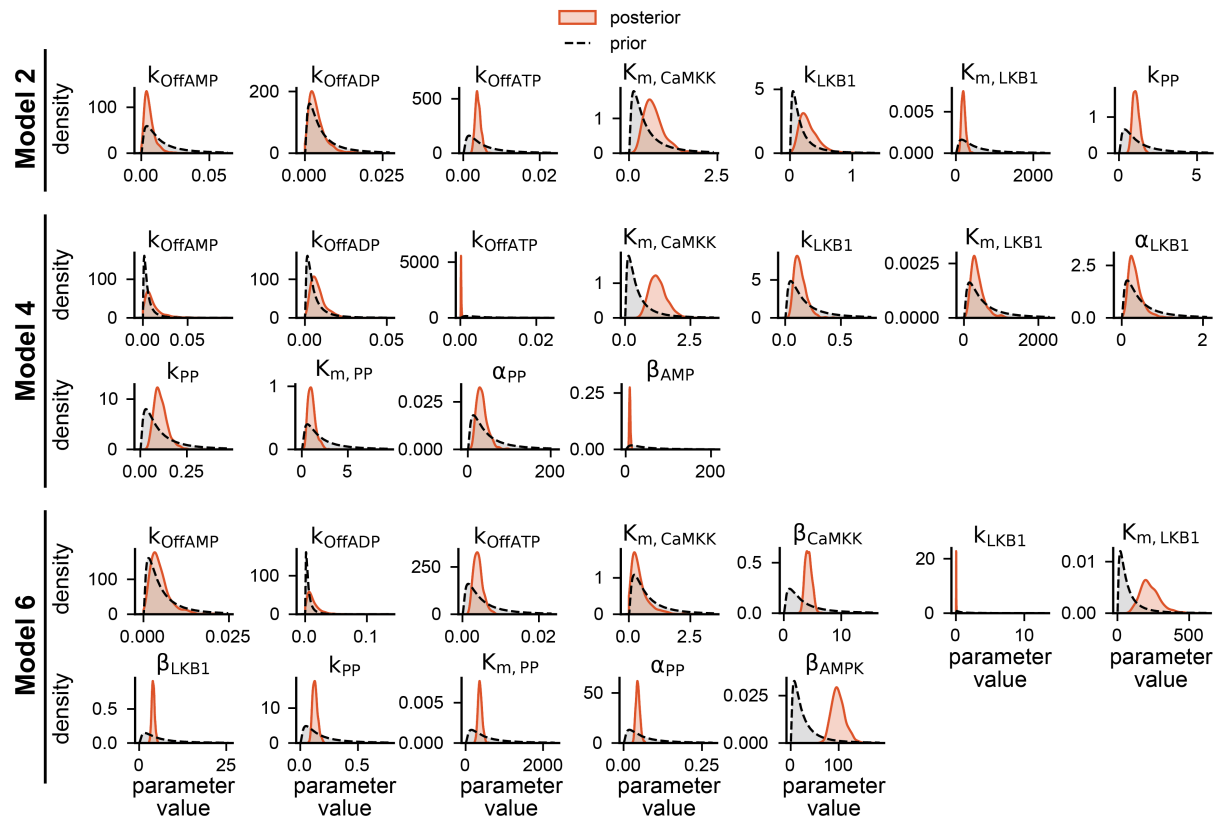


Figure C.6: Densities were estimated by fitting a kernel density estimator to 1000 posterior samples. Posterior was sampled using the Pathfinder variational inference algorithm [320].

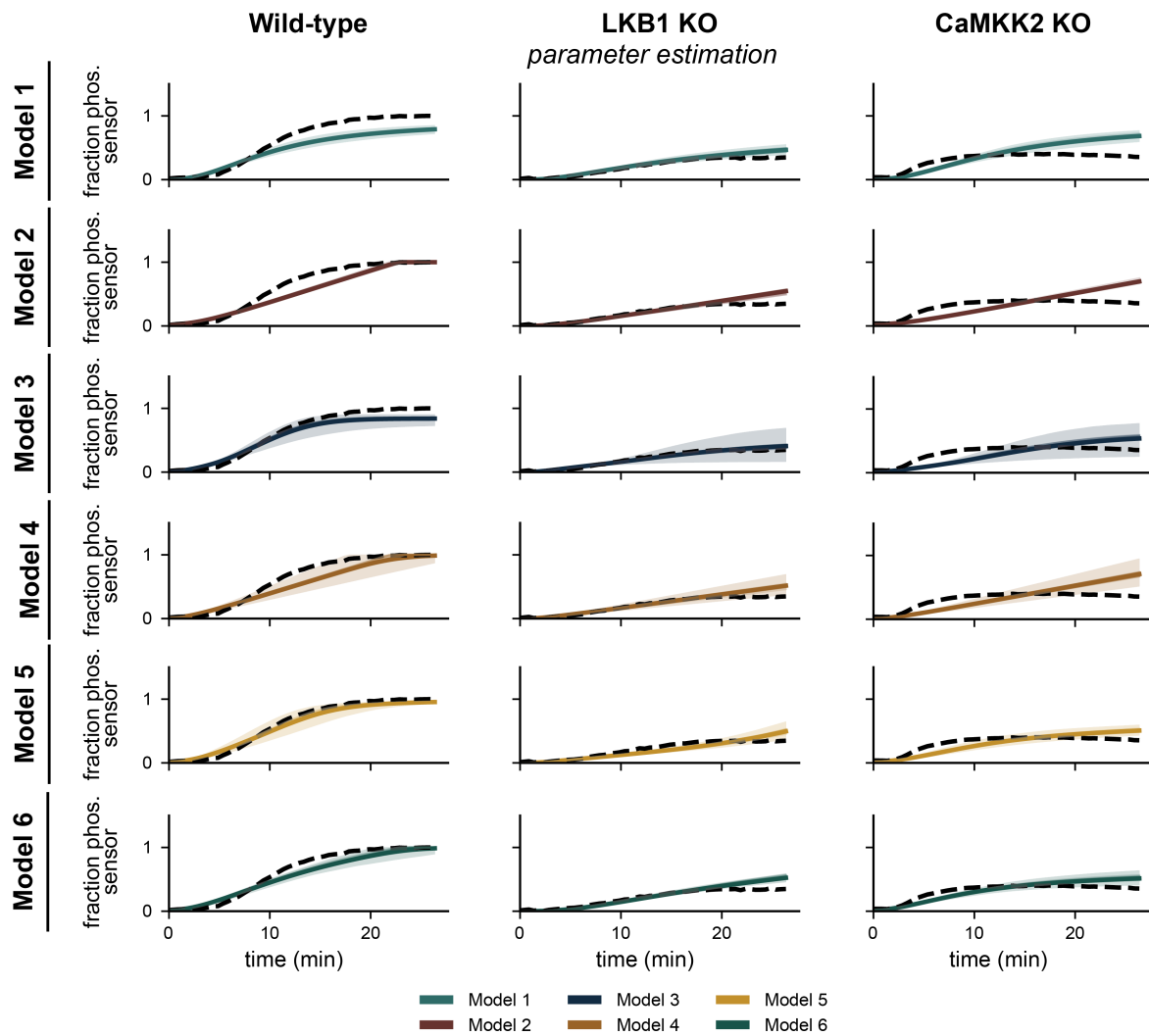


Figure C.7: Dashed black lines show the data mean. Solid colored lines show the posterior mean, the shaded band shows the 95% credible interval, and the transparent lines show 10 samples from the posterior density.

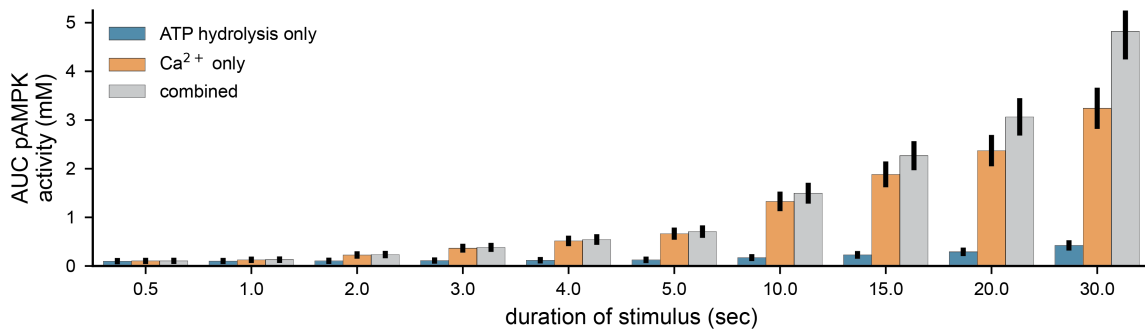


Figure C.8: Stimuli strengths were either $[\text{Ca}^{2+}] = 0.1 \mu\text{M}$ (Ca^{2+} only), ATP flux = 2.0 mM/s (ATP hydrolysis only), or both (combined).

Table C.1

Model	Parameter	Units	Prior
Mass action	k_f	1/(mM s)	Gamma ($\alpha = 3.14, \beta = 33.01$)
Mass action	k_r	1/s	Gamma ($\alpha = 3.14, \beta = 132.04$)
Mass action	k_{cat}	1/s	Gamma ($\alpha = 3.14, \beta = 33.01$)
Michaelis-Menten	V_{max}	1/s	Gamma ($\alpha = 4.45, \beta = 420.00$)
Michaelis-Menten	K_m	mM	Gamma ($\alpha = 4.45, \beta = 1680.00$)
Hill-type	V_{max}	1/s	Gamma ($\alpha = 4.45, \beta = 4200.00$)
Hill-type	K_m	mM	Gamma ($\alpha = 3.14, \beta = 132.04$)
Hill-type	n	unitless	TruncatedNormal* ($\mu = 1.39, \sigma = 0.36, \text{lb} = 1, \text{ub} = \infty$)

* truncated normal distribution is truncated at one, so support is over $[1, \infty)$

Table C.2: Naming indicates the kinetic formulation used for the first model and the second reaction, respectively. For example, mass action-mass action uses mass action kinetics for both reactions. Abbreviations: MA–mass action, MM–Michaelis-Menten, Hill–Hill-type.

Model	Locally identifiable $y_{\text{obs}} = [P]$	NOT identifiable $y_{\text{obs}} = [P]$	Locally identifiable $y_{\text{obs}} = \frac{[P]}{[S]_{\text{tot}}}$	NOT identifiable $y_{\text{obs}} = \frac{[P]}{[S]_{\text{tot}}}$
MA–MA	$k_{f,1}, k_{r,1}, k_{\text{cat},1}, k_{f,2}, k_{r,2}, k_{\text{cat},2}$	—	$k_{r,1}, k_{\text{cat},1}, k_{r,2}, k_{\text{cat},2}$	$k_{f,1}, k_{f,2}$
MA–MM	$k_{f,1}, k_{r,1}, k_{\text{cat},1}, V_{\text{max},2}, K_{m,2}$	—	$k_{r,1}, k_{\text{cat},1}$	$k_{f,1}, V_{\text{max},2}, K_{m,2}$
MA–Hill	$k_{f,1}, k_{r,1}, k_{\text{cat},1}, V_{\text{max},2}, K_{m,2}$	—	$k_{r,1}, k_{\text{cat},1}$	$k_{f,1}, V_{\text{max},2}, K_{m,2}$
MM–MM	$V_{\text{max},1}, K_{m,1}, V_{\text{max},2}, K_{m,2}$	—	—	$V_{\text{max},1}, K_{m,1}, V_{\text{max},2}, K_{m,2}$
MM–MA	$V_{\text{max},1}, K_{m,1}, k_{f,2}, k_{r,2}, k_{\text{cat},2}$	—	$V_{\text{max},1}, K_{m,1}, k_{f,2}, k_{r,2}, k_{\text{cat},2}$	—
MM–Hill	$V_{\text{max},1}, K_{m,1}, V_{\text{max},2}, K_{m,2}$	—	—	$V_{\text{max},1}, K_{m,1}, V_{\text{max},2}, K_{m,2}$
Hill–Hill	$V_{\text{max},1}, K_{m,1}, V_{\text{max},2}, K_{m,2}$	—	—	$V_{\text{max},1}, K_{m,1}, V_{\text{max},2}, K_{m,2}$
Hill–MA	$V_{\text{max},1}, K_{m,1}, k_{f,2}, k_{r,2}, k_{\text{cat},2}$	—	$k_{r,2}, k_{\text{cat},2}$	$V_{\text{max},1}, K_{m,1}, k_{f,2}$
Hill–MM	$V_{\text{max},1}, K_{m,1}, V_{\text{max},2}, K_{m,2}$	—	—	$V_{\text{max},1}, K_{m,1}, V_{\text{max},2}, K_{m,2}$

Table C.3

Reaction	Flux	Reference
Oxidative phos.	$J_{\text{OxPhos}} = \frac{V_{\text{max,OxPhos}} \left(\frac{[\text{ADP}]}{K_{\text{ADP}}} \right)^n}{1 + \left(\frac{[\text{ADP}]}{K_{\text{ADP}}} \right)^n}$	[388]
Glycolysis	$J_{\text{glycolysis}} = k_{\text{glycolysis}} [\text{ADP}]$	n/a
ATP Hydrolysis	$J_{\text{hydro}} = k_{\text{hydro}} [\text{ATP}]$	[49]
Adenylate kinase (forward)	$J_{\text{AK,for}} = \frac{\frac{V_{\text{AK,for}} [\text{ATP}] [\text{AMP}]}{k_{\text{mt}} \times k_{\text{mm}}}}{1 + \frac{[\text{ATP}]}{k_{\text{mt}}} + \frac{[\text{AMP}]}{k_{\text{mm}}} + \frac{[\text{ATP}] [\text{AMP}]}{k_{\text{mt}} \times k_{\text{mm}}} + \frac{2[\text{ADP}]}{k_{\text{md}}} + \frac{[\text{ADP}]^2}{k_{\text{md}}^2}}$	[389]
Adenylate kinase (reverse)*	$J_{\text{AK,rev}} = \frac{\frac{V_{\text{AK,rev}} [\text{ADP}]^2}{k_{\text{md}}^2}}{1 + \frac{[\text{ATP}]}{k_{\text{mt}}} + \frac{[\text{AMP}]}{k_{\text{mm}}} + \frac{[\text{ATP}] [\text{AMP}]}{k_{\text{mt}} \times k_{\text{mm}}} + \frac{2[\text{ADP}]}{k_{\text{md}}} + \frac{[\text{ADP}]^2}{k_{\text{md}}^2}}$	[389]
Adenylate kinase	$J_{\text{AK}} = J_{\text{AK,for}} - J_{\text{AK,rev}}$	[389]
Creatine kinase (forward)	$J_{\text{CK,for}} = \frac{\frac{V_{\text{CK,for}} K_{iq} K_p [\text{ADP}] [\text{PCr}]}{K_{ia} K_b}}{1 + \frac{[\text{ADP}]}{K_{ia}} + \frac{[\text{PCr}]}{K_{ib}} + \frac{[\text{ATP}]}{K_{iq}} + \frac{[\text{ADP}] [\text{PCr}]}{K_{ia} K_b} + \frac{([\text{Cr}]_{\text{total}} - [\text{PCr}]) [\text{ATP}]}{K_{iq} K_p}}$	[389]
Creatine kinase (reverse)**	$J_{\text{CK,rev}} = \frac{\frac{V_{\text{CK,rev}} ([\text{Cr}]_{\text{total}} - [\text{PCr}]) [\text{ATP}]}{K_{iq} K_p}}{1 + \frac{[\text{ADP}]}{K_{ia}} + \frac{[\text{PCr}]}{K_{ib}} + \frac{[\text{ATP}]}{K_{iq}} + \frac{[\text{ADP}] [\text{PCr}]}{K_{ia} K_b} + \frac{([\text{Cr}]_{\text{total}} - [\text{PCr}]) [\text{ATP}]}{K_{iq} K_p}}$	[389]
Creatine kinase	$J_{\text{CK}} = J_{\text{CK,rev}} - J_{\text{CK,for}}$	[389]

$$* V_{\text{AK,rev}} := \frac{V_{\text{AK,for}} \times k_{\text{md}}^2}{K_{\text{eqAK}} \times k_{\text{mt}} \times k_{\text{mm}}}$$

$$** V_{\text{CK,rev}} := \frac{V_{\text{CK,for}} K_{iq} K_p}{K_{\text{eqCK}} K_{ia} K_b}$$

Table C.4

Reaction	Parameter	Units	Nominal value	Reference/ Rationale
Oxidative phos.	$V_{\max, \text{OxPhos}}$	mM/s	0.5	[49], [388]
	K_{ADP}	mM	5.8×10^{-2}	[49], [388]
	n	unitless	2.568	[49], [388]
Glycolysis	k_{Gly}	mM/s	0.5	derived*
ATP Hydrolysis	k_{hydro}	1/s	1.5×10^{-1}	tuned
Adenylate kinase	$V_{\text{AK,for}}$	mM/s	14.66	[49], [389]
	k_{mt}	mM	0.27	[49], [389]
	k_{mm}	mM	0.32	[49], [389]
	k_{md}	mM	0.35	[49], [389]
	K_{eqAK}	unitless	2.221	[49], [389]
Creatine kinase	$V_{\text{CK,for}}$	mM/s	1×10^2	[49], [388]
	K_b	mM	1.11	[49], [388]
	K_{ia}	mM	0.135	[49], [388]
	K_{ib}	mM	3.9	[49], [388]
	K_{iq}	mM	3.5	[49], [388]
	K_p	mM	3.8	[49], [388]
	K_{eqCK}	unitless	$1.77 \times 10^{(9-\text{pH})}$	assume pH = 7 [49], [388]
	$[\text{Cr}]_{\text{total}}$	mM	39.0	[49], [388]

* See note in Supplemental Section C.1.2.

Table C.5: Nominal values and bounds.

