Universidade Estadual de Campinas
Instituto de Computação

Cristiano Gabriel de Souza Campos

Modeling cell signaling pathways through universal differential equations and joint inference of first-principle parameters and neural network weights

Modelando vias de sinalização celular por meio de equações diferenciais universais e inferência conjunta de parâmetros de modelo de primeiros princípios e pesos de rede neural

CAMPINAS
2024

Cristiano Gabriel de Souza Campos

Modeling cell signaling pathways through universal differential equations and joint inference of first-principle parameters and neural network weights

Modelando vias de sinalização celular por meio de equações diferenciais universais e inferência conjunta de parâmetros de modelo de primeiros princípios e pesos de rede neural

Dissertação apresentada ao Instituto de Computação da Universidade Estadual de Campinas como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação.

Dissertation presented to the Institute of Computing of the University of Campinas in partial fulfillment of the requirements for the degree of Master in Computer Science.

Supervisor/Orientador: Prof. Dr. Marcelo da Silva Reis

Este exemplar corresponde à versão final da Dissertação defendida por Cristiano Gabriel de Souza Campos e orientada pelo Prof. Dr. Marcelo da Silva Reis.

CAMPINAS
2024

Ficha catalográfica
Universidade Estadual de Campinas (UNICAMP)
Biblioteca do Instituto de Matemática, Estatística e Computação Científica
Ana Regina Machado - CRB 8/5467

C157m

Campos, Cristiano Gabriel de Souza, 1996-
Modeling cell signaling pathways through universal differential equations and joint inference of first-principle parameters and neural network weights / Cristiano Gabriel de Souza Campos. – Campinas, SP : [s.n.], 2024.

Orientador: Marcelo da Silva Reis.
Dissertação (mestrado) – Universidade Estadual de Campinas (UNICAMP), Instituto de Computação.

1. Sinalização celular. 2. Equações diferenciais universais. 3. Redes neurais (Computação). I. Reis, Marcelo da Silva, 1979-. II. Universidade Estadual de Campinas (UNICAMP). Instituto de Computação. III. Título.

Informações Complementares

Título em outro idioma: Modelagem de vias de sinalização celular por meio de equações diferenciais universais e inferência conjunta de parâmetros de modelo de primeiro princípio e peso de rede neural
Palavras-chave em inglês:
Cell signaling
Universal differential equations
Neural networks (Computer science)
Área de concentração: Ciência da Computação
Titulação: Mestre em Ciência da Computação
Banca examinadora:
Marcelo da Silva Reis
João Meidanis
Yayoi Natsume-Kitatani
Data de defesa: 16-05-2024
Programa de Pós-Graduação: Ciência da Computação

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0009-0002-2352-247X
- Currículo Lattes do autor: http://lattes.cnpq.br/1839718147467617

Universidade Estadual de Campinas
Instituto de Computação

Cristiano Gabriel de Souza Campos

Modeling cell signaling pathways through universal differential equations and joint inference of first-principle parameters and neural network weights

Modelando vias de sinalização celular por meio de equações diferenciais universais e inferência conjunta de parâmetros de modelo de primeiros princípios e pesos de rede neural

Banca Examinadora:

- Prof. Dr. Marcelo da Silva Reis
  Instituto de Computação/UNICAMP
- Prof. Dr. João Meidanis
  Instituto de Computação/UNICAMP
- Dra. Yayoi Natsume-Kitatani
  NIBIOHN/Japão

A ata da defesa, assinada pelos membros da Comissão Examinadora, consta no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria do Programa da Unidade.

Campinas, 16 de maio de 2024

Acknowledgements

Words cannot express my gratitude to my advisor, Marcelo Reis, whose profound knowledge and invaluable feedback have been essential throughout my journey. His expertise has been nothing short of immeasurable. I am equally thankful to my research group members—Ronaldo, Montoni, Marcelo, and Vivian—for their insightful exchanges, consistent feedback, and unwavering support. Additionally, this endeavor would not have been possible without the generous financial support from Santander, which funded my research endeavors (through a Santander BECAS scholarship).

Special thanks are due to Professor Ph.D. Vera, whose teachings have profoundly influenced my academic and personal development. Her guidance was pivotal in encouraging me to pursue this path.

Finally, I would be remiss in not mentioning my family—especially my father, Antônio, my sister, Beatriz, my brother-in-law, Junior, and my nephews, Caio and Matheus. Their unwavering belief in me has been a constant source of encouragement and motivation. Love you all.

Resumo

A orquestração dos processos celulares ocorre através de sequências de reações químicas conhecidas como vias de sinalização celular. Essas vias, fundamentais para a regulação do comportamento celular, enfrentam o desafio do "problema da falta de isolamento", que consiste na ausência de comunicação das reações contidas no modelo com as reações do restante da célula. Esse problema prejudica a modelagem de vias de sinalização celular usando abordagens baseadas em equações diferenciais ordinárias (ODE), pois pode levar à perda de informações contextuais críticas, dificultando assim a precisão da previsão. Para abordar esse problema, uma possibilidade é o uso de uma modelagem híbrida, na qual um modelo baseado em primeiro princípios e ODE é combinado com um modelo baseado em redes neurais e dados. Um framework matemático que implementa tal solução é a Equação Diferencial Universal (UDE). No entanto, em configurações reais de modelagem de vias de sinalização celular baseadas em UDE, deve-se inferir não apenas os pesos da rede neural, mas também parâmetros desconhecidos de primeiros princípios (por exemplo, constantes de taxa faltantes); até onde sabemos, nenhum método de inferência está disponível na literatura. Portanto, aqui propomos uma abordagem para a modelagem de vias de sinalização celular que aproveita o framework da UDE e também infere conjuntamente os parâmetros faltantes do modelo de primeiros princípios e os pesos da rede neural. Para avaliar essa abordagem, realizamos experimentos computacionais usando quatro modelos diferentes de sinalização celular e também a implementação da UDE disponível no ecossistema SciML. Nossos achados demonstram melhorias marcantes tanto na precisão da previsão quanto na interpretabilidade em comparação com a abordagem baseada em ODE, destacando assim a eficácia dos modelos híbridos baseados em UDE para estudos de vias de sinalização celular. Em conclusão, esta pesquisa forneceu algumas ferramentas promissoras para explorar a dinâmica complexa dos sistemas biológicos.

Abstract

The orchestration of cellular processes occurs through sequences of chemical reactions known as cell signaling pathways. These pathways, fundamental to the regulation of cellular behavior, face the challenge of the "lack of isolation problem", which consists of the absence of communication of the reactions contained in the model with the reactions of the remainder of the cell. This issue impairs the modeling of cell signaling pathways using ordinary differential equation (ODE)-based approaches, as it can lead to the loss of critical contextual information, thereby hindering the prediction accuracy. To address this problem, one possibility is the usage of a hybrid modeling, in which a first-principle, ODE-based model is combined with a data-driven, neural network-based model. One mathematical framework that implements such solution is the Universal Differential Equation (UDE). However, in real-world settings of UDE-based cell signaling pathway modeling, one should infer not only the neural network weights, but also first-principle unknown parameters (*e.g.*, missing rate constants); to the best of our knowledge, no such inference method is available in the literature. Therefore, here we propose an approach for the modeling of cell signaling pathways that leverages the UDE framework and also jointly infers missing parameters of the first-principle model and the neural network weights. To assess that approach, we carried out computational experiments using four different cell signaling models and also the UDE implementation available at the SciML ecosystem. Our findings demonstrate marked improvements in both prediction accuracy and interpretability in comparison with the ODE-based approach, thus highlighting the effectiveness of UDE-based hybrid models for cell signaling pathway studies. In conclusion, this research provided some promising tools for exploring the complex dynamics of biological systems.

List of Figures

1.1 The Lack of Isolation Problem in Modeling cell signaling Pathways. 15

2.1 Neural Network with a single Layer 27

3.1 Flowchart illustrating our systematic review process in accordance with PRISMA guidelines. 32

4.1 Flowchart depicting the step-by-step methodology for addressing the lack of isolation problem in cell signaling pathways. This includes the generation of time series data, definition and refinement of pathway cutouts, inference of the hybrid model, and the subsequent evaluation of model performance. 39
4.2 Diagram in systems biology graphical notation (SBGN), depicting the Toy Model cell signaling pathway. 47
4.3 SBGN diagram of the JAK-STAT cell signaling pathway. 49
4.4 SBGN diagram of the MAPK OrderedElementary cell signaling pathway. Nodes symbolize chemical species (such as proteins) and edges denote reactions. 50
4.5 SBGN diagram of the FeMetabolism FeDeficient cell signaling pathway. . . 51

5.1 Model performance for the Toy Model, where modeling approaches are compared based on the SMAPE for training. 54
5.2 Data expansion impact on UDE with neural network for the Toy Model. . 55
5.3 Predictive accuracy of the UDE with neural network for the Toy Model (10 observations). 56
5.4 Predictive accuracy of the UDE with neural network for the Toy Model (100 observations). 56
5.5 Impact of first-principle parameter removal on mean SMAPE for the Toy Model. 57
5.6 Model performance across noisy datasets for the Toy Model. 58
5.7 Predictive accuracy of the UDE with neural network in noisy conditions for the Toy Model (10 observations). 59
5.9 Data expansion impact on UDE with neural network for the JAK-STAT. Increasing the number of training observations affects the mean SMAPE for the UDE model integrated with a neural network. 59
5.8 Model performance for the JAK-STAT, where modeling approaches are compared based on the SMAPE for training, validation, and test sets (each set containing 10 observations). 60
5.10 Predictive accuracy of the UDE with linear regression (10 observations). . 61
5.11 Predictive accuracy of the UDE with linear regression (100 observations). . 61
5.12 Predictive accuracy of the UDE with linear regression (1000 observations). 62

5.13 Impact of first-principle parameter removal on mean SMAPE for the JAK-STAT pathway. 63
5.14 Model performance across noisy datasets for the JAK-STAT pathway. . . . 64
5.15 Predictive accuracy of the UDE with neural network in noisy conditions for the JAK-STAT pathway (10 observations). 64
5.16 Impact of first-principle parameter removal on mean SMAPE for the JAK-STAT pathway, in noisy conditions. 65
5.17 Model performance for the MAPK OrderedElement model, where modeling approaches are compared based on the SMAPE for training, validation, and test sets (each set containing 10 observations). 66
5.18 Data expansion impact on UDE with neural network for the MAPK orderedElemented. 67
5.19 Predictive accuracy of the UDE with neural network for the MAPK orderedElementary (10 observations). 67
5.20 Predictive accuracy of the UDE with neural network for the MAPK orderedElementary (100 observations). 68
5.21 Predictive accuracy of the UDE with neural network for the MAPK orderedElementary (1000 observations). 68
5.22 Impact of first-principle parameter removal on mean SMAPE for the MAPK orderedElementary. 69
5.23 Model performance across noisy datasets for the MAPK orderedELementar.This graph compares different modeling approaches under noisy conditions by showcasing the SMAPE values across training, validation, and test datasets, each with 10 observations. 70
5.24 Predictive accuracy of the UDE with neural network in noisy conditions for the MAPK orderedELementay (10 observations). 70
5.25 Impact of first-principle parameter removal on mean SMAPE for the MAPK orderedElementary, in noisy conditions. It is depicted here the effect on the mean SMAPE of iteratively removing sets of first-principle parameters, using 10 noisy observations for training set, 10 for validation set, and 10 for test set. 71
5.26 Model performance for the FeMetabolism FeDeficient Model, where modeling approaches are compared based on the SMAPE for training, validation, and test sets (each set containing 10 observations). 72
5.27 Data expansion impact on UDE with linear regression for the FeMetabolism FeDeficient Model. 73
5.28 Predictive accuracy of the UDE with linear regression for the FeMetabolism FeDeficient model (10 observations). 73
5.29 Predictive accuracy of the UDE with linear regression for the FeMetabolism FeDeficient model (100 observations). 74
5.30 Predictive accuracy of the UDE with linear regression for the FeMetabolism FeDeficient model (1000 observations). 74
5.31 Impact of first-principle parameter removal on mean SMAPE for the FeDeficient FeMetabolism model. 75
5.32 Model performance across noisy datasets for the FeMetabolism FeDeficient Model. 77
5.33 Predictive accuracy of the UDE with linear regression in noisy conditions for the JFeMetabolism FeDeficient model (10 observations). 77

5.34 Impact of first-principle parameter removal on mean SMAPE for the FeMetabolism FeDeficient model, in noisy conditions. It is depicted here the effect on the mean SMAPE of iteratively removing sets of first-principle parameters, using 10 noisy observations for training set, 10 for validation set, and 10 for test set. 78

List of Tables

3.1 Literature review of hybrid modeling, including UDE-bases ones. For each paper, we present the type of dataset used in the paper, the data-driven model couple in the first-principle model, and the number of evaluated case studies. 33
3.2 Literature review of first-principle parameter estimation. For each paper, we present the type of dataset used in the paper, the used estimation technique, and the number of evaluated case studies. 36

4.1 First-principle parameters of the Toy Model pathway cutout. 47
4.2 First-principle parameters of the JAK-STAT pathway cutout. 49
4.3 First-principle parameters of the MAPK orderedElementary pathway cutout. 51
4.4 First-principle parameters of the FeMetabolism FeDeficient pathway cutout. 52

A.1 Complete sequence of first-principle parameter removal sets for the MAPK orderedElementary model. 86
A.2 Complete sequence of first-principle parameter removal sets for the MAPK orderedElementary model using noisy data. 86
A.3 Complete sequence of first-principle parameter removal sets for the FeMetabolism FeDeficient model. 87
A.4 Complete sequence of first-principle parameter removal sets for the FeMetabolism FeDeficient model using noisy data. 87
A.5 Complete sequence of first-principle parameter removal sets for the JAK-STAT model. 88
A.6 Complete sequence of first-principle parameter removal sets for the JAK-STAT model using noisy data. 89
A.7 Full list of the sequence of the first-principle removal for Toy Model. 90

Contents

1 Introduction **14**
- 1.1 Research questions . . . 16
- 1.2 Contributions of this dissertation . . . 17
- 1.3 Structure of work . . . 18

2 Fundamental Concepts **19**
- 2.1 Cell signaling pathways: A brief overview . . . 19
- 2.2 Modeling of cell signaling pathways . . . 20
  - 2.2.1 First-principle modeling . . . 20
  - 2.2.2 Data-driven modeling . . . 21
- 2.3 Chemical kinetics . . . 22
  - 2.3.1 Reaction rate . . . 22
  - 2.3.2 Parallel reactions . . . 24
  - 2.3.3 Sequential reactions . . . 24
  - 2.3.4 Reversible reactions . . . 25
  - 2.3.5 Enzymatic reaction . . . 26
- 2.4 Neural networks . . . 27
- 2.5 Universal differential equations . . . 29

3 Literature Review **31**
- 3.1 Hybrid modeling in chemical reactions . . . 33
- 3.2 Advancements in UDE-based hybrid modeling . . . 34
  - 3.2.1 Estimation of First-Principle Parameters . . . 36
- 3.3 Identification of research gaps . . . 38
  - 3.3.1 Lack of generalization across multiple initial conditions . . . 38
  - 3.3.2 Joint inference of first-principle parameters and neural network weights 38

4 Methodology **39**
- 4.1 Proposed method . . . 40
  - 4.1.1 The inference of the hybrid model . . . 40
- 4.2 Experimental details . . . 42
  - 4.2.1 Data acquisition for modeling . . . 42
  - 4.2.2 Training process . . . 43
  - 4.2.3 Model assessment . . . 44
  - 4.2.4 Description of the experiments . . . 45
  - 4.2.5 Computational resources . . . 46
- 4.3 Case studies . . . 46
  - 4.3.1 Toy Model . . . 46
  - 4.3.2 JAK-STAT pathway . . . 48

4.3.3 MAPK OrderedElementary 50
4.3.4 FeMetabolism FeDeficient 51

5 Results and Discussion **53**
5.1 Results with the Toy model 54
5.1.1 Experiments with noisy data for the Toy Model 56
5.2 Results with JAK-STAT pathway 57
5.2.1 Experimentation with noisy data for the JAK-STAT pathway . . . 61
5.3 Results with MAPK orderedElementary 63
5.3.1 Experimentation with noisy data for the MAPK orderedElementary 68
5.4 Results with FeMetabolism FeDeficient 71
5.4.1 Experimentation with noisy data for the FeDeficient and FeMetabolism pathway model 75
5.5 Discussion 76

6 Conclusion **79**
6.1 Answers to the research questions 79
6.2 List of publications and other contributions 80
6.3 Suggestions for future research 80

Bibliography **82**

A Lists of removed sets of rate constants **86**

Chapter 1

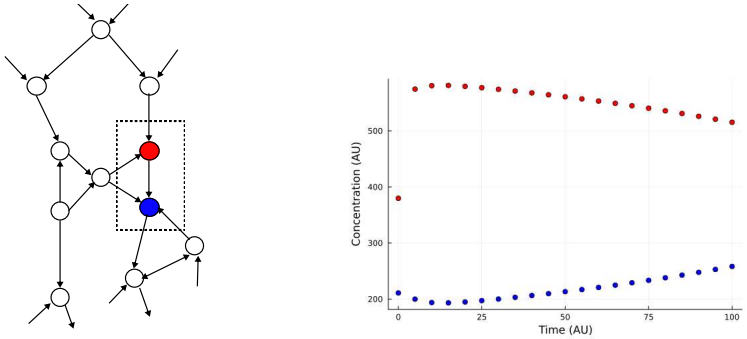
Introduction

The intricate dynamics of cell signaling pathways is fundamental for the understanding of the myriad processes that govern cellular behavior. These pathways consist of sets of chemical species, often proteins, that are intertwined with each other through chains of chemical reactions. Information is propagated along those pathways through changes of concentration in the involved chemical species along time, often through protein-protein interactions and/or protein post-translational modifications (*e.g.*, phosphorylation). Cell signaling pathways, characterized by their complex interactions and regulatory mechanisms, present a significant modeling challenge within the field of systems biology.

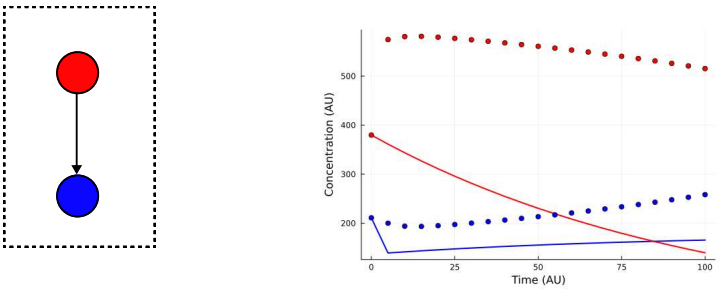
Traditionally, ordinary differential equations (ODEs) have been employed to model these pathways, offering a deterministic framework to elucidate the temporal evolution of cellular processes [30]. To this end, chemical reactions of the pathway are modeled using differential equations that describe the rate of those reactions. Kinetic constants of the chemical reactions that were previously experimentally measured are gathered from the literature or from repositories such as SABIO-RK [37] [1]; for the unknown kinetic constants, time series measurements must be done for one or more chemical species in the cellular system of interest, and their values must be inferred through an optimization method (*e.g.*, simulated annealing [35]).

However, a persistent challenge in accurately modeling cell signaling pathways through ODEs is the "lack of isolation" problem, wherein the interconnected nature of cellular processes often prevents the possibility of modeling pathways in isolation without losing critical contextual information [29]. More specifically, when one selects a set of chemical reactions to compose a cell signaling pathway model, it is left aside reactions which involves chemical species not included in the model (Figure 1.1(a), left). This can lead to a model whose fit does not explain the behaviour of the actual cellular system (Figure 1.1(b)). Common approaches to this problem includes: the inclusion of more reactions into the model (Figure 1.1(c), left), which might lead to overfitting; the inclusion of putative species (Figure 1.1(c), right), which might lead to an unrealistic model. Both approaches might yield a model that fails to predict the behaviour of the cell signaling pathway for different initial conditions. Therefore, it is desirable the development of alternative approaches to deal with the lack of isolation problem.

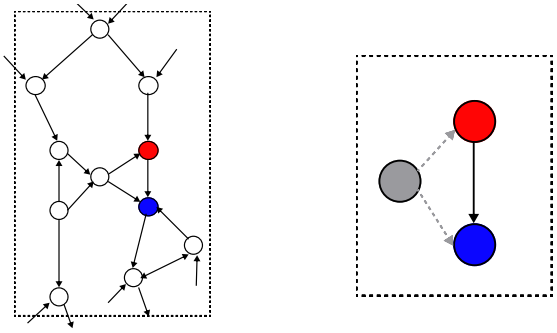
[1] sabio.h-its.org/.



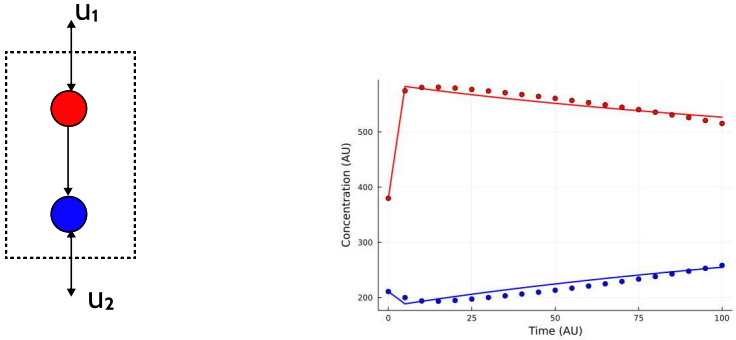
(a)



(b)



(c)



(d)

Figure 1.1: **The Lack of Isolation Problem in Modeling cell signaling Pathways.** Fig. 1.1a: A signaling pathway with two chemical species (orange and blue) measured over a period of time. A set of reactions is selected for modeling. Fig. 1.1b: The model fails to explain the observed data. Fig. 1.1c: Common approaches include expanding the model or introducing hypothetical species (in gray), whose existence is uncertain. Fig. 1.1d: By providing missing inputs, the model reproduces the observed data. Adapted from Reis and Liepe [29].

In recent years, the advent of data-driven models, particularly those leveraging the power of machine learning, has offered new avenues for modeling complex biological systems [5]. These models perform well in handling vast datasets, learning patterns, and making predictions where traditional mechanistic (first-principle) models may falter due to their their tendency to oversimplify or omit complex network interactions. Nonetheless, a significant drawback of purely data-driven approaches is their lack of interpretability, a critical aspect when aiming to understand the underlying biological phenomena.

To address these challenges, hybrid models have emerged as a promising solution, integrating the interpretability of mechanistic models with the flexibility and learning capabilities of neural networks. A notable attempt by Lee and colleagues in 2020 proposed a complex hybrid method to tackle the lack of isolation problem [15]. However, this method was criticized for its complexity and reliance on ad hoc procedures, highlighting the need for a more streamlined and effective approach.

The introduction of universal differential equations (UDEs) in 2021 marked a significant milestone in hybrid modeling, providing a framework that combines the strengths of mechanistic models and neural networks [27]. UDEs have since been applied in various contexts, such as the work by Bangi and colleagues in 2022, which demonstrated their superior performance in modeling the production of beta-carotene in yeast [1]. More recently, in 2023, Santana et al. explored a novel UDE-based hybrid modeling technique for the kinetics of sorption uptake, resulting in an effectively fitted model that underscores the potential of UDEs in addressing complex biological modeling challenges [31].

Recently, our research group reported the application of Universal Differential Equations (UDEs) to address the lack of isolation problem in cell signaling pathways [32]. The methodology of this work involved the integration of a neural network with a system of ordinary differential equations that delineates the dynamics of the investigated signaling pathway. The primary objective of this integration was to enable the neural network to learn the interactions between proteins within the designated pathway and those external to it. This configuration allows for the identification of extrinsic protein interactions that influence the dynamics of the pathway (refer to Figure 1.1(d), left; in our model, the u_i is represented by the output of a neural network).

However, one notable drawback of the approach is the presupposition that all first-principle parameters are known, which seldom holds true in real-word scenarios. Therefore, there is a need for the development of UDE-based models of cell signaling pathways where one or more first-principle parameters, such as rate constants, are jointly estimated from data along with the weights of the neural network. As far we know, this is still an open problem.

1.1 Research questions

The endeavor to enhance the fidelity and applicability of hybrid models for cell signaling pathways in scenarios involving missing first-principle parameters necessitates a meticulous investigation into the methodologies and techniques employed in their development. Central to this investigation are two pivotal research questions that aim to dissect and

address the core challenges of modeling such complex biological systems. These questions are instrumental in guiding the direction of our research and in formulating the methodologies that will be employed. The research questions (RQs) are as follows:

RQ 1 *How well can we identify hybrid models of cell signaling pathways by employing time series data of species present in the first-principle part of the model?*

This question addresses the practical applicability and robustness of hybrid models in capturing the dynamics of cell signaling pathways, with a particular focus on addressing the lack of isolation problem. The performance and adaptability of the hybrid models are assessed using the Symmetric Mean Absolute Percentage Error (SMAPE) metric, providing a quantitative measure of the fidelity of the model in capturing the dynamics of the cellular processes under study.