Parameter	units	nominal value	bounds
k_{OnAMP}	$1/(\text{mM} \times \text{s})$	1.0	–
k_{OffAMP}	1/s	2.5 ± 0.6 [293] ^a	[2.5e-9, 25]
k_{OnADP}	$1/(\text{mM} \times \text{s})$	1.0	–
k_{OffADP}	1/s	1.5 ± 0.4 [293]	[1.5e-9, 15]
k_{OnATP}	$1/(\text{mM} \times \text{s})$	1.0	–
k_{OffATP}	1/s	1.7 ± 0.5 [293]	[1.7e-9, 17]
$k_{\text{on,LKB1}}$	$1/(\text{mM} \times \text{s})$	1.0	–
$k_{\text{off,LKB1}}$	1/s	approx. from Mich. Ment.	[0.001396, 1396.0]
$k_{\text{phos,LKB1}}$	1/s	3.92×10^{-2}	[3.92e-06, 0.392]
k_{LKB1}	1/s	$3.92 \times 10^{-2} \pm 3.4 \times 10^{-3}$ [49], [325]**	[3.92e-06, 0.392]
$K_{\text{m,LKB1}}$	mM	1.4 ± 0.24 [49], [325] ^b	[0.00139992, 1,396.392]
α_{LKB1}	–	–	[1e-4, 1.0]
β_{LKB1}	–	–	[0.95, 113.03]
$k_{\text{on,CaMKK}}$	$1/(\text{mM} \times \text{s})$	1.0	–
$k_{\text{off,CaMKK}}$	1/s	approx. from Mich. Ment.	[1.11e-05, 1.11]
$k_{\text{phos,CaMKK}}$	1/s	$3.92\text{e-}2^{\text{c}}$	[3.92e-05, 0.392]
k_{CaMKK}	1/s	$3.92\text{e-}2^{\text{c}}$	[3.92e-05, 0.392]
$K_{\text{m,CaMKK}}$	μM	15	[5.03e-5, 1.502]
β_{CaMKK}	–	–	[0.95, 113.03]
$k_{\text{on,PP}}$	$1/(\text{mM} \times \text{s})$	1.0	–
$k_{\text{off,PP}}$	1/s	approx. from Mich. Ment.	[5.6e-06, 5.6]
$k_{\text{dephos,PP}}$	1/s	$1.1\text{e-}1^{\text{d}}$	[1.1e-06, 0.11]
k_{PP}	1/s	$1.1 \times 10^{-1} \pm 4.0 \times 10^{-3}$ [49], [333] ^d	[1.1e-06, 0.11]
$K_{\text{m,PP}}$	mM	$6.7 \times 10^{-2} \pm 1.3 \times 10^{-2}$ [49], [333]	[6.7e-6, 5.71]
α_{PP}	–	–	[1.0, 100.0]
$k_{\text{on,PP1}}$	$1/(\text{mM} \times \text{s})$	1.0	–
$k_{\text{off,PP1}}$	1/s	approx. from Mich. Ment. ^e	[5.6e-06, 5.6]
$k_{\text{dephos,PP1}}$	1/s	$1.1\text{e-}1^{\text{e}}$	[1.1e-06, 0.11]
k_{PP1}	1/s	^e	[1.1e-06, 0.11]
$K_{\text{m,PP1}}$	mM	^e	[6.7e-6, 5.71]
$k_{\text{on,AMPK}}$	$1/(\text{mM} \times \text{s})$	1.0	–
$k_{\text{off,AMPK}}$	1/s	approx. from Mich. Ment.	[0.000301, 301.0]
$k_{\text{phos,AMPK}}$	1/s	6.33 ± 0.29 [303] ^f	[0.00633, 6330.0]
k_{AMPK}	1/s	6.33 ± 0.29 [303]	[0.00633, 6330.0]
$K_{\text{m,AMPK}}$	μM	4.67 ± 1.48 [303] ^f	[0.006631, 6,631]
β_{AMPK}	–	–	[1.0, 100.0]
β_{AMP}	–	–	[0.95, 113.03]

^a K_d for strong binding site from [293]^b Set to V_{max} (assume enzyme concentration of 1.0 mM). Originally from [325]. Converted by [49].^c Assume same as LKB1.^d Set to V_{max} (assume enzyme concentration of 1.0 mM). The original nominal value had units of $\mu\text{M}/\text{min}/\text{mg}$ [333]. The conversion was done by [49] using unknown conversion factors.^e Same as PP.^f Values based on Table 1 of [303].

Table C.6

Model	Locally Identifiable	NOT Locally Identifiable
1	kOnAMP, kOffAMP, kOnADP, kOffADP, kOnATP, kOffATP, kOnCaMKK, kOffCaMKK, kPhosCaMKK, kOnLKB1, kOffLKB1, kPhosLKB1, kOnPP, kOffPP, kDephosPP, kOnAMPK, kOffAMPK, kPhosAMPK, kOnPP1, kOffPP1, kDephosPP1	
2	kOnAMP, kOffAMP, kOnADP, kOffADP, kOnATP, kOffATP, KmCaMKK, kLKB1, KmLKB1, kPP, KmPP	kCaMKK, kAMPK, KmAMPK, kPP1, KmPP1
3	kOnAMP, kOffAMP, kOnADP, kOffADP, kOnATP, kOffATP, kOnCaMKK, kOffCaMKK, kPhosCaMKK, kOnLKB1, kOffLKB1, alphaLKB1, kPhosLKB1, kOnPP, kOffPP, alphaPP, kDephosPP, kOnAMPK, kOffAMPK, kPhosAMPK, betaAMP, kOnPP1, kOffPP1, kDephosPP1	
4	kOnAMP, kOffAMP, kOnADP, kOffADP, kOnATP, kOffATP, KmCaMKK, kLKB1, KmLKB1, alphaLKB1, kPP, KmPP, alphaPP, betaAMP	kCaMKK, kAMPK, KmAMPK, kPP1, KmPP1
5	kOnAMP, kOffAMP, kOnADP, kOffADP, kOnATP, kOffATP, kOnCaMKK, kOffCaMKK, kPhosCaMKK, kOnLKB1, kOffLKB1, kPhosLKB1, kOnPP, kOffPP, kDephosPP, kOnAMPK, kOffAMPK, kPhosAMPK, kOnPP1, kOffPP1, kDephosPP1, alphaPP, betaAMP, betaCaMKK, betaLKB1	
6	kOnAMP, kOffAMP, kOnADP, kOffADP, kOnATP, kOffATP, KmCaMKK, kLKB1, KmLKB1, kPP, KmPP, betaAMPK, betaLKB1, betaCaMKK, alphaPP	kCaMKK, kAMPK, KmAMPK, kPP1, KmPP1

Table C.7: Bold values indicate influential parameters with a mean index greater than or equal to 0.01.

Parameter	Function	max fraction act. AMPKAR			Time to half-max		
		WT	LKB1 KO	CaMKK2 KO	WT	LKB1 KO	CaMKK2 KO
k_{OffAMP}	AMP binding	0.135 $\pm$ 0.043	0.151 $\pm$ 0.044	0.151 $\pm$ 0.048	0.067 $\pm$ 0.017	0.056 $\pm$ 0.012	0.073 $\pm$ 0.015
k_{OffADP}	ADP binding	0.087 $\pm$ 0.035	0.082 $\pm$ 0.034	0.085 $\pm$ 0.037	0.059 $\pm$ 0.018	0.095 $\pm$ 0.033	0.056 $\pm$ 0.018
k_{OffATP}	ATP binding	0.219 $\pm$ 0.071	0.207 $\pm$ 0.059	0.184 $\pm$ 0.062	0.407 $\pm$ 0.118	0.342 $\pm$ 0.095	0.359 $\pm$ 0.098
k_{OffCaMKK}	NaN	0.001 $\pm$ 0.001	0.010 $\pm$ 0.004	0.000 $\pm$ 0.000	0.001 $\pm$ 0.001	0.019 $\pm$ 0.007	0.000 $\pm$ 0.000
$k_{\text{PhosCaMKK}}$	AMPK phos.	0.003 $\pm$ 0.003	0.084 $\pm$ 0.045	0.000 $\pm$ 0.000	0.004 $\pm$ 0.002	0.155 $\pm$ 0.051	0.000 $\pm$ 0.000
k_{OffLKB1}	AMPK phos.	0.001 $\pm$ 0.001	0.000 $\pm$ 0.000	0.007 $\pm$ 0.003	0.005 $\pm$ 0.001	0.000 $\pm$ 0.000	0.023 $\pm$ 0.010
k_{PhosLKB1}	AMPK phos.	0.004 $\pm$ 0.003	0.000 $\pm$ 0.000	0.076 $\pm$ 0.044	0.021 $\pm$ 0.008	0.000 $\pm$ 0.000	0.126 $\pm$ 0.057
k_{OffPP}	AMPK dephos.	0.004 $\pm$ 0.003	0.005 $\pm$ 0.003	0.002 $\pm$ 0.002	0.002 $\pm$ 0.001	0.004 $\pm$ 0.002	0.000 $\pm$ 0.000
k_{DephosPP}	AMPK dephos.	0.003 $\pm$ 0.003	0.003 $\pm$ 0.002	0.001 $\pm$ 0.001	0.001 $\pm$ 0.001	0.001 $\pm$ 0.001	0.001 $\pm$ 0.000
k_{OffAMPK}	AMPK kinase act.	0.183 $\pm$ 0.055	0.166 $\pm$ 0.050	0.177 $\pm$ 0.053	0.108 $\pm$ 0.023	0.085 $\pm$ 0.020	0.106 $\pm$ 0.028
k_{PhosAMPK}	AMPK kinase act.	0.516 $\pm$ 0.108	0.454 $\pm$ 0.084	0.480 $\pm$ 0.096	0.545 $\pm$ 0.120	0.514 $\pm$ 0.122	0.493 $\pm$ 0.101
k_{OffPP1}	AMKPAR dephos.	0.002 $\pm$ 0.001	0.002 $\pm$ 0.001	0.002 $\pm$ 0.001	0.003 $\pm$ 0.003	0.003 $\pm$ 0.003	0.002 $\pm$ 0.001
$k_{\text{Dephos,PP1}}$	AMKPAR dephos.	0.007 $\pm$ 0.002	0.006 $\pm$ 0.001	0.007 $\pm$ 0.002	0.003 $\pm$ 0.001	0.004 $\pm$ 0.002	0.003 $\pm$ 0.001

Table C.8: Bold values indicate influential parameters with a mean index greater than or equal to 0.01.

Parameter	Function	max fraction act. AMPKAR			Time to half-max		
		WT	LKB1 KO	CaMKK2 KO	WT	LKB1 KO	CaMKK2 KO
k_{OffAMP}	AMP binding	0.498 ± 0.089	0.509 ± 0.098	0.724 ± 0.158	0.806 ± 0.135	0.776 ± 0.131	0.478 ± 0.144
k_{OffADP}	ADP binding	0.174 ± 0.063	0.163 ± 0.054	0.038 ± 0.022	0.144 ± 0.056	0.158 ± 0.062	0.218 ± 0.091
k_{OffATP}	ATP binding	0.327 ± 0.086	0.310 ± 0.081	0.155 ± 0.076	0.255 ± 0.085	0.237 ± 0.088	0.242 ± 0.115
$K_{m,\text{CaMKK}}$	AMPK phos.	0.127 ± 0.043	0.147 ± 0.046	0.000 ± 0.000	0.149 ± 0.048	0.197 ± 0.068	0.000 ± 0.000
k_{LKB1}	AMPK phos.	0.004 ± 0.003	0.000 ± 0.000	0.215 ± 0.114	0.007 ± 0.004	0.000 ± 0.000	0.260 ± 0.093
$K_{m,\text{LKB1}}$	AMPK phos.	0.011 ± 0.005	0.000 ± 0.000	0.233 ± 0.079	0.018 ± 0.011	0.000 ± 0.000	0.557 ± 0.286
k_{PP}	AMPK dephos.	0.035 ± 0.016	0.037 ± 0.015	0.012 ± 0.007	0.017 ± 0.009	0.020 ± 0.011	0.085 ± 0.044
$K_{m,\text{PP}}$	AMPK dephos.	0.024 ± 0.009	0.024 ± 0.010	0.009 ± 0.005	0.022 ± 0.011	0.025 ± 0.014	0.036 ± 0.018

Table C.9: Bold values indicate influential parameters with a mean index greater than or equal to 0.01.

Parameter	Function	max fraction act. AMPKAR			Time to half-max		
		WT	LKB1 KO	CaMKK2 KO	WT	LKB1 KO	CaMKK2 KO
k_{OffAMP}	AMP binding	0.198 ± 0.053	0.227 ± 0.064	0.208 ± 0.055	0.416 ± 0.108	0.454 ± 0.133	0.415 ± 0.125
k_{OffADP}	ADP binding	0.086 ± 0.038	0.072 ± 0.040	0.081 ± 0.031	0.057 ± 0.024	0.028 ± 0.012	0.063 ± 0.031
k_{OffATP}	ATP binding	0.075 ± 0.036	0.082 ± 0.046	0.074 ± 0.032	0.584 ± 0.210	0.348 ± 0.145	0.517 ± 0.152
k_{OffCaMKK}	NaN	0.001 ± 0.000	0.007 ± 0.004	0.000 ± 0.000	0.002 ± 0.001	0.023 ± 0.013	0.000 ± 0.000
$k_{\text{PhosCaMKK}}$	AMPK phos.	0.002 ± 0.001	0.055 ± 0.026	0.000 ± 0.000	0.025 ± 0.027	0.248 ± 0.102	0.000 ± 0.000
k_{OffLKB1}	AMPK phos.	0.002 ± 0.001	0.000 ± 0.000	0.007 ± 0.005	0.010 ± 0.004	0.000 ± 0.000	0.047 ± 0.052
k_{PhosLKB1}	AMPK phos.	0.007 ± 0.004	0.000 ± 0.000	0.026 ± 0.015	0.045 ± 0.023	0.000 ± 0.000	0.150 ± 0.087
k_{OffPP}	AMPK dephos.	0.002 ± 0.002	0.004 ± 0.002	0.002 ± 0.002	0.005 ± 0.005	0.016 ± 0.014	0.005 ± 0.005
k_{DephosPP}	AMPK dephos.	0.001 ± 0.000	0.003 ± 0.002	0.001 ± 0.000	0.002 ± 0.002	0.011 ± 0.010	0.002 ± 0.002
k_{OffAMPK}	AMPK kinase act.	0.036 ± 0.012	0.036 ± 0.015	0.036 ± 0.013	0.052 ± 0.037	0.081 ± 0.065	0.044 ± 0.029
k_{PhosAMPK}	AMPK kinase act.	0.135 ± 0.070	0.123 ± 0.075	0.134 ± 0.071	0.036 ± 0.028	0.020 ± 0.014	0.034 ± 0.026
k_{OffPP1}	AMKPAR dephos.	0.030 ± 0.008	0.029 ± 0.010	0.031 ± 0.011	0.022 ± 0.012	0.014 ± 0.009	0.023 ± 0.013
$k_{\text{Dephos,PP1}}$	AMKPAR dephos.	0.501 ± 0.109	0.471 ± 0.128	0.508 ± 0.107	0.087 ± 0.054	0.057 ± 0.048	0.072 ± 0.037
α_{LKB1}	AMPK phos.	0.001 ± 0.000	0.000 ± 0.000	0.004 ± 0.004	0.004 ± 0.002	0.000 ± 0.000	0.015 ± 0.009
α_{PP}	AMPK dephos.	0.002 ± 0.001	0.003 ± 0.002	0.002 ± 0.001	0.002 ± 0.002	0.005 ± 0.003	0.002 ± 0.001
β_{AMP}	AMPK kinase act.	0.106 ± 0.046	0.090 ± 0.044	0.105 ± 0.050	0.056 ± 0.050	0.037 ± 0.034	0.050 ± 0.038

Table C.10: Bold values indicate influential parameters with a mean index greater than or equal to 0.01.

Parameter	Function	max fraction act. AMPKAR			Time to half-max		
		WT	LKB1 KO	CaMKK2 KO	WT	LKB1 KO	CaMKK2 KO
k_{OffAMP}	AMP binding	2.38e9 $\pm$ 2.79e9	2.69e9 $\pm$ 2.91e9	0.111 $\pm$ 0.112	0.070 $\pm$ 0.055	0.044 $\pm$ 0.033	0.139 $\pm$ 0.042
k_{OffADP}	ADP binding	0.014 $\pm$ 0.010	0.014 $\pm$ 0.011	0.025 $\pm$ 0.034	0.020 $\pm$ 0.022	0.006 $\pm$ 0.006	0.019 $\pm$ 0.010
k_{OffATP}	ATP binding	0.860 $\pm$ 3.147	0.852 $\pm$ 2.769	0.240 $\pm$ 0.179	0.601 $\pm$ 0.980	0.604 $\pm$ 1.048	0.263 $\pm$ 0.094
$K_{m,\text{CaMKK}}$	AMPK phos.	0.341 $\pm$ 0.372	0.344 $\pm$ 0.343	0.000 $\pm$ 0.000	0.166 $\pm$ 0.176	0.210 $\pm$ 0.165	0.000 $\pm$ 0.000
k_{LKB1}	AMPK phos.	0.000 $\pm$ 0.000	0.000 $\pm$ 0.000	0.324 $\pm$ 0.220	0.039 $\pm$ 0.046	0.000 $\pm$ 0.000	0.404 $\pm$ 0.113
$K_{m,\text{LKB1}}$	AMPK phos.	0.000 $\pm$ 0.000	0.000 $\pm$ 0.000	0.223 $\pm$ 0.184	0.053 $\pm$ 0.060	0.000 $\pm$ 0.000	0.184 $\pm$ 0.060
k_{PP}	AMPK dephos.	0.000 $\pm$ 0.000	0.000 $\pm$ 0.000	0.179 $\pm$ 0.141	0.046 $\pm$ 0.053	0.000 $\pm$ 0.000	0.177 $\pm$ 0.083
$K_{M,\text{PP}}$	AMPK dephos.	0.014 $\pm$ 0.010	0.014 $\pm$ 0.009	0.074 $\pm$ 0.070	0.010 $\pm$ 0.005	0.007 $\pm$ 0.004	0.023 $\pm$ 0.011
α_{LKB1}	AMPK phos.	0.475 $\pm$ 0.455	0.479 $\pm$ 0.412	0.107 $\pm$ 0.107	0.212 $\pm$ 0.228	0.224 $\pm$ 0.254	0.051 $\pm$ 0.034
α_{PP}	AMPK dephos.	0.457 $\pm$ 0.431	0.461 $\pm$ 0.392	0.129 $\pm$ 0.135	0.200 $\pm$ 0.206	0.204 $\pm$ 0.232	0.039 $\pm$ 0.023
β_{AMP}	AMPK kinase act.	0.882 $\pm$ 0.471	0.887 $\pm$ 0.416	0.385 $\pm$ 0.284	0.573 $\pm$ 0.261	0.618 $\pm$ 0.289	0.206 $\pm$ 0.054

Table C.11: Bold values indicate influential parameters with a mean index greater than or equal to 0.01.

Parameter	Function	max fraction act. AMPKAR			Time to half-max		
		WT	LKB1 KO	CaMKK2 KO	WT	LKB1 KO	CaMKK2 KO
k_{OffAMP}	AMP binding	0.092 $\pm$ 0.027	0.089 $\pm$ 0.023	0.090 $\pm$ 0.026	0.181 $\pm$ 0.102	0.136 $\pm$ 0.111	0.169 $\pm$ 0.081
k_{OffADP}	ADP binding	0.127 $\pm$ 0.042	0.054 $\pm$ 0.023	0.125 $\pm$ 0.044	0.195 $\pm$ 0.146	0.082 $\pm$ 0.059	0.186 $\pm$ 0.118
k_{OffATP}	ATP binding	0.246 $\pm$ 0.059	0.131 $\pm$ 0.048	0.249 $\pm$ 0.076	0.447 $\pm$ 0.269	0.148 $\pm$ 0.100	0.425 $\pm$ 0.188
$k_{OffCaMKK}$	NaN	0.000 $\pm$ 0.000	0.002 $\pm$ 0.001	0.000 $\pm$ 0.000	0.001 $\pm$ 0.001	0.005 $\pm$ 0.003	0.000 $\pm$ 0.000
$k_{PhosCaMKK}$	AMPK phos.	0.002 $\pm$ 0.002	0.011 $\pm$ 0.007	0.000 $\pm$ 0.000	0.004 $\pm$ 0.002	0.014 $\pm$ 0.009	0.000 $\pm$ 0.000
$k_{OffLKB1}$	AMPK phos.	0.003 $\pm$ 0.002	0.000 $\pm$ 0.000	0.006 $\pm$ 0.003	0.014 $\pm$ 0.007	0.000 $\pm$ 0.000	0.021 $\pm$ 0.010
$k_{PhosLKB1}$	AMPK phos.	0.012 $\pm$ 0.008	0.000 $\pm$ 0.000	0.023 $\pm$ 0.010	0.037 $\pm$ 0.012	0.000 $\pm$ 0.000	0.060 $\pm$ 0.025
k_{OffPP}	AMPK dephos.	0.036 $\pm$ 0.019	0.036 $\pm$ 0.012	0.035 $\pm$ 0.018	0.457 $\pm$ 0.268	0.549 $\pm$ 0.367	0.375 $\pm$ 0.219
$k_{DephosPP}$	AMPK dephos.	0.026 $\pm$ 0.019	0.070 $\pm$ 0.023	0.027 $\pm$ 0.020	0.084 $\pm$ 0.038	0.193 $\pm$ 0.177	0.085 $\pm$ 0.049
$k_{OffAMPK}$	AMPK kinase act.	0.024 $\pm$ 0.013	0.031 $\pm$ 0.019	0.024 $\pm$ 0.013	0.156 $\pm$ 0.081	0.183 $\pm$ 0.193	0.155 $\pm$ 0.090
$k_{PhosAMPK}$	AMPK kinase act.	0.197 $\pm$ 0.058	0.118 $\pm$ 0.041	0.187 $\pm$ 0.050	0.321 $\pm$ 0.115	0.061 $\pm$ 0.040	0.318 $\pm$ 0.140
k_{OffPP1}	AMKPAR dephos.	0.095 $\pm$ 0.025	0.084 $\pm$ 0.024	0.090 $\pm$ 0.025	0.191 $\pm$ 0.096	0.050 $\pm$ 0.052	0.196 $\pm$ 0.110
$k_{DephosPP1}$	AMKPAR dephos.	0.398 $\pm$ 0.094	0.500 $\pm$ 0.098	0.387 $\pm$ 0.088	0.328 $\pm$ 0.144	0.066 $\pm$ 0.049	0.311 $\pm$ 0.130
α_{PP}	AMPK dephos.	0.040 $\pm$ 0.018	0.113 $\pm$ 0.041	0.043 $\pm$ 0.019	0.187 $\pm$ 0.194	0.369 $\pm$ 0.232	0.143 $\pm$ 0.136
β_{AMP}	AMPK kinase act.	0.071 $\pm$ 0.030	0.046 $\pm$ 0.024	0.069 $\pm$ 0.033	0.086 $\pm$ 0.041	0.012 $\pm$ 0.011	0.085 $\pm$ 0.043
β_{LKB1}	AMPK phos.	0.022 $\pm$ 0.011	0.010 $\pm$ 0.009	0.044 $\pm$ 0.027	0.028 $\pm$ 0.023	0.004 $\pm$ 0.004	0.038 $\pm$ 0.026
β_{CaMKK}	AMPK phos.	0.000 $\pm$ 0.000	0.000 $\pm$ 0.000	0.000 $\pm$ 0.000	0.000 $\pm$ 0.000	0.000 $\pm$ 0.000	0.000 $\pm$ 0.000

Table C.12: Bold values indicate influential parameters with a mean index greater than or equal to 0.01.

		max fraction act. AMPKAR			Time to half-max		
Parameter	Function	WT	LKB1 KO	CaMKK2 KO	WT	LKB1 KO	CaMKK2 KO
k_{OffAMP}	AMP binding	0.196 $\pm$ 0.071	0.261 $\pm$ 0.075	0.198 $\pm$ 0.072	0.271 $\pm$ 0.060	0.401 $\pm$ 0.171	0.270 $\pm$ 0.055
k_{OffADP}	ADP binding	0.174 $\pm$ 0.080	0.075 $\pm$ 0.035	0.173 $\pm$ 0.062	0.114 $\pm$ 0.037	0.076 $\pm$ 0.050	0.114 $\pm$ 0.034
k_{OffATP}	ATP binding	0.271 $\pm$ 0.084	0.139 $\pm$ 0.059	0.273 $\pm$ 0.086	0.414 $\pm$ 0.126	0.243 $\pm$ 0.160	0.412 $\pm$ 0.117
$K_{m,\text{CaMKK}}$	AMPK phos.	0.000 $\pm$ 0.000	0.311 $\pm$ 0.110	0.000 $\pm$ 0.000	0.000 $\pm$ 0.000	0.403 $\pm$ 0.249	0.000 $\pm$ 0.000
k_{LKB1}	AMPK phos.	0.018 $\pm$ 0.029	0.000 $\pm$ 0.000	0.020 $\pm$ 0.037	0.016 $\pm$ 0.010	0.000 $\pm$ 0.000	0.016 $\pm$ 0.010
$K_{m,\text{LKB1}}$	AMPK phos.	0.017 $\pm$ 0.029	0.000 $\pm$ 0.000	0.020 $\pm$ 0.037	0.006 $\pm$ 0.007	0.000 $\pm$ 0.000	0.006 $\pm$ 0.007
k_{PP}	AMPK dephos.	0.000 $\pm$ 0.000	0.003 $\pm$ 0.002	0.000 $\pm$ 0.001	0.000 $\pm$ 0.000	0.054 $\pm$ 0.038	0.000 $\pm$ 0.000
$K_{M,\text{PP}}$	AMPK dephos.	0.004 $\pm$ 0.004	0.006 $\pm$ 0.003	0.003 $\pm$ 0.004	0.001 $\pm$ 0.001	0.233 $\pm$ 0.116	0.001 $\pm$ 0.001
α_{PP}	AMPK dephos.	0.598 $\pm$ 0.127	0.271 $\pm$ 0.108	0.595 $\pm$ 0.130	0.595 $\pm$ 0.110	0.250 $\pm$ 0.161	0.595 $\pm$ 0.097
β_{AMP}	AMPK kinase act.	0.001 $\pm$ 0.001	0.000 $\pm$ 0.000	0.001 $\pm$ 0.001	0.002 $\pm$ 0.002	0.000 $\pm$ 0.000	0.002 $\pm$ 0.002
β_{LKB1}	AMPK phos.	0.000 $\pm$ 0.000	0.258 $\pm$ 0.095	0.000 $\pm$ 0.000	0.000 $\pm$ 0.000	0.392 $\pm$ 0.237	0.000 $\pm$ 0.000
β_{CaMKK}	AMPK phos.	0.000 $\pm$ 0.000	0.000 $\pm$ 0.000	0.000 $\pm$ 0.000	0.000 $\pm$ 0.000	0.012 $\pm$ 0.007	0.000 $\pm$ 0.000

Bibliography

- [1] U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits* (Chapman & Hall/CRC Mathematical and Computational Biology). Taylor & Francis, 2006, ISBN: 9781584886426.
- [2] J. Keener and J. Sneyd, *Mathematical physiology: II: Systems physiology*. Springer, 2009.
- [3] L. Edelstein-Keshet, *Mathematical models in biology*. SIAM, 2005.
- [4] S. T. Vittadello and M. P. H. Stumpf, “Open problems in mathematical biology,” *Math. Biosci.*, vol. 354, p. 108 926, Dec. 2022.
- [5] R. C. Smith, *Uncertainty quantification: Theory, implementation, and applications*. SIAM, 2013, vol. 12.
- [6] S. Mehta and J. Zhang, “Illuminating the Cell’s Biochemical Activity Architecture,” *Biochemistry*, vol. 56, no. 39, pp. 5210–5213, Oct. 3, 2017, ISSN: 0006-2960, 1520-4995.
- [7] D. L. Schmitt, S. Mehta, and J. Zhang, “Illuminating the kinome: Visualizing real-time kinase activity in biological systems using genetically encoded fluorescent protein-based biosensors,” *Current Opinion in Chemical Biology*, vol. 54, pp. 63–69, Feb. 2020, ISSN: 13675931.
- [8] A. C. Lyons, S. Mehta, and J. Zhang, “Fluorescent biosensors illuminate the spatial regulation of cell signaling across scales,” *Biochemical Journal*, vol. 480, no. 20, pp. 1693–1717, Oct. 31, 2023, ISSN: 0264-6021, 1470-8728.
- [9] H. Kitano, “Systems biology: A brief overview,” *Science*, vol. 295, no. 5560, pp. 1662–1664, Mar. 2002.
- [10] T. Ideker, T. Galitski, and L. Hood, “A NEW APPROACH TO DECODING LIFE: Systems Biology,” *Annual Review of Genomics and Human Genetics*, vol. 2, pp. 343–372, Volume 2, 2001 Sep. 1, 2001, ISSN: 1527-8204, 1545-293X.
- [11] A. Ma’ayan, “Complex systems biology,” *Journal of The Royal Society Interface*, vol. 14, no. 134, p. 20 170 391, 2017.

- [12] H. Kitano, "Systems biology: A brief overview," *Science*, vol. 295, no. 5560, pp. 1662–1664, Mar. 2002.
- [13] Z. Ji, K. Yan, W. Li, H. Hu, and X. Zhu, "Mathematical and computational modeling in complex biological systems," *BioMed Research International*, vol. 2017, no. 1, p. 5958321, 2017.
- [14] L. Qiao, A. Khalilimeybodi, N. J. Linden-Santangeli, and P. Rangamani, "The Evolution of Systems Biology and Systems Medicine: From Mechanistic Models to Uncertainty Quantification," Feb. 19, 2025.
- [15] E. D. Mitra, R. Suderman, J. Colvin, A. Ionkov, A. Hu, H. M. Sauro, R. G. Posner, and W. S. Hlavacek, "PyBioNetFit and the biological property specification language," *iScience*, vol. 19, pp. 1012–1036, Sep. 2019.
- [16] M. K. Bell and P. Rangamani, "Design decisions for incorporating spatial and mechanical aspects in models of signaling networks," *Current Opinion in Systems Biology*, vol. 25, pp. 70–77, Mar. 1, 2021, ISSN: 2452-3100.
- [17] E. A. Francis, J. G. Laughlin, J. S. Dokken, H. N. T. Finsberg, C. T. Lee, M. E. Rognes, and P. Rangamani, "Spatial modeling algorithms for reactions and transport in biological cells," *Nature Computational Science*, vol. 5, no. 1, pp. 76–89, Jan. 2025, ISSN: 2662-8457.
- [18] J. Walpole, J. A. Papin, and S. M. Peirce, "Multiscale computational models of complex biological systems," *Annual Review of Biomedical Engineering*, vol. 15, no. Volume 15, 2013, pp. 137–154, 2013, ISSN: 1545-4274.
- [19] H. A. Kestler, C. Wawra, B. Kracher, and M. Köhl, "Network modeling of signal transduction: Establishing the global view," *BioEssays*, vol. 30, no. 11-12, pp. 1110–1125, 2008.
- [20] H. de Jong, "Modeling and simulation of genetic regulatory systems: A literature review," *Journal of Computational Biology*, vol. 9, no. 1, pp. 67–103, 2002, PMID: 11911796.
- [21] G. M. Cooper, "Pathways of intracellular signal transduction," in *The Cell: A Molecular Approach*, 2nd ed. Sunderland, MA: Sinauer Associates, 2000, Available online.
- [22] R. Sever and J. S. Brugge, "Signal Transduction in Cancer," *Cold Spring Harbor Perspectives in Medicine*, vol. 5, no. 4, a006098–a006098, Apr. 1, 2015, ISSN: 2157-1422.
- [23] N. Borisov, E. Aksamitiene, A. Kiyatkin, S. Legewie, J. Berkhout, T. Maiwald, N. P. Kaimachnikov, J. Timmer, J. B. Hoek, and B. N. Kholodenko, "Systems-level interactions between insulin–EGF networks amplify mitogenic signaling," *Molecular Systems Biology*, vol. 5, no. 1, p. 256, Jan. 2009, ISSN: 1744-4292, 1744-4292.
- [24] A. Bertuzzi, S. Salinari, and G. Mingrone, "Insulin granule trafficking in β -cells: Mathematical model of glucose-induced insulin secretion," *American Journal of Physiology-Endocrinology and Metabolism*, vol. 293, no. 1, E396–E409, Jul. 2007, ISSN: 0193-1849, 1522-1555.

- [25] C. Y. Huang and J. E. Ferrell, “Ultrasensitivity in the mitogen-activated protein kinase cascade,” *Proceedings of the National Academy of Sciences*, vol. 93, no. 19, pp. 10 078–10 083, Sep. 17, 1996, ISSN: 0027-8424, 1091-6490.
- [26] M. Ordyan, T. Bartol, M. Kennedy, P. Rangamani, and T. Sejnowski, “Interactions between calmodulin and neurogranin govern the dynamics of CaMKII as a leaky integrator,” *PLOS Computational Biology*, vol. 16, no. 7, e1008015, Jul. 17, 2020, ISSN: 1553-7358.
- [27] E. M. Shockley, C. A. Rouzer, L. J. Marnett, E. J. Deeds, and C. F. Lopez, “Signal integration and information transfer in an allosterically regulated network,” *npj Systems Biology and Applications*, vol. 5, no. 1, p. 23, Jul. 18, 2019, ISSN: 2056-7189.
- [28] L. Qiao, W. Zhao, C. Tang, Q. Nie, and L. Zhang, “Network Topologies That Can Achieve Dual Function of Adaptation and Noise Attenuation,” *Cell Systems*, vol. 9, no. 3, 271–285.e7, Sep. 2019, ISSN: 24054712.
- [29] U. S. Bhalla and R. Iyengar, “Robustness of the bistable behavior of a biological signaling feedback loop,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 11, no. 1, pp. 221–226, 2001.
- [30] H. Ryu, M. Chung, M. Dobrzyński, D. Fey, Y. Blum, S. S. Lee, M. Peter, B. N. Kholodenko, N. L. Jeon, and O. Pertz, “Frequency modulation of ERK activation dynamics rewires cell fate,” *Molecular Systems Biology*, vol. 11, no. 11, p. 838, Nov. 2015, ISSN: 1744-4292, 1744-4292.
- [31] F. A. Brightman and D. A. Fell, “Differential feedback regulation of the MAPK cascade underlies the quantitative differences in EGF and NGF signalling in PC12 cells,” *FEBS Letters*, vol. 482, no. 3, pp. 169–174, Oct. 6, 2000, ISSN: 0014-5793, 1873-3468.
- [32] H. Lavoie, J. Gagnon, and M. Therrien, “ERK signalling: A master regulator of cell behaviour, life and fate,” *Nature Reviews Molecular Cell Biology*, vol. 21, no. 10, pp. 607–632, Oct. 2020, ISSN: 1471-0080.
- [33] L. Qiao, A. Khalilimeybodi, N. J. Linden-Santangeli, and P. Rangamani. “The evolution of systems biology and systems medicine: From mechanistic models to uncertainty quantification.” (Aug. 9, 2024), [Online]. Available: <http://arxiv.org/abs/2408.05395> (visited on 08/17/2024), pre-published.
- [34] B. N. Kholodenko, O. V. Demin, G. Moehren, and J. B. Hoek, “Quantification of Short Term Signaling by the Epidermal Growth Factor Receptor,” *Journal of Biological Chemistry*, vol. 274, no. 42, pp. 30 169–30 181, Oct. 1999, ISSN: 00219258.
- [35] B. N. Kholodenko, J. F. Hancock, and W. Kolch, “Signalling ballet in space and time,” *Nat. Rev. Mol. Cell Biol.*, vol. 11, no. 6, pp. 414–426, Jun. 2010.
- [36] U. S. Bhalla and R. Iyengar, “Emergent Properties of Networks of Biological Signaling Pathways,” *Science*, vol. 283, no. 5400, pp. 381–387, Jan. 1999.

- [37] E. Klipp and W. Liebermeister, “Mathematical modeling of intracellular signaling pathways,” *BMC Neuroscience*, vol. 7, no. 1, S10, Oct. 2006, ISSN: 1471-2202.
- [38] B. Schoeberl, C. Eichler-Jonsson, E. D. Gilles, and G. Müller, “Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors,” *Nature Biotechnology*, vol. 20, no. 4, pp. 370–375, Apr. 2002, ISSN: 1087-0156, 1546-1696.
- [39] T. Lipniacki, P. Paszek, A. R. Brasier, B. Luxon, and M. Kimmel, “Mathematical model of nf- κ b regulatory module,” *Journal of Theoretical Biology*, vol. 228, no. 2, pp. 195–215, 2004, ISSN: 0022-5193.
- [40] S. Basak, M. Behar, and A. Hoffmann, “Lessons from mathematically modeling the NF- κ B pathway,” *Immunological Reviews*, vol. 246, no. 1, pp. 221–238, 2012, ISSN: 1600-065X.
- [41] G. Dupont, L. Combettes, G. S. Bird, and J. W. Putney, “Calcium oscillations,” *Cold Spring Harbor Perspectives in Biology*, vol. 3, no. 3, 2011.
- [42] D. Ohadi and P. Rangamani, “Geometric Control of Frequency Modulation of cAMP Oscillations due to Calcium in Dendritic Spines,” *Biophysical Journal*, vol. 117, no. 10, pp. 1981–1994, Nov. 19, 2019, ISSN: 0006-3495, 1542-0086.
- [43] F. Pappalardo, G. Russo, S. Candido, M. Pennisi, S. Cavalieri, S. Motta, J. A. McCubrey, F. Nicoletti, and M. Libra, “Computational modeling of PI3K/AKT and MAPK signaling pathways in melanoma cancer,” *PLOS ONE*, vol. 11, no. 3, pp. 1–10, Mar. 2016.
- [44] R. J. Orton, O. E. Sturm, V. Vyshemirsky, M. Calder, D. R. Gilbert, and W. Kolch, “Computational modelling of the receptor-tyrosine-kinase-activated MAPK pathway,” *Biochemical Journal*, vol. 392, no. 2, pp. 249–261, Dec. 1, 2005, ISSN: 0264-6021, 1470-8728.
- [45] A. Von Kriegsheim, D. Baiocchi, M. Birtwistle, D. Sumpton, W. Bienvenut, N. Morrice, K. Yamada, A. Lamond, G. Kalna, R. Orton, D. Gilbert, and W. Kolch, “Cell fate decisions are specified by the dynamic ERK interactome,” *Nature Cell Biology*, vol. 11, no. 12, pp. 1458–1464, Dec. 2009, ISSN: 1465-7392, 1476-4679.
- [46] A. Khalilimeybodi, S. Fraley, and P. Rangamani, “Mechanisms underlying divergent relationships between Ca^{2+} and YAP/TAZ signalling,” *The Journal of Physiology*, vol. 601, no. 3, pp. 483–515, 2023.
- [47] K. E. Scott, S. I. Fraley, and P. Rangamani, “A spatial model of YAP/TAZ signaling reveals how stiffness, dimensionality, and shape contribute to emergent outcomes,” *Proceedings of the National Academy of Sciences*, vol. 118, no. 20, e2021571118, May 18, 2021.
- [48] E. A. Francis and P. Rangamani, “Computational modeling establishes mechanotransduction as a potent modulator of the mammalian circadian clock,” *Journal of Cell Science*, vol. 137, no. 17, 2024.