RQ 2 *How can we efficiently infer the parameters of the hybrid model, encompassing both the mechanistic (first-principle) and data-driven components?*

This question probes the challenge of parameter inference within hybrid models, seeking strategies that effectively balance the interpretability of mechanistic models with the predictive power of data-driven approaches. The inquiry is focused on identifying methods that not only optimize the fit of the model to simulated data but also ensure that the parameters reflect biologically plausible values. This involves exploring novel computational techniques, optimization algorithms, and model architectures that can integrate the distinct nature of data derived from mechanistic understandings and empirical observations.

Together, these research questions aim to unravel the complexities of hybrid modeling for cell signaling pathways, seeking to improve the accuracy of the model, interpretability, and practical utility. By addressing these questions, the research seeks to contribute significantly to the field of computational biology, paving the way for more sophisticated and reliable models that can aid in the understanding and manipulation of biological systems.

1.2 Contributions of this dissertation

The main scientific contribution of this dissertation is a method, rooted in the principles of universal differential equations, to overcome the 'lack of isolation' challenge in modeling cell signaling pathways, in scenarios where one or more first-principle parameters (in this case, rate constants) are missing. Unlike previous methods that primarily rely on a single initial condition time series for model training and validation, our methodology employs multiple initial conditions to train the model, followed by distinct conditions for validation and testing. This not only enhances the robustness of the model but also ensures its applicability across a wider range of biological scenarios. Furthermore, recognizing that first-principle parameters are not always known *a priori*, we introduce a novel method for simultaneously training the weights of the neural network and estimating the

first-principle parameters. This dual approach represents a significant departure from conventional strategies, promising to deliver models with greater accuracy, interpretability, and relevance to biological research.

1.3 Structure of work

After this introduction, this dissertation is organized into five main chapters, each of them designed to systematically guide the reader through the research undertaken.

Chapter 2 lays the foundation by introducing the key concepts necessary for understanding the context and framework of our study. It provides a comprehensive overview of the theoretical foundations that form the basis of our research.

Chapter 3 offers an literature review, setting the stage for our research by examining existing studies and theories related to our topic. This chapter not only contextualizes our work within the broader academic discourse but also identifies gaps in current knowledge, paving the way for our investigation.

Chapter 4 details the methodology employed in our work. This section elaborates on the research design, data acquisition, and techniques used to address our research questions, offering insight into the procedural aspects of our study.

Chapter 5 presents the findings of our experiments. It discusses the outcomes of our research, providing a detailed analysis and highlighting significant observations and patterns identified during the study.

Finally, Chapter 6 concludes the dissertation by summarizing the key findings, discussing their implications, and answering the research questions. This chapter also outlines the contributions of our work to the field and suggests directions for future research, pointing towards potential areas for further investigation.

Chapter 2

Fundamental Concepts

In this chapter, we establish the essential concepts that underlie our research methodology, beginning with an overview of cell signaling pathways in Section 2.1. This sets the stage for understanding the modeling techniques discussed in Section 2.2, crucial for representing the cascades of chemical reactions inherent to these pathways. We then highlight the principles of chemical reaction kinetics in Section 2.3, essential for comprehending the dynamic behaviors within cell signaling. Section 2.4 recalls the fundamentals of the neural network, a data-driven model. Concluding the chapter, Section 2.5 brings the basic principles of universal differential equations (UDEs), the formalism behind our hybrid modeling approach.

2.1 Cell signaling pathways: A brief overview

The cell represents the fundamental metabolic unit of any living organism. Bridging this concept with cellular communication, cell signaling pathways are sophisticated communication systems that cells utilize to transmit and receive vital information. These pathways enable cells to respond to environmental stimuli, adapt to external changes, and execute specific functions.

Cellular function is grounded in various processes, including growth, proliferation, migration, and programmed cell death, among others. These processes are orchestrated by messages transmitted through cell signaling pathways. Comprising a series of chemical reactions, such as enzymatic reactions, the products of one reaction often serve as substrates for subsequent reactions. For a cell to maintain its health and functionality, it is imperative that these cellular processes are meticulously coordinated. Disruptions in cell signaling pathways are implicated in the pathophysiology of several diseases, including cancer [9, 11].

Imagine the intricate workings of cell signaling pathways akin to the multifaceted technology of a smartphone. Just as a smartphone connects with a network, receiving and processing a myriad of signals – be it a call, a text, or an app notification – it promptly responds with specific actions like ringing, vibrating, or displaying messages. In a similar vein, cells in our body are equipped with their own sophisticated "communication technology". They continually intercept and interpret diverse signals from their surroundings

or other cells. These signals, akin to incoming calls or messages, trigger a series of complex, well-orchestrated intracellular reactions. These reactions, much like the apps on a smartphone, are tailored to perform specific functions, ultimately guiding the response of the cell to its dynamic environment.

This concept is exemplified vividly in the process of glucose regulation, a critical aspect of the internal communication network of our body. Take, for instance, the process of glucose regulation. Upon consuming food, glucose levels rise in the bloodstream, serving as a stimulus. In response, the pancreas secretes insulin, acting as the "messenger". This hormone traverses the body, binding to specific receptors on the surface of muscle cells, liver cells, and adipose tissues. The binding of insulin initiates a series of intracellular events, akin to a domino effect, where one event triggers the subsequent one. Ultimately, this cascade leads to enhanced glucose uptake from the blood, ensuring homeostasis of glucose levels.

Thus, the insulin-mediated cell signaling pathway exemplifies how cells detect environmental alterations and generate appropriate responses. It is imperative to understand that various distinct cell signaling pathways operate in the human body, each tailored to respond to unique stimuli and elicit specific cellular responses. These pathways are pivotal in orchestrating and modulating various physiological processes.

2.2 Modeling of cell signaling pathways

Owing to the intricacies of regulatory feedback loops and signal transduction branches, cell signaling pathways are complex nonlinear systems whose dynamics are influenced by the temporal variations in the concentration of chemical species, such as proteins [28]. Given the inherent nonlinearity of these systems, human intuition alone is insufficient to predict the behavior of a specific cell signaling pathway based on alterations in initial conditions. For instance, understanding how a particular pathway responds to various cellular stimuli, such as the introduction of different compounds into a cell culture, requires more than just observational insights. In this context, modeling the dynamic of cell signaling pathways becomes an indispensable tool for mechanistic studies of cellular processes. Two predominant approaches for modeling cell signaling pathways are the first-principle models and data-driven models.

2.2.1 First-principle modeling

First-principle modeling involves describing the signaling pathway based on the physicochemical principles underlying its operation. In this context, a classic approach assumes that, for each chemical species, the number of copies of the species molecule (i.e., its concentration within the cell) is sufficiently large and that species' copies comprise a homogeneous mixture. Under these assumptions, it is feasible to model a signaling pathway using systems of ordinary differential equations (ODEs) [30]. These equations describe the temporal changes in concentration according to the principles of chemical kinetics.

For instance, a cell signaling pathway can be characterized as follows:

$$\frac{d\mathbf{x}(t)}{dt} = f(\mathbf{x}(t), \mathbf{u}(t); \boldsymbol{\theta}) \tag{2.1a}$$

$$\mathbf{y}(t) = h(\mathbf{x}(t)) + \epsilon(t) \tag{2.1b}$$

$$\mathbf{x}(0) = \mathbf{x}_0, \tag{2.1c}$$

where $\mathbf{x}(t) \in \mathbb{R}^n_+$ is a state vector at time t, $\frac{d\mathbf{x}(t)}{dt}$ is the rate of change of the state vector at t, $\mathbf{x}_0$ is the initial value of the system, and $\mathbf{u}(t)$ is the vector of known inputs belonging to $\mathbb{R}^n$. The output $\mathbf{y}(t)$ of the system is composed of the sum of a function h applied to the state vector $\mathbf{x}(t)$ and an error function $\epsilon(t)$. The function f describes the dynamic model using arguments $\mathbf{x}(t)$ and $\mathbf{u}(t)$.

First-principle parameters $\boldsymbol{\theta}$ presumed to be well-defined are often fixed, while those that are either not well-understood or whose alterations lead to significant variations in model outputs are targeted for estimation via optimization techniques [24]. Additionally, certain scenarios necessitate the simultaneous adjustment of all first-principle parameters. In this context, first-principle parameters can be determined through the application of optimization techniques, such as simulated annealing [19].

2.2.2 Data-driven modeling

Data-driven modeling offers a powerful toolkit for unraveling the complexities of cell signaling pathways, utilizing both unsupervised and supervised learning approaches to extract meaningful insights from biological data.

In unsupervised learning, principal component analysis (PCA) is a widely adopted technique for the dimensional reduction of phosphoprotein data, facilitating the stratification of drug-responsive melanoma cell lines and the discernment of distinctive pathways in patients with primary Sjögren's syndrome compared to healthy individuals [24]. Beyond PCA, nonlinear projection methods such as t-distributed stochastic neighbor embedding (t-SNE) and uniform manifold approximation and projection (UMAP) have demonstrated superior performance in segregating single-cell measurements, including mass cytometry and single-cell RNA sequencing data [24]. These advanced methods provide a more nuanced separation of cellular states, enabling a deeper exploration of cellular heterogeneity and pathway activation.

The realm of supervised learning encompasses a diverse array of methodologies, including linear models, tree-based models, neural networks (detailed in Section 2.4), and ensemble methods, each offering unique advantages for modeling cell signaling dynamics. A notable application involved the integration of t-SNE with tree-based classifiers to delineate various clinically significant epithelial-mesenchymal transition states in lung cancer, utilizing mass cytometry data [24]. Furthermore, a comprehensive study undertook the challenge of employing machine learning models to predict missing markers, novel conditions, and the dynamic responses of single cells to stimuli, with or without kinase inhibitors [5]. This ambitious project leveraged a mass cytometry dataset encompassing 36 markers across more than 4000 conditions, collecting data on 80 million single

cells from 67 breast cancer cell lines [5]. The top three teams in this challenge achieved notable results using an ensemble of linear and tree-based methods, convolutional neural networks, and gradient boosting technique.

2.3 Chemical kinetics

The examples and concepts of this subsection were mainly obtained from various resources [7, 14, 23].

As previously mentioned, an effective approach to modeling cell signaling pathways is achieved through chemical kinetics. Chemical reactions are understood as interactive processes between substances, which often result in the formation of new substances. Chemical kinetics plays a crucial role in studying the rates of these reactions and the factors influencing these rates.

In cell signaling pathways, the concept of chemical kinetics is fundamental. Cells use signaling pathways to process information and respond to external and internal stimuli. These pathways function through a series of chemical reactions, where molecules such as proteins and small metabolites interact in a specific and regulated manner. Each reaction within a signaling pathway can be viewed as an individual step in a broader chemical process, where the speed and efficiency of the reaction are vital for an appropriate cellular response.

For instance, the binding of a hormone to a receptor on the cell surface triggers a cascade of reactions, each following principles of chemical kinetics. The efficiency with which these reactions occur and the speed at which the products are formed or degraded determine the intensity and duration of the cellular response. Factors such as enzyme concentration, the presence of co-factors, and the physical and chemical conditions of the cellular environment can alter the dynamics of these reactions, thus influencing the overall behavior of the signaling pathway.

Therefore, the study of chemical kinetics in cell signaling pathways is not limited to understanding isolated reactions but encompasses an integrative analysis of how these interconnected reactions modulate complex cellular processes. This approach offers valuable insights into how cells regulate their vital functions and adapt to changes in their environment, proving crucial for advancements in fields such as pharmacology, regenerative medicine, and biotechnology.

2.3.1 Reaction rate

The rate of a chemical reaction is determined by the speed at which reactants are consumed and converted into products. Therefore, the rate of a reaction can be represented either by the rate of consumption of the reactants or by the rate of formation of the products.

Consider the chemical the following chemical reaction:

$$\mathrm{aA} + \mathrm{bB} \xrightarrow{k} \mathrm{cC} + \mathrm{dD}. \tag{2.2}$$

In Reaction 2.2, A and B are represent the reactants, while C and D are referred to as products. The variables a, b, c, and d denote the stoichiometric coefficients, which represent the proportional quantities in which reactants and products participate in the reaction. The symbol k denotes the rate constant, a parameter that quantifies the rate at which the reaction proceeds under given conditions.

The rate of a chemical reaction, denoted as v, can be mathematically represented by the following equation [3]:

$$v = k \prod_{i=1}^{m} [R_i]^{r_i} . \tag{2.3}$$

Here, k represents the rate constant of the reaction, which is a measure of the intrinsic speed at which the reaction occurs. The term $[R_i]$, where i ranges from 1 to m, refers to the molar concentration of the i^{th} reactant. The exponent r_i is known as the reaction order with respect to the reactant R_i and dictates the dependency of the reaction rate on the concentration of R_i.

This expression is a generalized rate law for a chemical reaction, illustrating that the reaction rate is proportional to the product of the concentrations of the reactants, each raised to a power corresponding to the stoichiometric coefficient in the balanced chemical equation, which is also the reaction order for that reactant [3]. The reaction order can be any real number and reflects the sensitivity of the reaction rate to changes in the concentration of that particular reactant. Thus, the rate of the chemical reaction 2.2 can be calculated by:

$$v = k \, [A]^a \, [B]^b . \tag{2.4}$$

Fundamentally, the rate of a chemical reaction dendictates the rapidity at which reactant concentrations deplete and product concentrations accumulate. Using the expression 2.4, we can describe the rate of variation of concentration of each element in 2.2 by:

$$\frac{d[B]}{dt} = -k[A]^a[B]^b \tag{2.5a}$$

$$\frac{d[C]}{dt} = k[A]^a[B]^b \tag{2.5b}$$

$$\frac{d[D]}{dt} = k[A]^a[B]^b, \tag{2.5c}$$

where $[A]$, $[B]$, $[C]$ and $[D]$ are, respectively, the concentrations of species A, B, C and D at time t (for the sake of simplicity, we write "$[A]$" instead of "$[A](t)$"). These differential equations delineate how the concentrations of reactants and products change over time. The negative sign in the first two equations indicates that the concentrations of A and B are decreasing, which is expected, as they are reactants being consumed in the reaction. The last two equations have positive signs, indicating that the concentrations of C and D are increasing, as would be expected for products of a reaction.

These rates of change are derived from the rate law. In the generic reaction, the concentrations of A and B decrease as the reaction progresses, which is reflected by the negative sign in the rate of change. The rate of consumption of A and B is proportional

to the rate law. For the products C and D, the formation rates are positive and also proportional to the rate law, reflecting the increase in their concentrations as the reaction proceeds.

2.3.2 Parallel reactions

Parallel reactions are those where a single reactant undergoes different pathways to yield multiple products. Consider the following example:

$$\mathrm{A} \xrightarrow{k_1} \mathrm{B} \tag{2.6a}$$

$$\mathrm{A} \xrightarrow{k_2} \mathrm{C}. \tag{2.6b}$$

In these reactions, the reactant A can be converted into either product B or C, each reaction pathway characterized by distinct rate constants (k_1 and k_2). This scenario is typical in systems where a reactant can be transformed through various reaction types or mechanisms. The rate of consumption or formation of each species in the reactions 2.6a and 2.6b can be mathematically described as:

$$\frac{d[A]}{dt} = -k_1[A] - k_2[A] \tag{2.7a}$$

$$\frac{d[B]}{dt} = k_1[A] \tag{2.7b}$$

$$\frac{d[C]}{dt} = k_2[A]. \tag{2.7c}$$

Here, the rate of change in the concentrations of B and C is dependent solely on the concentration of A and their respective rate constants. This is because these reactions are first-order with respect to A. The negative sign in the rate expression for A indicates its consumption, while the positive signs in the expressions for B and C indicate their formation.

2.3.3 Sequential reactions

Sequential reactions occur when the product of one reaction serves as the reactant for a subsequent reaction. Consider the following example:

$$\mathrm{A} \xrightarrow{k_1} \mathrm{B} \xrightarrow{k_2} \mathrm{C}. \tag{2.8}$$

Reaction 2.8 illustrates a series of chemical reactions where the reactant A is first transformed into B with a rate constant k_1, and then B is further converted into C with a rate constant k_2. The dynamics of the concentrations of A, B, and C over time, as governed by the kinetics of these sequential reactions, can be described by the following differential

equations:

$$\frac{d[A]}{dt} = -k_1[A], \tag{2.9a}$$

$$\frac{d[B]}{dt} = k_1[A] - k_2[B], \tag{2.9b}$$

$$\frac{d[C]}{dt} = k_2[B]. \tag{2.9c}$$

The rate of change in the concentration of A is given by $-k_1[A]$, indicating the consumption of A to produce B, where the rate is directly proportional to the concentration of A. The concentration of B increases as A is converted into B, at a rate of $k_1[A]$, but B is concurrently being consumed to produce C at a rate of $k_2[B]$. Therefore, the net rate of change in the concentration of B is the balance between its formation from A and its consumption to form C.

Lastly, the increase in the concentration of C is attributed to the conversion of B into C, with the formation rate of C being $k_2[B]$. This indicates that the rate at which C is formed is directly contingent upon the concentration of B.

2.3.4 Reversible reactions

Reversible reactions are characterized by the ability of the products to convert back into reactants, thereby establishing an equilibrium state. Consider the following reaction as an example:

$$\mathrm{A} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathrm{B}. \tag{2.10}$$

Reaction 2.10 depicts a reversible chemical reaction where the reactant A is converted into product B with a forward rate constant k_1, and conversely, B is converted back into A with a reverse rate constant k_2. At equilibrium, the rate of conversion from A to B is equal to the rate of reconversion from B to A, resulting in constant concentrations of A and B over time. The ratio of the rate constants k_1 to k_2 determines the position of the chemical equilibrium, indicating the relative concentrations of the reactants and products at equilibrium. This relationship is crucial for understanding how changes in conditions or rate constants can shift the equilibrium position. The rates of change in the concentrations of A and B due to the rate law can be mathematically described as follows:

$$\frac{d[A]}{dt} = -k_1[A] + k_2[B], \tag{2.11a}$$

$$\frac{d[B]}{dt} = k_1[A] - k_2[B]. \tag{2.11b}$$

Equation 2.11a indicates that the rate of decrease in the concentration of A is proportional to its concentration multiplied by the forward rate constant k_1, and is offset by the rate at which B is converted back to A, which is proportional to the concentration of B and the reverse rate constant k_2. Conversely, Equation 2.11b demonstrates that the rate of increase in the concentration of B is proportional to the concentration of A and the

forward rate constant k_1, diminished by the rate at which B converts back to A, reflecting the influence of the concentration of B and the reverse rate constant k_2.

2.3.5 Enzymatic reaction

Enzymes are proteins that function as biological catalysts, speeding up chemical reactions within an organism without being consumed by the reaction. An enzymatic reaction involves the chemical transformation of substrates into products, facilitated by a specific enzyme. Consider the following reaction as an example:

$$\mathrm{E} + \mathrm{S} \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \mathrm{ES} \xrightarrow{k_3} \mathrm{E} + \mathrm{P}.$$

In this reaction, an enzyme (E) binds to a substrate (S) to form an enzyme-substrate complex (ES), as denoted by the reversible arrows with rate constants k_1 for the forward reaction and k_2 for the reverse reaction. This ES complex can either dissociate back into E and S or be converted to the product (P), with the enzyme being released to catalyze subsequent reactions, indicated by the irreversible conversion with the rate constant k_3. This step signifies the substrate's conversion to the product, showcasing the enzyme's role in facilitating this transformation without being consumed, thereby being available for further catalytic cycles. This mechanism is crucial for the regulation of metabolic pathways and highlights the dynamic nature of enzyme activity. The rate of change of each element due to the rate law can be mathematically described as follows:

$$\frac{d[E]}{dt} = -k_1[E][S] + k_2[ES] + k_3[ES], \tag{2.12a}$$

$$\frac{d[S]}{dt} = -k_1[E][S] + k_2[ES], \tag{2.12b}$$

$$\frac{d[ES]}{dt} = k_1[E][S] - k_2[ES] - k_3[ES], \tag{2.12c}$$

$$\frac{d[P]}{dt} = k_3[ES]. \tag{2.12d}$$

The enzyme concentration $[E]$ changes due to the formation of the ES complex and the release of E either upon the conversion of the complex to P or its dissociation back into E and S. The binding of E to S decreases $[E]$ (hence the negative term), while the dissociation of ES and the conversion to P release E back into the system, reflected by positive terms in the equation for $\frac{d[E]}{dt}$. The substrate concentration $[S]$ diminishes as it binds to E to form ES, with the potential to increase if the ES complex dissociates back into E and S, as shown in the equation for $\frac{d[S]}{dt}$ with a negative term for the formation of ES and a positive term for its dissociation. The concentration of the ES complex is governed by its formation through the binding of E and S and its reduction via conversion to P or dissociation back to E and S, captured in the equation for $\frac{d[ES]}{dt}$ with a positive term for its formation and negative terms for its decrease. Finally, the product P is generated exclusively through the conversion of the ES complex into P, as illustrated by the equation for $\frac{d[P]}{dt}$, indicating the direct relationship between the rate of product

formation and the concentration of the ES complex.

2.4 Neural networks

The concepts in this subsection were primarily derived from the book "An Introduction to Statistical Learning with Applications in Python," written by James and colleagues [10].

A neural network can be defined as a nonlinear real function, f, which takes as input a vector of size p, $X = (X_1, \ldots, X_p)$, where $X_i \in \mathbb{R}$ is known as a feature. The output of this function, $Y \in \mathbb{R}^n$, is the predicted response. In our case, the response is quantitative, as we aim to predict the chemical concentration of the interactions between proteins.

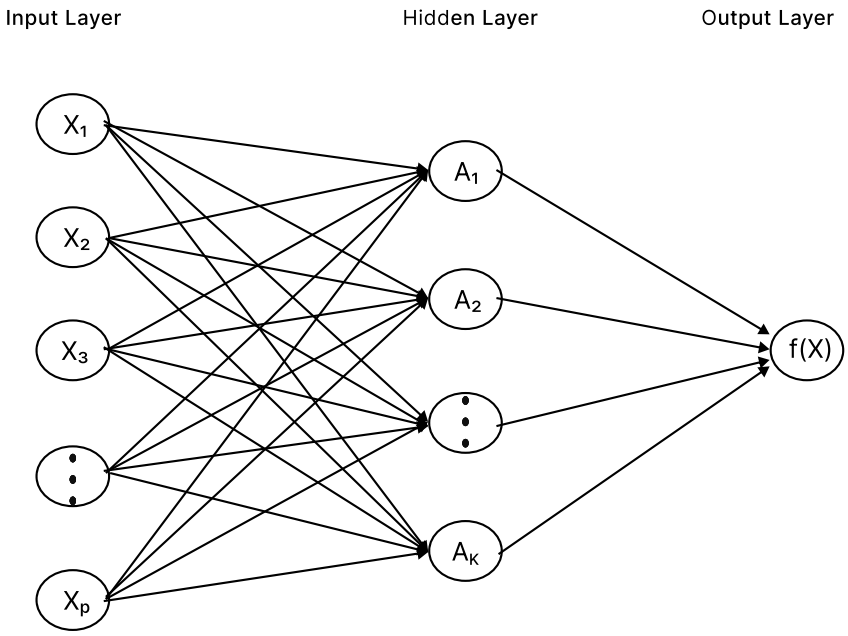


Figure 2.1: Neural network architecture with a single Layer. The neural network architecture consists of three layers: The input layer has p neurons, each corresponding to one input data value X_1, X_2 ..., X_p, where $X_i \in \mathbb{R}$. The hidden layer has K neurons, where each neuron applies a nonlinear transformation to the inputs, producing activation values A_1, A_2, ..., A_3. The output layer has one neuron, which produces an output value by applying a linear function to the activation values from the hidden layer.