- [49] I. F. Coccimiglio and D. C. Clarke, “ADP is the dominant controller of AMP-activated protein kinase activity dynamics in skeletal muscle during exercise,” *PLOS Computational Biology*, vol. 16, no. 7, D. A. Beard, Ed., e1008079, Jul. 30, 2020, ISSN: 1553-7358.
- [50] N. M. C. Connolly, B. D’Orsi, N. Monsefi, H. J. Huber, and J. H. M. Prehn, “Computational Analysis of AMPK-Mediated Neuroprotection Suggests Acute Excitotoxic Bioenergetics and Glucose Dynamics Are Regulated by a Minimal Set of Critical Reactions,” *PLOS ONE*, vol. 11, no. 2, M. Jekabsons, Ed., e0148326, Feb. 3, 2016, ISSN: 1932-6203.
- [51] A. Leung and P. Rangamani, “Computational modeling of AMPK and mTOR crosstalk in glutamatergic synapse calcium signaling,” *npj Systems Biology and Applications*, vol. 9, no. 1, pp. 1–15, Jul. 17, 2023, ISSN: 2056-7189.
- [52] J. Z. Zhang, T.-W. Lu, L. M. Stolerman, B. Tenner, J. R. Yang, J.-F. Zhang, M. Falcke, P. Rangamani, S. S. Taylor, S. Mehta, *et al.*, “Phase separation of a pka regulatory subunit controls camp compartmentation and oncogenic signaling,” *Cell*, vol. 182, no. 6, pp. 1531–1544, 2020.
- [53] R. J. Orton, M. E. Adriaens, A. Gormand, O. E. Sturm, W. Kolch, and D. R. Gilbert, “Computational modelling of cancerous mutations in the EGFR/ERK signalling pathway,” *BMC Systems Biology*, vol. 3, no. 1, p. 100, Dec. 2009, ISSN: 1752-0509.
- [54] C. Dalla Man, R. A. Rizza, and C. Cobelli, “Meal Simulation Model of the Glucose-Insulin System,” *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 10, pp. 1740–1749, Oct. 2007, ISSN: 0018-9294.
- [55] C. Dalla Man, F. Micheletto, A. Sathananthan, R. A. Rizza, A. Vella, and C. Cobelli, “A model of GLP-1 action on insulin secretion in nondiabetic subjects,” *American Journal of Physiology-Endocrinology and Metabolism*, vol. 298, no. 6, E1115–E1121, Jun. 2010, ISSN: 0193-1849, 1522-1555.
- [56] R. D. Mortlock, S. K. Georgia, and S. D. Finley, “Dynamic Regulation of JAK-STAT Signaling Through the Prolactin Receptor Predicted by Computational Modeling,” *Cellular and Molecular Bioengineering*, vol. 14, no. 1, pp. 15–30, Feb. 2021, ISSN: 1865-5025, 1865-5033.
- [57] S. R. Neves, P. Tsokas, A. Sarkar, E. A. Grace, P. Rangamani, S. M. Taubenfeld, C. M. Alberini, J. C. Schaff, R. D. Blitzer, I. I. Moraru, *et al.*, “Cell shape and negative links in regulatory motifs together control spatial information flow in signaling networks,” *Cell*, vol. 133, no. 4, pp. 666–680, 2008.
- [58] M. K. Bell and P. Rangamani, “Crosstalk between biochemical signalling network architecture and trafficking governs AMPAR dynamics in synaptic plasticity,” *The Journal of Physiology*, vol. 601, no. 15, pp. 3377–3402, 2023, ISSN: 1469-7793.
- [59] M. Bell, T. Bartol, T. Sejnowski, and P. Rangamani, “Dendritic spine geometry and spine apparatus organization govern the spatiotemporal dynamics of calcium,” *Journal of General Physiology*, vol. 151, no. 8, pp. 1017–1034, Jul. 19, 2019, ISSN: 0022-1295.

- [60] K. M. Pearce, M. Bell, W. H. Linthicum, Q. Wen, J. Srinivasan, P. Rangamani, and S. Scarlata, "G α q-mediated calcium dynamics and membrane tension modulate neurite plasticity," *Molecular Biology of the Cell*, vol. 31, no. 7, J. York, Ed., pp. 683–694, Mar. 19, 2020, ISSN: 1059-1524, 1939-4586.
- [61] O. Wolkenhauer, P. Wellstead, K.-H. Cho, P. Rangamani, and R. Iyengar, "Modelling cellular signalling systems," *Essays in Biochemistry*, vol. 45, pp. 83–94, 2008.
- [62] Z. Zi, "A tutorial on mathematical modeling of biological signaling pathways," *Computational Modeling of Signaling Networks*, pp. 41–51, 2012.
- [63] U. S. Bhalla and R. Iyengar, "Emergent properties of networks of biological signaling pathways," *Science*, vol. 283, no. 5400, pp. 381–387, Jan. 1999.
- [64] I. H. Segel, *Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems*. Wiley, May 1975, ISBN: 978-0-471-77425-9.
- [65] G. Borrego-Yaniz, L. C. Terrón-Camero, M. Kerick, E. Andrés-León, and J. Martin, "A holistic approach to understanding immune-mediated inflammatory diseases: Bioinformatic tools to integrate omics data," *Comput. Struct. Biotechnol. J.*, vol. 23, pp. 96–105, Dec. 2024.
- [66] A. Baysoy, Z. Bai, R. Satija, and R. Fan, "The technological landscape and applications of single-cell multi-omics," *Nature Reviews Molecular Cell Biology*, vol. 24, no. 10, pp. 695–713, 2023.
- [67] R. Y. Tsien, "The green fluorescent protein," *Annual review of biochemistry*, vol. 67, no. 1, pp. 509–544, 1998.
- [68] J. Zhang, R. E. Campbell, A. Y. Ting, and R. Y. Tsien, "Creating new fluorescent probes for cell biology," *Nature reviews Molecular cell biology*, vol. 3, no. 12, pp. 906–918, 2002.
- [69] M. S. Frei, S. Mehta, and J. Zhang, "Next-Generation Genetically Encoded Fluorescent Biosensors Illuminate Cell Signaling and Metabolism," Feb. 2024.
- [70] E. C. Greenwald, S. Mehta, and J. Zhang, "Genetically Encoded Fluorescent Biosensors Illuminate the Spatiotemporal Regulation of Signaling Networks," *Chemical Reviews*, vol. 118, no. 24, pp. 11 707–11 794, Dec. 2018, ISSN: 0009-2665, 1520-6890.
- [71] J. Keyes, A. Ganesan, O. Molinar-Inglis, A. Hamidzadeh, J. Zhang, M. Ling, J. Trejo, A. Levchenko, and J. Zhang, "Signaling diversity enabled by Rap1-regulated plasma membrane ERK with distinct temporal dynamics," *eLife*, vol. 9, e57410, May 26, 2020, ISSN: 2050-084X.
- [72] D. L. Schmitt, S. D. Curtis, A. C. Lyons, J.-f. Zhang, M. Chen, C. Y. He, S. Mehta, R. J. Shaw, and J. Zhang, "Spatial regulation of AMPK signaling revealed by a sensitive kinase activity reporter," *Nature Communications*, vol. 13, no. 1, p. 3856, Jul. 5, 2022, ISSN: 2041-1723.

- [73] E. C. Greenwald, R. K. Polanowska-Grabowska, and J. J. Saucerman, “Integrating Fluorescent Biosensor Data Using Computational Models,” in *Fluorescent Protein-Based Biosensors*, J. Zhang, Q. Ni, and R. H. Newman, Eds., vol. 1071, Totowa, NJ: Humana Press, 2014, pp. 227–248, ISBN: 978-1-62703-621-4 978-1-62703-622-1.
- [74] C. Posner, S. Mehta, and J. Zhang, “Fluorescent biosensor imaging meets deterministic mathematical modelling: Quantitative investigation of signalling compartmentalization,” *The Journal of Physiology*, vol. 601, no. 19, pp. 4227–4241, 2023, ISSN: 1469-7793.
- [75] S. T. Vittadello and M. P. Stumpf, “Open problems in mathematical biology,” *Mathematical Biosciences*, vol. 354, p. 108 926, Dec. 2022, ISSN: 00255564.
- [76] M. Stumpf, D. J. Balding, and M. Girolami, *Handbook of statistical systems biology*, M. P. H. Stumpf, D. J. Balding, and M. Girolami, Eds. Hoboken, NJ: Wiley-Blackwell, Oct. 2011.
- [77] N. Linden-Santangeli, J. Zhang, B. Kramer, and P. Rangamani. “Increasing certainty in systems biology models using Bayesian multimodel inference.” (Jun. 17, 2024), [Online]. Available: <https://www.biorxiv.org/content/10.1101/2024.06.16.599231v1> (visited on 07/01/2024), pre-published.
- [78] J. J. Saucerman, J. Zhang, J. C. Martin, L. X. Peng, A. E. Stenbit, R. Y. Tsien, and A. D. McCulloch, “Systems analysis of PKA-mediated phosphorylation gradients in live cardiac myocytes,” *Proceedings of the National Academy of Sciences*, vol. 103, no. 34, pp. 12 923–12 928, Aug. 22, 2006, ISSN: 0027-8424, 1091-6490.
- [79] M. C. Kennedy and A. O’Hagan, “Bayesian calibration of computer models,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 3, pp. 425–464, 2001.
- [80] A. C. Babbie, P. Kirk, and M. P. H. Stumpf, “Topological sensitivity analysis for systems biology,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 111, no. 52, pp. 18 507–18 512, Dec. 2014.
- [81] R. E. Morrison, T. A. Oliver, and R. D. Moser, “Representing model inadequacy: A stochastic operator approach,” *SIAM/ASA Journal on Uncertainty Quantification*, vol. 6, no. 2, pp. 457–496, 2018.
- [82] N. Galioto and A. A. Gorodetsky, “Bayesian system ID: Optimal management of parameter, model, and measurement uncertainty,” *Nonlinear Dyn.*, vol. 102, no. 1, pp. 241–267, Sep. 2020.
- [83] A. Raue, M. Schilling, J. Bachmann, A. Matteson, M. Schelker, D. Kaschek, S. Hug, C. Kreutz, B. D. Harms, F. J. Theis, U. Klingmüller, and J. Timmer, “Lessons learned from quantitative dynamical modeling in systems biology,” *PLoS One*, vol. 8, no. 9, e74335, Sep. 2013.
- [84] A. Gelman, J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin, *Bayesian data analysis*, third. CRC Press, 2013.
- [85] E. D. Mitra and W. S. Hlavacek, “Parameter estimation and uncertainty quantification for systems biology models,” *Curr Opin Syst Biol*, vol. 18, pp. 9–18, Dec. 2019.

- [86] A. C. Babbie and M. P. H. Stumpf, “How to deal with parameters for whole-cell modelling,” *J. R. Soc. Interface*, vol. 14, no. 133, Aug. 2017.
- [87] A. Yazdani, L. Lu, M. Raissi, and G. E. Karniadakis, “Systems biology informed deep learning for inferring parameters and hidden dynamics,” *PLoS Comput. Biol.*, vol. 16, no. 11, e1007575, Nov. 2020.
- [88] R. N. Gutenkunst, J. J. Waterfall, F. P. Casey, K. S. Brown, C. R. Myers, and J. P. Sethna, “Universally sloppy parameter sensitivities in systems biology models,” *PLoS Comput. Biol.*, vol. 3, no. 10, pp. 1871–1878, Oct. 2007.
- [89] F.-G. Wieland, A. L. Hauber, M. Rosenblatt, C. Tönsing, and J. Timmer, “On structural and practical identifiability,” *Current Opinion in Systems Biology*, vol. 25, pp. 60–69, Mar. 2021.
- [90] G. Lillacci and M. Khammash, “Parameter estimation and model selection in computational biology,” *PLoS Comput. Biol.*, vol. 6, no. 3, e1000696, Mar. 2010.
- [91] M. B. Stone, S. A. Shelby, and S. L. Veatch, “Super-Resolution Microscopy: Shedding Light on the Cellular Plasma Membrane,” *Chemical Reviews*, vol. 117, no. 11, pp. 7457–7477, Jun. 2017, ISSN: 0009-2665.
- [92] M. Thattai and A. Van Oudenaarden, “Intrinsic noise in gene regulatory networks,” *Proceedings of the National Academy of Sciences*, vol. 98, no. 15, pp. 8614–8619, 2001.
- [93] A. M. Stuart, “Inverse problems: A Bayesian perspective,” *Acta numerica*, vol. 19, pp. 451–559, 2010.
- [94] “Large-Scale Inverse Problems and Quantification of Uncertainty | Wiley,” Wiley.com. (), [Online]. Available: <https://www.wiley.com/en-us/Large-Scale+Inverse+Problems+and+Quantification+of+Uncertainty-p-9780470697436> (visited on 02/27/2025).
- [95] D. Blatter, M. Morzfeld, K. Key, and S. Constable, “Uncertainty quantification for regularized inversion of electromagnetic geophysical data—Part I: Motivation and theory,” *Geophysical Journal International*, vol. 231, no. 2, pp. 1057–1074, Nov. 1, 2022, ISSN: 0956-540X.
- [96] J. Hobbs, A. Braverman, N. Cressie, R. Granat, and M. Gunson, “Simulation-Based Uncertainty Quantification for Estimating Atmospheric CO₂ from Satellite Data,” *SIAM/ASA Journal on Uncertainty Quantification*, vol. 5, no. 1, pp. 956–985, Jan. 2017, ISSN: 2166-2525.
- [97] S. M. Hirsh, D. A. Barajas-Solano, and J. N. Kutz, “Sparsifying priors for Bayesian uncertainty quantification in model discovery,” *Royal Society Open Science*, vol. 9, no. 2, p. 211 823, Feb. 2022, ISSN: 2054-5703.
- [98] M. Heinkenschloss, B. Kramer, T. Takhtaganov, and K. Willcox, “Conditional-Value-at-Risk Estimation via Reduced-Order Models,” *SIAM/ASA Journal on Uncertainty Quantification*, vol. 6, no. 4, pp. 1395–1423, Jan. 2018, ISSN: 2166-2525.

- [99] B. Peherstorfer, B. Kramer, and K. Willcox, “Multifidelity Preconditioning of the Cross-Entropy Method for Rare Event Simulation and Failure Probability Estimation,” *SIAM/ASA Journal on Uncertainty Quantification*, vol. 6, no. 2, pp. 737–761, Jan. 2018, ISSN: 2166-2525.
- [100] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola, *Global Sensitivity Analysis. The Primer*, 1st ed. Wiley, Dec. 2007.
- [101] G. Qian and A. Mahdi, “Sensitivity analysis methods in the biomedical sciences,” *Mathematical Biosciences*, vol. 323, p. 108 306, May 1, 2020, ISSN: 0025-5564.
- [102] M. Clyde and E. I. George, “Model Uncertainty,” *Statistical Science*, vol. 19, no. 1, Feb. 1, 2004, ISSN: 0883-4237.
- [103] M. Chung, M. Binois, R. B. Gramacy, J. M. Bardsley, D. J. Moquin, A. P. Smith, and A. M. Smith, “Parameter and Uncertainty Estimation for Dynamical Systems Using Surrogate Stochastic Processes,” *SIAM Journal on Scientific Computing*, vol. 41, no. 4, A2212–A2238, Jan. 2019, ISSN: 1064-8275, 1095-7197.
- [104] I.-G. Farcas, B. Peherstorfer, T. Neckel, F. Jenko, and H.-J. Bungartz. “Context-aware learning of hierarchies of low-fidelity models for multi-fidelity uncertainty quantification.” (Nov. 19, 2022), [Online]. Available: <http://arxiv.org/abs/2211.10835> (visited on 05/04/2024), pre-published.
- [105] M. Renardy, T.-M. Yi, D. Xiu, and C.-S. Chou, “Parameter uncertainty quantification using surrogate models applied to a spatial model of yeast mating polarization,” *PLOS Computational Biology*, vol. 14, no. 5, M. Johnson, Ed., e1006181, May 29, 2018, ISSN: 1553-7358.
- [106] W. A. Pruett and R. L. Hester, “The Creation of Surrogate Models for Fast Estimation of Complex Model Outcomes,” *PLOS ONE*, vol. 11, no. 6, J. Boltze, Ed., e0156574, Jun. 3, 2016, ISSN: 1932-6203.
- [107] E. Qian, B. Kramer, B. Peherstorfer, and K. Willcox, “Lift & Learn: Physics-informed machine learning for large-scale nonlinear dynamical systems,” *Physica D: Nonlinear Phenomena*, vol. 406, p. 132 401, May 2020, ISSN: 01672789.
- [108] A. Gelman and X.-L. Meng, “Simulating normalizing constants: From importance sampling to bridge sampling to path sampling,” *Stat. Sci.*, vol. 13, no. 2, pp. 163–185, May 1998.
- [109] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes 3rd Edition: The Art of Scientific Computing*, 3rd ed. USA: Cambridge University Press, 2007, ISBN: 0521880688.
- [110] A. B. Owen, *Monte Carlo theory, methods and examples*. 2013.
- [111] G. M. Martin, D. T. Frazier, and C. P. Robert. “Computing Bayes: Bayesian Computation from 1763 to the 21st Century.” (Dec. 5, 2020), [Online]. Available: <http://arxiv.org/abs/2004.06425> (visited on 05/04/2024), pre-published.

- [112] A. Gelman, J. Hill, and A. Vehtari, *Regression and Other Stories* (Analytical Methods for Social Research). Cambridge University Press, 2021, ISBN: 978-1-107-02398-7.
- [113] N. J. Eungdamrong and R. Iyengar, “Computational approaches for modeling regulatory cellular networks,” *Trends in Cell Biology*, vol. 14, no. 12, pp. 661–669, 2004.
- [114] A. Lipshtat, G. Jayaraman, J. C. He, and R. Iyengar, “Design of versatile biochemical switches that respond to amplitude, duration, and spatial cues,” *Proceedings of the National Academy of Sciences*, vol. 107, no. 3, pp. 1247–1252, 2010.
- [115] A. Ma’ayan, R. D. Blitzer, and R. Iyengar, “Toward predictive models of mammalian cells,” *Annu. Rev. Biophys. Biomol. Struct.*, vol. 34, pp. 319–349, 2005.
- [116] M. Song, D. Li, S. Z. Makaryan, and S. D. Finley, “Quantitative modeling to understand cell signaling in the tumor microenvironment,” *Current Opinion in Systems Biology*, vol. 27, p. 100 345, Sep. 2021.
- [117] A. K. Chakraborty, “A perspective on the role of computational models in immunology,” *Annu. Rev. Immunol.*, vol. 35, pp. 403–439, Apr. 2017.
- [118] S. Haney, L. Bardwell, and Q. Nie, “Ultrasensitive responses and specificity in cell signaling,” *BMC systems biology*, vol. 4, no. 1, pp. 1–14, 2010.
- [119] L. Qiao, W. Zhao, C. Tang, Q. Nie, and L. Zhang, “Network topologies that can achieve dual function of adaptation and noise attenuation,” *Cell systems*, vol. 9, no. 3, pp. 271–285, 2019.
- [120] L. Geris and D. Gomez-Cabrero, Eds., *Uncertainty in biology: A computational modeling approach*. Springer, Cham, 2016.
- [121] M. Ashyraliyev, Y. Fomekong-Nanfack, J. A. Kaandorp, and J. G. Blom, “Systems biology: Parameter estimation for biochemical models,” *FEBS J.*, vol. 276, no. 4, pp. 886–902, Feb. 2009.
- [122] G. I. Valderrama-Bahamóndez and H. Fröhlich, “MCMC techniques for parameter estimation of ODE based models in systems biology,” *Frontiers in Applied Mathematics and Statistics*, vol. 5, p. 55, 2019.
- [123] J. T. Oden, I. Babuška, and D. Faghihi, “Predictive computational science: Computer predictions in the presence of uncertainty,” *Encyclopedia of Computational Mechanics Second Edition*, pp. 1–26, 2017.
- [124] T. J. Sullivan, *Introduction to uncertainty quantification*. Springer, 2015, vol. 63.
- [125] L. Schmiester, Y. Schälte, F. T. Bergmann, T. Camba, E. Dudkin, J. Egert, F. Fröhlich, L. Fuhrmann, A. L. Hauber, S. Kemmer, *et al.*, “PETab–interoperable specification of parameter estimation problems in systems biology,” *PLoS computational biology*, vol. 17, no. 1, e1008646, 2021.

- [126] E. Kent, S. Neumann, U. Kummer, and P. Mendes, “What can we learn from global sensitivity analysis of biochemical systems?” *PLoS One*, vol. 8, no. 11, e79244, Nov. 2013.
- [127] F. Anstett-Collin, L. Denis-Vidal, and G. Millérioux, “A priori identifiability: An overview on definitions and approaches,” *Annu. Rev. Control*, vol. 50, pp. 139–149, Jan. 2020.
- [128] 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.*, vol. 7, no. 5, pp. 1593–1602, May 2011.
- [129] 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*, vol. 25, no. 15, pp. 1923–1929, Aug. 2009.
- [130] H. Hong, A. Ovchinnikov, G. Pogudin, and C. Yap, “Global identifiability of differential models,” *Commun. Pure Appl. Math.*, vol. 73, no. 9, pp. 1831–1879, Sep. 2020.
- [131] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola, *Global sensitivity analysis. The primer*. Wiley, 2008.
- [132] A. Varma, M. Morbidelli, and H. Wu, *Parametric sensitivity in chemical systems* (Cambridge Series in Chemical Engineering). Cambridge University Press, 1999.
- [133] S. Marino, I. B. Hogue, C. J. Ray, and D. E. Kirschner, “A methodology for performing global uncertainty and sensitivity analysis in systems biology,” *J. Theor. Biol.*, vol. 254, no. 1, pp. 178–196, Sep. 2008.
- [134] R. D. Mortlock, S. K. Georgia, and S. D. Finley, “Dynamic regulation of JAK-STAT signaling through the prolactin receptor predicted by computational modeling,” *Cell. Mol. Bioeng.*, vol. 14, no. 1, pp. 15–30, Feb. 2021.
- [135] S. M. Kay, *Fundamentals of statistical signal processing: estimation theory*. Prentice-Hall, Inc., 1993.
- [136] T. K. Moon and W. C. Stirling, *Mathematical methods and algorithms for signal processing*. Prentice Hall, 2000.
- [137] J. Liepe, P. Kirk, S. Filippi, C. P. Toni Tinaand Barnes, and M. P. H. Stumpf, “A framework for parameter estimation and model selection from experimental data in systems biology using approximate Bayesian computation,” *Nat. Protoc.*, vol. 9, no. 2, pp. 439–456, Feb. 2014.
- [138] T. Toni, D. Welch, N. Strelkowa, A. Ipsen, and M. P. H. Stumpf, “Approximate Bayesian computation scheme for parameter inference and model selection in dynamical systems,” *J. R. Soc. Interface*, vol. 6, no. 31, pp. 187–202, Feb. 2009.

- [139] A. Golightly and D. J. Wilkinson, “Bayesian parameter inference for stochastic biochemical network models using particle markov chain monte carlo,” *Interface Focus*, vol. 1, no. 6, pp. 807–820, Dec. 2011.
- [140] O. Ghasemi, M. L. Lindsey, T. Yang, N. Nguyen, Y. Huang, and Y.-F. Jin, “Bayesian parameter estimation for nonlinear modelling of biological pathways,” *BMC Syst. Biol.*, vol. 5 Suppl 3, S9, Dec. 2011.
- [141] F. Bianconi, C. Antonini, L. Tomassoni, and P. Valigi, “Application of conditional robust calibration to ordinary differential equations models in computational systems biology: A comparison of two sampling strategies,” *IET Syst. Biol.*, vol. 14, no. 3, pp. 107–119, Jun. 2020.
- [142] D. J. Wilkinson, “Bayesian methods in bioinformatics and computational systems biology,” *Brief. Bioinform.*, vol. 8, no. 2, pp. 109–116, Mar. 2007.
- [143] D. J. Klinken 2nd, “An empirical Bayesian approach for model-based inference of cellular signaling networks,” *BMC Bioinformatics*, vol. 10, p. 371, Nov. 2009.
- [144] M. Renardy, T.-M. Yi, D. Xiu, and C.-S. Chou, “Parameter uncertainty quantification using surrogate models applied to a spatial model of yeast mating polarization,” *PLoS Comput. Biol.*, vol. 14, no. 5, e1006181, May 2018.
- [145] K. Erazo and S. Nagarajaiah, “An offline approach for output-only Bayesian identification of stochastic nonlinear systems using unscented Kalman filtering,” *J. Sound Vib.*, vol. 397, pp. 222–240, Jun. 2017.
- [146] H. Hong, A. Ovchinnikov, G. Pogudin, and C. Yap, “SIAN: Software for structural identifiability analysis of ODE models,” *Bioinformatics*, vol. 35, no. 16, pp. 2873–2874, Aug. 2019.
- [147] I. M. Sobol, “Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates,” *Math. Comput. Simul.*, vol. 55, pp. 271–280, 2001.
- [148] B. Teixeira, L. A. B. Torres, L. A. Aguirre, and D. S. Bernstein, “Unscented filtering for interval-constrained nonlinear systems,” in *Proceedings of the 47th IEEE Conference on Decision and Control, CDC 2008, December 9-11, 2008, Cancún, México*, Institute of Electrical and Electronics Engineers, Jan. 2008, pp. 5116–5121.
- [149] A. F. Villaverde, N. D. Evans, M. J. Chappell, and J. R. Banga, “Input-Dependent structural identifiability of nonlinear systems,” *IEEE Control Systems Letters*, vol. 3, no. 2, pp. 272–277, Apr. 2019.
- [150] L. K. Nguyen, A. Degasperis, P. Cotter, and B. N. Kholodenko, “DYVIPAC: An integrated analysis and visualisation framework to probe multi-dimensional biological networks,” *Sci. Rep.*, vol. 5, p. 12 569, Jul. 2015.

- [151] H. J. Pi and J. E. Lisman, “Coupled phosphatase and kinase switches produce the tristability required for long-term potentiation and long-term depression,” *J. Neurosci.*, vol. 28, no. 49, pp. 13 132–13 138, Dec. 2008.
- [152] H. Haario, M. Laine, A. Mira, and E. Saksman, “DRAM: Efficient adaptive MCMC,” *Stat. Comput.*, vol. 16, no. 4, pp. 339–354, Dec. 2006.
- [153] J. Goodman and J. Weare, “Ensemble samplers with affine invariance,” *Communications in Applied Mathematics and Computational Science*, vol. 5, no. 1, pp. 65–80, 2010.
- [154] J. Norton, E. Walter, and L. Pronzato, *Identification of parametric models from experimental data* (Communications and Control Engineering). Springer London, 1997, ISBN: 9781849969963.
- [155] A. F. Villaverde, “Observability and structural identifiability of nonlinear biological systems,” *Complexity*, vol. 2019, Jan. 2019.
- [156] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, “Julia: A fresh approach to numerical computing,” *SIAM review*, vol. 59, no. 1, pp. 65–98, 2017.
- [157] I. Ilia, A. Ovchinnikov, and G. Pogudin, *Sian.jl-Implementation of SIAN in Julia*, <https://github.com/alexeyovchinnikov/SIAN-Julia>, 2022.
- [158] S. Marelli and B. Sudret, “Uqlab: A framework for uncertainty quantification in matlab,” in *Vulnerability, uncertainty, and risk: quantification, mitigation, and management*, 2014, pp. 2554–2563.
- [159] S. Marelli, C. Lamas, K. Konakli, C. Mylonas, P. Wiederkehr, and B. Sudret, “UQLab user manual – Sensitivity analysis,” Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich, Switzerland, Tech. Rep., 2022, Report UQLab-V2.0-106.
- [160] C. Rackauckas and Q. Nie, “DifferentialEquations.jl—a performant and feature-rich ecosystem for solving differential equations in julia,” *Journal of Open Research Software*, vol. 5, no. 1, 2017.
- [161] S. Särkkä, *Bayesian filtering and smoothing*. Cambridge University Press, 2013.
- [162] S. J. Julier and J. K. Uhlmann, “Unscented filtering and nonlinear estimation,” *Proc. IEEE*, vol. 92, no. 3, pp. 401–422, Mar. 2004.
- [163] P. Vachhani, S. Narasimhan, and R. Rengaswamy, “Robust and reliable estimation via unscented recursive nonlinear dynamic data reconciliation,” *J. Process Control*, vol. 16, no. 10, pp. 1075–1086, Dec. 2006.
- [164] D. Simon, “Kalman filtering with state constraints: A survey of linear and nonlinear algorithms,” *IET Control Theory Appl.*, vol. 4, no. 8, pp. 1303–1318, Aug. 2010.

- [165] S. J. Julier and J. K. Uhlmann, “New extension of the kalman filter to nonlinear systems,” in *Signal processing, sensor fusion, and target recognition VI*, International Society for Optics and Photonics, vol. 3068, 1997, pp. 182–193.
- [166] A. Tsigkinopoulou, A. Hawari, M. Uttley, and R. Breitling, “Defining informative priors for ensemble modeling in systems biology,” *Nat. Protoc.*, vol. 13, no. 11, pp. 2643–2663, Nov. 2018.
- [167] A. Gelman, “Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper),” *Bayesian Analysis*, vol. 1, no. 3, pp. 515–534, Sep. 2006.
- [168] A. Gelman, G. Roberts, and W. Gilks, “Efficient Metropolis jumping rules,” *Bayesian statistics*, 1996.
- [169] A. Sokal, “Monte carlo methods in statistical mechanics: Foundations and new algorithms,” in *Functional integration*, Springer, 1997, pp. 131–192.
- [170] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, “Equation of state calculations by fast computing machines,” *The journal of chemical physics*, vol. 21, no. 6, pp. 1087–1092, 1953.
- [171] W. K. Hastings, “Monte Carlo sampling methods using Markov chains and their applications,” *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.
- [172] L. Tierney, “Markov chains for exploring posterior distributions,” *the Annals of Statistics*, pp. 1701–1728, 1994.
- [173] H. Haario, E. Saksman, and J. Tamminen, “An adaptive metropolis algorithm,” *Bernoulli*, pp. 223–242, 2001.
- [174] P.-R. Wagner, J. Nagel, S. Marelli, and B. Sudret, “UQLab user manual–Bayesian inversion for model calibration and validation,” Chair of Risk, Safety and Uncertainty Quantification, ETH Zurich, Switzerland, Tech. Rep., 2022, Report UQLab-V2.0-113.
- [175] U. Wolff, “Monte carlo errors with less errors,” *Comput. Phys. Commun.*, vol. 156, no. 2, pp. 143–153, Jan. 2004.
- [176] L. Mišković and V. Hatzimanikatis, “Modeling of uncertainties in biochemical reactions,” *Biotechnol. Bioeng.*, vol. 108, no. 2, pp. 413–423, Feb. 2011.
- [177] A. W. Bowman and A. Azzalini, *Applied smoothing techniques for data analysis: The kernel approach with S-Plus illustrations*. OUP Oxford, 1997, vol. 18.
- [178] H. Shankaran and H. S. Wiley, “Oscillatory dynamics of the extracellular signal-regulated kinase pathway,” *Curr. Opin. Genet. Dev.*, vol. 20, no. 6, pp. 650–655, Dec. 2010.

- [179] Y. D. Shaul and R. Seger, “The mek/erk cascade: From signaling specificity to diverse functions,” *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, vol. 1773, no. 8, pp. 1213–1226, 2007.
- [180] J. Vera, O. Rath, E. Balsa-Canto, J. R. Banga, W. Kolch, and O. Wolkenhauer, “Investigating dynamics of inhibitory and feedback loops in ERK signalling using power-law models,” *Mol. Biosyst.*, vol. 6, no. 11, pp. 2174–2191, Nov. 2010.
- [181] B. N. Kholodenko, “Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades,” *Eur. J. Biochem.*, vol. 267, no. 6, pp. 1583–1588, Mar. 2000.
- [182] S. Lunderman, M. Morzfeld, F. Glassmeier, and G. Feingold, “Estimating parameters of the nonlinear cloud and rain equation from a large-eddy simulation,” *Physica D*, vol. 410, p. 132 500, Sep. 2020.
- [183] M. T. Wentworth, R. C. Smith, and H. T. Banks, “Parameter selection and verification techniques based on global sensitivity analysis illustrated for an HIV model,” *SIAM/ASA J. Uncertainty Quantification*, vol. 4, no. 1, pp. 266–297, Jan. 2016.
- [184] S. T. Vittadello and M. P. H. Stumpf, “Open problems in mathematical biology,” *arXiv*, Jun. 2022.
- [185] K. D. Coleman, K. Schmidt, and R. C. Smith, “Frequentist and bayesian lasso techniques for parameter selection in nonlinearly parameterized models,” *IFAC-PapersOnLine*, vol. 49, no. 18, pp. 416–421, Jan. 2016.
- [186] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, “Variational inference: A review for statisticians,” *J. Am. Stat. Assoc.*, vol. 112, no. 518, pp. 859–877, Apr. 2017.
- [187] J. M. Bardsley, A. Solonen, H. Haario, and M. Laine, “Randomize-then-optimize: A method for sampling from posterior distributions in nonlinear inverse problems,” *SIAM Journal on Scientific Computing*, vol. 36, no. 4, A1895–A1910, 2014.
- [188] E. M. Shockley, J. A. Vrugt, and C. F. Lopez, “PyDREAM: High-dimensional parameter inference for biological models in python,” *Bioinformatics*, vol. 34, no. 4, pp. 695–697, Feb. 2018.
- [189] S. Gupta, L. Hainsworth, J. S. Hogg, R. E. C. Lee, and J. R. Faeder, “Evaluation of parallel tempering to accelerate bayesian parameter estimation in systems biology,” *Proc Euromicro Int Conf Parallel Distrib Netw Based Process*, vol. 2018, pp. 690–697, Mar. 2018.
- [190] T. Cui, K. J. H. Law, and Y. M. Marzouk, “Dimension-independent likelihood-informed MCMC,” *J. Comput. Phys.*, vol. 304, pp. 109–137, Jan. 2016.
- [191] T. Äijö and R. Bonneau, “Biophysically motivated regulatory network inference: Progress and prospects,” *Hum. Hered.*, vol. 81, no. 2, pp. 62–77, 2016.

- [192] J. Casadiego, M. Nitzan, S. Hallerberg, and M. Timme, “Model-free inference of direct network interactions from nonlinear collective dynamics,” *Nat. Commun.*, vol. 8, no. 1, p. 2192, Dec. 2017.
- [193] M. Schmidt and H. Lipson, “Distilling free-form natural laws from experimental data,” *Science*, vol. 324, no. 5923, pp. 81–85, Apr. 2009.
- [194] S. L. Brunton, J. L. Proctor, and J. N. Kutz, “Discovering governing equations from data by sparse identification of nonlinear dynamical systems,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 113, no. 15, pp. 3932–3937, Apr. 2016.
- [195] N. M. Mangan, S. L. Brunton, J. L. Proctor, and J. N. Kutz, “Inferring biological networks by sparse identification of nonlinear dynamics,” *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, vol. 2, no. 1, pp. 52–63, Jun. 2016.
- [196] M. Hoffmann, C. Fröhner, and F. Noé, “Reactive SINDy: Discovering governing reactions from concentration data,” *J. Chem. Phys.*, vol. 150, no. 2, p. 025 101, Jan. 2019.
- [197] K. Kaheman, J. N. Kutz, and S. L. Brunton, “SINDy-PI: A robust algorithm for parallel implicit sparse identification of nonlinear dynamics,” *Proc. Math. Phys. Eng. Sci.*, vol. 476, no. 2242, p. 20 200 279, Oct. 2020.
- [198] S. M. Hirsh, D. A. Barajas-Solano, and J. N. Kutz, “Sparsifying priors for Bayesian uncertainty quantification in model discovery,” *arXiv preprint arXiv:2107.02107*, 2021.
- [199] H. Gasca-Aragon, *Data combination from multiple sources under measurement error*. University of Massachusetts Amherst, 2012.
- [200] F. Marin-Martinez and J. Sánchez-Meca, “Weighting by inverse variance or by sample size in random-effects meta-analysis,” *Educational and Psychological Measurement*, vol. 70, no. 1, pp. 56–73, 2010.
- [201] R. T. Clemen and R. L. Winkler, “Combining probability distributions from experts in risk analysis,” *Risk analysis*, vol. 19, no. 2, pp. 187–203, 1999.
- [202] L. Schermelleh, A. Ferrand, T. Huser, C. Eggeling, M. Sauer, O. Biehlmaier, and G. P. C. Drummen, “Super-resolution microscopy demystified,” *Nature Cell Biology*, vol. 21, no. 1, pp. 72–84, Jan. 2019, ISSN: 1465-7392, 1476-4679.
- [203] B. N. Kholodenko, J. F. Hancock, and W. Kolch, “Signalling ballet in space and time,” *Nat. Rev. Mol. Cell Biol.*, vol. 11, no. 6, pp. 414–426, Jun. 2010.
- [204] M. Glont, T. V. N. Nguyen, M. Graesslin, R. Hälke, R. Ali, J. Schramm, S. M. Wimalaratne, V. B. Kothamachu, N. Rodriguez, M. J. Swat, J. Eils, R. Eils, C. Laibe, R. S. Malik-Sheriff, V. Chelliah, N. Le Novère, and H. Hermjakob, “Biomodels: Expanding horizons to include more modelling approaches and formats,” *Nucleic Acids Research*, vol. 46, no. D1, pp. D1248–D1253, 2018.