Neural networks consist of an input layer, hidden layers (which can be more than one), and an output layer. Each layer is composed of neurons. The input layer neurons hold the values of the input vector, while the hidden layer neurons contain values obtained from an activation function, g. The input to this activation function is the weighted sum of the previous neurons' values plus the bias. Figure 2.1 illustrates the architecture of a neural network with a single hidden layer. The arrows in that figure indicate that each

neuron feeds into the next neuron. The neural network illustrated in Figure 2.1 can be mathematically described as follows:

$$f(X) = \beta_0 + \sum_{k=1}^{K} \beta_k g(w_{k0} + \sum_{j=1}^{p} w_{kj} X_k),$$

where g is a nonlinear function called the activation function; the parameters β_k and w_{kj} represent the weights of the output and hidden layer, respectively; and the parameters β_0 and w_{k0} represent the bias. All these parameters need to be estimated from data. The vectorized form of this equations is:

$$f(X) = \beta_0 + \beta^T g(WX + w_0),$$

where X is the input vector, W is the weight matrix for the hidden layer, w_0 is the bias vector for the hidden layer, β is the weight vector for the output layer, and g is the activation function applied element-wise. The most common activation functions are sigmoid and ReLU. The sigmoid function is defined as:

$$g(z) = \frac{e^z}{1 + e^z} = \frac{1}{1 + e^{-z}}.$$

The output of the sigmoid function ranges from 0 to 1. On the other hand, the ReLU function is defined as:

$$g(z) = \max(0, z).$$

In our work, we will use the leaky ReLU, a variation of ReLU, defined as:

$$g(z) = \begin{cases} z & \text{if } z > 0 \\ \alpha z & \text{if } z \leq 0. \end{cases}$$

The slope coefficient $\alpha \in \mathbb{R}$ is defined before training. It is important for the activation function to be nonlinear because if the activation function is linear, the neural network function f would collapse into a linear regression model.

Modern deep learning techniques require neural network to have multiple layers. For a neural network with multiple hidden layers, the mathematical representation can be extended to include all layers. Suppose we have L hidden layers. The output of each layer is computed in a vectorized form as follows:

For the first hidden layer:

$$H^{[1]} = g^{[1]}\left(W^{[1]}X + b^{[1]}\right).$$

For the second hidden layer:

$$H^{[2]} = g^{[2]}\left(W^{[2]}H^{[1]} + b^{[2]}\right).$$

Continuing in this fashion, for the ℓ-th hidden layer:

$$H^{[\ell]} = g^{[\ell]}\left(W^{[\ell]}H^{[\ell-1]} + b^{[\ell]}\right).$$

Finally, the output layer produces the final prediction Y:

$$Y = f(X) = g^{[L+1]}\left(W^{[L+1]}H^{[L]} + b^{[L+1]}\right).$$

Here, the output Y is computed using the activation function $g^{[L+1]}$ applied to the weighted sum of the final hidden layer's output $H^{[L]}$ and the biases $b^{[L+1]}$.

Neural networks are trained using a process called backpropagation, combined with an optimization algorithm like gradient descent. During training, the network adjusts its weights and biases to minimize the error between the predicted output and the actual output. The error is calculated using a loss function, such as mean squared error for regression tasks. The training process involves iteratively updating these weights $W^{[\ell]}$ and biases $b^{[\ell]}$ using the gradients of the loss function with respect to these parameters. This is done using backpropagation, which calculates the gradients, and gradient descent, which updates the parameters to minimize the loss.

2.5 Universal differential equations

A pivotal advancement in machine learning has been achieved by incorporating more structural knowledge into models [27]. Embedding additional structure into a model enhances its ability to fit data more accurately and efficiently, even with limited datasets. However, the application of deep learning methods is constrained in many scientific domains due to their limited accuracy when trained on small datasets. In these fields, mechanistic models continue to dominate due to their reliability and foundational basis in scientific principles [27].

Rackauckas and colleagues introduced the concept of universal differential equations (UDEs), a sophisticated mathematical framework that synergizes the strengths of both mechanistic and data-driven models while addressing their individual shortcomings [27]. The mechanistic component of a UDE encapsulates the structured scientific knowledge, including all recognized scientific laws and numerical methods. In contrast, the neural network component is tasked with "filling in the gaps", thereby covering aspects of the system not explicitly defined by the mechanistic model.

A UDE is characterized by a differential equation that is defined, wholly or in part, by a universal approximator [27]. A universal approximator is a parametric system capable of approximating any function $f : \mathbb{R}^n \to \mathbb{R}^m$ to any desired degree of accuracy, assuming sufficient resources are available [27].

According to the universal approximation theorem, neural networks stand out as a prime example of universal approximators. It is crucial to recognize that, although a universal approximator has the theoretical capability to approximate any function, its practical ability to do so from a specific dataset is not guaranteed. The success of learning an accurate approximation is influenced by various factors, including the design of the network architecture, the choice of optimization algorithm for training, the quality

and volume of the training data, and the implementation of regularization strategies to mitigate overfitting.

Let C be a cost function defined as:

$$C(\boldsymbol{\omega}) = \sum_{i=1}^{n} \|x(t_i, \boldsymbol{\omega}) - d_i\|, \tag{2.13}$$

where $x(t, \boldsymbol{\omega})$ represents the solution to the universal differential equation with respect to the parameters $\boldsymbol{\omega}$, and d_i is the true solution at time t_i. Here n denotes the number of data points, and $\|\cdot\|$ represents the Euclidean norm. Training a UDE involves minimizing that cost function. By applying the chain rule, the gradient of $C(\boldsymbol{\omega})$ is given by:

$$\nabla C(\boldsymbol{\omega}) = 2 \sum_{i=1}^{n} (x(t_i, \boldsymbol{\omega}) - d_i) \nabla x(t_i, \boldsymbol{\omega}). \tag{2.14}$$

Efficiently training a UDE, therefore, depends on the accurate computation of the gradient of the differential equation solution with respect to the parameter vector $\boldsymbol{\omega}$. This approach not only integrates the predictive power of neural networks with the deterministic nature of mechanistic models but also opens a new era in modeling complex systems.

Chapter 3

Literature Review

This chapter presents a literature review that systematically explores the domain of hybrid modeling, particularly within the context of chemical kinetics and cell signaling pathways. Our discussion is structured into three sections, each addressing distinct aspects of the current research landscape and emerging methodologies. In Section 3.1, we delve into the evolution and application of hybrid models in chemical kinetics, elucidating how these models synergize the predictive power of data-driven approaches with the foundational principles of traditional kinetic modeling. This section lays the groundwork for understanding the significance of hybrid modeling in advancing our comprehension of complex chemical processes. Section 3.2 focuses on the recent advancements in UDE-based hybrid models. Universal Differential Equations (UDEs) represent a cutting-edge approach that marries the flexibility of neural networks with the structured framework of differential equations, offering new possibilities for modeling dynamic systems with enhanced accuracy and interpretability. In Section 3.3, we identify and discuss the prevailing research gaps within the field. Despite significant progress, certain challenges remain unaddressed, particularly regarding the generalization of models across varying initial conditions and the simultaneous inference of first-principle parameters and neural network weights. Highlighting these gaps not only provides direction for future research but also underscores the need for innovative solutions to enhance the utility and applicability of hybrid models.

Our final literature search spanned from February 8, 2024, to February 16, 2024, leveraging the Scopus database to compile relevant studies. We employed a series of queries focusing on "hybrid modeling" and "chemical reaction" or "cell signaling pathway", "universal differential equations", and "parameter estimation" in chemical reactions or cell signaling pathways, specifically targeting publications from 2022 to 2024. Despite experimenting with various queries and exploring other databases such as PubMed and IEEE, our most pertinent findings were derived from Scopus, owing to its comprehensive coverage and relevance to our research focus. The selection process and the criteria for inclusion followed the PRISMA guidelines [26], and are visually summarized in a flow diagram (Figure 3.1), which elucidates the systematic approach adopted in identifying and selecting the papers for this review. In a total, we selected 16 papers to carefully assess and discuss in this literature review. Some main properties of those papers are summarized in Tables 3.1 and 3.2. In the following, we will critically discuss those papers.

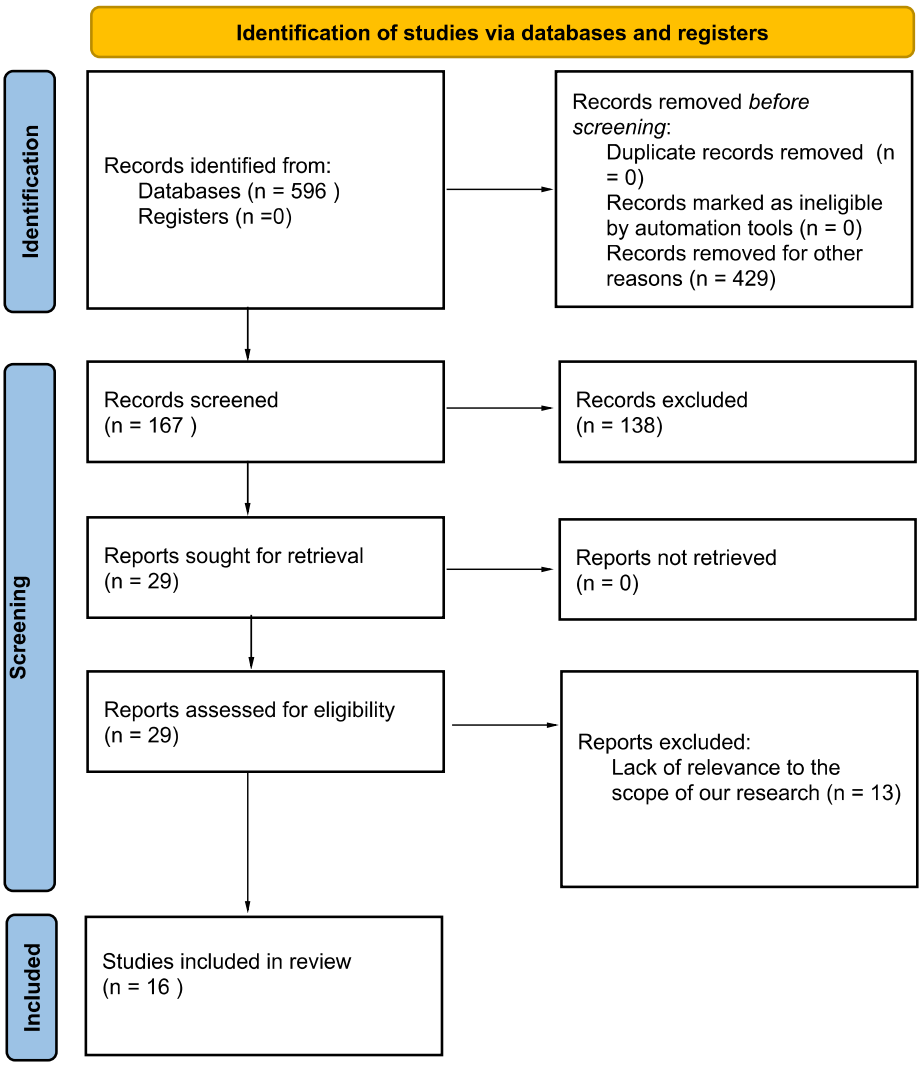


Figure 3.1: Flowchart illustrating our systematic review process in accordance with PRISMA guidelines.

Paper	Dataset	Data-Driven part	Case studies
Zander et al. [39]	Real	Neural Network (NN)	2
Wouver et al. [34]	Real	RBF Network	1
Narayanan et al. [25]	Simulation	Functional Transformations	6
Kaili et al. [16]	Real	LSTM	1
Dong et al. [4]	N/A	LSTM	1
Bang et al. [1]	Real	UDE with NN	1
Liam et al. [17]	Real and Simulation	UDE with NN	1
Santana et al. [31]	Simulation	UDE with NN	1
Koch et al. [12]	Simulation	UDE with NN	1
Kuwahara et al. [13]	Real	UDE with NN	1
Heimerson et al. [8]	Real	UDE with NN	1
Vortmeyer-Kley [36]	Simulation	UDE with NN	4

Table 3.1: Literature review of hybrid modeling, including UDE-bases ones. For each paper, we present the type of dataset used in the paper, the data-driven model couple in the first-principle model, and the number of evaluated case studies.

3.1 Hybrid modeling in chemical reactions

The advent of hybrid modeling in chemical kinetics marks a significant milestone in the evolution of computational modeling, blending traditional kinetic modeling with the computational power of neural networks. This interdisciplinary approach, initiated by the pioneering work of Zander and colleagues in 1999, has sought to leverage the strengths of both physical-based and data-driven models to enhance the prediction and generalization capabilities in dynamic modeling of chemical reaction systems [39]. The exploration of Zander and colleagues into hybrid models for ethane pyrolysis and dodecanol ethoxylation underscored the potential for such models to reduce dependency on voluminous experimental data while improving model generalization [39]. Despite these advantages, the work also highlighted inherent challenges, notably the intricate balance required between incorporating empirical insights and maintaining the integrity of physical modeling principles. A critical aspect of their methodology was the evaluation of extrapolation capabilities, pointing to the significant role of physical modeling components in extending predictions beyond the confines of training datasets.

Expanding upon these foundational insights, Wouwer and colleagues in 2004 applied hybrid modeling to bio-processes, specifically targeting batch CHO animal cell cultures [34]. By integrating radial basis function (RBF) networks with physical models, this study aimed to address non-linearities and prediction discrepancies inherent in biological reaction systems [34]. This approach detailed a comprehensive weights estimation process, blending unsupervised and supervised learning to enhance neural network performance. This approach, however, was not without its challenges. The method proposed by Wouwer and colleagues introduced an intricate parameter estimation process that, while innovative, raised concerns regarding its complexity and the potential for increased computational demand.

A significant leap forward was made by Narayanan and colleagues, who introduced a functional-hybrid modeling framework that combines domain-specific functional transformations with symbolic regression [25]. This innovative approach, emphasizing model interpretability, has demonstrated commendable performance across various biochemical processes. Notably, it offers advantages in data-scarce environments, showcasing enhanced extrapolation capabilities compared to traditional hybrid artificial neural network models. However, while advancing model interpretability, symbolic regression introduces new challenges, including the potential for increased computational demand and the complexity of deriving suitable symbolic expressions that accurately reflect system dynamics.

As the field evolves, hybrid models have increasingly incorporated advanced techniques such as deep learning to enhance their predictive capabilities. A notable contribution by Kaili and colleagues represents a significant advancement in environmental modeling, particularly in the prediction of nitrous oxide (N2O) emissions from wastewater treatment plants (WWTPs) [16]. Their approach integrates a first-principle model based on the activated sludge model (ASM), modified to encompass major nitrogen species transformations, with a deep learning model employing long short-term memory (LSTM) architecture. This blend of mechanistic insight and data-driven pattern recognition yields predictions of N2O emissions with higher accuracy than either standalone mechanistic or deep learning models. Validated against data from a full-scale WWTP, this hybrid model underscores the potential of combining first-principle models with LSTM to achieve superior performance across diverse operational conditions and configurations. Despite its promising results, the effectiveness of the model hinges on data quality, highlighting the need for broader validation across various WWTPs.

In a similar vein, Dong and colleagues developed a hybrid model that merges first-principle mechanistic models with LSTM neural networks to predict the gold leaching rate accurately, crucial for environmental and economic optimization in hydrometallurgy [4]. This model synergizes the conservation principles and chemical dynamics of gold cyanidation leaching with an LSTM network, addressing unmodeled dynamics and prediction errors inherent in first-principle models. Although this hybrid model shows superior predictive performance for the leaching rate compared to individual first-principle or LSTM approaches, it faces limitations related to data quality and computational complexity, particularly in parameter optimization. The model's applicability to other leaching processes and operational conditions also warrants further investigation, illustrating the ongoing challenges and opportunities in hybrid modeling for chemical reactions.

3.2 Advancements in UDE-based hybrid modeling

The development of universal differential equations (UDEs) by Rackauckas and colleagues represents a paradigm shift in hybrid modeling, facilitating the integration of symbolic expressions with neural networks [27]. UDEs have emerged as a versatile tool, enabling the creation of models that are both interpretable and adaptable to limited data scenarios. This advancement heralds a new era in hybrid modeling, combining the precision of symbolic regression with the flexibility of neural networks to forge models of unprece-

dented efficacy and efficiency. Recent years have witnessed significant advancements in the application of UDEs for hybrid modeling, marking a pivotal shift towards integrating data-driven approaches with traditional differential equation frameworks. This section presents recent usage of UDE-based hybrid modeling within various scientific domains.

In 2022, Bangi and Kwon employed a UDE-based hybrid model to elucidate the batch kinetics of aerobic carotenoid production in *Saccharomyces cerevisiae*. By embedding a deep neural network within the structure of ODEs, this model adeptly captures the complex dynamics of the bio-fermentation process. The hybrid model demonstrated superior predictive accuracy for key biochemical components, including biomass, glucose, ethanol, acetic acid, and β-carotene, showcasing a marked improvement over traditional first-principles models [1]. However, a critical limitation emerged from its dependence on training data, suggesting potential challenges in generalizing predictions to scenarios not encompassed by the training dataset. This dependence underscores the need for enhanced methodologies or diversified datasets to bolster the extrapolation capabilities of the model.

Building on this momentum, Lima and colleagues, in 2023, introduced a novel UDE-based hybrid model tailored for the crystallization process of potassium sulfate [17]. This model combines population balance models (PBM) with neural networks to predict nucleation, growth, and dissolution rates, trained on experimental data and validated against PBM simulations [17]. The UDE approach not only matched the performance of conventional PBMs in terms of error metrics but also achieved this with a significantly smaller dataset, spotlighting its efficiency and reduced data requirement for model development.

Furthermore, in the same year, Santana and colleagues ventured into the realm of sorption processes, employing a UDE-based hybrid model to investigate sorption uptake dynamics within non-linear advection-diffusion-sorption systems [31]. Utilizing sparse and symbolic regression to integrate missing functions into an artificial neural network, this model excelled in predicting sorption breakthrough curves from noisy in-silico data. Despite its success in modeling and discovering sorption kinetic laws, the approach raised concerns about the computational demands of symbolic regression and the interpretability and simplicity of the resultant models. The challenge lies in balancing the desire for model simplicity with the need for accuracy and interpretability, particularly when extrapolating beyond the confines of available data.

Integrating the concept of UDEs into broader scientific inquiries, Koch and colleagues lead efforts in data-driven modeling for networked dynamical systems, employing UDEs to decipher the complex interrelations within such systems based on nodal observations [12]. Their work innovatively combined first-principles modeling and neural networks within the UDE framework to approximate unknown dynamics, notably enhancing the understanding of individual unit physics, graphical structures, and coupling physics in networked systems. This methodology demonstrated its prowess in modeling non-linear coupled oscillators, offering insights into future state predictions and system behaviors across varied network topologies.

Extending UDE applications, Kuwahara and Bauch harnessed this framework for pandemic forecasting, specifically COVID-19, blending compartmental models with real-world data [13]. Despite its novel integration, the dependency of the model on timely

and accurate data collection highlighted challenges in applying UDEs across fluctuating data landscapes, emphasizing the necessity for model refinement to aid in public health decision-making.

Further exploration by Heimerson and Ruuskanen into microservices modeling showcased the capability of the UDEs to capture dynamics unaccounted for by simplified assumptions, through a hybrid model integrating neural networks with fluid models [8]. This approach, while enhancing model adaptability, underscored the importance of an accurate base fluid model for effective extension through neural networks.

Moreover, the development of a trajectory-based LDA loss function by Vortmeyer-Kley and colleagues for systems experiencing bifurcations exemplified the integration of physics-informed differential equations with machine learning within the UDE framework [36]. This innovation proved effective in capturing dynamical behaviors around bifurcation points, despite potential limitations in data quality and the complexity of hyperparameter tuning.

3.2.1 Estimation of First-Principle Parameters

The ongoing evolution of methodologies for first-principle parameter estimation in chemical reaction networks is a pivotal aspect of the interplay between theoretical modeling and experimental validation in the field of systems biology. As researchers seek to unravel the complexities of biochemical processes, the accuracy and reliability of these methods are paramount. In Table 3.2, we present the reviewed papers on first-principle parameter estimation.

Paper	Dataset	Technique	Case studies
Gasparyan et al. [6]	Simulation	Bézier curve	2
Tang et al. [33]	Simulation	Moment estimation	1
Barrows et al. [2]	Simulation	Particle filtering	3
Gasparyan et al. [6]	Simulation	Kron reduction	2

Table 3.2: Literature review of first-principle parameter estimation. For each paper, we present the type of dataset used in the paper, the used estimation technique, and the number of evaluated case studies.

In 2021, Gasparyan and colleagues introduced an innovative approach for parameter estimation in enzymatic chemical reaction networks based on time-series experimental data of reaction rates [21]. The core of their method is the utilization of parametric Bézier curves to transform available reaction rate data into species concentration data. This key transformation enables the application of the least squares method for estimating the parameters of a mathematical model describing the chemical reaction networks. This technique effectively bridges the gap between experimental observations and theoretical modeling, providing a more accurate and comprehensive understanding of enzymatic behaviors. The accuracy of the method of Gasparyan in retrieving known parameter values from synthetic experimental data highlighted its potential for practical applications in systems biology. The reliance of the method on Bézier curves introduces a novel and

flexible approach for dealing with the non-linear dynamics typical of enzymatic chemical reaction networks. The automated nature of the procedure simplifies the parameter estimation process, making sophisticated modeling accessible to researchers without extensive expertise in computational modeling. However, the efficacy of this approach is contingent upon the quality and resolution of the experimental data, with sparse or noisy data potentially undermining the reliability of the estimated parameters. The issue of model identifiability also presents a challenge, emphasizing the need for comprehensive data to ensure unique and accurate parameter estimation.

Transitioning to 2022, Han Tang [33] introduced a pioneering approach that incorporates time delays into the modeling of chemical reactions using uncertain differential equations. This method extends the traditional chemical kinetics models by providing a more nuanced understanding of reactions, acknowledging that products may emerge after certain time delays. The work of Tang is anchored in the established theory of uncertain differential equations, applying it to chemical kinetics to offer innovative solutions for parameter estimation and hypothesis testing. This methodology allows for the estimation of unknown parameters and the validation of chemical reaction models against observed data, addressing some of the limitations found in instantaneous reaction assumptions. Despite its innovative contributions, the methodology of Tang shares similar limitations with previous models, particularly regarding the dependency on high-quality and granular observed data. The generalizability of this model to reactions without significant delays or where deterministic models suffice also poses questions, indicating a need for further exploration into the applicability of uncertain differential equations in diverse chemical reaction contexts.

In 2023, Barrows and colleagues proposed a methodological advancement by leveraging iterated particle filtering to estimate parameters within the context of the reaction-diffusion master equation [2]. This approach, tested across various systems, demonstrates an impressive ability to recover known rate parameters accurately. It marks a significant step forward in modeling stochastic biochemical systems, especially in the realm of systems biology where spatial heterogeneity and low species copy numbers are common challenges. The methodology introduced by Barrows and colleagues offers a sophisticated alternative to traditional parameter estimation methods, though with increased computational complexity and a requirement for specialized knowledge in particle filtering techniques. The sensitivity of this approach to initial conditions and data quality underscores the continuing challenge of balancing model accuracy with the practicalities of experimental data availability.

Further advancements in 2023 by Gasparyan and colleagues address the intricacies of estimating parameters from partial experimental data, employing the Kron reduction technique combined with least squares optimization [6]. This method facilitates parameter estimation even when not all species concentrations are measurable, offering a solution to a common issue in biochemical experiments. While this approach mitigates some of the challenges associated with high-dimensional parameter spaces, it also introduces new considerations regarding the accuracy and identifiability of parameter estimates, highlighting the ongoing need for comprehensive and high-quality experimental data in the field of systems biology.

3.3 Identification of research gaps

Despite the remarkable advancements in hybrid modeling, particularly those leveraging universal differential equations (UDEs), our literature review has unveiled crucial gaps in the current state of research. These gaps underscore limitations in the generalization capabilities of existing models and highlight a reliance on predefined first-principle knowledge that may restrict their broader applicability and efficacy.

3.3.1 Lack of generalization across multiple initial conditions

A significant research gap identified is the absence of methodologies tailored to train hybrid models across a diverse array of initial conditions. Current approaches primarily focus on modeling specific scenarios or datasets, often overlooking the variability inherent in biological and chemical systems. This limitation is particularly pronounced in models that do not account for the wide range of initial states these systems can exhibit, potentially restricting the ability of the model to generalize and accurately predict outcomes under previously unencountered conditions. Addressing this gap by developing methods that incorporate a myriad of initial conditions could significantly enhance the robustness and predictive power of hybrid models, making them more versatile and applicable to a broader spectrum of scientific inquiries.

3.3.2 Joint inference of first-principle parameters and neural network weights

Another critical gap in the existing body of research is the lack of approaches that simultaneously infer both the first-principle parameters and neural network weights within hybrid models. The prevailing methodologies often assume that first-principle parameters are known *a priori*, which may not always be the case. This assumption places a considerable constraint on the flexibility of the model, rendering it dependent on the availability and accuracy of first-principle knowledge. A reliance of the model on such predefined parameters can limit its application, especially in systems where first-principle knowledge is incomplete or evolving. Developing a framework that allows for the joint inference of both sets of parameters could significantly diminish this dependency, thereby enhancing the adaptability of the model and reducing the need for extensive prior knowledge.