- [205] R. S. Malik-Sheriff, M. Glont, T. V. N. Nguyen, K. Tiwari, M. G. Roberts, A. Xavier, M. T. Vu, J. Men, M. Maire, S. Kananathan, E. L. Fairbanks, J. P. Meyer, C. Arankalle, T. M. Varusai, V. Knight-Schrijver, L. Li, C. Dueñas-Roca, G. Dass, S. M. Keating, Y. M. Park, N. Buso, N. Rodriguez, M. Hucka, and H. Hermjakob, “BioModels — 15 years of sharing computational models in life science,” *Nucleic Acids Research*, vol. 48, no. D1, pp. D407–D415, Jan. 2020, gkz1055, ISSN: 0305-1048.
- [206] R. C. Smith, *Uncertainty Quantification: Theory, Implementation, and Applications* (Computational Science and Engineering). Society for Industrial and Applied Mathematics, 2013.
- [207] L. Geris and D. Gomez-Cabrero, Eds., *Uncertainty in Biology: A Computational Modeling Approach*. Springer, Cham, 2016.
- [208] N. J. Linden, B. Kramer, and P. Rangamani, “Bayesian parameter estimation for dynamical models in systems biology,” *PLoS Comput. Biol.*, vol. 18, no. 10, e1010651, Oct. 2022.
- [209] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis*. CRC Press, 2014.
- [210] K. P. Burnham and D. R. Anderson, *Model selection and multimodel inference: A practical information-theoretic approach*, 2nd ed., K. P. Burnham and D. R. Anderson, Eds. New York, NY: Springer, Sep. 2002.
- [211] H. Akaike, “A new look at the statistical model identification,” *IEEE Trans. Automat. Contr.*, vol. 19, no. 6, pp. 716–723, Dec. 1974.
- [212] A. E. Raftery, M. Kárný, and P. Ettler, “Online Prediction Under Model Uncertainty via Dynamic Model Averaging: Application to a Cold Rolling Mill,” *Technometrics*, vol. 52, no. 1, pp. 52–66, Feb. 2010, ISSN: 0040-1706, 1537-2723.
- [213] P. Kirk, T. Thorne, and M. P. H. Stumpf, “Model selection in systems and synthetic biology,” *Curr. Opin. Biotechnol.*, vol. 24, no. 4, pp. 767–774, Aug. 2013.
- [214] A. Vehtari, A. Gelman, and J. Gabry, “Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC,” *Statistics and Computing*, vol. 27, no. 5, pp. 1413–1432, Aug. 2016.
- [215] J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky, “Bayesian model averaging: A tutorial,” *SSO Schweiz. Monatsschr. Zahnheilkd.*, vol. 14, no. 4, pp. 382–417, Nov. 1999.
- [216] M. P. H. Stumpf and T. Thorne, “Multi-model inference of network properties from incomplete data,” *J. Integr. Bioinform.*, vol. 3, no. 2, pp. 123–136, Dec. 2006.
- [217] M. P. H. Stumpf, “Multi-model and network inference based on ensemble estimates: Avoiding the madness of crowds,” *J. R. Soc. Interface*, vol. 17, no. 171, p. 20200419, Oct. 2020.

- [218] Y. Yao, A. Vehtari, D. Simpson, and A. Gelman, “Using stacking to average Bayesian predictive distributions (with discussion),” *Bayesian Anal.*, vol. 13, no. 3, pp. 917–1007, Sep. 2018.
- [219] R. T. Clemen and R. L. Winkler, “Combining probability distributions from experts in risk analysis,” *Risk Anal.*, vol. 19, no. 2, pp. 187–203, Apr. 1999.
- [220] R. L. Winkler, “Combining probability distributions from dependent information sources,” *Manage. Sci.*, vol. 27, no. 4, pp. 479–488, 1981.
- [221] M. Stone, “The opinion pool,” *Ann. Math. Stat.*, vol. 32, no. 4, pp. 1339–1342, 1961.
- [222] S. P. Beik, L. A. Harris, M. A. Kochen, J. Sage, V. Quaranta, and C. F. Lopez, “Unified tumor growth mechanisms from multimodel inference and dataset integration,” *PLoS Comput. Biol.*, vol. 19, no. 7, e1011215, Jul. 2023.
- [223] J. Keyes, A. Ganesan, O. Molinar-Inglis, A. Hamidzadeh, J. Zhang, M. Ling, J. Trejo, A. Levchenko, and J. Zhang, “Signaling diversity enabled by Rap1-regulated plasma membrane ERK with distinct temporal dynamics,” *Elife*, vol. 9, May 2020.
- [224] National Center for Biotechnology Information, *PubChem compound summary for CID 56841751, epidermal growth factor*, <https://pubchem.ncbi.nlm.nih.gov/compound/Epidermal-growth-factor>, Accessed: 2023-10-3, 2023.
- [225] R. Milo, P. Jorgensen, U. Moran, G. Weber, and M. Springer, “BioNumbers—the database of key numbers in molecular and cell biology,” *Nucleic Acids Res.*, vol. 38, no. Database issue, pp. D750–3, Jan. 2010.
- [226] M. R. Birtwistle, M. Hatakeyama, N. Yumoto, B. A. Ogunnaike, J. B. Hoek, and B. N. Kholodenko, “Ligand-dependent responses of the ErbB signaling network: Experimental and modeling analyses,” *Mol. Syst. Biol.*, vol. 3, no. 1, p. 144, Nov. 2007.
- [227] A. von Kriegsheim, D. Baiocchi, M. Birtwistle, D. Sumpton, W. Bienvenut, N. Morrice, K. Yamada, A. Lamond, G. Kalna, R. Orton, D. Gilbert, and W. Kolch, “Cell fate decisions are specified by the dynamic ERK interactome,” *Nat. Cell Biol.*, vol. 11, no. 12, pp. 1458–1464, Dec. 2009.
- [228] C. Y. Huang and J. E. Ferrell Jr, “Ultrasensitivity in the mitogen-activated protein kinase cascade,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 93, no. 19, pp. 10 078–10 083, Sep. 1996.
- [229] S. D. M. Santos, P. J. Verveer, and P. I. H. Bastiaens, “Growth factor-induced MAPK network topology shapes erk response determining PC-12 cell fate,” *Nat. Cell Biol.*, vol. 9, no. 3, pp. 324–330, Mar. 2007.
- [230] J. Ching and Y.-C. Chen, “Transitional Markov Chain Monte Carlo Method for Bayesian Model Updating, Model Class Selection, and Model Averaging,” *Journal of Engineering Mechanics*, vol. 133, no. 7, pp. 816–832, Jul. 2007, ISSN: 0733-9399, 1943-7889.

- [231] S. E. Minson, M. Simons, and J. L. Beck, “Bayesian inversion for finite fault earthquake source models I—theory and algorithm,” *Geophysical Journal International*, vol. 194, no. 3, pp. 1701–1726, Sep. 2013, ISSN: 1365-246X, 0956-540X.
- [232] A.-P. Oriol, A. Virgile, C. Colin, D. Larry, F. C. J., K. Maxim, K. Ravin, L. Jupeng, L. C. C., M. O. A., O. Michael, V. Ricardo, W. Thomas, and Z. Robert, “PyMC: A modern and comprehensive probabilistic programming framework in python,” *PeerJ Computer Science*, vol. 9, e1516, 2023.
- [233] O. A. Martin, R. Kumar, and J. Lao, *Bayesian Modeling and Computation in Python* (Chapman & Hall/CRC Texts in Statistical Science), 1st ed. London, England: CRC Press, Dec. 2021.
- [234] R. E. Kass and A. E. Raftery, “Bayes factors,” *J. Am. Stat. Assoc.*, vol. 90, no. 430, pp. 773–795, Jun. 1995.
- [235] Q. F. Gronau, A. Sarafoglou, D. Matzke, A. Ly, U. Boehm, M. Marsman, D. Leslie, J. Forster, E.-J. Wagenmakers, and H. Steingroever, “A tutorial on bridge sampling,” *BITSS*, Dec. 2017.
- [236] E. I. George, “Bayesian model selection,” in *Wiley StatsRef: Statistics Reference Online*. Chichester, UK: John Wiley & Sons, Ltd, Sep. 2014.
- [237] D. B. Rubin, “The Bayesian bootstrap,” *aos*, vol. 9, no. 1, pp. 130–134, Jan. 1981.
- [238] S. Watanabe, “Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory,” *Journal of Machine Learning Research*, vol. 11, no. 116, pp. 3571–3594, 2010.
- [239] A. Vehtari, D. Simpson, A. Gelman, Y. Yao, and J. Gabry, “Pareto smoothed importance sampling,” *Journal of Machine Learning Research*, vol. 25, no. 72, pp. 1–58, 2024.
- [240] R. Kumar, C. Carroll, A. Hartikainen, and O. Martin, “ArviZ a unified library for exploratory analysis of bayesian models in python,” *Journal of Open Source Software*, vol. 4, no. 33, p. 1143, 2019.
- [241] A. Kværnø, “Singly diagonally implicit Runge–Kutta methods with an explicit first stage,” *BIT Numerical Mathematics*, vol. 44, no. 3, pp. 489–502, 2004.
- [242] P. Kidger, “On Neural Differential Equations,” Ph.D. dissertation, University of Oxford, 2021.
- [243] G. Söderlind, “Digital Filters in Adaptive Time-Stepping,” *ACM Transactions on Mathematical Software*, vol. 20, no. 1, pp. 1–26, 2003.
- [244] G. Söderlind, “Automatic control and adaptive time-stepping,” *Numerical Algorithms*, vol. 31, pp. 281–310, 2002.
- [245] E. Hairer and G. Wanner, *Solving Ordinary Differential Equations II Stiff and Differential-Algebraic Problems*, Second Revised Edition. Berlin: Springer, 2002.

- [246] J. Rader, T. Lyons, and P. Kidger, “Optimistix: Modular optimisation in jax and equinox,” *arXiv preprint arXiv:2402.09983*, 2024.
- [247] F.-G. Wieland, A. L. Hauber, M. Rosenblatt, C. Tönsing, and J. Timmer, “On structural and practical identifiability,” *Current Opinion in Systems Biology*, vol. 25, pp. 60–69, Mar. 2021.
- [248] R. Dong, C. Goodbrake, H. A. Harrington, and G. Pogudin, “Differential elimination for dynamical models via projections with applications to structural identifiability,” *SIAM J. Appl. Algebr. Geom.*, vol. 7, no. 1, pp. 194–235, Mar. 2023.
- [249] J. J. Hornberg, B. Binder, F. J. Bruggeman, B. Schoeberl, R. Heinrich, and H. V. Westerhoff, “Control of MAPK signalling: From complexity to what really matters,” *Oncogene*, vol. 24, no. 36, pp. 5533–5542, Aug. 2005.
- [250] M. D. Morris, “Factorial sampling plans for preliminary computational experiments,” *Technometrics*, vol. 33, no. 2, pp. 161–174, May 1991.
- [251] F. Campolongo, J. Cariboni, and A. Saltelli, “An effective screening design for sensitivity analysis of large models,” *Environ. Model. Softw.*, vol. 22, no. 10, pp. 1509–1518, Oct. 2007.
- [252] T. Iwanaga, W. Usher, and J. Herman, “Toward SALib 2.0: Advancing the accessibility and interpretability of global sensitivity analyses,” *Socio-Environmental Systems Modelling*, vol. 4, p. 18 155, May 2022.
- [253] J. Herman and W. Usher, “SALib: An open-source python library for sensitivity analysis,” *The Journal of Open Source Software*, vol. 2, no. 9, May 2017.
- [254] J. Geweke and G. Amisano, “Optimal prediction pools,” *J. Econom.*, vol. 164, no. 1, pp. 130–141, Sep. 2011.
- [255] M. A. Clyde, “Bayesian model averaging and model search strategies,” in *Bayesian Statistics 6*, Oxford University PressOxford, Aug. 1999, pp. 157–186.
- [256] A. Vehtari and J. Ojanen, “A survey of Bayesian predictive methods for model assessment, selection and comparison,” *Statistics Surveys*, vol. 6, no. none, pp. 142–228, 2012.
- [257] R. Roskoski, “ERK1/2 MAP kinases: Structure, function, and regulation,” *Pharmacological Research*, vol. 66, no. 2, pp. 105–143, Aug. 2012, ISSN: 1043-6618.
- [258] R. J. Orton, O. E. Sturm, V. Vyshemirsky, M. Calder, D. R. Gilbert, and W. Kolch, “Computational modelling of the receptor-tyrosine-kinase-activated MAPK pathway,” *Biochem. J.*, vol. 392, no. Pt 2, pp. 249–261, Dec. 2005.
- [259] A. Levchenko, J. Bruck, and P. W. Sternberg, “Scaffold proteins may biphasically affect the levels of mitogen-activated protein kinase signaling and reduce its threshold properties,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 97, no. 11, pp. 5818–5823, May 2000.

- [260] B. N. Kholodenko, “Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades,” *Eur. J. Biochem.*, vol. 267, no. 6, pp. 1583–1588, Mar. 2000.
- [261] R. J. Orton, O. E. Sturm, A. Gormand, W. Wolch, and D. R. Gilbert, “Computational modelling reveals feedback redundancy within the epidermal growth factor receptor/extracellular-signal regulated kinase signalling pathway,” *IET Syst. Biol.*, vol. 2, no. 4, pp. 173–183, Jul. 2008.
- [262] S.-Y. Shin, T. Kim, H.-S. Lee, J. H. Kang, J. Y. Lee, K.-H. Cho, and D. H. Kim, “The switching role of β -adrenergic receptor signalling in cell survival or death decision of cardiomyocytes,” *Nat. Commun.*, vol. 5, no. 1, p. 5777, Dec. 2014.
- [263] H. Ryu, M. Chung, M. Dobrzynski, D. Fey, Y. Blum, S. S. Lee, M. Peter, B. N. Kholodenko, N. L. Jeon, and O. Pertz, “Frequency modulation of ERK activation dynamics rewires cell fate,” *Mol. Syst. Biol.*, vol. 11, no. 11, p. 838, Nov. 2015.
- [264] M. Kochańczyk, P. Kocieniewski, E. Kozłowska, J. Jaruszewicz-Błońska, B. Sparta, M. Pargett, J. G. Albeck, W. S. Hlavacek, and T. Lipniacki, “Relaxation oscillations and hierarchy of feedbacks in MAPK signaling,” *Sci. Rep.*, vol. 7, no. 1, p. 38 244, Jan. 2017.
- [265] Y. Yao, A. Vehtari, D. Simpson, and A. Gelman, “Using Stacking to Average Bayesian Predictive Distributions (with Discussion),” *Bayesian Analysis*, vol. 13, no. 3, Sep. 1, 2018, ISSN: 1936-0975.
- [266] R. J. Orton, M. E. Adriaens, A. Gormand, O. E. Sturm, W. Kolch, and D. R. Gilbert, “Computational modelling of cancerous mutations in the EGFR/ERK signalling pathway,” *BMC Syst. Biol.*, vol. 3, p. 100, Oct. 2009.
- [267] 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*, vol. 3, no. 10, A. P. Arkin, Ed., Oct. 2007.
- [268] F. A. Brightman and D. A. Fell, “Differential feedback regulation of the MAPK cascade underlies the quantitative differences in EGF and NGF signalling in PC12 cells,” *FEBS Lett.*, vol. 482, no. 3, pp. 169–174, Oct. 2000.
- [269] Y. Yao, G. Pirš, A. Vehtari, and A. Gelman, “Bayesian hierarchical stacking: Some models are (somewhere) useful,” *Bayesian Analysis*, vol. 17, no. 4, pp. 1043–1071, 2022.
- [270] B. J. Reich and S. K. Ghosh, *Bayesian statistical methods*. Chapman and Hall/CRC, 2019.
- [271] N. Galagali and Y. M. Marzouk, “Exploiting network topology for large-scale inference of nonlinear reaction models,” *Journal of The Royal Society Interface*, vol. 16, no. 152, p. 20 180 766, Mar. 2019, ISSN: 1742-5689, 1742-5662.
- [272] T. G. Dietterich, “Ensemble methods in machine learning,” in *Multiple Classifier Systems*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2000, pp. 1–15, ISBN: 978-3-540-45014-6.

- [273] Z. M. Laubach, E. J. Murray, K. L. Hoke, R. J. Safran, and W. Perng, “A biologist’s guide to model selection and causal inference,” *Proceedings of the Royal Society B: Biological Sciences*, vol. 288, no. 1943, p. 20202815, Jan. 2021.
- [274] S. L. Brunton, J. L. Proctor, and J. N. Kutz, “Discovering governing equations from data by sparse identification of nonlinear dynamical systems,” *Proceedings of the National Academy of Sciences*, vol. 113, no. 15, pp. 3932–3937, Apr. 2016.
- [275] N. M. Mangan, S. L. Brunton, J. L. Proctor, and J. N. Kutz, “Inferring Biological Networks by Sparse Identification of Nonlinear Dynamics,” *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, vol. 2, no. 1, pp. 52–63, Jun. 2016, ISSN: 2372-2061, 2332-7804.
- [276] R. Jiang, P. Singh, F. Wrede, A. Hellander, and L. Petzold, “Identification of dynamic mass-action biochemical reaction networks using sparse Bayesian methods,” *PLOS Computational Biology*, vol. 18, no. 1, P. Mendes, Ed., e1009830, Jan. 2022, ISSN: 1553-7358.
- [277] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, “Variational Inference: A Review for Statisticians,” *Journal of the American Statistical Association*, vol. 112, no. 518, pp. 859–877, Apr. 2017, ISSN: 0162-1459, 1537-274X.
- [278] A. Narayan, Y. Marzouk, and D. Xiu, “Sequential data assimilation with multiple models,” *J. Comput. Phys.*, vol. 231, no. 19, pp. 6401–6418, Aug. 2012.
- [279] D. Garcia and R. J. Shaw, “AMPK: Mechanisms of Cellular Energy Sensing and Restoration of Metabolic Balance,” *Molecular Cell*, vol. 66, no. 6, pp. 789–800, Jun. 2017, ISSN: 10972765.
- [280] S.-C. Lin and D. G. Hardie, “AMPK: Sensing Glucose as well as Cellular Energy Status,” *Cell Metabolism*, vol. 27, no. 2, pp. 299–313, Feb. 2018, ISSN: 15504131.
- [281] S. Herzig and R. J. Shaw, “AMPK: Guardian of metabolism and mitochondrial homeostasis,” *Nature Reviews Molecular Cell Biology*, vol. 19, no. 2, pp. 121–135, Feb. 2018, ISSN: 1471-0072, 1471-0080.
- [282] D. G. Hardie, “Keeping the home fires burning: AMP-activated protein kinase,” *Journal of The Royal Society Interface*, vol. 15, no. 138, p. 20170774, Jan. 2018, ISSN: 1742-5689, 1742-5662.
- [283] G. R. Steinberg and D. G. Hardie, “New insights into activation and function of the AMPK,” *Nature Reviews Molecular Cell Biology*, vol. 24, no. 4, pp. 255–272, Apr. 2023, ISSN: 1471-0072, 1471-0080.
- [284] E. Trefts and R. J. Shaw, “AMPK: Restoring metabolic homeostasis over space and time,” *Molecular Cell*, vol. 81, no. 18, pp. 3677–3690, Sep. 16, 2021, ISSN: 1097-4164.
- [285] D. G. Hardie, “AMP-activated protein kinase — a journey from 1 to 100 downstream targets,” *Biochemical Journal*, vol. 479, no. 22, pp. 2327–2343, Nov. 30, 2022, ISSN: 0264-6021, 1470-8728.

- [286] D. G. Hardie, B. E. Schaffer, and A. Brunet, “AMPK: An Energy-Sensing Pathway with Multiple Inputs and Outputs,” *Trends in Cell Biology*, vol. 26, no. 3, pp. 190–201, Mar. 2016, ISSN: 09628924.
- [287] A. Fowler, K. R. Knaus, S. Khuu, A. Khalilimeybodi, S. Schenk, S. R. Ward, A. C. Fry, P. Rangamani, and A. D. McCulloch, “Network model of skeletal muscle cell signalling predicts differential responses to endurance and resistance exercise training,” *Experimental Physiology*, vol. 109, no. 6, pp. 939–955, 2024, ISSN: 1469-445X.
- [288] A. Khalilimeybodi, J. J. Saucerman, and P. Rangamani, “Modeling cardiomyocyte signaling and metabolism predicts genotype-to-phenotype mechanisms in hypertrophic cardiomyopathy,” *Computers in Biology and Medicine*, vol. 175, p. 108 499, Jun. 1, 2024, ISSN: 0010-4825.
- [289] A. González, M. N. Hall, S.-C. Lin, and D. G. Hardie, “AMPK and TOR: The Yin and Yang of Cellular Nutrient Sensing and Growth Control,” *Cell Metabolism*, vol. 31, no. 3, pp. 472–492, Mar. 3, 2020, ISSN: 1932-7420.
- [290] S. A. Hawley, D. A. Pan, K. J. Mustard, L. Ross, J. Bain, A. M. Edelman, B. G. Frenguelli, and D. G. Hardie, “Calmodulin-dependent protein kinase kinase- β is an alternative upstream kinase for AMP-activated protein kinase,” *Cell Metabolism*, vol. 2, no. 1, pp. 9–19, Jul. 1, 2005, ISSN: 1550-4131.
- [291] R. L. Hurley, K. A. Anderson, J. M. Franzone, B. E. Kemp, A. R. Means, and L. A. Witters, “The Ca²⁺/Calmodulin-dependent Protein Kinase Kinases Are AMP-activated Protein Kinase Kinases*,” *Journal of Biological Chemistry*, vol. 280, no. 32, pp. 29 060–29 066, Aug. 12, 2005, ISSN: 0021-9258.
- [292] A. Woods, K. Dickerson, R. Heath, S.-P. Hong, M. Momcilovic, S. R. Johnstone, M. Carlson, and D. Carling, “Ca²⁺/calmodulin-dependent protein kinase kinase- β acts upstream of AMP-activated protein kinase in mammalian cells,” *Cell Metabolism*, vol. 2, no. 1, pp. 21–33, Jul. 1, 2005, ISSN: 1550-4131.
- [293] B. Xiao, M. J. Sanders, E. Underwood, R. Heath, F. V. Mayer, D. Carmena, C. Jing, P. A. Walker, J. F. Eccleston, L. F. Haire, P. Saiu, S. A. Howell, R. Aasland, S. R. Martin, D. Carling, and S. J. Gamblin, “Structure of mammalian AMPK and its regulation by ADP,” *Nature*, vol. 472, no. 7342, pp. 230–233, Apr. 14, 2011, ISSN: 0028-0836, 1476-4687.
- [294] Y. Yan, S. Mukherjee, K. G. Harikumar, T. S. Strutzenberg, X. E. Zhou, K. Suino-Powell, T.-H. Xu, R. D. Sheldon, J. Lamp, J. S. Brunzelle, K. Radziwon, A. Ellis, S. J. Novick, I. E. Vega, R. G. Jones, L. J. Miller, H. E. Xu, P. R. Griffin, A. A. Kossiakoff, and K. Melcher, “Structure of an AMPK complex in an inactive, ATP-bound state,” *Science*, vol. 373, no. 6553, pp. 413–419, Jul. 23, 2021.
- [295] D. Carling, P. R. Clarke, V. A. Zammit, and D. G. Hardie, “Purification and characterization of the AMP-activated protein kinase,” *European Journal of Biochemistry*, vol. 186, no. 1–2, pp. 129–136, 1989, ISSN: 1432-1033.

- [296] D. F. Egan, D. B. Shackelford, M. M. Mihaylova, S. Gelino, R. A. Kohnz, W. Mair, D. S. Vasquez, A. Joshi, D. M. Gwinn, R. Taylor, J. M. Asara, J. Fitzpatrick, A. Dillin, B. Viollet, M. Kundu, M. Hansen, and R. J. Shaw, "Phosphorylation of ULK1 (hATG1) by AMP-Activated Protein Kinase Connects Energy Sensing to Mitophagy," *Science*, vol. 331, no. 6016, pp. 456–461, Jan. 28, 2011.
- [297] D. M. Gwinn, D. B. Shackelford, D. F. Egan, M. M. Mihaylova, A. Mery, D. S. Vasquez, B. E. Turk, and R. J. Shaw, "AMPK Phosphorylation of Raptor Mediates a Metabolic Checkpoint," *Molecular Cell*, vol. 30, no. 2, pp. 214–226, Apr. 25, 2008, ISSN: 1097-2765.
- [298] G. R. Steinberg and D. Carling, "AMP-activated protein kinase: The current landscape for drug development," *Nature Reviews Drug Discovery*, vol. 18, no. 7, pp. 527–551, Jul. 2019, ISSN: 1474-1784.
- [299] M. Foretz, B. Guigas, L. Bertrand, M. Pollak, and B. Viollet, "Metformin: From Mechanisms of Action to Therapies," *Cell Metabolism*, vol. 20, no. 6, pp. 953–966, Dec. 2, 2014, ISSN: 1550-4131.
- [300] G. Zhou, R. Myers, Y. Li, Y. Chen, X. Shen, J. Fenyk-Melody, M. Wu, J. Ventre, T. Doebber, N. Fujii, N. Musi, M. F. Hirshman, L. J. Goodyear, and D. E. Moller, "Role of AMP-activated protein kinase in mechanism of metformin action," *The Journal of Clinical Investigation*, vol. 108, no. 8, pp. 1167–1174, Oct. 15, 2001, ISSN: 0021-9738.
- [301] S. A. Hawley, M. D. Fullerton, F. A. Ross, J. D. Schertzer, C. Chevtzoff, K. J. Walker, M. W. Pegg, D. Zibrova, K. A. Green, K. J. Mustard, B. E. Kemp, K. Sakamoto, G. R. Steinberg, and D. G. Hardie, "The Ancient Drug Salicylate Directly Activates AMP-Activated Protein Kinase," *Science*, vol. 336, no. 6083, pp. 918–922, May 18, 2012.
- [302] R. Blazev, C. S. Carl, Y.-K. Ng, J. Molendijk, C. T. Voldstedlund, Y. Zhao, D. Xiao, A. J. Kueh, P. M. Miotto, V. R. Haynes, J. P. Hardee, J. D. Chung, J. W. McNamara, H. Qian, P. Gregorevic, J. S. Oakhill, M. J. Herold, T. E. Jensen, L. Lisowski, G. S. Lynch, G. T. Dodd, M. J. Watt, P. Yang, B. Kiens, E. A. Richter, and B. L. Parker, "Phosphoproteomics of three exercise modalities identifies canonical signaling and C18ORF25 as an AMPK substrate regulating skeletal muscle function," *Cell Metabolism*, vol. 34, no. 10, 1561–1577.e9, Oct. 4, 2022, ISSN: 1550-4131.
- [303] J. W. Scott, D. G. Norman, S. A. Hawley, L. Kontogiannis, and D. Hardie, "Protein kinase substrate recognition studied using the recombinant catalytic domain of AMP-activated protein kinase and a model substrate," *Journal of Molecular Biology*, vol. 317, no. 2, pp. 309–323, Mar. 2002, ISSN: 00222836.
- [304] D. L. Schmitt, S. Mehta, and J. Zhang, "Study of spatiotemporal regulation of kinase signaling using genetically encodable molecular tools," *Current Opinion in Chemical Biology*, vol. 71, p. 102 224, Dec. 2022, ISSN: 13675931.
- [305] P. Tsou, B. Zheng, C.-H. Hsu, A. T. Sasaki, and L. C. Cantley, "A Fluorescent Reporter of AMPK Activity and Cellular Energy Stress," *Cell Metabolism*, vol. 13, no. 4, pp. 476–486, Apr. 2011, ISSN: 15504131.

- [306] S. Chenna, W. J. H. Koopman, J. H. M. Prehn, and N. M. C. Connolly, “Mechanisms and mathematical modeling of ROS production by the mitochondrial electron transport chain,” *American Journal of Physiology-Cell Physiology*, vol. 323, no. 1, pp. C69–C83, Jul. 2022, ISSN: 0363-6143.
- [307] R. C. Smith, *Uncertainty Quantification: Theory, Implementation, and Applications, Second Edition*. Philadelphia, PA: Society for Industrial and Applied Mathematics, Jan. 2024, ISBN: 978-1-61197-783-7 978-1-61197-784-4.
- [308] N. J. Linden, B. Kramer, and P. Rangamani, “Bayesian parameter estimation for dynamical models in systems biology,” *PLOS Computational Biology*, vol. 18, no. 10, J. J. Saucerman, Ed., e1010651, Oct. 21, 2022, ISSN: 1553-7358.
- [309] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, “Bayesian Data Analysis Third edition (with errors fixed as of 15 February 2021),”
- [310] E. D. Mitra and W. S. Hlavacek, “Parameter estimation and uncertainty quantification for systems biology models,” *Current Opinion in Systems Biology*, vol. 18, pp. 9–18, Dec. 2019, ISSN: 24523100.
- [311] F.-G. Wieland, A. L. Hauber, M. Rosenblatt, C. Tönsing, and J. Timmer, “On structural and practical identifiability,” *Current Opinion in Systems Biology*, vol. 25, pp. 60–69, Mar. 2021, ISSN: 24523100.
- [312] A. Vehtari and J. Lampinen, “Bayesian Model Assessment and Comparison Using Cross-Validation Predictive Densities,” *Neural Computation*, vol. 14, no. 10, pp. 2439–2468, Oct. 1, 2002, ISSN: 0899-7667, 1530-888X.
- [313] K. P. Burnham, D. R. Anderson, and K. P. Burnham, *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*, 2nd ed. New York: Springer, 2002, 488 pp., ISBN: 978-0-387-95364-9.
- [314] H. Hong, A. Ovchinnikov, G. Pogudin, and C. Yap, “Global Identifiability of Differential Models,” *Communications on Pure and Applied Mathematics*, vol. 73, no. 9, pp. 1831–1879, Sep. 2020, ISSN: 0010-3640, 1097-0312.
- [315] L. Geris and D. Gomez-Cabrero, Eds., *Uncertainty in Biology: A Computational Modeling Approach* (Studies in Mechanobiology, Tissue Engineering and Biomaterials). Cham: Springer International Publishing, 2016, vol. 17, ISBN: 978-3-319-21295-1 978-3-319-21296-8.
- [316] S. P. Beik, L. A. Harris, M. A. Kochen, J. Sage, V. Quaranta, and C. F. Lopez, “Unified tumor growth mechanisms from multimodel inference and dataset integration,” *PLOS Computational Biology*, vol. 19, no. 7, K. R. Patil, Ed., e1011215, Jul. 5, 2023, ISSN: 1553-7358.
- [317] F. Anstett-Collin, L. Denis-Vidal, and G. Millérioux, “A priori identifiability: An overview on definitions and approaches,” *Annual Reviews in Control*, vol. 50, pp. 139–149, 2020, ISSN: 13675788.

- [318] R. Dong, C. Goodbrake, H. A. Harrington, and G. Pogudin, “Differential Elimination for Dynamical Models via Projections with Applications to Structural Identifiability,” *SIAM Journal on Applied Algebra and Geometry*, vol. 7, no. 1, pp. 194–235, Mar. 31, 2023, ISSN: 2470-6566.
- [319] P. Mikkola, O. A. Martin, S. Chandramouli, M. Hartmann, O. A. Pla, O. Thomas, H. Pesonen, J. Corander, A. Vehtari, S. Kaski, P.-C. Bürkner, and A. Klami. “Prior knowledge elicitation: The past, present, and future.” (May 9, 2023), [Online]. Available: <http://arxiv.org/abs/2112.01380> (visited on 05/04/2024), pre-published.
- [320] L. Zhang, B. Carpenter, A. Gelman, and A. Vehtari, “Pathfinder: Parallel quasi-Newton variational inference,” *Journal of Machine Learning Research*, vol. 23, no. 306, pp. 1–49, 2022.
- [321] A. Vehtari, A. Gelman, and J. Gabry, “Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC,” *Statistics and Computing*, vol. 27, no. 5, pp. 1413–1432, Sep. 2017, ISSN: 0960-3174, 1573-1375.
- [322] A. Vehtari, D. Simpson, A. Gelman, Y. Yao, and J. Gabry. “Pareto Smoothed Importance Sampling.” (Mar. 13, 2024), [Online]. Available: <http://arxiv.org/abs/1507.02646> (visited on 05/04/2024), pre-published.
- [323] P. Kidger. “On Neural Differential Equations.” (Feb. 4, 2022), [Online]. Available: <http://arxiv.org/abs/2202.02435> (visited on 05/07/2025), pre-published.
- [324] C. Posner, S. Mehta, and J. Zhang, “Fluorescent biosensor imaging meets deterministic mathematical modelling: Quantitative investigation of signalling compartmentalization,” *The Journal of Physiology*, vol. 601, no. 19, pp. 4227–4241, 2023, ISSN: 1469-7793.
- [325] J. M. Lizcano, O. Göransson, R. Toth, M. Deak, N. A. Morrice, J. Boudeau, S. A. Hawley, L. Udd, T. P. Mäkelä, D. G. Hardie, and D. R. Alessi, “LKB1 is a master kinase that activates 13 kinases of the AMPK subfamily, including MARK/PAR-1,” *The EMBO Journal*, vol. 23, no. 4, pp. 833–843, Feb. 25, 2004, ISSN: 0261-4189, 1460-2075.
- [326] R. G. Duggleby and R. B. Clarke, “Experimental designs for estimating the parameters of the Michaelis-Menten equation from progress curves of enzyme-catalyzed reactions,” *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology*, vol. 1080, no. 3, pp. 231–236, Nov. 15, 1991, ISSN: 0167-4838.
- [327] G. J. Gowans, S. A. Hawley, F. A. Ross, and D. G. Hardie, “AMP Is a True Physiological Regulator of AMP-Activated Protein Kinase by Both Allosteric Activation and Enhancing Net Phosphorylation,” *Cell Metabolism*, vol. 18, no. 4, pp. 556–566, Oct. 2013, ISSN: 15504131.
- [328] J. S. Oakhill, R. Steel, Z.-P. Chen, J. W. Scott, N. Ling, S. Tam, and B. E. Kemp, “AMPK Is a Direct Adenylate Charge-Regulated Protein Kinase,” *Science*, vol. 332, no. 6036, pp. 1433–1435, Jun. 17, 2011.