The identified research gaps highlight an opportunity for significant contributions to the field of hybrid modeling. By addressing these gaps, future research can pave the way for the development of more generalized and self-sufficient hybrid models. Such models would not only advance our understanding of complex systems but also broaden the scope of their applications, ultimately contributing to the advancement of scientific knowledge and technology.

Chapter 4

Methodology

The methodology of our work, depicted in the diagram presented in Figure 4.1, was designed to address the lack of isolation problem in cell signaling pathways. This methodological framework is in direct response to our research questions. It begins by tackling RQ 1, which concerns the effectiveness of hybrid models in identifying cell signaling pathways using simulated time series data. To this end, we simulate a comprehensive first-principle model across a spectrum of initial conditions, generating an extensive dataset that reflects the multifaceted nature of cell signaling.

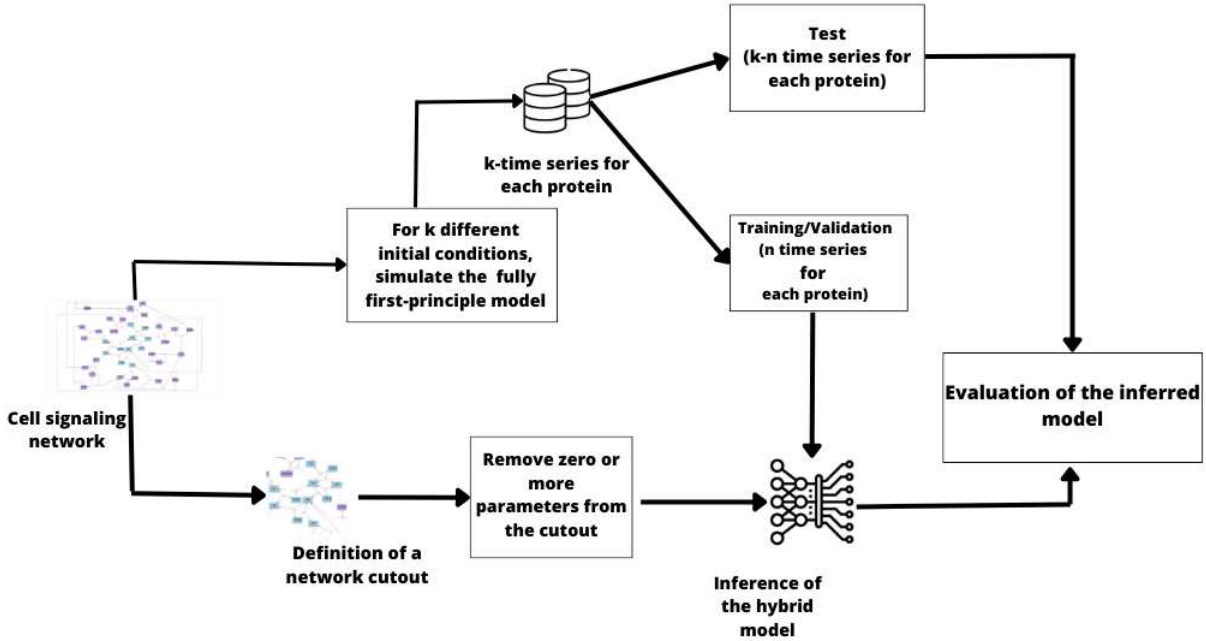


Figure 4.1: Flowchart depicting the step-by-step methodology for addressing the lack of isolation problem in cell signaling pathways. This includes the generation of time series data, definition and refinement of pathway cutouts, inference of the hybrid model, and the subsequent evaluation of model performance.

Subsequent to the simulation phase, we move to a critical step that directly engages with RQ 2: the iterative removal of parameters from the mechanistic part of the model. Our objective here is to test the robustness and predictive accuracy of the hybrid model. Initially, we retain all first-principle parameters, building a pathway cutout that encap-

sulates the complete set ofcel known parameters. Then, in a stepwise manner, we remove parameters—one at a time, then two at a time, and so forth—until we have iteratively tested the performance of the model with each possible subset of parameters. This iterative parameter removal serves a dual purpose. First, it evaluates the capacity of the model to infer the weights of the neural network and the parameters of the first-principle component simultaneously. Second, it assesses the resilience of the model and predictive capabilities as the known first-principle parameters becomes increasingly incomplete. This process ensures that our hybrid model is not only capable of accurately representing the system when all first-principle parameters is present but also retains its predictive power in scenarios where such knowledge is partial or uncertain.

4.1 Proposed method

4.1.1 The inference of the hybrid model

Our proposed methodology aims to use the predictive power of first-principle models and the adaptability of neural networks using the universal differential equation (UDE) formalism. This approach allows us to create a hybrid model, integrating the structured, mechanistic understanding with the flexibility of machine learning. The hybrid model is mathematically expressed as:

$$\dot{\mathbf{x}} = f(\mathbf{x}(t), \mathbf{u}(t), U(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\omega}); \boldsymbol{\theta}), \tag{4.1}$$

where U represents the neural network, $\boldsymbol{\omega}$ denotes the weights of the neural network, and $\boldsymbol{\theta}$ symbolizes the parameter vector of the first-principle model. As discussed in Chapter 2, training the hybrid model involves efficiently computing the gradients of the solution to Equation 2.5 with respect to ω. To this end, we will initially employ the ADAM optimization algorithm, followed by the BFGS algorithm. The rationale behind this sequence is that while ADAM excels at rapidly converging to local minima, BFGS is more effective in fine-tuning the solution within the local minima, providing a more accurate final parameter set [27]. In our model, we incorporate a diverse array of time series data generated from different initial conditions. To manage this, we have designed a loss function tailored to handle multiple time series within the training process. Algorithm 1 illustrates this loss function.

To facilitate the simultaneous inference of first-principle model parameters and neural network weights, we leverage the gradient of the loss function, which involves calculating the gradient of the hybrid model solution with respect to the neural network weights. We approach this by treating the first-principle parameters as additional parameters to be optimized in conjunction with the neural network during the training phase. This is achieved by concatenating the first-principle parameters and neural network weights into a single vector, allowing the optimizer to update all parameters simultaneously. However, this method may encounter two potential issues: firstly, the neural network could overpower the hybrid model, leading the optimizer to ascribe zero values to the first-principle parameters; secondly, the optimizer might return negative values for the first-principle

Algorithm 1 UDE-based model approach

Input: n distinct time series, each describing the dynamics of a cell signaling pathway
Output: Optimized weights for a UDE-based hybrid model and the corresponding loss value

function LOSS(Θ)
    $simulations \leftarrow$ SOLVEENSEMBLE($ensembleUDEModels, \Theta, trajectories$)
    $error \leftarrow$ CALCULATEREGULARIZATION(Θ)
    $instabilityPenalty \leftarrow 10^6$
    for $i \leftarrow 1$ **to** N **do**
        $currentCondition \leftarrow$ GETCONDITION($step + i$)
        $solution \leftarrow simulations[i]$
        if ISSUCCESSFUL($solution$) **then**
            $predictedData \leftarrow$ EXTRACTSOLUTIONDATA($solution$)
            $errorIncrement \leftarrow$ CALCULATEMAE($predictedData, currentCondition$)
            $error \leftarrow error + errorIncrement$
        else
            $error \leftarrow error + instabilityPenalty$
    return $error$

function OPTIMIZER(f, Θ_0, $maxIter$, ε_Θ, ε_F)
    $k \leftarrow 0$
    $convergence \leftarrow$ False
    while $k < maxIter$ **and not** $convergence$ **do**
        $k \leftarrow k + 1$
        $\Delta\Theta_k \leftarrow$ COMPUTEDELTA(f, Θ_{k-1})
        $\Theta_k \leftarrow \Theta_{k-1} + \Delta\Theta_k$
        $F_k \leftarrow f(\Theta_k)$
        $\Delta F_k \leftarrow F_{k-1} - F_k$
        $convergence \leftarrow$ CHECKCONVERGENCE($\Delta\Theta_k, \varepsilon_\Theta, \Delta F_k, \varepsilon_F$)
    $\Theta^* \leftarrow \Theta_k$
    $F^* \leftarrow f(\Theta^*)$
    return Θ^*, F^*

$\Theta^{est}, F^{est} \leftarrow$ OPTIMIZER(Loss, Θ_0, maxIter, ε_Θ, ε_F)

parameters, which is not feasible in many biological contexts. To mitigate these concerns, we apply regularization exclusively to the neural network weights and impose penalties on negative values of first-principle parameters by employing a rectified linear unit (ReLU) function. The Algorithm 2 reflects the adjustments made to enable the joint inference of neural network weights and first-principle model parameters.

Algorithm 2 Modification in Loss Function

function LossFunction(Θ)
  $simulations \leftarrow$ SolveEnsemble($ensembleODEModels, \Theta, trajectories$)
  $regularizationTerm \leftarrow$ CalculateRegularization($\Theta[neuralNetwork]$)
  $penaltyForNegatives \leftarrow$ PenalizeNegativeParameters($\Theta[firstPrinciple]$)
  $error \leftarrow regularizationTerm + penaltyForNegatives$
  $instabilityPenalty \leftarrow 10^6$
  for $i \leftarrow 1$ **to** $totalTrajectories$ **do**
    $currentCondition \leftarrow$ GetCondition($step + i$)
    $solution \leftarrow simulations[i]$
    if IsSuccessful($solution$) **then**
      $predictedData \leftarrow$ ExtractSolutionData($solution$)
      $errorIncrement \leftarrow$ CalculateMAE($predictedData, currentCondition$)
      $error \leftarrow error + errorIncrement$
    else
      $error \leftarrow error + instabilityPenalty$
  return $error$

By integrating these techniques, we aim to construct a model that is both robust in the face of diverse initial conditions and capable of yielding biologically plausible parameter estimations, thereby overcoming the critical challenges of the lack of isolation problem.

4.2 Experimental details

4.2.1 Data acquisition for modeling

The data supporting our study was generated through *in silico* simulations, designed to reflect the dynamic interactions of a complete cell signaling pathway, designated as $\mathcal{E}$. This pathway is initially assumed to be free from the lack of isolation problem, providing a fully connected and idealized network for our analysis. For the purposes of our research, cell signaling pathways were predominantly retrieved from Odebase, a repository of ODE systems associated with biological systems [20]. Alongside these, a toy model was crafted by our research group to serve as a simplified yet informative complement to the existing datasets from Odebase.

To initiate the simulations, we generated a range of random initial values that conform to the fundamental principles of chemical reactions. Using a numerical solver, we then resolved the resulting initial value problems (IVPs) across a specified time interval. This process yielded the state vector $\mathbf{x}(t)$, capturing the concentration changes of each species over time.

The temporal resolution of our model—how frequently we sample the state vector $\mathbf{x}(t)$—is a critical factor in our simulation. In our case, the time interval $[t_0, t_{end}]$ was partitioned into n equal segments, resulting in a time vector $t = [t_0, t_1, \ldots, t_n]$. For all our studies, we standardized the simulation interval to $[0, 100]$ and chose to sample at 101 equally spaced time points, effectively setting our timestep at 1 unit (i.e., $t_0 = 0, t_1 = 1, \ldots, t_{100} = 100$).

The final step in our data acquisition was to define the pathway cutout $\mathcal{F}$. This involved selecting a subset of $\mathcal{E}$, focusing on the time series of the chemical species common to both $\mathcal{E}$ and $\mathcal{F}$. This selective process was crucial for establishing the dataset from which we would infer the missing parameter of our hybrid model.

Recognizing that biological data typically exhibits inherent noise, we also incorporated noisy data into our experiments to simulate more realistic conditions. To introduce noise into our synthetic data, each sample was perturbed with a Gaussian distribution having a mean (μ) of zero, and a standard deviation calculated as follows:

$$\sigma = 0.05\overline{x_i}, \tag{4.2}$$

where $\overline{x_i}$ represents the mean value of the state variable x_i over the observed time period. This adjustment ensured that our model training, validation, and testing phases considered the typical variability found in experimental biological data.

This structured approach to data generation provided us with a foundation for the subsequent stages of our methodology, ensuring that the synthetic data used for model training, validation, and testing were relevant to the dynamics of cell signaling pathways we aimed to capture.

4.2.2 Training process

The training process is a pivotal stage in the development of our hybrid model, where we aim to achieve a balance between fitting the data accurately and avoiding overfitting. Our approach is structured into two training phases: initially, we focus on training the neural network weights assuming known first-principle parameters; subsequently, we engage in joint training to simultaneously infer both the first-principle parameters and neural network weights.

During the training phase, the dataset is divided into three parts: training, validation, and testing, with each set having an equal share of the data. The training set is used to adjust the model parameters, while the validation set is employed to monitor for overfitting and guide the early stopping mechanism. Early stopping is a crucial technique to prevent the model from learning noise in the training data. Our early stopping criterion is specifically tied to the mean absolute error (MAE) of the validation set. If the MAE does not improve over a span of 100 iterations, the training is halted. This approach ensures that the model generalizes well to new data and does not merely memorize the training set.

To mitigate the risk of overfitting, we have incorporated L2 regularization into our loss function. For a set of initial conditions $\{1, \ldots, N\}$, let $\boldsymbol{X}^{(n)}$ be the $I \times J$ matrix

for the nth initial condition, where a matrix entry $x_{ij}^{(n)}$ is the concentration of chemical species i at time point j. From now on, we will call such matrix as an *observation*. The loss function is then defined as follows:

$$C(\mathbf{w}) = \frac{1}{NIJ} \sum_{n=1}^{N} \sum_{i=1}^{I} \sum_{j=1}^{J} |\hat{x}_{ij}^{(n)} - x_{ij}^{(n)}| + \lambda \sum_{k=0}^{W} w_k^2. \tag{4.3}$$

Here, $\hat{x}_{ij}^{(n)}$ and $x_{ij}^{(n)}$ are, respectively, the predicted and the actual matrix entry for the nth initial condition, the ith species and jth time point, and $\mathbf{w}$ is the parameter vector of the neural network, with weights $w_0, \ldots, w_W$. λ is a regularization constant, which in our experiments is set to 10^{-3}.

The optimization process begins with the ADAM algorithm, which is selected for its efficiency in converging to local minima. We use a learning rate of 0.1 over 2000 iterations, with each iteration encompassing the entire dataset (equivalent to 2000 epochs). Following the initial optimization with ADAM, we transition to the BFGS algorithm for a further 1000 iterations, employing a backtracking line search to avoid numerical instabilities. To enhance stability from the outset, the neural network weights are initialized with values close to zero.

When performing joint inference, we maintain the same technical details as in the initial training phase. However, we introduce a penalty mechanism to prevent the optimizer from assigning negative values to the first-principle parameters, which are not biologically plausible. The penalty for negative estimations of first-principle parameters is set at 100,000, ensuring their non-negativity.

4.2.3 Model assessment

The evaluation of the performance of our hybrid model is critical for understanding its effectiveness and reliability in capturing the dynamics of cell signaling pathways, especially under conditions of uncertainty such as the lack of isolation problem or unknown first-principle parameters. To assess the robustness of our model, we employ two widely recognized regression metrics: the mean absolute error (MAE) and the symmetric mean absolute percentage error (SMAPE).

MAE offers a straightforward measure of the average magnitude of errors between predicted and actual values, without considering their direction. MAE is defined mathematically as:

$$MAE = \frac{1}{NIJ} \sum_{n=1}^{N} \sum_{i=1}^{I} \sum_{j=1}^{J} |\hat{x}_{ij}^{(n)} - x_{ij}^{(n)}|. \tag{4.4}$$

The MAE was primarily employed to train the UDE-based model rather than to assess the performance of the model.

The SMAPE, on the other hand, provides a normalized measure of prediction accuracy, making it particularly useful for comparing model performance across different scales. It

is expressed as:

$$SMAPE = \frac{100}{NIJ} \sum_{n=1}^{N} \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{|\hat{x}_{ij}^{(n)} - x_{ij}^{(n)}|}{\hat{x}_{ij}^{(n)} + x_{ij}^{(n)}}. \tag{4.5}$$

Both MAE and SMAPE are instrumental in providing insights into the predictive accuracy of the model: while MAE gives an absolute scale of error magnitudes, SMAPE offers a percentage-based comparison, which is invaluable for assessing the performance of the model in a relative sense, especially when dealing with varying magnitudes of the target variables. Utilizing these two metrics simultaneously allows for a comprehensive evaluation of the model, ensuring that we can accurately gauge its efficacy in predicting the dynamics of cell signaling pathways under diverse experimental conditions and constraints.

4.2.4 Description of the experiments

Our experimental design unfolds in two phases, each crafted to test the resilience and accuracy of the hybrid model, particularly when the cell signaling pathway suffers from the lack of isolation problem and/or the first-principle parameters are unknown.

First phase. Here, we focus on the inference of the neural network weights, under the premise that the first-principle parameters are already established. The primary goal at this stage is to evaluate the ability of the hybrid model to compensate for missing system inputs. We begin with a set of 30 time series, divided equally into training, validation, and testing subsets, each containing 10 time series. This set serves as the foundation for our initial model assessments. Subsequently, we expand the dataset to encompass 300 time series to scrutinize the impact of dataset size on the learning capacity of the model. An increase in the volume of time series data provides insight into the scalability of the model and its ability to generalize across a broader range of input variations.

Second phase. This involves the joint inference of the first-principle parameters and neural network weights. Here, we implement an iterative approach where we systematically remove one first-principle parameter at a time, gauging the adaptability of the model in simultaneously inferring the first-principle parameter removed with the neural network weights. This iterative removal continues, advancing to the exclusion of two parameters at a time, and so forth, until all first-principle parameters have been considered. The objective is to discern the degree to which the performance of the hybrid model – as measured by the Symmetric Mean Absolute Percentage Error (SMAPE) – is influenced by the omission of specific first-principle parameters. This process also helps to reveal any potential dependencies that the model may have on particular parameters. Owing to computational constraints, this phase utilizes a smaller dataset of 30 time series, again evenly split across training, validation, and testing sets. This allows us to maintain computational efficiency while still rigorously testing the parameter inference capabilities of the model.

4.2.5 Computational resources

The development and execution of our experiments were carried out using the Julia programming language, renowned for its performance in scientific computing. Specifically, we leveraged the SciML ecosystem, a comprehensive suite of Julia packages designed for scientific machine learning applications. This ecosystem provides a robust framework for implementing UDEs and includes efficient numerical solvers for ODEs, which were instrumental in simulating our first-principle models and generating the requisite datasets.

Our computational experiments were conducted on two main platforms:

- **Recod.AI Cluster**: This platform provided the bulk of computational resources, featuring a processing node equipped with an Intel(R) Xeon(R) CPU E5620 @ 2.40GHz, encompassing 16 cores and 48 GB of RAM. This environment was pivotal for running large-scale simulations and training sessions that required substantial computational power;
- **Legion Notebook**: For tasks necessitating portability and those of a smaller scale, we utilized a notebook with an Intel Core i7-10750H CPU and 16GB of RAM. This setup allowed for flexibility in development and preliminary testing of our algorithms and models.

The combination of these computational resources, coupled with the advanced capabilities of the Julia programming language and the SciML ecosystem, enabled us to efficiently implement our algorithms and conduct the extensive experiments required for this study.

4.3 Case studies

4.3.1 Toy Model

Our research group has developed a toy model (to which we will refer as "the Toy Model"), which consists in a simplified representation designed to elucidate the dynamics within a cascade of enzymatic reactions. Detailed information on this model is available in our conference paper [32]. The Toy Model is a comprehensive system comprising 38 chemical species and 51 chemical reactions, governed by 12 first-principle parameters. A specific pathway cutout, prominently featured in Figure 4.2 (marked in blue), has been carefully selected to demonstrate the challenges associated with the lack of isolation problem. This particular segment of the model includes 12 chemical species and 8 chemical reactions, all regulated by 12 first-principle parameters.

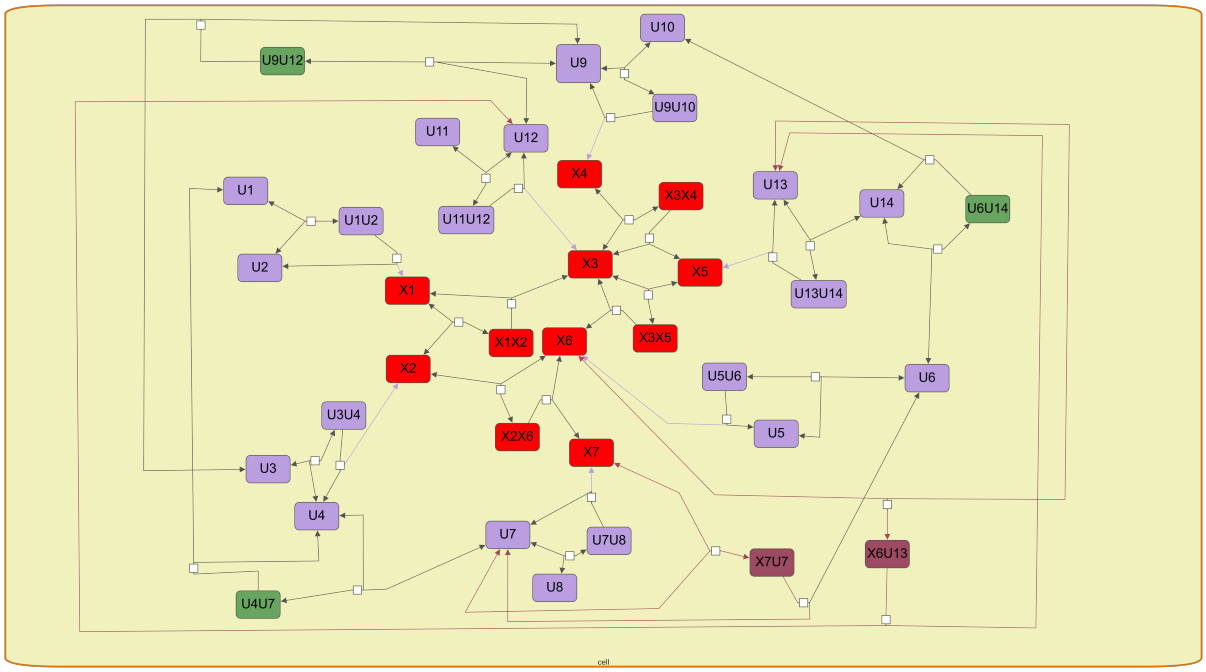


Figure 4.2: Diagram in systems biology graphical notation (SBGN), depicting the Toy Model cell signaling pathway. Nodes symbolize chemical species (such as proteins) and edges denote reactions. The pathway cutout is highlighted by red nodes and all connecting edges, emphasizing the specific segments and interactions under study.

The first-principle model, which describes the dynamics of each chemical species within the pathway cutout, is presented in Equations 4.6a–4.6l. The corresponding first-principle parameters are listed in Table 4.1.

k_{f_1}	k_{r_1}	k_{cat_1}
0.0150	0.1000	0.0030
k_{f_2}	k_{r_2}	k_{cat_2}
0.0990	0.1150	0.0850
k_{f_3}	k_{r_3}	k_{cat_3}
0.0890	0.0500	0.1500
k_{f_4}	k_{r_4}	k_{cat_4}
0.2500	0.4325	0.0150

Table 4.1: First-principle parameters of the Toy Model pathway cutout.

$$\frac{d[X_1]}{dt} = k_{cat_1}[X_1X_2] + k_{r_1}[X_1X_2] - k_{f_1}[X1][X2] \tag{4.6a}$$
$$\frac{d[X_2]}{dt} = k_{r_1}[X_1X_2] - k_{f_1}[X1][X2] - k_{f_4}[X_2][X_6] \tag{4.6b}$$
$$\frac{d[X_1X_2]}{dt} = k_{f_1}[X1][X2] - k_{r_1}[X_1X_2] - k_{cat_1}[X_1X_2] \tag{4.6c}$$
$$\frac{d[X_3]}{dt} = k_{r_2}[X_3X_4] + k_{r_3}[X_3X_5] + k_{cat_2}[X_3X_4] + k_{cat_3}[X_3X_5] \tag{4.6d}$$
$$- k_{f_3}[X_3][X_5] - k_{f_2}[X_3][X_4] \tag{4.6e}$$
$$\frac{d[X_4]}{dt} = k_{r_2}[X_3X4] - k_{f_2}[X_3][X_4] \tag{4.6f}$$
$$\frac{d[X_3X_4]}{dt} = k_{f_2}[X_3][X_4] - k_{r_2}[X_3X4] - k_{cat_2}[X_3X_4] \tag{4.6g}$$
$$\frac{d[X_3X_5]}{dt} = k_{f_3}[X_3][X_5] - k_{r_3}[X_3X5] - k_{cat_3}[X_3X_5] \tag{4.6h}$$
$$\frac{d[X_5]}{dt} = k_{r_3}[X_3X5] - k_{f_3}[X_3][X_5] \tag{4.6i}$$
$$\frac{d[X_6]}{dt} = k_{r_4}[X_2X6] - k_{f_4}[X_2][X_6] + k_{cat_4}[X_2X_6] \tag{4.6j}$$
$$\frac{d[X_5X_6]}{dt} = k_{f_4}[X_2][X_6] - k_{r_4}[X_2X6] - k_{cat_4}[X_2X_6] \tag{4.6k}$$
$$\frac{d[X_7]}{dt} = k_{cat_4}[X_2X_6]. \tag{4.6l}$$

4.3.2 JAK-STAT pathway

The JAK-STAT signaling pathway is a quintessential model that exemplifies the intricate control mechanisms governing cellular processes [38]. This pathway, encompassing 36 chemical species and 46 irreversible reactions, is underpinned by a complex network of 86 first-principle parameters. Such a comprehensive assembly underscores the pivotal role of the pathway in transmitting signals from extracellular cues to the nucleus, influencing gene expression patterns crucial for cell growth, differentiation, and immune response [38]. In Figure 4.3, we present a graphical representation of the JAK-STAT pathway.