- [329] S. P. Davies, N. R. Helps, P. T. W. Cohen, and D. G. Hardie, "5-AMP inhibits dephosphorylation, as well as promoting phosphorylation, of the AMP-activated protein kinase. Studies using bacterially expressed human protein phosphatase-2C α and native bovine protein phosphatase-2Ac," *FEBS Letters*, vol. 377, no. 3, pp. 421–425, Dec. 27, 1995, ISSN: 0014-5793.
- [330] F. A. Ross, T. E. Jensen, and D. G. Hardie, "Differential regulation by AMP and ADP of AMPK complexes containing different γ subunit isoforms," *Biochemical Journal*, vol. 473, p. 189, Pt 2 Jan. 5, 2016.
- [331] J. S. Oakhill, J. W. Scott, and B. E. Kemp, "AMPK functions as an adenylate charge-regulated protein kinase," *Trends in endocrinology and metabolism: TEM*, vol. 23, no. 3, pp. 125–132, Mar. 2012, ISSN: 1879-3061.
- [332] P. Rangamani, M. G. Levy, S. Khan, and G. Oster, "Paradoxical signaling regulates structural plasticity in dendritic spines," *Proceedings of the National Academy of Sciences*, vol. 113, no. 36, E5298–E5307, Sep. 6, 2016.
- [333] A. E. Marley, J. E. Sullivan, D. Carling, W. M. Abbott, G. J. Smith, I. W. F. Taylor, F. Carey, and R. K. Beri, "Biochemical characterization and deletion analysis of recombinant human protein phosphatase 2C α ," *Biochemical Journal*, vol. 320, no. 3, pp. 801–806, Dec. 15, 1996, ISSN: 0264-6021, 1470-8728.
- [334] W. Ament, G. J. J. Bonga, A. L. Hof, and G. J. Verkerke, "EMG median power frequency in an exhausting exercise," *Journal of Electromyography and Kinesiology*, vol. 3, no. 4, pp. 214–220, Jan. 1, 1993, ISSN: 1050-6411.
- [335] E. M. Shockley, J. A. Vrugt, and C. F. Lopez, "PyDREAM: High-dimensional parameter inference for biological models in python," *Bioinformatics*, vol. 34, no. 4, A. Valencia, Ed., pp. 695–697, Feb. 15, 2018, ISSN: 1367-4803, 1367-4811.
- [336] C. G. Cess and S. D. Finley, "Data-driven analysis of a mechanistic model of CAR T cell signaling predicts effects of cell-to-cell heterogeneity," *Journal of Theoretical Biology*, vol. 489, p. 110 125, Mar. 2020, ISSN: 00225193.
- [337] T. Toni, Y.-i. Ozaki, P. Kirk, S. Kuroda, and M. P. H. Stumpf, "Elucidating the in vivo phosphorylation dynamics of the ERK MAP kinase using quantitative proteomics data and Bayesian model selection," *Molecular BioSystems*, vol. 8, no. 7, p. 1921, 2012, ISSN: 1742-206X, 1742-2051.
- [338] S. L. Brunton, J. L. Proctor, and J. N. Kutz, "Sparse Identification of Nonlinear Dynamics with Control (SINDYc)**SLB acknowledges support from the U.S. Air Force Center of Excellence on Nature Inspired Flight Technologies and Ideas (FA9550-14-1-0398). JLP thanks Bill and Melinda Gates for their active support of the Institute of Disease Modeling and their sponsorship through the Global Good Fund. JNK acknowledges support from the U.S. Air Force Office of Scientific Research (FA9550-09-0174).," *IFAC-PapersOnLine*, vol. 49, no. 18, pp. 710–715, 2016, ISSN: 24058963.

- [339] C. Rackauckas, Y. Ma, J. Martensen, C. Warner, K. Zubov, R. Supekar, D. Skinner, A. Ramadhan, and A. Edelman. “Universal Differential Equations for Scientific Machine Learning.” (Nov. 2, 2021), [Online]. Available: <http://arxiv.org/abs/2001.04385> (visited on 03/07/2025), pre-published.
- [340] C. for Disease Control and Prevention, *National diabetes statistics report website*, <https://www.cdc.gov/diabetes/data/statistics-report/index.html>, Accessed: 2023-10-23.
- [341] “IDF diabetes atlas, 10th edn,” International Diabetes Federation, Tech. Rep., 2021.
- [342] S. L. Aronoff, K. B. A. B. C. Cde, and B. S. Rn, “Glucose metabolism and regulation: Beyond insulin and glucagon,” *Diabetes Spectr.*, vol. 17, no. 3, 2004.
- [343] M. C. Riddle, H. C. Gerstein, and P. D. Home, “Lingering effects of hyperglycemia in recently diagnosed diabetes during long-term follow-up of the DCCT/EDIC and UKPDS cohorts: More evidence that early control matters,” *Diabetes Care*, vol. 44, no. 10, pp. 2212–2215, Sep. 2021.
- [344] N. A. ElSayed, G. Aleppo, V. R. Aroda, R. R. Bannuru, F. M. Brown, D. Bruemmer, B. S. Collins, M. E. Hilliard, D. Isaacs, E. L. Johnson, S. Kahan, K. Khunti, J. Leon, S. K. Lyons, M. L. Perry, P. Prahalad, R. E. Pratley, J. J. Seley, R. C. Stanton, R. A. Gabbay, and on behalf of the American Diabetes Association, “2. classification and diagnosis of diabetes: Standards of care in diabetes-2023,” *Diabetes Care*, vol. 46, no. Suppl 1, S19–S40, Jan. 2023.
- [345] W. C. Knowler, E. Barrett-Connor, S. E. Fowler, R. F. Hamman, J. M. Lachin, E. A. Walker, D. M. Nathan, and Diabetes Prevention Program Research Group, “Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin,” *N. Engl. J. Med.*, vol. 346, no. 6, pp. 393–403, Feb. 2002.
- [346] J. Tuomilehto, J. Lindström, J. G. Eriksson, T. T. Valle, H. Hämäläinen, P. Ilanne-Parikka, S. Keinänen-Kiukaanniemi, M. Laakso, A. Louheranta, M. Rastas, V. Salminen, M. Uusitupa, and Finnish Diabetes Prevention Study Group, “Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance,” *N. Engl. J. Med.*, vol. 344, no. 18, pp. 1343–1350, May 2001.
- [347] J. M. Lachin, D. M. Nathan, and DCCT/EDIC Research Group, “Understanding metabolic memory: The prolonged influence of glycemia during the diabetes control and complications trial (DCCT) on future risks of complications during the study of the epidemiology of diabetes interventions and complications (EDIC),” *Diabetes Care*, vol. 44, no. 10, pp. 2216–2224, Sep. 2021.
- [348] M. Lind, H. Imberg, R. L. Coleman, O. Nerman, and R. R. Holman, “Historical HbA1c values may explain the type 2 diabetes legacy effect: UKPDS 88,” *Diabetes Care*, vol. 44, no. 10, pp. 2231–2237, Jul. 2021.
- [349] A. Gedeberg, T. P. Almdal, K. Berencsi, J. Rungby, J. S. Nielsen, D. R. Witte, S. Friborg, I. Brandslund, A. Vaag, H. Beck-Nielsen, H. T. Sørensen, and R. W. Thomsen, “Prevalence of micro- and macrovascular diabetes complications at time of type 2 diabetes diagnosis and associated

clinical characteristics: A cross-sectional baseline study of 6958 patients in the danish DD2 cohort,” *J. Diabetes Complications*, vol. 32, no. 1, pp. 34–40, Jan. 2018.

- [350] I. Ajmera, M. Swat, C. Laibe, N. Le Novère, and V. Chelliah, “The impact of mathematical modeling on the understanding of diabetes and related complications,” *CPT Pharmacometrics Syst Pharmacol*, vol. 2, e54, Jul. 2013.
- [351] American Diabetes Association Consensus Panel, “Guidelines for computer modeling of diabetes and its complications,” *Diabetes Care*, vol. 27, no. 9, pp. 2262–2265, Sep. 2004.
- [352] E. Nyman, G. Cedersund, and P. Strålfors, “Insulin signaling - mathematical modeling comes of age,” *Trends Endocrinol. Metab.*, vol. 23, no. 3, pp. 107–115, Mar. 2012.
- [353] T. S. Hannon, S. E. Kahn, K. M. Utzschneider, T. A. Buchanan, K. J. Nadeau, P. S. Zeitler, D. A. Ehrmann, S. A. Arslanian, S. Caprio, S. L. Edelstein, P. J. Savage, K. J. Mather, and for the RISE Consortium, “Review of methods for measuring β -cell function: Design considerations from the restoring insulin secretion,” *Diabetes Obes. Metab.*, vol. 20, no. 1, pp. 14–24, Jan. 2018.
- [354] E. Ahlqvist, P. Storm, A. Käräjämäki, M. Martinell, M. Dorkhan, A. Carlsson, P. Vikman, R. B. Prasad, D. M. Aly, P. Almgren, Y. Wessman, N. Shaat, P. Spéigel, H. Mulder, E. Lindholm, O. Melander, O. Hansson, U. Malmqvist, Å. Lernmark, K. Lahti, T. Forsén, T. Tuomi, A. H. Rosengren, and L. Groop, “Novel subgroups of adult-onset diabetes and their association with outcomes: A data-driven cluster analysis of six variables,” *Lancet Diabetes Endocrinol.*, vol. 6, no. 5, pp. 361–369, May 2018.
- [355] C. Cobelli, C. Dalla Man, G. Toffolo, R. Basu, A. Vella, and R. Rizza, “The oral minimal model method,” *Diabetes*, vol. 63, no. 4, pp. 1203–1213, 2014.
- [356] C. Cobelli and C. Dalla Man, “Minimal and maximal models to quantitate glucose metabolism: Tools to measure, to simulate and to run in silico clinical trials,” *J. Diabetes Sci. Technol.*, vol. 16, no. 5, pp. 1270–1298, Sep. 2022.
- [357] A. Mari, A. Tura, E. Grespan, and R. Bizzotto, “Mathematical modeling for the physiological and clinical investigation of glucose homeostasis and diabetes,” *Front. Physiol.*, vol. 11, p. 575 789, Nov. 2020.
- [358] P. Palumbo, S. Ditlevsen, A. Bertuzzi, and A. De Gaetano, “Mathematical modeling of the glucose-insulin system: A review,” *Math. Biosci.*, vol. 244, no. 2, pp. 69–81, Aug. 2013.
- [359] D. Bogdanet, P. O’Shea, C. Lyons, A. Shafat, and F. Dunne, “The oral glucose tolerance test-is it time for a change?-a literature review with an emphasis on pregnancy,” *J. Clin. Med.*, vol. 9, no. 11, p. 3451, Oct. 2020.
- [360] K. M. Utzschneider, R. L. Prigeon, J. Tong, F. Gerchman, D. B. Carr, S. Zraika, J. Udayasankar, B. Montgomery, A. Mari, and S. E. Kahn, “Within-subject variability of measures of beta cell

function derived from a 2 h OGTT: Implications for research studies,” *Diabetologia*, vol. 50, no. 12, pp. 2516–2525, Dec. 2007.

- [361] K. Schousboe, J. E. Henriksen, K. O. Kyvik, T. I. A. Sørensen, and P. Hyltoft Petersen, “Reproducibility of S-insulin and B-glucose responses in two identical oral glucose tolerance tests,” *Scand. J. Clin. Lab. Invest.*, vol. 62, no. 8, pp. 623–630, 2002.
- [362] A. Bagust and P. McEwan, “Guidelines for computer modeling of diabetes and its complications: Response to american diabetes association consensus panel,” *Diabetes Care*, vol. 28, no. 2, pp. 500–501, Feb. 2005.
- [363] R. Kahn, “Guidelines for computer modeling of diabetes and its complications: Response to bagust and McEwan,” *Diabetes Care*, vol. 28, no. 2, pp. 500–501, Feb. 2005.
- [364] R. Hovorka, L. Chassin, S. D. Luzio, R. Playle, and D. R. Owens, “Pancreatic beta-cell responsiveness during meal tolerance test: Model assessment in normal subjects and subjects with newly diagnosed noninsulin-dependent diabetes mellitus,” *J. Clin. Endocrinol. Metab.*, vol. 83, no. 3, pp. 744–750, Mar. 1998.
- [365] A. Cretti, M. Lehtovirta, E. Bonora, B. Brunato, M. G. Zenti, F. Tosi, M. Caputo, B. Caruso, L. C. Groop, M. Muggeo, and R. C. Bonadonna, “Assessment of beta-cell function during the oral glucose tolerance test by a minimal model of insulin secretion,” *Eur. J. Clin. Invest.*, vol. 31, no. 5, pp. 405–416, May 2001.
- [366] E. Breda, M. K. Cavaghan, G. Toffolo, K. S. Polonsky, and C. Cobelli, “Oral glucose tolerance test minimal model indexes of beta-cell function and insulin sensitivity,” *Diabetes*, vol. 50, no. 1, pp. 150–158, Jan. 2001.
- [367] H. Hall, D. Perelman, A. Breschi, P. Limcaoco, R. Kellogg, T. McLaughlin, and M. Snyder, “Glucotypes reveal new patterns of glucose dysregulation,” *PLoS Biol.*, vol. 16, no. 7, e2005143, Jul. 2018.
- [368] A. H. Tjaden, S. L. Edelstein, S. Arslanian, E. Barengolts, S. Caprio, M. Cree-Green, A. Lteif, K. J. Mather, M. Savoye, A. H. Xiang, and S. E. Kahn, “Reproducibility of glycemic measures among dysglycemic youth and adults in the RISE study,” *J. Clin. Endocrinol. Metab.*, dgad135, Mar. 2023.
- [369] RISE Consortium, “Restoring insulin secretion (RISE): Design of studies of β -cell preservation in prediabetes and early type 2 diabetes across the life span,” *Diabetes Care*, vol. 37, no. 3, pp. 780–788, Feb. 2014.
- [370] S. Cao, Y. Aboelkassem, A. Wang, D. Valdez-Jasso, J. J. Saucerman, J. H. Omens, and A. D. McCulloch, “Quantification of model and data uncertainty in a network analysis of cardiac myocyte mechanosignalling,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 378, no. 2173, p. 20190336, Jun. 12, 2020, ISSN: 1364-503X, 1471-2962.

- [371] C. G. Moles, P. Mendes, and J. R. Banga, "Parameter estimation in biochemical pathways: A comparison of global optimization methods," vol. 13, no. 11, pp. 2467–2474, Nov. 2003.
- [372] A. Raue, J. Karlsson, M. P. Saccomani, M. Jirstrand, and J. Timmer, "Comparison of approaches for parameter identifiability analysis of biological systems," *Bioinformatics*, vol. 30, no. 10, pp. 1440–1448, May 15, 2014, ISSN: 1367-4811, 1367-4803.
- [373] P. Kirk, T. Thorne, and M. P. Stumpf, "Model selection in systems and synthetic biology," *Current Opinion in Biotechnology*, vol. 24, no. 4, pp. 767–774, Aug. 2013, ISSN: 09581669.
- [374] R. E. Morrison and A. Cunha, "Embedded model discrepancy: A case study of Zika modeling," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 30, no. 5, p. 051 103, May 1, 2020, ISSN: 1054-1500, 1089-7682.
- [375] M. P. H. Stumpf, "Multi-model and network inference based on ensemble estimates: Avoiding the madness of crowds," *Journal of The Royal Society Interface*, vol. 17, no. 171, p. 20 200 419, Oct. 2020, ISSN: 1742-5689, 1742-5662.
- [376] D. Ohadi and P. Rangamani, "Geometric control of frequency modulation of cAMP oscillations due to calcium in dendritic spines," *Biophys. J.*, vol. 117, no. 10, pp. 1981–1994, Oct. 2019.
- [377] P. Rangamani, A. Lipshtat, E. U. Azeloglu, R. C. Calizo, M. Hu, S. Ghassemi, J. Hone, S. Scarlata, S. R. Neves, and R. Iyengar, "Decoding Information in Cell Shape," *Cell*, vol. 154, no. 6, pp. 1356–1369, Sep. 12, 2013, ISSN: 0092-8674, 1097-4172.
- [378] Y. Liu, K. Suh, P. K. Maini, D. J. Cohen, and R. E. Baker. "Parameter identifiability and model selection for partial differential equation models of cell invasion." (Oct. 18, 2023), [Online]. Available: <http://arxiv.org/abs/2309.01476> (visited on 05/04/2024), pre-published.
- [379] B. Adams, W. Bohnhoff, R. Canfield, W. Coomber, K. Dalbey, M. Ebeida, J. Eddy, M. Eldred, G. Geraci, R. Hooper, P. Hough, K. Hu, J. Jakeman, C. Kent, M. Khalil, K. Maupin, J. Monschke, E. Ridgway, A. Rushdi, D. Seidl, J. Stephens, L. Swiler, A. Tran, D. Vigil, T. Wildey, J. Winokur, F. Menhorn, and X. Zeng, "Dakota, a multilevel parallel object-oriented framework for design optimization, parameter estimation, uncertainty quantification, and sensitivity analysis: Version 6.16.0 reference manual," Sandia National Laboratories, Tech. Rep., May 2022.
- [380] K.-T. Kim, U. Villa, M. Parno, Y. Marzouk, O. Ghattas, and N. Petra, "hIPPYlib-MUQ: A Bayesian Inference Software Framework for Integration of Data with Complex Predictive Models under Uncertainty," *ACM Trans. Math. Softw.*, vol. 49, no. 2, 17:1–17:31, Jun. 15, 2023, ISSN: 0098-3500.
- [381] L. Biegler, G. Biros, O. Ghattas, M. Heinkenschloss, D. Keyes, B. Mallick, L. Tenorio, B. van Bloemen Waanders, K. Willcox, and Y. Marzouk, *Large-scale inverse problems and quantification of uncertainty*. Wiley Online Library, 2010, vol. 712.
- [382] P. L. Houtekamer and H. L. Mitchell, "A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation," Jan. 1, 2001, ISSN: 1520-0493.

- [383] B. Kramer, A. Marques, B. Peherstorfer, U. Villa, and K. Willcox, “Multifidelity probability estimation via fusion of estimators,” *Journal of Computational Physics*, vol. 392, pp. 385–402, 2019.
- [384] M. Heinkenschloss, B. Kramer, T. Takhtaganov, and K. Willcox, “Conditional-value-at-risk estimation via reduced-order models,” *SIAM/ASA Journal on Uncertainty Quantification*, vol. 6, no. 4, pp. 1395–1423, 2018.
- [385] R. Swischuk, B. Kramer, C. Huang, and K. Willcox, “Learning physics-based reduced-order models for a single-injector combustion process,” *AIAA Journal*, vol. 58:6, pp. 2658–2672, 2020.
- [386] B. Kramer, B. Peherstorfer, and K. Willcox, “Learning nonlinear reduced models from data with operator inference,” *Annual Review of Fluid Mechanics*, vol. 56, pp. 521–548, 1 2024.
- [387] R. Bommasani, D. A. Hudson, E. Adeli, R. Altman, S. Arora, S. von Arx, M. S. Bernstein, J. Bohg, A. Bosselut, E. Brunskill, *et al.*, “On the opportunities and risks of foundation models,” *arXiv preprint arXiv:2108.07258*, 2021.
- [388] P. Vicini and M. J. Kushmerick, “Cellular energetics analysis by a mathematical model of energy balance: Estimation of parameters in human skeletal muscle,” *American Journal of Physiology-Cell Physiology*, vol. 279, no. 1, pp. C213–C224, Jul. 1, 2000, ISSN: 0363-6143, 1522-1563.
- [389] M. J. Lambeth and M. J. Kushmerick, “A Computational Model for Glycogenolysis in Skeletal Muscle,” *Annals of Biomedical Engineering*, vol. 30, no. 6, pp. 808–827, Jun. 2002, ISSN: 0090-6964.
- [390] A. Salti, S. Etemad, M. S. Cubero, E. Albertini, B. Kovacs-Szalka, M. Holzknecht, E. Cappuccio, M. Cavinato, F. Edenhofer, and P. Jansen Dürr, “High Glycolytic Activity Enhances Stem Cell Reprogramming of Fahd1-KO Mouse Embryonic Fibroblasts,” *Cells*, vol. 10, no. 8, p. 2040, Aug. 10, 2021, ISSN: 2073-4409.

The mountains are calling & I must go & I will work on while I can, studying incessantly.

—John Muir