A specific subset of the JAK-STAT pathway was selected for our study. This cutout, highlighted in red in Figure 4.3, focuses on a segment of the pathway that particularly exemplifies this issue. The chosen cutout encompasses 5 chemical species involved in 4 irreversible reactions, delineated by 6 first-principle parameters. This reduced complexity allows for a targeted investigation into the dynamics of the pathway, enabling us to assess the efficacy of our hybrid modeling approach under conditions where traditional modeling techniques may falter. The chemical species of the cutout are represented as follows: x_1 for IFNRJ2_star_STAT1c, x_2 for SHP2, x_3 for IFNRJ2_star_SHP2_SOCS1_STAT1c, x_4 for IFNRJ2_star_SOCS1_STAT1c, and x_5 for IFNRJ2_star_SHP2_STAT1c. The first-principle model that describes the dynamics of each species is outlined in Equations 4.7a–4.7e. Corresponding first-principle parameters are detailed in Table 4.2.

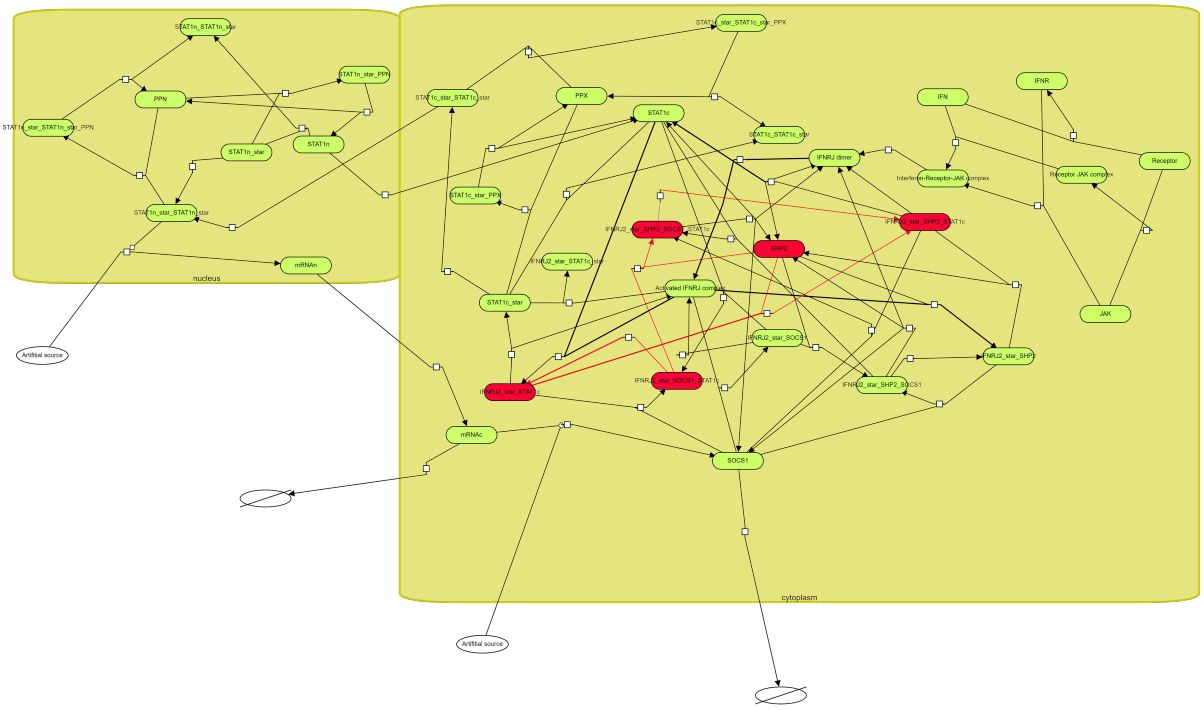


Figure 4.3: SBGN diagram of the JAK-STAT cell signaling pathway. Nodes symbolize chemical species (such as proteins) and edges denote reactions. The pathway cutout is highlighted by red nodes and red edges, emphasizing the specific segments and interactions under study.

$$\frac{d[x_1]}{dt} = -(k_{58}[x_2][x_1] - k_{59}[x_5]) + k_{61}x_4 \tag{4.7a}$$

$$\frac{d[x_2]}{dt} = -(k_{50}[x_2][x_4] - k_{51}[x_3]) - (k_{58}[x_2][x_1] - k_{59}[x_5]) \tag{4.7b}$$

$$\frac{d[x_3]}{dt} = k_{50}[x_2][x_4] - k_{51}[x_3] - k_{53}[x_3] \tag{4.7c}$$

$$\frac{d[x_4]}{dt} = -(k_{50}[x_2][x_4] - k_{51}[x_3]) - k_{61}[x_4] \tag{4.7d}$$

$$\frac{d[x_5]}{dt} = k_{53}[x_3] + (k_{58}[x_2][x_1] - k_{59}[x_5]). \tag{4.7e}$$

k_{50}	k_{51}	k_{53}
0.001	0.2	0.0005
k_{58}	k_{59}	k_{61}
0.001	0.2	0.0005

Table 4.2: First-principle parameters of the JAK-STAT pathway cutout.

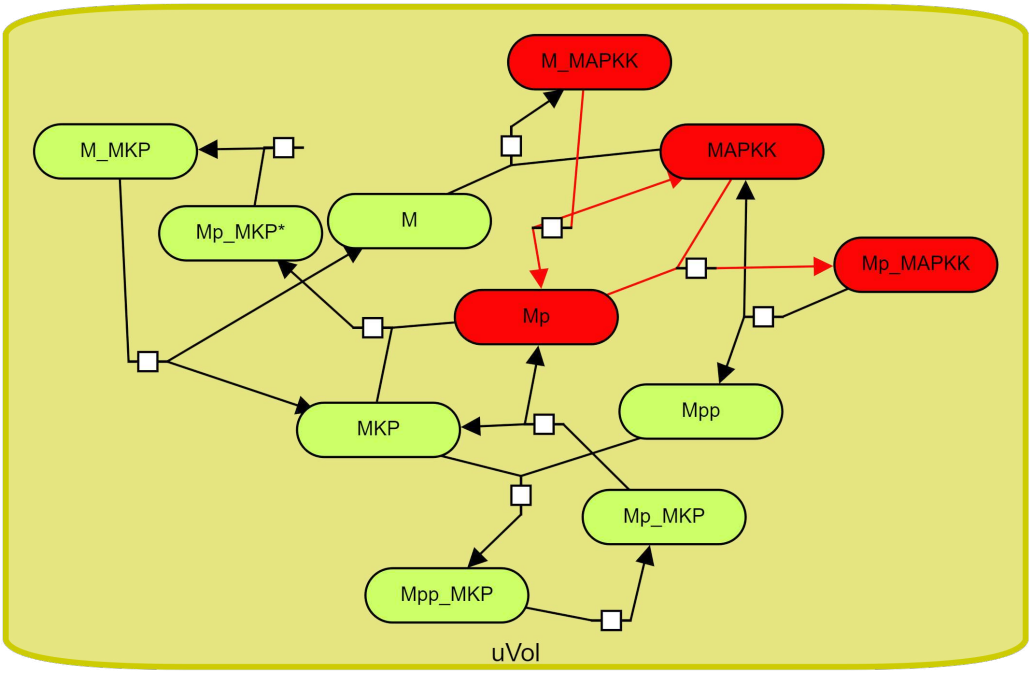


Figure 4.4: SBGN diagram of the MAPK OrderedElementary cell signaling pathway. Nodes symbolize chemical species (such as proteins) and edges denote reactions. The pathway cutout is highlighted by red nodes and red edges, emphasizing the specific segments and interactions under study.

4.3.3 MAPK OrderedElementary

The MAP kinase (MAPK) OrderedElementary model originates from the work by Markevich and colleagues, which delineates the MAPK activation pathway [22]. This model encapsulates the dynamics between MAPKK and MKP, underscoring the dual phosphorylation-dephosphorylation cycle of MAPK [22]. The model encompasses 11 distinct chemical species, 20 parameters grounded in first-principle knowledge, alongside 4 irreversible and 6 reversible reactions. A diagram of that model is depicted in Figure 4.4.

The pathway cutout is highlighted in red in Figure 4.4; this cutout distills the pathway to a core subset of interactions for focused study. This reduced model comprises 4 chemical species and 3 first-principle parameters. The chemical species of the cutout are represented as follows: x_1 for species Mp, x_2 for MAPKK, x_3 for M_MPAPKK, and x_4 for Mp_MAPKK. The first-principle model that describes the dynamics of each chemical species within the pathway cutout is detailed in Equations 4.8a–4.8d. The associated first-principle parameters are provided in Table 4.3.

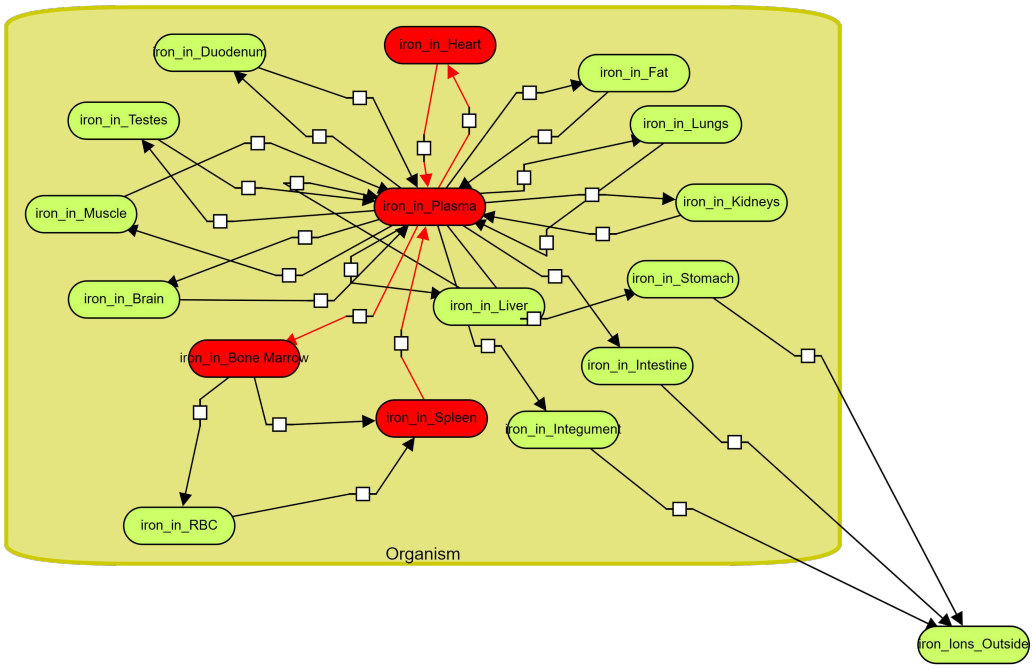


Figure 4.5: SBGN diagram of the FeMetabolism FeDeficient cell signaling pathway. Nodes symbolize chemical species (such as proteins) and edges denote reactions. The pathway cutout is highlighted by red nodes and red edges, emphasizing the specific segments and interactions under study.

$$\frac{d[x_1]}{dt} = k_3[x_3] - (k_4[x_1][x_2] - k_5[x_4]) \tag{4.8a}$$
$$\frac{d[x_2]}{dt} = k_3[x_3] - (k_4[x_1][x_2] - k_5[x_4]) \tag{4.8b}$$
$$\frac{d[x_3]}{dt} = -k_3[x_3] \tag{4.8c}$$
$$\frac{d[x_4]}{dt} = k_4[x_1][x_2] - k_5[x_4]. \tag{4.8d}$$

k_3	k_4	k_5
0.01	0.032	1

Table 4.3: First-principle parameters of the MAPK orderedElementary pathway cutout.

4.3.4 FeMetabolism FeDeficient

The FeMetabolism FeDeficient model is a sophisticated kinematic representation of iron metabolism, detailing the dynamics of iron pools and fluxes based on comprehensive ferrokinetic data and chemical analyses [18]. The model draws from studies on C57BL6 wild-type mice subjected to varying dietary conditions: iron-deficient, iron-adequate, and iron-loaded diets [18]. A diagram of that model is showed in Figure 4.5.

The FeMetabolism FeDeficient model encompasses 17 chemical species and 33 first-principle parameters, highlighting the network of 29 irreversible reactions that govern iron metabolism. The pathway cutout, marked in red in Figure 4.5, includes 4 chemical species and 4 first-principle parameters, underpinning 3 critical reactions. The chemical species of the cutout are represented as follows: x_1 for iron_in_Plasma, x_2 for iron_in_Bone_Marrow, x_3 for iron_in_Spleen and x_4 for iron_in_Heart. The first-principle model that describes the dynamic of each species is given in Equation 4.9a–4.9d. The first-principle parameters are listed in the Table 4.4.

$$\frac{d[x_1]}{dt} = -[x_1]k_3 - [x_1]k_{17} + [x_4]k_{18} \tag{4.9a}$$

$$\frac{[dx_2]}{dt} = [x_1]k_3 \tag{4.9b}$$

$$\frac{d[x_3]}{dt} = -[x_3]k_6 \tag{4.9c}$$

$$\frac{d[x_4]}{dt} = [x_1]k_{17}. \tag{4.9d}$$

k_3	k_6
13.22	14.61
k_{17}	k_{18}
0.11	0.06

Table 4.4: First-principle parameters of the FeMetabolism FeDeficient pathway cutout.

Chapter 5

Results and Discussion

In this chapter, we present the findings from our investigation into four cell signaling pathways using the methodology outlined in Chapter 4. The analysis is structured into four distinct sections, where we initially compared the performance of traditional first-principle-based models against our hybrid modeling approach. This comparison includes four models: a linear regression (baseline), a ODE-based first-principle model (baseline), an UDE-based model using a linear regression in the place of neural network, and an UDE-based model using a neural network designed specifically for each pathway. These neural networks are feedforward with variations in their architecture tailored to each model: the Toy model features 2 layers with 7 input neurons and 11 output neurons, the JAK-STAT model employs 4 layers with 5 input and output neurons, and the MAPK orderedElementary and FeMetabolism FeDeficient models use 4 layers with 4 input neurons and similarly structured output layers. All models use 5 neurons in the hidden layers (except the Toy model which uses 11), and RELU as the activation function in the hidden layers, transitioning to a linear combination in the output layer.

Following this initial evaluation, we selected the most promising model for further training on an expanded dataset $\{\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(100)}\}$ composed of 100 observations, generated for 100 different initial conditions, aiming to demonstrate the learning capabilities of the model. This step was taken with consideration of the computational constraints associated with training more complex models. The progression of the analysis involved systematically removing one first-principle parameter at a time and performing simultaneous inference with neural network weights, continuing this process until all first-principle parameters have been evaluated. This included scenarios where two and eventually all first-principle parameters are removed, examining the impact of such removals on the SMAPE metric on the test set.

Each of the following sections are dedicated to one of our four case studies presented in Section 4.3: Toy Model (Section 5.1), JAK-STAT pathway (Section 5.2), MAPK OrderedElementary (Section 5.3), and FeMetabolism FeDeficient (Section 5.4). The final section, Section 5.5, is dedicated to discussing these findings in detail.

5.1 Results with the Toy model

Our initial analysis of the Toy Model involved three distinct modeling approaches, each trained using 10 observations. Validation and testing datasets were equivalently sized to the training set. The first-principle model, used as a baseline, achieved mean SMAPE scores of 54.96, 53.83, and 56.08 across the training, validation, and test sets, respectively. In comparison, the purely linear regression model demonstrated overfitting, with SMAPE scores of 26.87 on training but escalating to 49.37 and 46.14 on validation and testing.

The UDE integrated with linear regression outperformed the purely linear model, recording mean SMAPE of 20.65, 31.74, and 33.00, respectively, across the datasets. Meanwhile, the UDE model utilizing a neural network architecture showed the most consistent performance across all models, with scores of 21.80, 27.50, and 31.15, thereby marking a significant improvement over the first-principle baseline. These comparisons are highlighted in Figure 5.1.

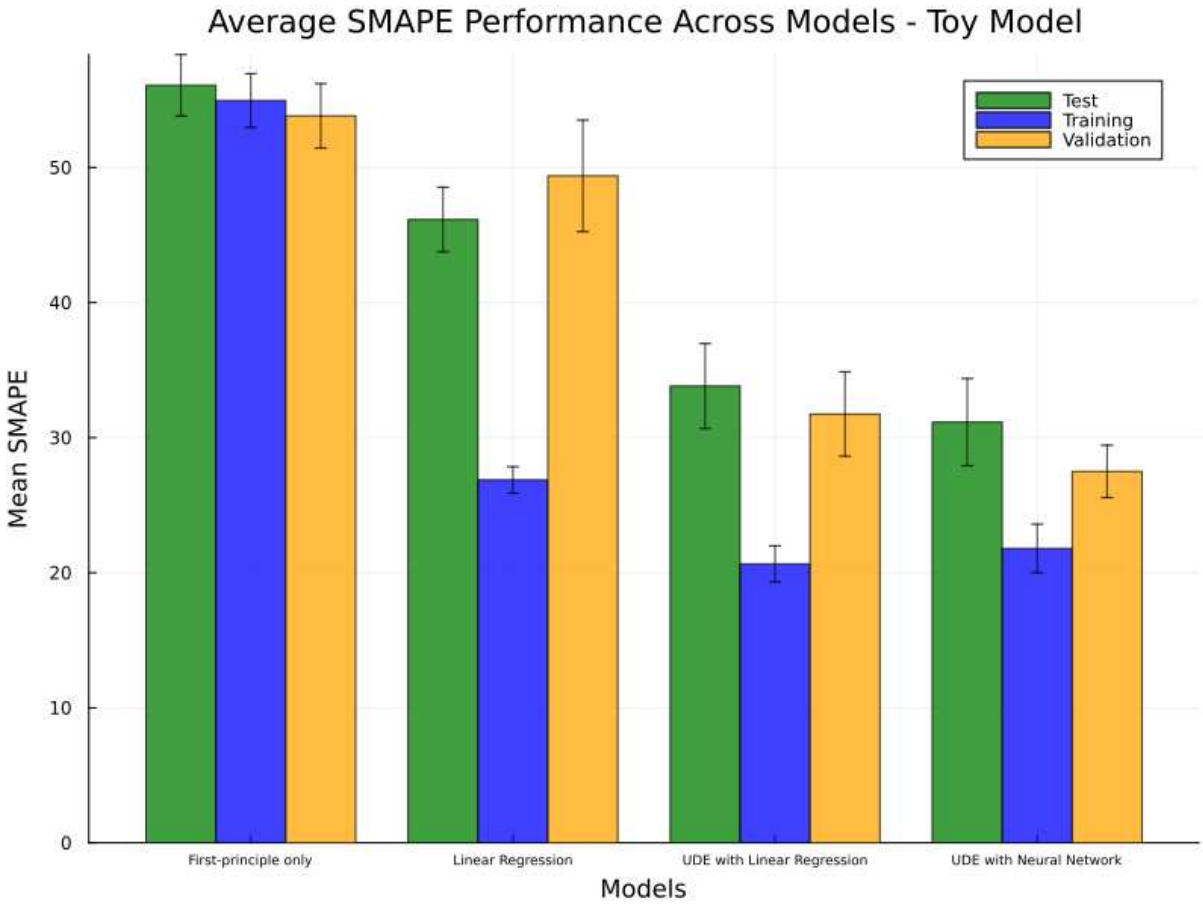


Figure 5.1: Model performance for the Toy Model, where modeling approaches are compared based on the SMAPE for training, validation, and test sets (each set containing 10 observations). The error bars represent the standard deviation.

To evaluate the scalability of the UDE-based model, the neural network version was retrained with an expanded dataset of 100 observations. An attempt to scale up to 1000 observations per species was computationally prohibitive and was not pursued further. The UDE with neural network, when retrained, achieved marginally better mean SMAPE

scores of 23.28, 26.24, and 25.46, respectively, suggesting slight improvements over the initial 100-observation training. These results are illustrated in Figure 5.2.

The finest prediction by the UDE with a neural network, trained on 10 observations, achieved a SMAPE of 20.50, while the least accurate prediction registered at 51.74. Figure 5.3 contrasts the best and the worst predictions, showing the best predictions closely approximating the true dynamics across all chemical species, whereas the poorest predictions failed significantly.

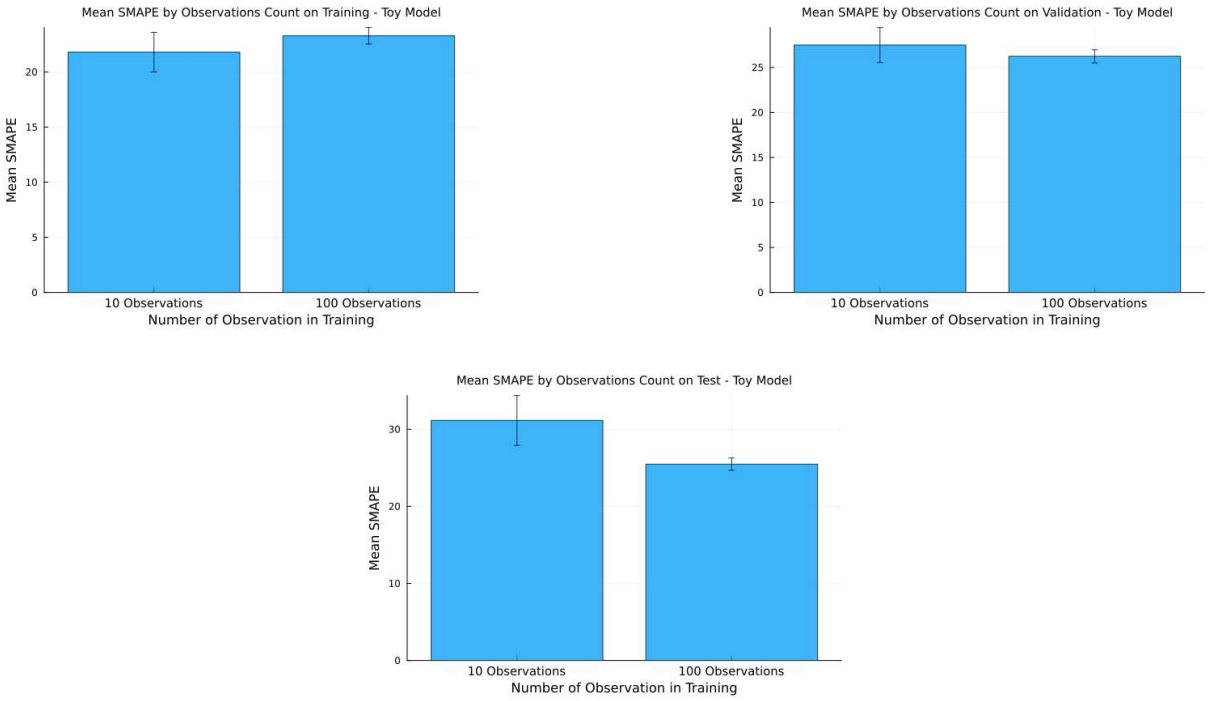


Figure 5.2: Data expansion impact on UDE with neural network for the Toy Model. Increasing the number of training observations affects the mean SMAPE for the UDE model integrated with a neural network. The x-axis indicates the number of observations, and the y-axis shows the mean SMAPE across corresponding datasets. The error bars represent the standard deviation.

When the neural network-based UDE was trained with 100 observations, the best and worst SMAPE scores recorded were 11.54 and 48.18, respectively. Figure 5.4 illustrates these outcomes, with the best prediction successfully approximating the true dynamics, while the worst struggled considerably across most chemical species.

Lastly, Figure 5.5 explores the impact of removing and subsequently estimating first-principle parameters alongside the weights of the neural network, using 10 observations for training, validation, and testing. This analysis revealed that parameter removal significantly influenced model performance; in some instances, the SMAPE deteriorated beyond the first-principle model, while in others, it remained stable or improved, indicating the model's sensitivity to parameter configuration and estimation strategies.

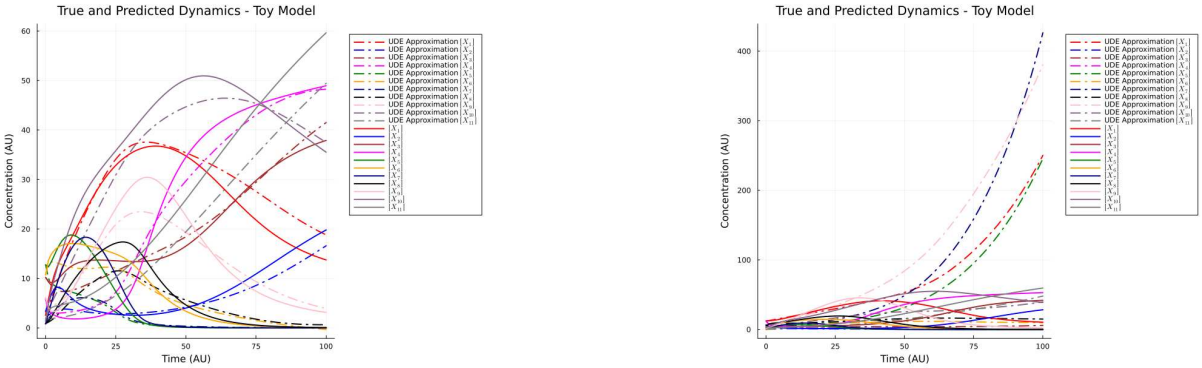


Figure 5.3: Predictive accuracy of the UDE with neural network for the Toy Model (10 observations): The best (left) and worst (right) predictions from the UDE model trained with 10 observations against the true dynamics of the Toy Model. Solid lines denote true dynamics, while dashed lines represent model predictions.

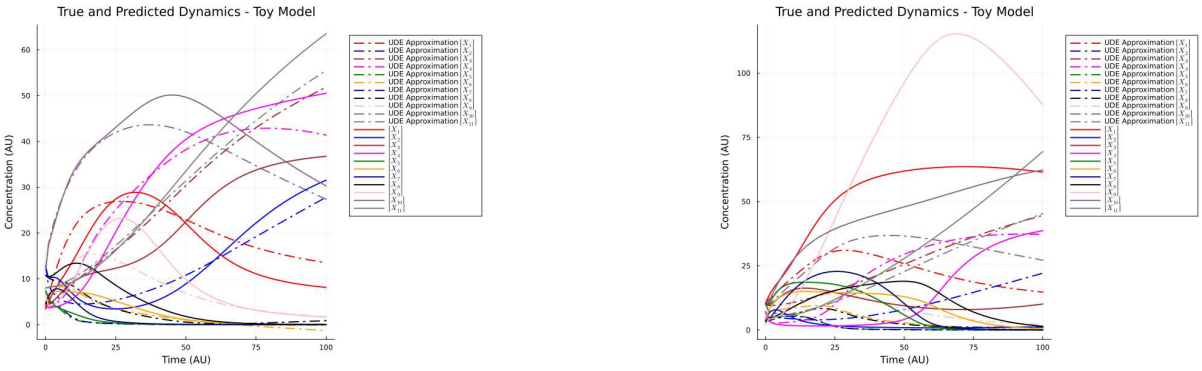


Figure 5.4: Predictive accuracy of the UDE with neural network for the Toy Model (100 observations). The best (left) and worst (right) predictions from the UDE model trained with 100 observations against the true dynamics of the Toy Model. Solid lines denote true dynamics, while dashed lines represent model predictions.

5.1.1 Experiments with noisy data for the Toy Model

To evaluate the robustness of our modeling approach under noisy conditions, we replicated the experimental setup previously used for the Toy Model. Due to the high computational demand, we did not conduct experiments involving the iterative removal and estimation of first-principle parameters alongside neural network weights for this dataset.

The first-principle model yielded mean SMAPE scores of 56.28, 54.84, and 57.37 for the training, validation, and test sets, respectively. In contrast, the purely linear regression model exhibited signs of overfitting, with SMAPE scores of 26.59, 50.67, and 49.10 across the same sets, indicating a significant gap between training and validation/test performance.

The UDE-based models demonstrated superior performance to the purely linear regression model. Specifically, the UDE integrated with linear regression recorded mean SMAPE scores of 21.47, 31.58, and 34.04, while the UDE coupled with a neural network achieved almost identical scores of 21.47, 31.58, and 34.02 across training, validation, and

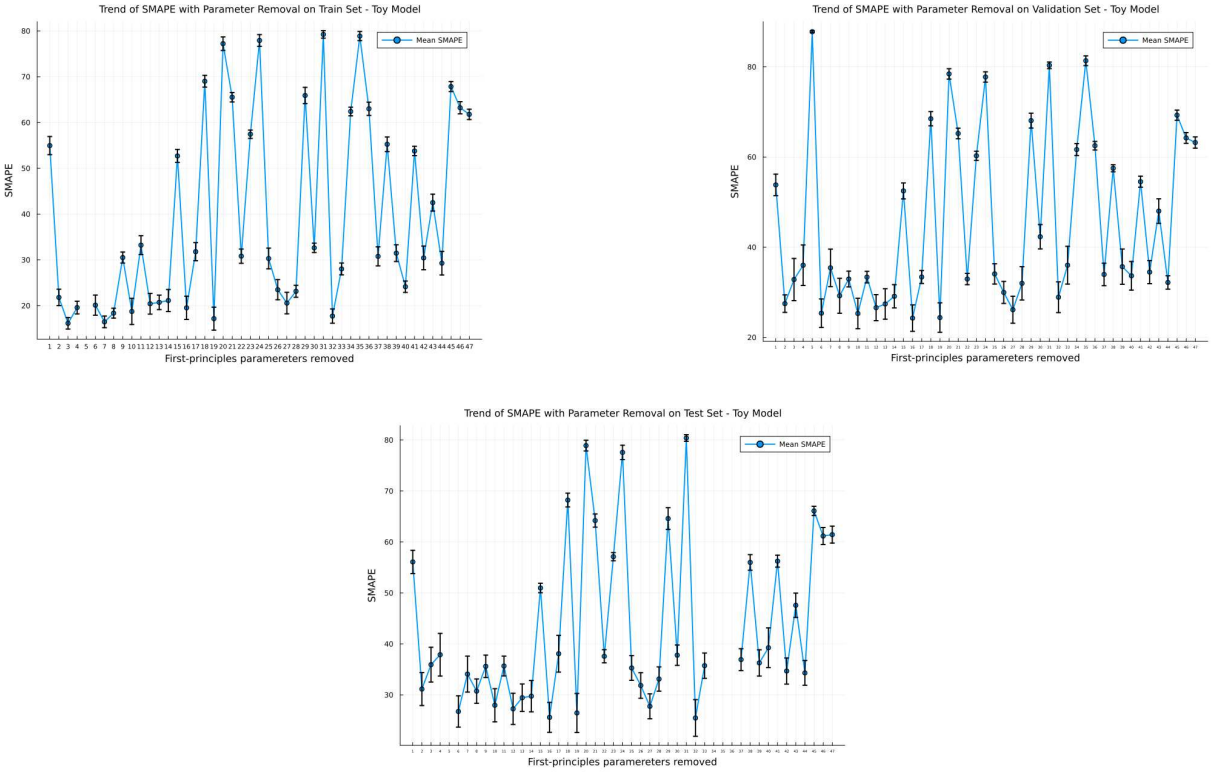


Figure 5.5: Impact of first-principle parameter removal on mean SMAPE for the Toy Model. It is depicted here the effect on the mean SMAPE of iteratively removing sets of first-principle parameters, using 10 observations for training set, 10 for validation set, and 10 for test set. The y-axis denotes the mean SMAPE of the respective dataset, and the x-axis lists the sets of parameters removed. The error bars represent the standard deviation (the full list of the sequence of removed rate constants is available in Table A.7). The error bars represent the standard deviation.

test sets respectively. Figure 5.6 presents these performance metrics.

Further analysis of the predictive accuracy of the UDE-based model is depicted in Figure 5.7. The best prediction from the UDE model integrated with a neural network, trained with 10 observations, attained a SMAPE score of 17.80, closely approximating the true dynamics of most chemical species but struggling with a few. Conversely, the worst prediction yielded a SMAPE score of 49.01, showcasing a significant deviation from expected dynamics, particularly in accurately modeling the behavior of some chemical species.

5.2 Results with JAK-STAT pathway

Our exploration of the JAK-STAT signaling pathway commenced with the training of three models, each utilizing three 10-observation sized datasets for training, validation, and test. Among the evaluated models, the UDE integrated with linear regression emerged as the most effective, achieving mean SMAPE scores of 12.93, 17.49, and 21.40 for the training, validation, and test sets, respectively. This performance represents a substantial

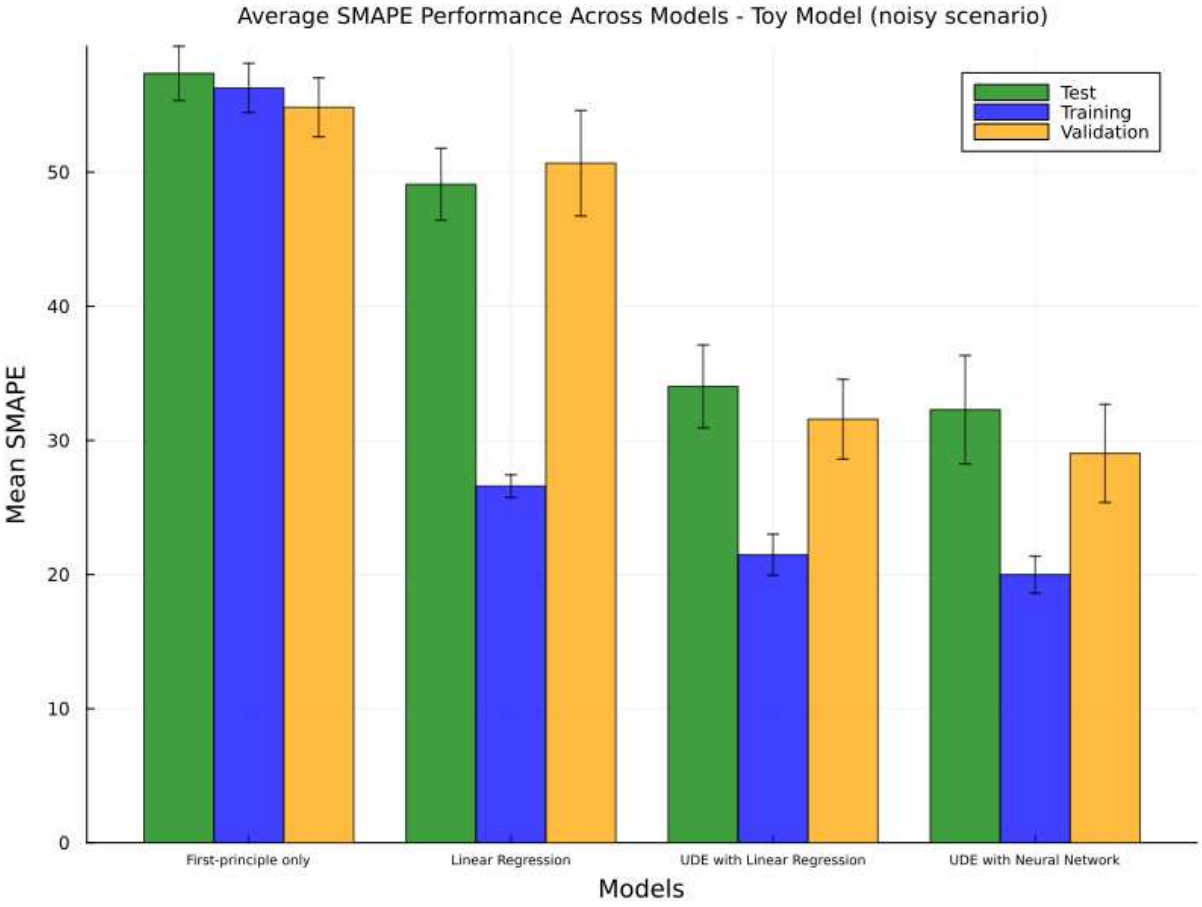


Figure 5.6: Model performance across noisy datasets for the Toy Model. This graph compares different modeling approaches under noisy conditions by showcasing the SMAPE values across training, validation, and test datasets, each with 10 observations. The y-axis displays the mean SMAPE for noisy data, while the x-axis categorizes the models. The error bars represent the standard deviation.

enhancement over the first-principle model, which recorded mean SMAPE scores of 40.50, 36.76, and 43.31 for the training, validation and test sets, respectively. In contrast, the purely linear regression model yielded mean SMAPE scores of 15.64, 27.23, and 29.83, while the UDE coupled with a neural network exhibited scores of 12.35, 25.04, and 32.60, respectively. Notably, both the linear regression and the neural network-based UDE models demonstrated a tendency towards overfitting, with marginal improvement on the test set compared to the first-principle model alone (Figure 5.8).

To further evaluate the performance of the hybrid model, the UDE with linear regression was subsequently trained on datasets expanded to 100 and then 1000 observations, with the validation and test sets sized equivalently to the training dataset. Training the UDE with linear regression on 100 observations resulted in mean SMAPE scores of 16.01, 17.55, and 18.83 for the train, validation, and test sets, respectively. Training on a larger dataset of 1000 observations slightly increased the SMAPE on train set to 16.32 but improved the SMAPE on validation and test set to 16.34 and 16.74, respectively. Interestingly, enlarging the dataset led to an increase in the SMAPE on training set but enhanced the performance of the model on the test set for both the 100 and 1000

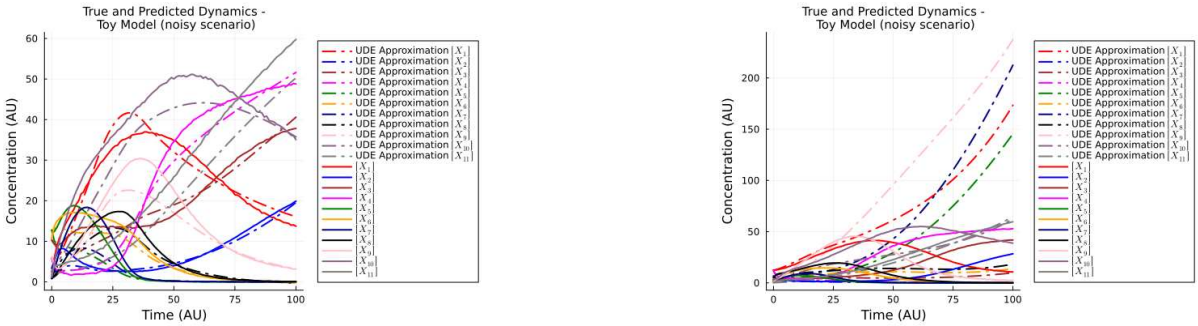


Figure 5.7: Predictive accuracy of the UDE with neural network in noisy conditions for the Toy Model (10 observations). The best (left) and worst (right) predictions from the UDE model trained with 10 observations under noisy conditions, this figure contrasts the model predictions (dashed lines) against the true dynamics (solid lines) of the Toy Model.

observations training scenarios, indicating improved generalizability (Figure 5.9).

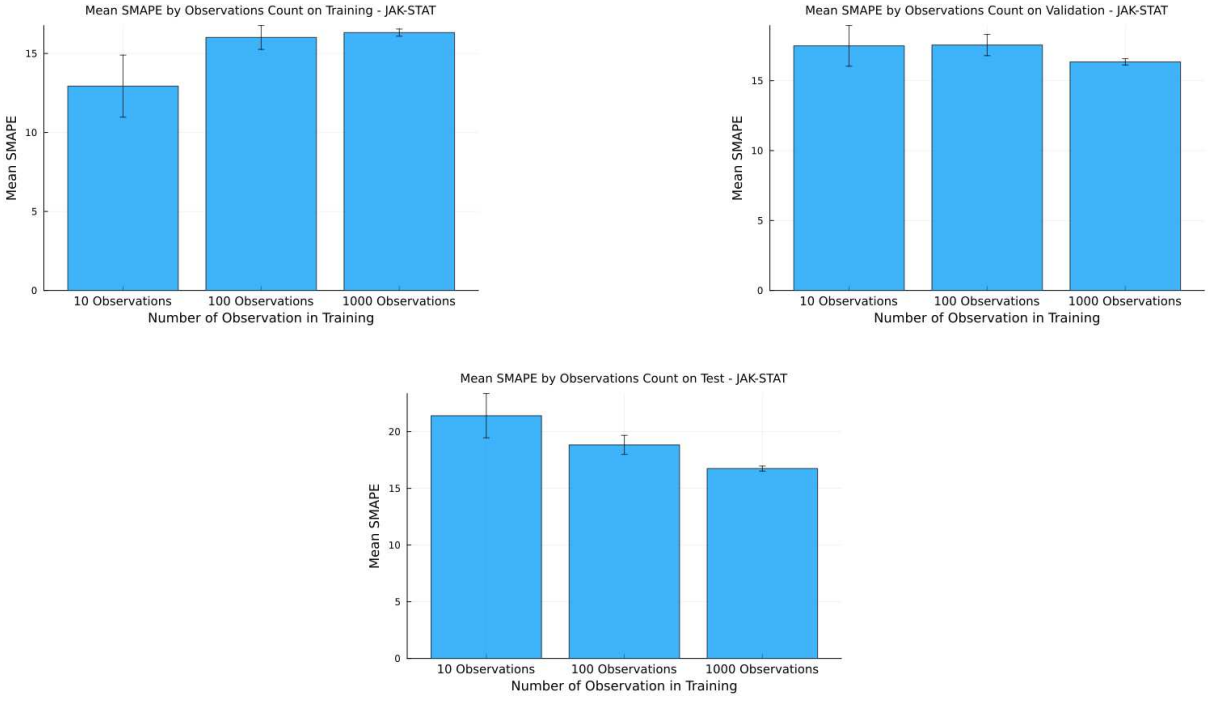


Figure 5.9: Data expansion impact on UDE with neural network for the JAK-STAT. Increasing the number of training observations affects the mean SMAPE for the UDE model integrated with a neural network. The x-axis indicates the number of observations, and the y-axis shows the mean SMAPE across corresponding datasets. The error bars represent the standard deviation.

The better prediction achieved by the UDE model, integrated with linear regression and trained on a dataset of 10 observations, recorded a SMAPE of 13.37. Conversely, the least accurate prediction by this model configuration recorded a SMAPE of 31.61. As depicted in Figure 5.10, both the best and worst prediction scenarios reveal challenges in accurately modeling the dynamics of chemical species x_3 and x_4. Notably, in the least

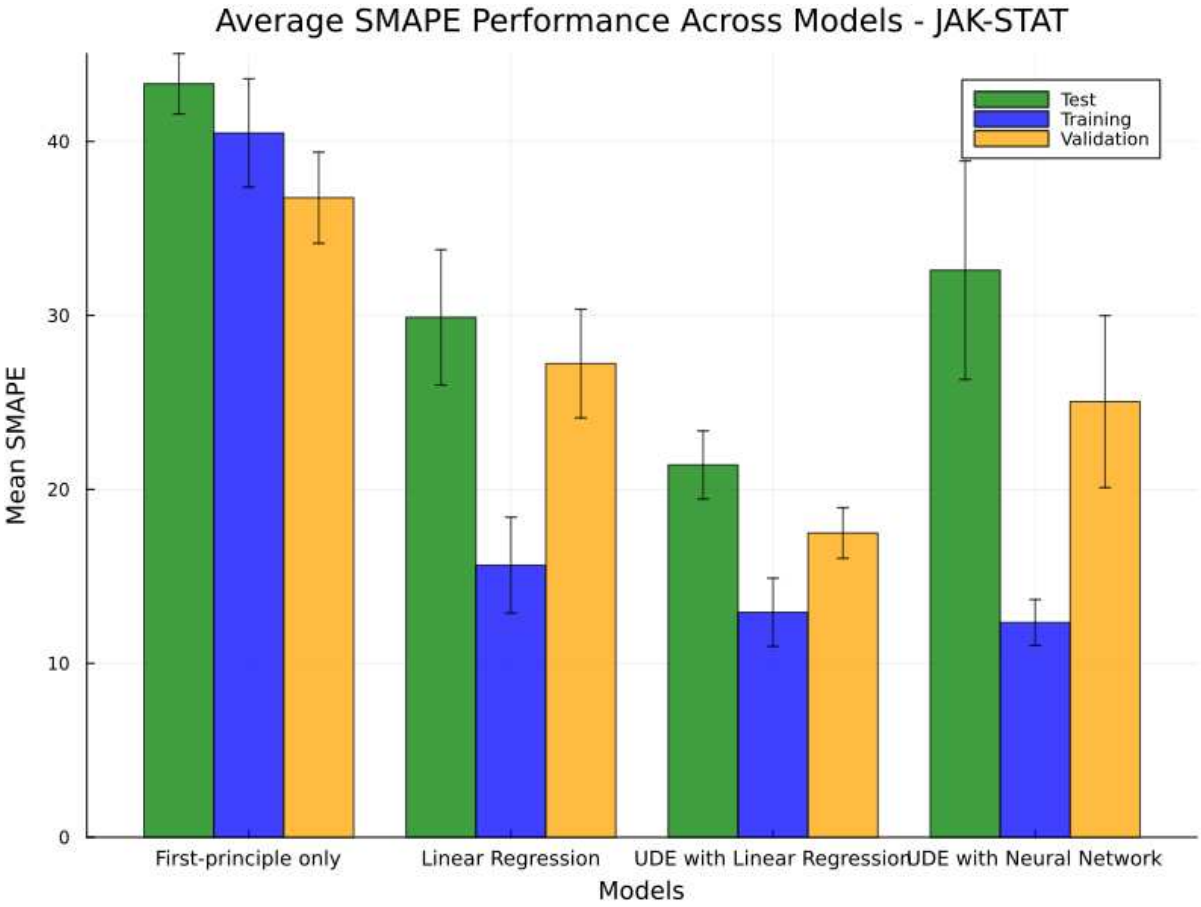


Figure 5.8: Model performance for the JAK-STAT, where modeling approaches are compared based on the SMAPE for training, validation, and test sets (each set containing 10 observations). The y-axis represents the mean SMAPE, while the x-axis identifies the model type. The error bars represent the standard deviation.

accurate predictions, the divergence between the predicted and true dynamics of these species was significantly discrepant, displaying disparate trends.

Progressing to the UDE model integrated with linear regression but trained with an expanded dataset of 100 observations, the range of prediction accuracy widened. The most accurate prediction within this set achieved a SMAPE of 4.67, while the least accurate prediction escalated to a SMAPE of 49.21. The corresponding dynamics, illustrated in Figure 5.11, indicate that despite the discrepancies in the less accurate prediction, the model captured the general trends of each dynamics of the chemical species.

Upon further expanding the training dataset to 1000 observations, the prediction accuracy for the UDE model integrated with linear regression showed the best and worst SMAPEs of 4.20 and 56.12, respectively. Remarkably, the least accurate prediction for this model surpassed the inaccuracy observed in the 100-observation training scenario. Figure 5.12 highlights that; similar to previous results, the model struggled to precisely replicate the dynamics of chemical species x_3 and x_4, especially in the least accurate prediction scenario, where the predicted dynamics diverged significantly from the true data.

The analysis of first-principal parameters removal within the JAK-STAT model is

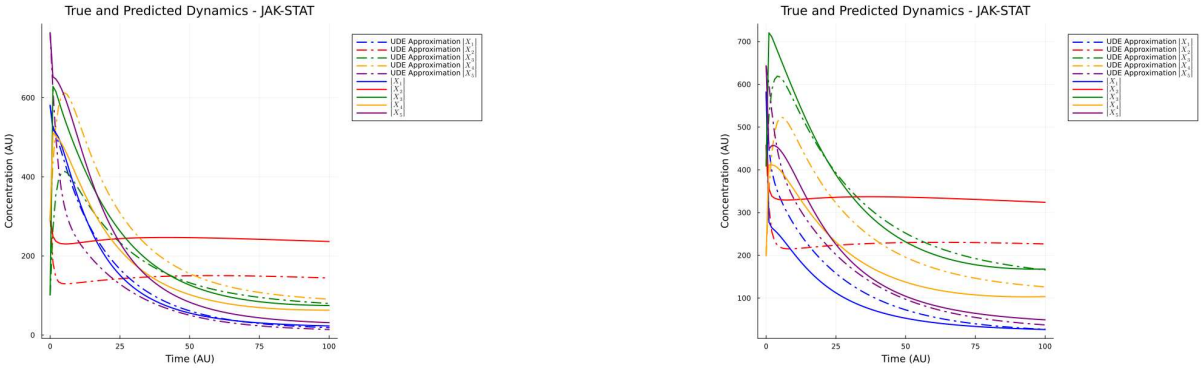


Figure 5.10: Predictive accuracy of the UDE with linear regression (10 observations): The best (left) and worst (right) predictions from the UDE model trained with 10 observations against the true dynamics of the JAK-STAT pathway. Solid lines denote true dynamics, while dashed lines represent model predictions.

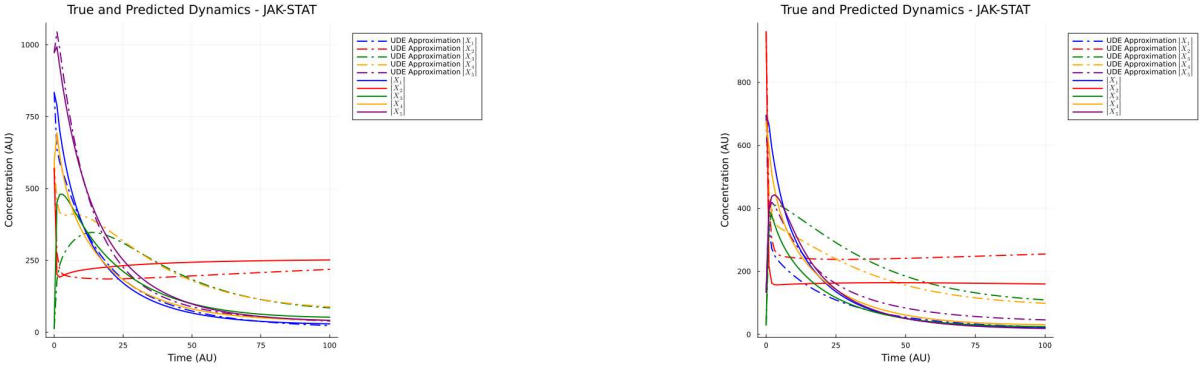


Figure 5.11: Predictive accuracy of the UDE with linear regression (100 observations): The best (left) and worst (right) predictions from the UDE model trained with 100 observations against the true dynamics of the JAK-STAT pathway. Solid lines denote true dynamics, while dashed lines represent model predictions.

illustrated in Figure 5.13, which delineates the effects of removing and subsequently estimating first-principle parameters in conjunction with the weights of the linear regression. This experiment utilized 10 observations each for training, validation, and testing. The findings reveal that the performance of the model is particularly sensitive to the removal and estimation of three specific sets of parameters. In contrast, the removal of most other parameter sets did not significantly impact model stability, indicating robustness when compared to training the UDE-based model under the assumption that all first-principle parameters are accurately known.

5.2.1 Experimentation with noisy data for the JAK-STAT pathway

To evaluate the robustness of our modeling approach, we replicated the experimental setup for the JAK-STAT pathway, incorporating noisy data this time. We trained four

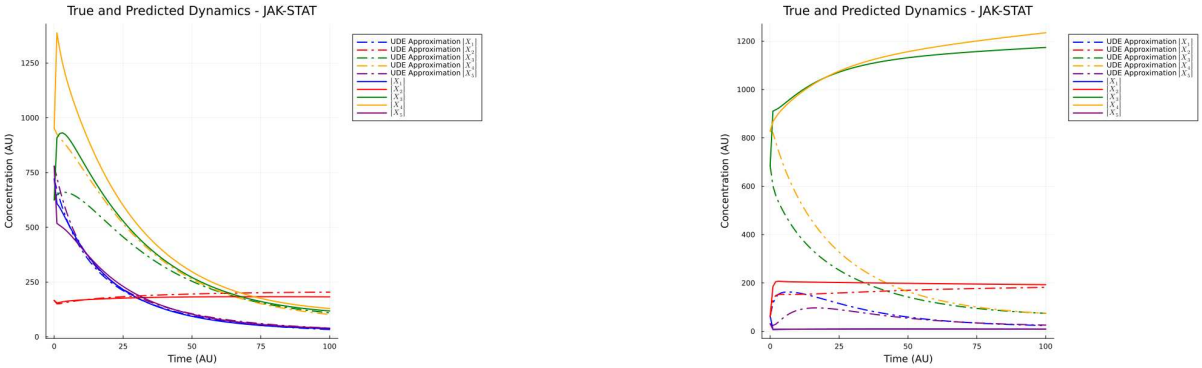


Figure 5.12: Predictive accuracy of the UDE with linear regression (1000 observations): The best (left) and worst (right) predictions from the UDE model trained with 1000 observations against the true dynamics of the JAK-STAT pathway. Solid lines denote true dynamics, while dashed lines represent model predictions.

distinct models, each using 10 observations, with validation and test sets sized equally. The first-principle model under noisy conditions achieved mean SMAPE scores of 40.50 on training, 44.81 on validation, and 34.86 on the test sets. In contrast, the purely linear regression model demonstrated improved performance, recording mean SMAPE of 15.20, 31.90, and 22.11 across the respective datasets, indicating a substantial enhancement over the first-principle model alone. The UDE-based models showed signs of overfitting; the UDE integrated with linear regression achieved SMAPE scores of 13.35, 19.90, and 20.81, while the UDE coupled with a neural network recorded SMAPE scores of 8.95, 18.89, and 27.29 across the training, validation, and test sets, respectively. Among all models, the UDE integrated with linear regression exhibited the best overall performance. These comparative results are illustrated in Figure 5.14.

Further analysis involved examining the best and worst predictions from the UDE integrated with linear regression, trained on 10 observations. The best prediction achieved a SMAPE score of 9.51, providing a satisfactory approximation to the true dynamics of some chemical species, yet struggled with species x_3 and x_4. The worst prediction, however, resulted in a SMAPE score of 39.12, presenting dynamics significantly divergent from the true data, and performing worse than the first-principle model. These outcomes are depicted in Figure 5.15.

Additionally, Figure 5.16 illustrates the effects of systematically removing and subsequently estimating sets of first-principle parameters along with the weights of the linear regression. This analysis, conducted with 10 observations for each dataset, indicated that the performance of the model remained stable when up to three parameters were removed. However, as more parameters were excluded, the model became increasingly sensitive to these removals, highlighting the critical influence of specific first-principle parameters on model accuracy and robustness.

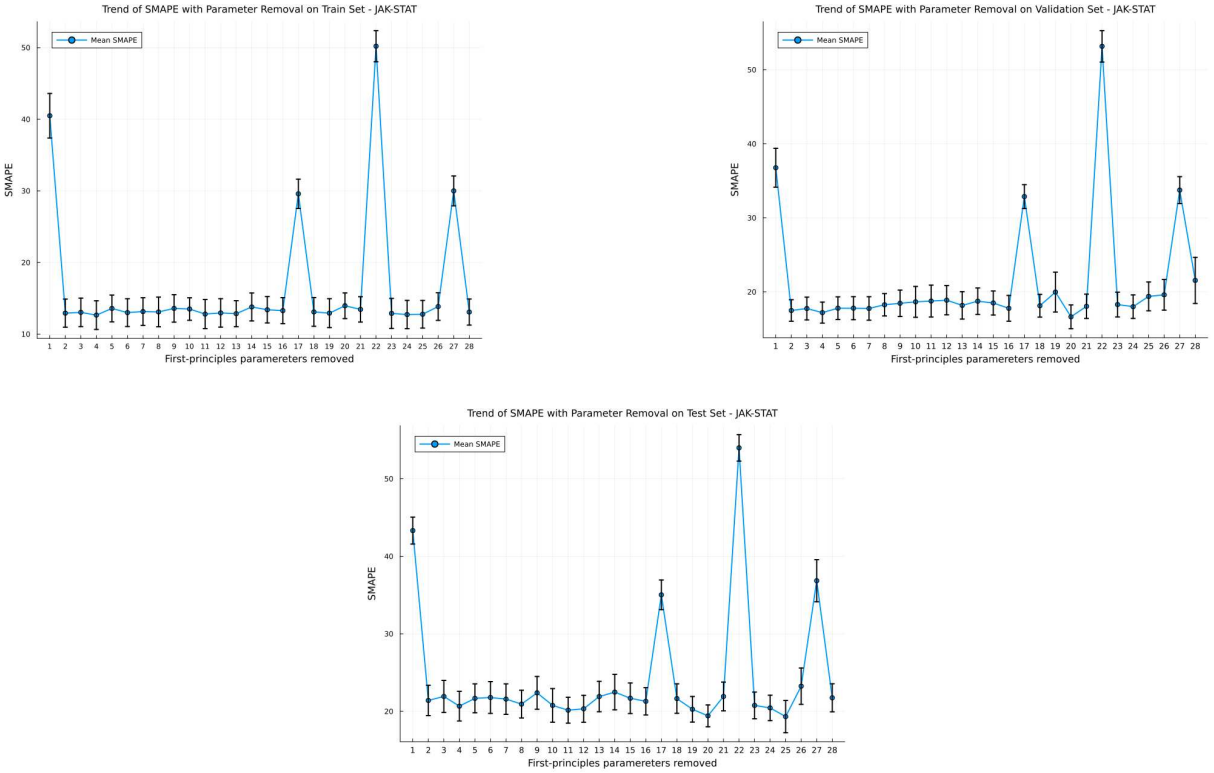


Figure 5.13: Impact of first-principle parameter removal on mean SMAPE for the JAK-STAT pathway. It is depicted here the effect on the mean SMAPE of iteratively removing sets of first-principle parameters, using 10 observations for training set, 10 for validation set, and 10 for test set. The y-axis denotes the mean SMAPE of the respective dataset, and the x-axis lists the sets of parameters removed (the full list of the sequence of removed rate constants is available in Table A.5). The error bars represent the standard deviation.

5.3 Results with MAPK orderedElementary

In our investigation of the MAPK OrderedElement model, the performance of different modeling approaches was quantitatively assessed using the SMAPE metric. The first-principle model only yielded a mean SMAPE value of 65.50 for the training set and 67.05 for the validation set. Enhancement of model performance was observed upon the application of our hybrid modeling approach. Initially, a linear regression-based hybrid model demonstrated a mean SMAPE of 30.72 on the training set and 39.06 on the validation set. Further improvements were achieved with our neural network-based hybrid model, which recorded a mean SMAPE of 26.95 for the training set and 32.75 for the validation set, utilizing a dataset comprising 10 observations. The neural network-based hybrid model emerged as the superior model. Figure 5.17 presents a comparative analysis of these models, highlighting the superior performance of the neural network-based hybrid approach.

We expanded our dataset to include 100 and subsequently 1000 observations to train the UDE integrated with a neural network. This increase in data volume significantly improved the performance of the model, with the model trained on 100 observations achieving mean SMAPE scores of 13.74 for training, 15.04 for validation, and 15.04 for

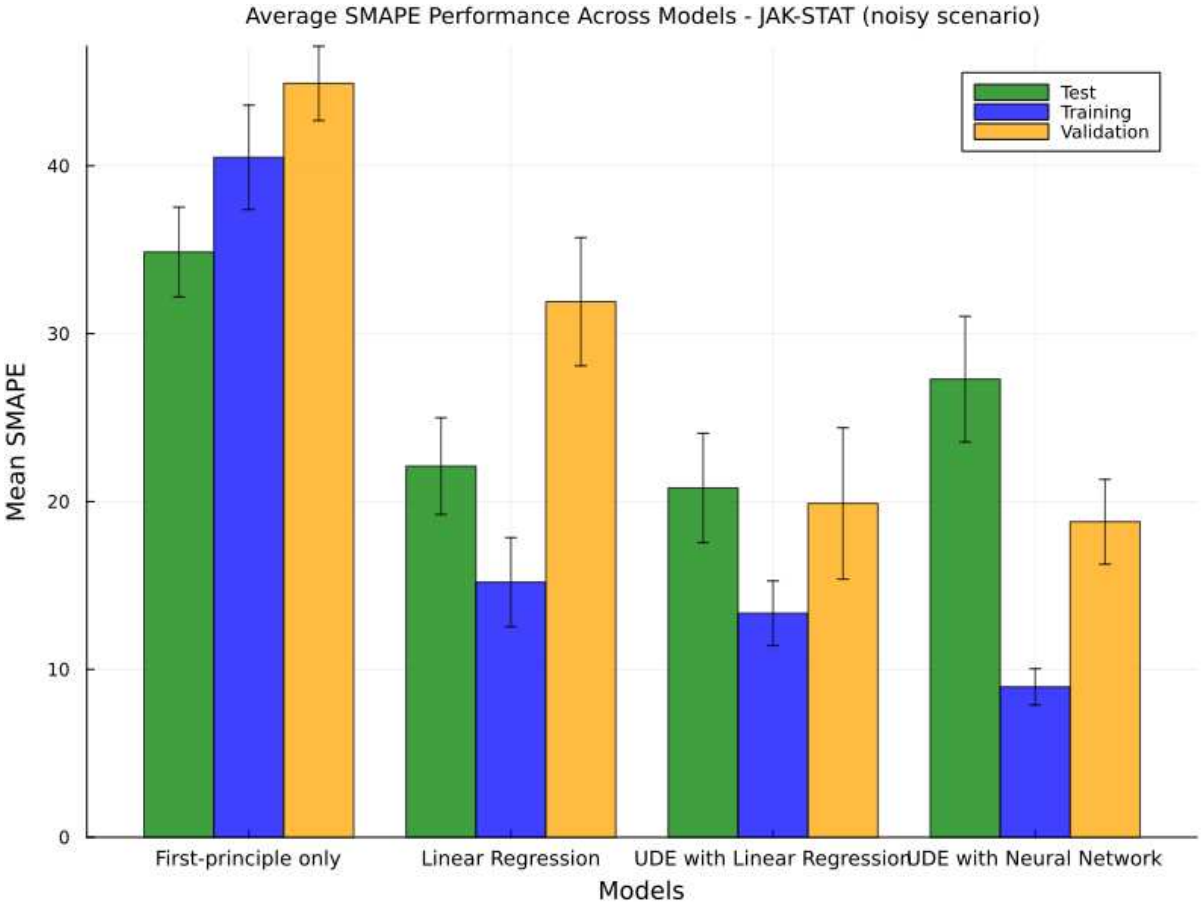


Figure 5.14: Model performance across noisy datasets for the JAK-STAT pathway. This graph compares different modeling approaches under noisy conditions by showcasing the SMAPE values across training, validation, and test datasets, each with 10 observations. The y-axis displays the mean SMAPE for noisy data, while the x-axis categorizes the models. The error bars represent the standard deviation.

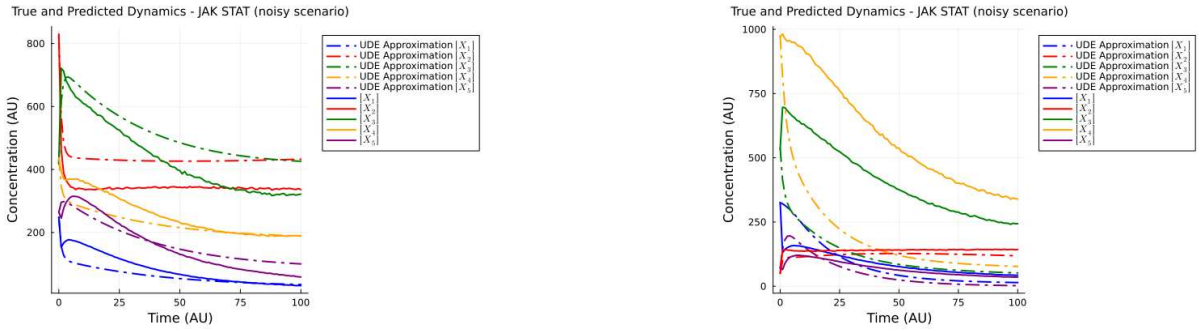


Figure 5.15: Predictive accuracy of the UDE with neural network in noisy conditions for the JAK-STAT pathway (10 observations). The best (left) and worst (right) predictions from the UDE model trained with 10 observations under noisy conditions, this figure contrasts the model predictions (dashed lines) against the true dynamics (solid lines) of the JAK-STAT pathway.

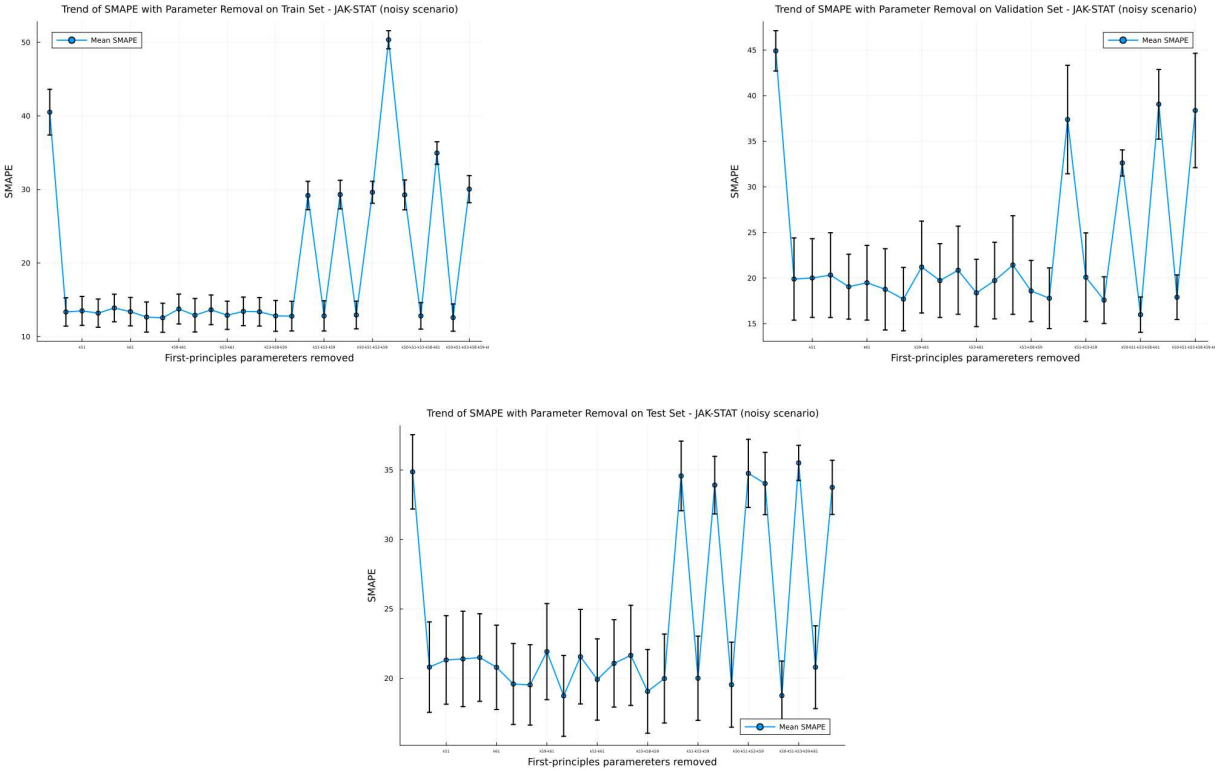


Figure 5.16: Impact of first-principle parameter removal on mean SMAPE for the JAK-STAT pathway, in noisy conditions. It is depicted here the effect on the mean SMAPE of iteratively removing sets of first-principle parameters, using 10 observations for training set, 10 for validation set, and 10 for test set. The y-axis denotes the mean SMAPE of the respective dataset, and the x-axis lists the sets of parameters removed(the full list of the sequence of removed rate constants is available in Table A.6). The error bars represent the standard deviation.

the test sets, demonstrating the model's enhanced learning capability. Furthermore, the model trained with 1000 observations recorded slightly better generalization, with SMAPE scores of 13.99 on training, 14.63 on validation, and 14.42 on the test set. Figure 5.18 depicts these performance enhancements.

To evaluate the prediction quality of the UDE-based model, we analyzed the best and worst predictions from the model integrated with the neural network. For the model trained with 10 observations, the best prediction achieved a SMAPE of 24.75, while the worst logged a SMAPE of 53.47. As shown in Figure 5.19, the best prediction accurately approximated most dynamics but faced challenges with chemical species x_2 and x_3. Conversely, the worst prediction only reliably approximated the dynamics of chemical species x_4, with significant discrepancies observed for other species.

When trained with 100 observations, the best prediction of the UDE integrated with a neural network achieved a SMAPE of 5.62, while the worst was 36.69. Figure 5.20 illustrates that while the best prediction closely matched the true dynamics, it struggled with species x_2 and x_3, similar to the smaller dataset. The dynamics of these species in the worst-case scenario were further from the true dynamics than those observed in the best case.

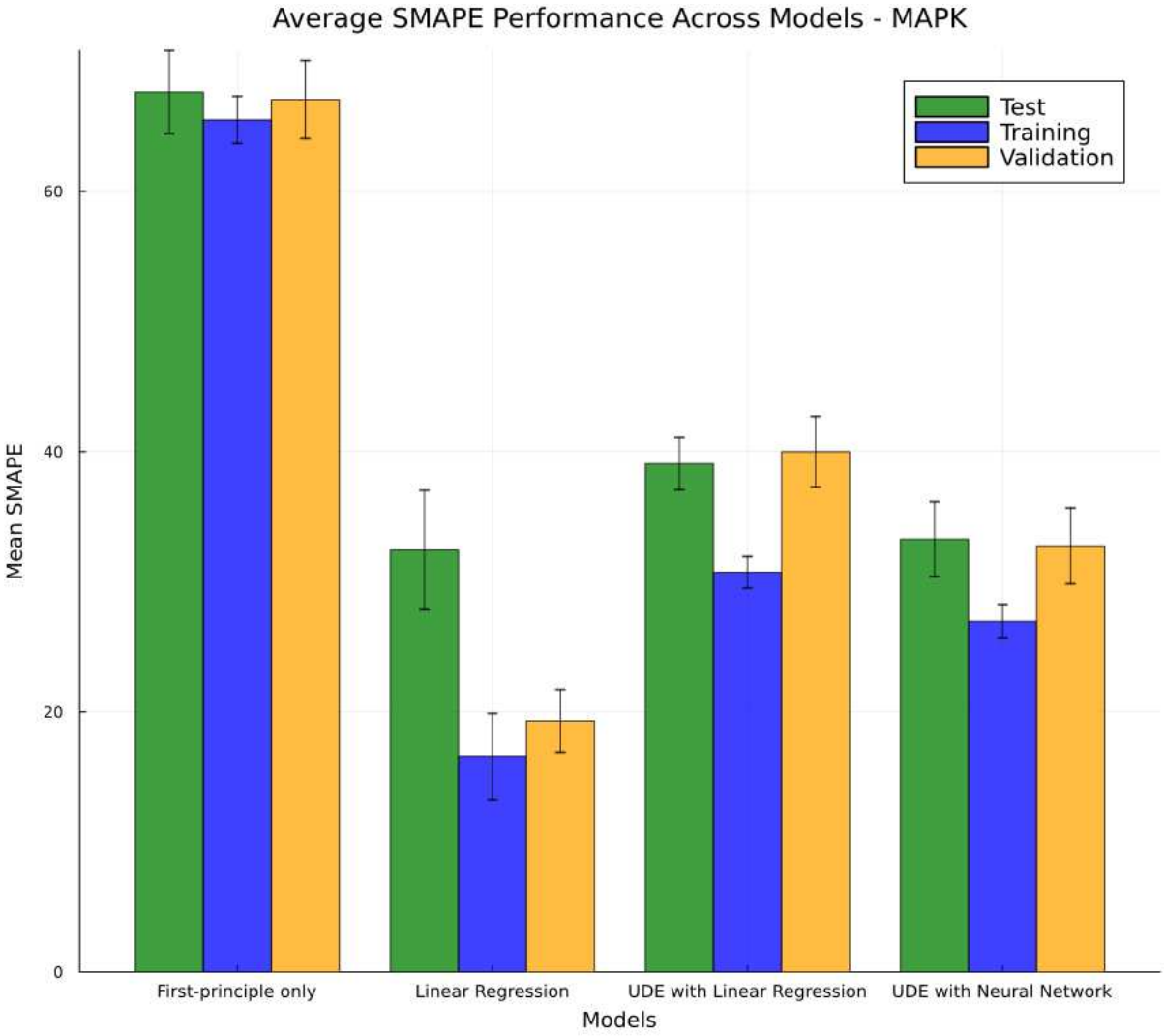


Figure 5.17: Model performance for the MAPK OrderedElement model, where modeling approaches are compared based on the SMAPE for training, validation, and test sets (each set containing 10 observations). The y-axis represents the mean SMAPE, while the x-axis identifies the model type. The error bars represent the standard deviation.

Finally, the model trained with 1000 observations exhibited a best prediction SMAPE of 2.23, showcasing a high degree of accuracy in capturing the true dynamics. The worst prediction for this training setup recorded a SMAPE of 44.18, demonstrating that while overall model performance improved, challenges remain in accurately modeling the behavior of certain chemical species under less ideal conditions, particularly species x_2 and x_3 as detailed in Figure 5.21. This indicates the sensitivity of the model to the variability inherent in larger datasets and the complexity of the system being modeled.

Concerning the experimentation involving the removal and subsequent joint estimation of first-principle parameters, within the MAPK OrderedElementary model, Figure 5.22 illustrates the relationship between the removal of first-principle parameters and the mean SMAPE on the test set. Interestingly, the mean SMAPE does not deteriorate as first-principle parameters are removed and subsequently estimated, in comparison to the scenario where all first-principle parameters are presumed known. This trend holds even

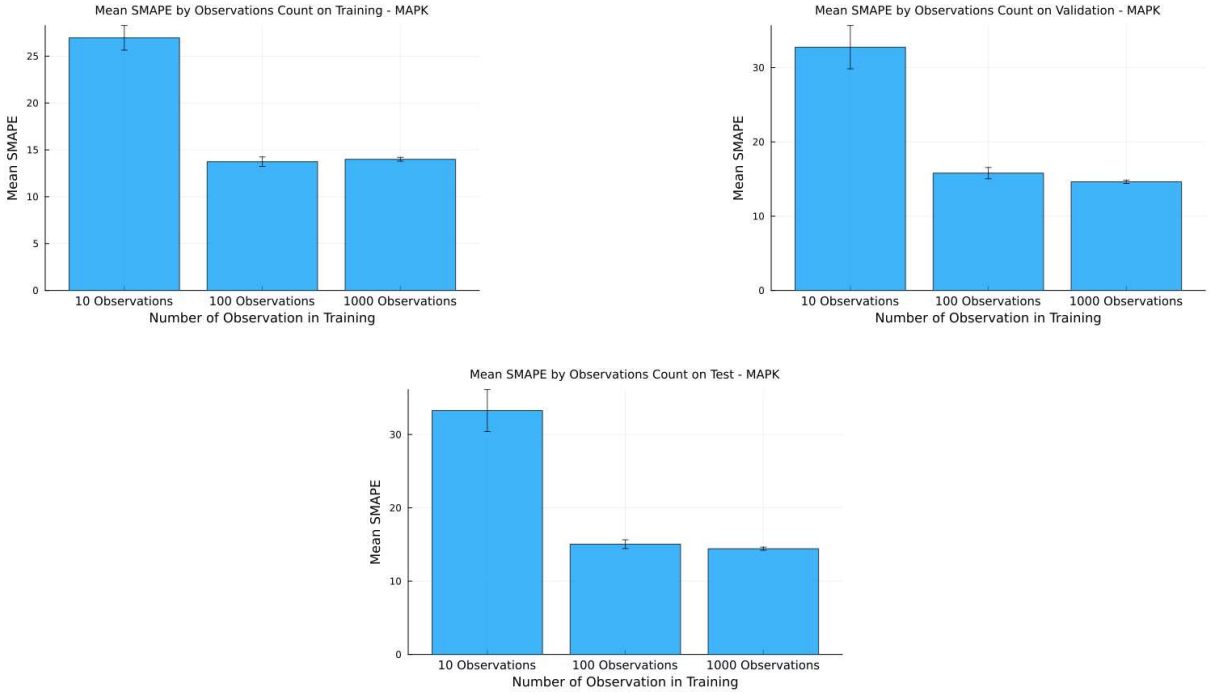


Figure 5.18: Data expansion impact on UDE with neural network for the MAPK orderedElemented. Increasing the number of training observations affects the mean SMAPE for the UDE model integrated with a neural network. The x-axis indicates the number of observations, and the y-axis shows the mean SMAPE across corresponding datasets. The error bars represent the standard deviation.

when the estimates for parameters $k4$ and $k3$—particularly in scenarios where $k4$ is removed alone, $k3$ and $k4$ are removed together, and $k4$ and $k5$ are removed simultaneously—are significantly inaccurate. The exception to this observation is the removal of $k3$ alone, which is the only instance where the mean SMAPE of our hybrid model worsens

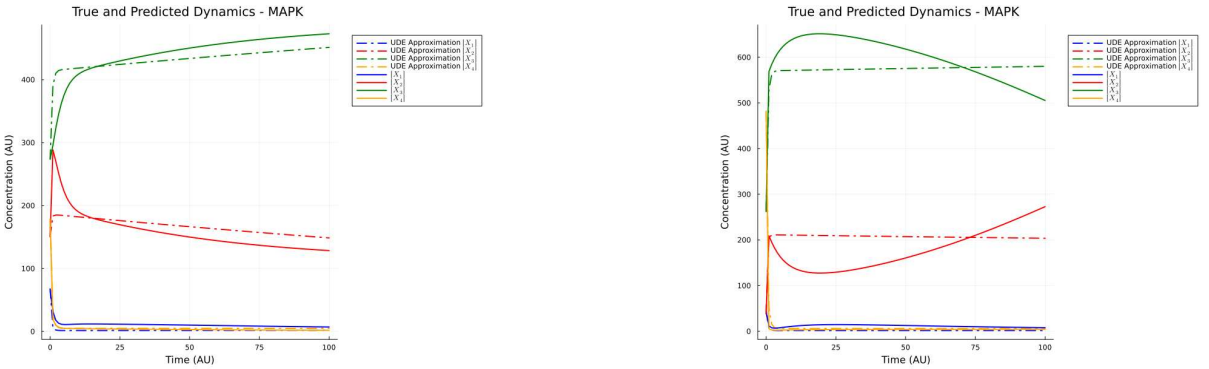


Figure 5.19: Predictive accuracy of the UDE with neural network for the MAPK orderedElementary (10 observations): The best (left) and worst (right) predictions from the UDE model trained with 10 observations against the true dynamics of the MAPK orderedElementary. Solid lines denote true dynamics, while dashed lines represent model predictions.

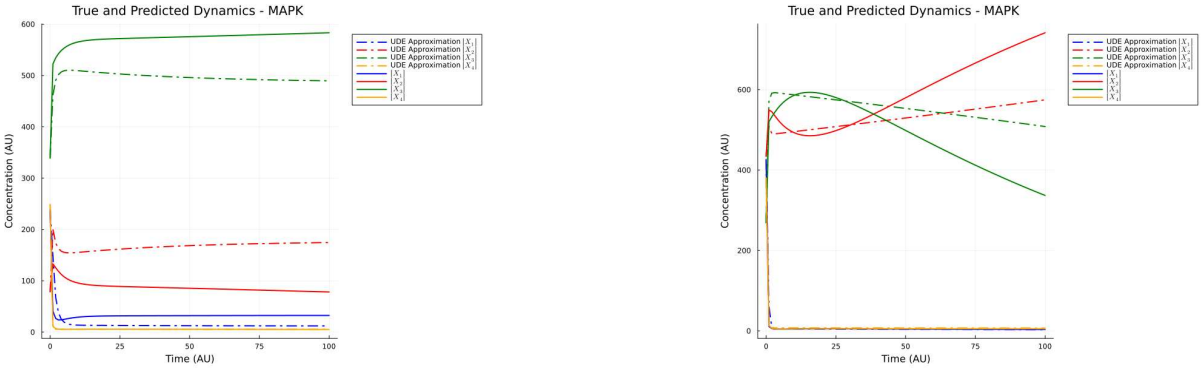


Figure 5.20: Predictive accuracy of the UDE with neural network for the MAPK orderedElementary (100 observations): The best (left) and worst (right) predictions from the UDE model trained with 100 observations against the true dynamics of the MAPK orderedElementary. Solid lines denote true dynamics, while dashed lines represent model predictions.

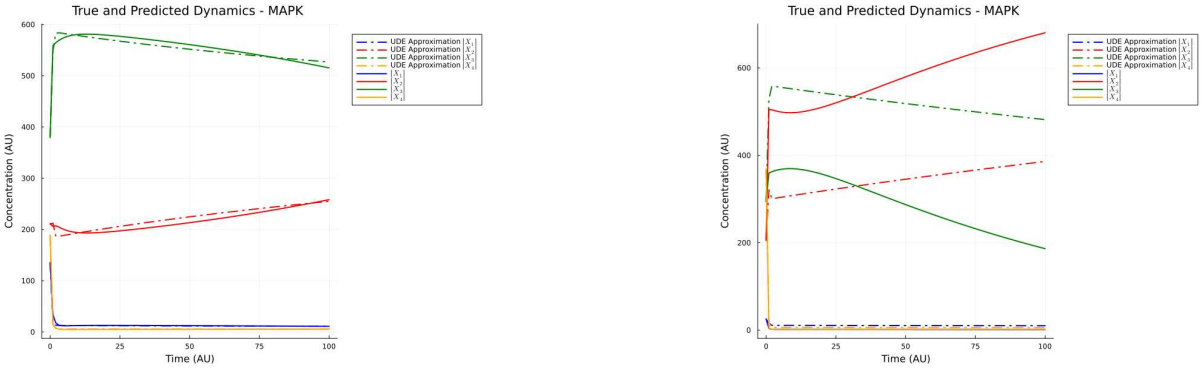


Figure 5.21: Predictive accuracy of the UDE with neural network for the MAPK orderedElementary (1000 observations): The best (left) and worst (right) predictions from the UDE model trained with 1000 observations against the true dynamics of the MAPK orderedElementary. Solid lines denote true dynamics, while dashed lines represent model predictions.

compared to the baseline scenario where all first-principle parameters are assumed to be accurately known.

5.3.1 Experimentation with noisy data for the MAPK orderedElementary

This subsection revisits the MAPK OrderedElementary model, this time evaluating its performance under noisy data conditions. We replicated the previous experimental setup with the same four models but introduced noise to assess robustness.The first-principle model under noisy conditions yielded mean SMAPE scores of 65.50, 56.23, and 63.38 for the training, validation, and test sets, respectively. The purely linear regression model demonstrated a notable improvement over the first-principle model, with SMAPE scores of 16.30, 18.63, and 31.98, respectively. In comparison, the UDE-based hybrid models

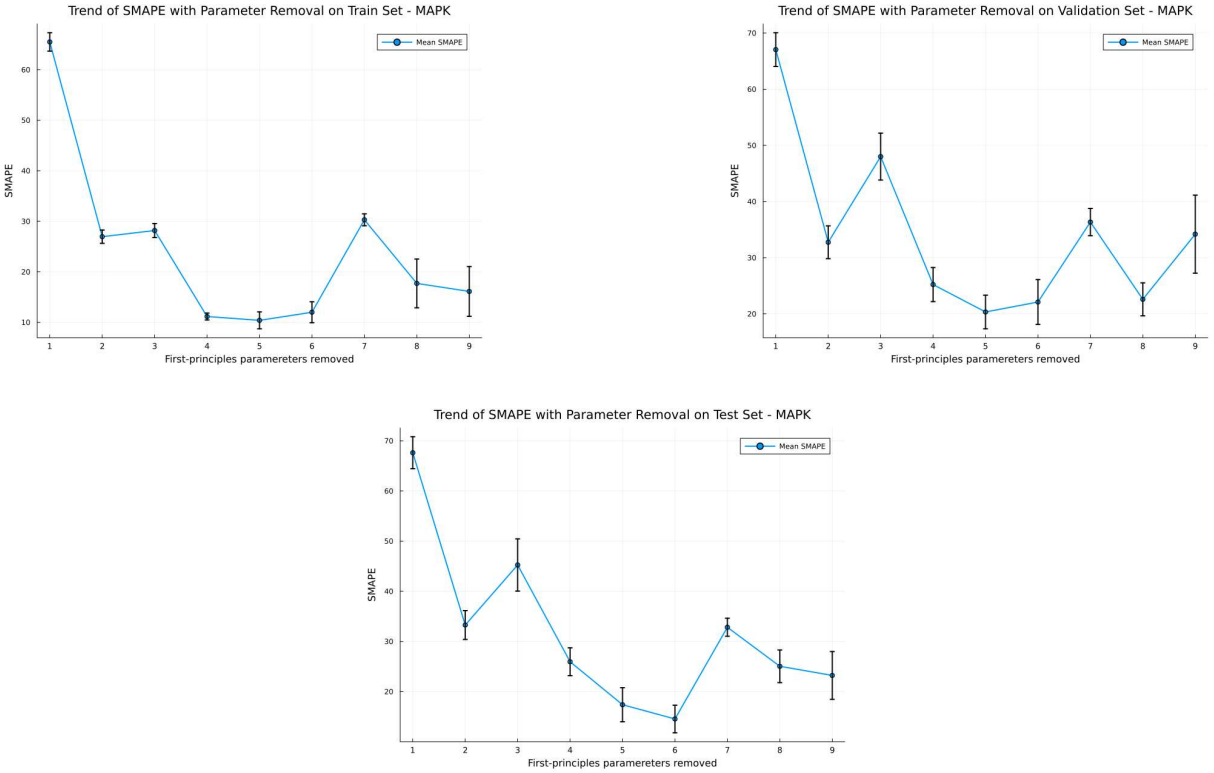


Figure 5.22: Impact of first-principle parameter removal on mean SMAPE for the MAPK orderedElementary. It is depicted here the effect on the mean SMAPE of iteratively removing sets of first-principle parameters, using 10 observations for training set, 10 for validation set, and 10 for test set. The y-axis denotes the mean SMAPE of the respective dataset, and the x-axis lists the sets of parameters removed(the full list of the sequence of removed rate constants is available in Table A.1. The error bars represent the standard deviation.

showed mixed results. The UDE integrated with linear regression recorded SMAPE scores of 30.60, 33.75, and 35.58 across the training, validation, and test datasets. Meanwhile, the UDE model incorporating a neural network achieved better results, with SMAPE scores of 21.46, 21.24, and 28.77, indicating superior performance over the linear regression UDE model (Figure 5.23).

To evaluate the predictive capabilities of the UDE-based models, Figure 5.24 displays the best and worst predictions for the UDE integrated with a neural network trained on 10 observations. The best prediction recorded a SMAPE score of 15.04, offering a reasonable approximation of the actual dynamics. Conversely, the worst prediction resulted in a SMAPE score of 73.06, with outputs that diverged significantly from the expected dynamics and even produced biologically implausible negative states.

Figure 5.25 illustrates the impact of parameter removal and estimation on the mean SMAPE. The model demonstrated sensitivity to the removal of specific parameter sets. Remarkably, similar to the experiments without noise, removing the parameter k_3 and estimating it along with the neural network weights yielded better performance compared to scenarios where all first-principle parameters were known and only the neural network weights were adjusted.

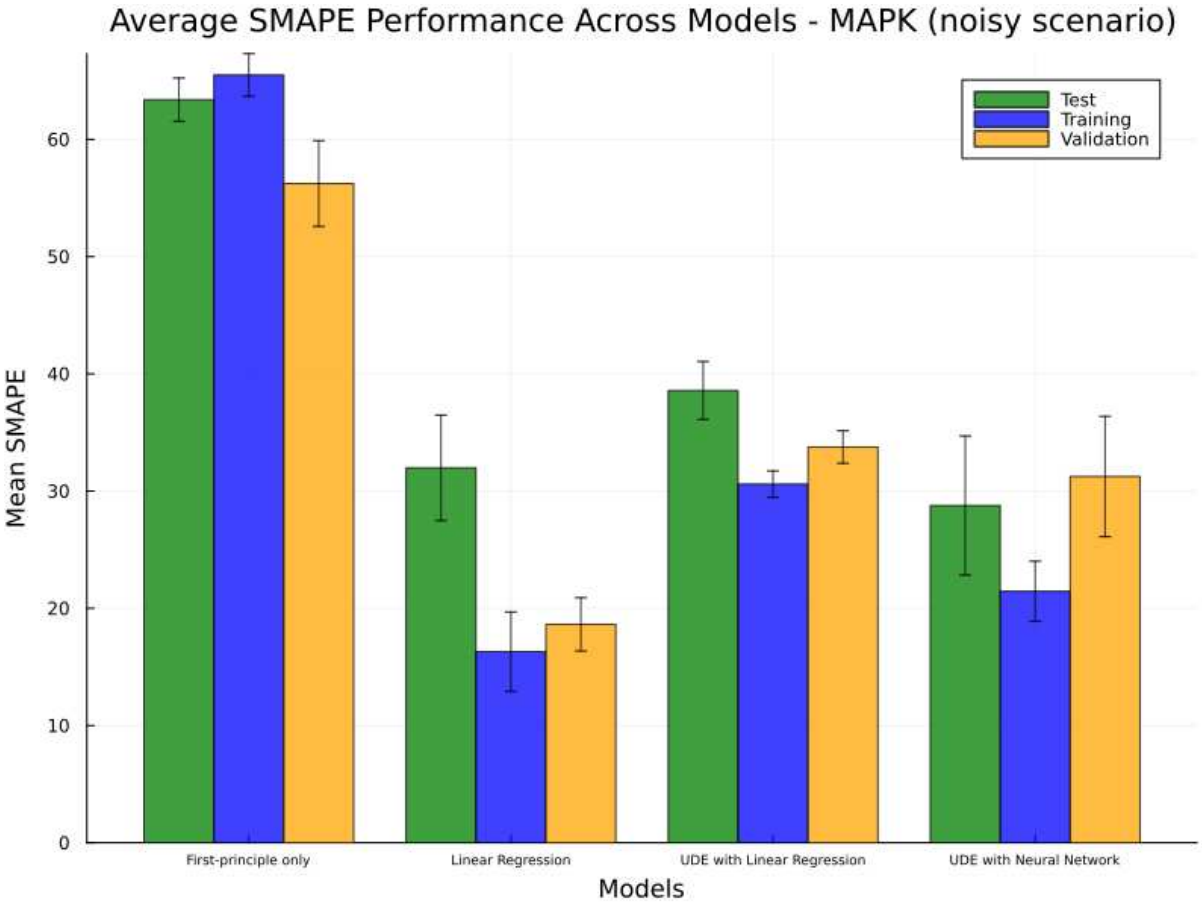


Figure 5.23: Model performance across noisy datasets for the MAPK orderedELementar.This graph compares different modeling approaches under noisy conditions by showcasing the SMAPE values across training, validation, and test datasets, each with 10 observations. The y-axis displays the mean SMAPE for noisy data, while the x-axis categorizes the models.The error bars represent the standard deviation.

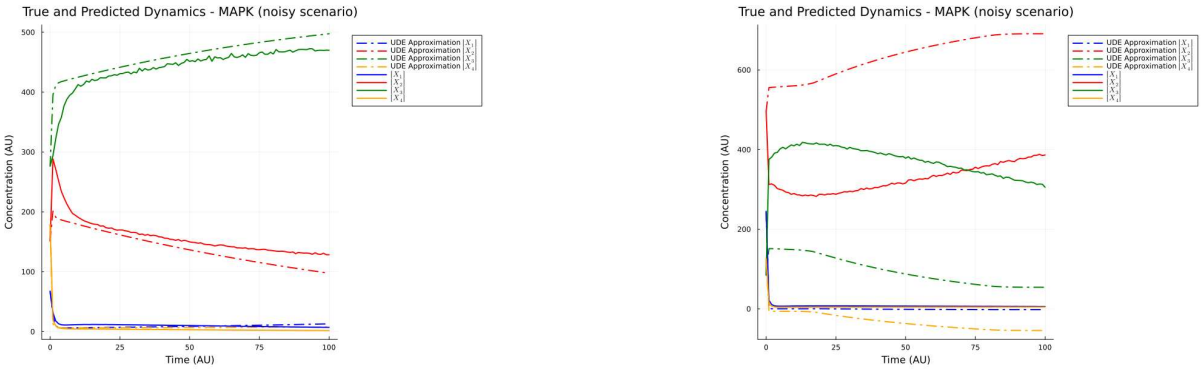


Figure 5.24: Predictive accuracy of the UDE with neural network in noisy conditions for the MAPK orderedELementay (10 observations). The best (left) and worst (right) predictions from the UDE model trained with 10 observations under noisy conditions, this figure contrasts the model predictions (dashed lines) against the true dynamics (solid lines) of the MAPK orderedELementay.

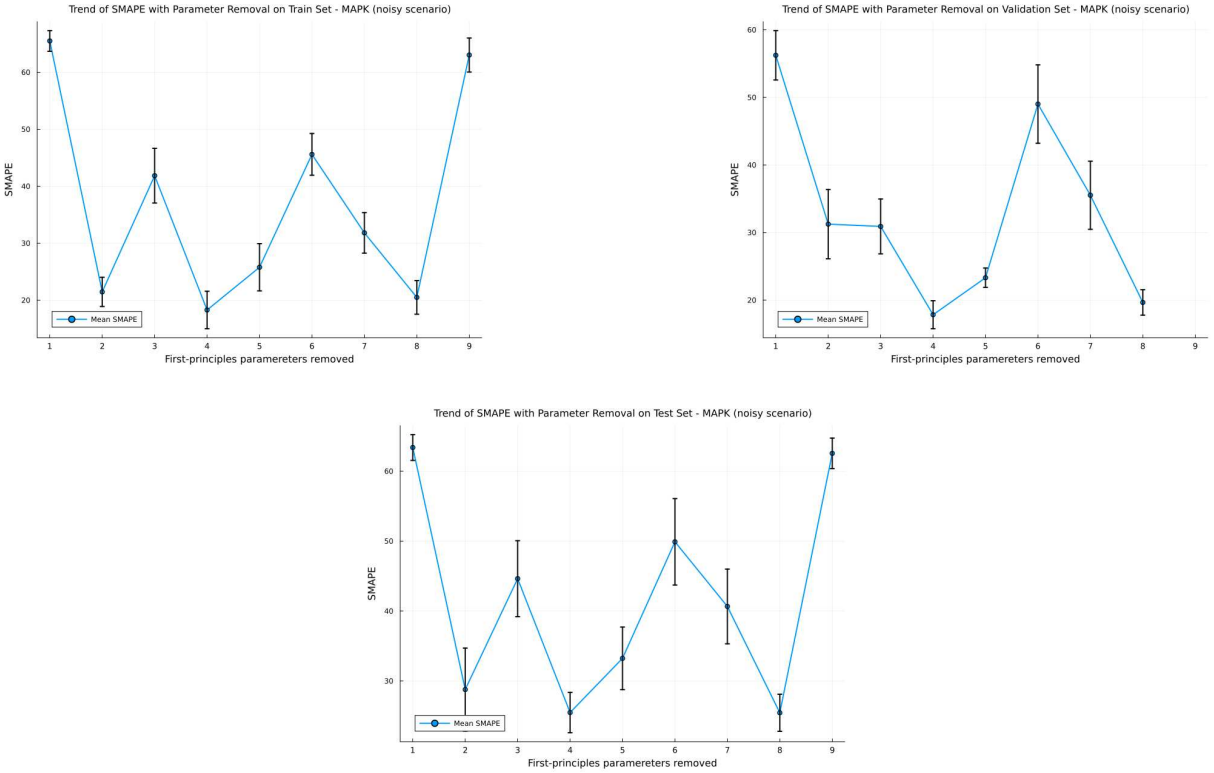


Figure 5.25: Impact of first-principle parameter removal on mean SMAPE for the MAPK orderedElementary, in noisy conditions. It is depicted here the effect on the mean SMAPE of iteratively removing sets of first-principle parameters, using 10 noisy observations for training set, 10 for validation set, and 10 for test set. The y-axis denotes the mean SMAPE of the respective dataset, and the x-axis lists the sets of parameters removed (the full list of the sequence of removed rate constants is available in Table A.2). The error bars represent the standard deviation.

5.4 Results with FeMetabolism FeDeficient

Our initial assessment of the FeMetabolism and FeDeficient pathway modeling involved training four distinct models, each utilizing a dataset of 10 observations for the training set, with matching numbers for validation and test sets. The first-principle only model achieved mean SMAPE of 81.84, 82.13, and 81.47 across the training, validation, and test sets, respectively. In contrast, the purely linear regression model, serving as our baseline, demonstrated significantly better performance with mean SMAPE of 6.70, 7.44, and 7.68 for the respective datasets.

Enhancements were further observed with the UDE models. The UDE integrated with linear regression outperformed other models, achieving a mean SMAPE of 6.22, 6.37, and 6.50 on the training, validation, and test sets, respectively. Meanwhile, the UDE model employing a neural network architecture yielded a mean SMAPE of 8.15, 8.92, and 8.95, indicating a modest decrease in performance compared to the linear regression-based UDE but still a significant improvement over the first-principle model. These comparative performance metrics are visually represented in Figure 5.26.

To explore the scalability and robustness of the UDE-based hybrid model, further

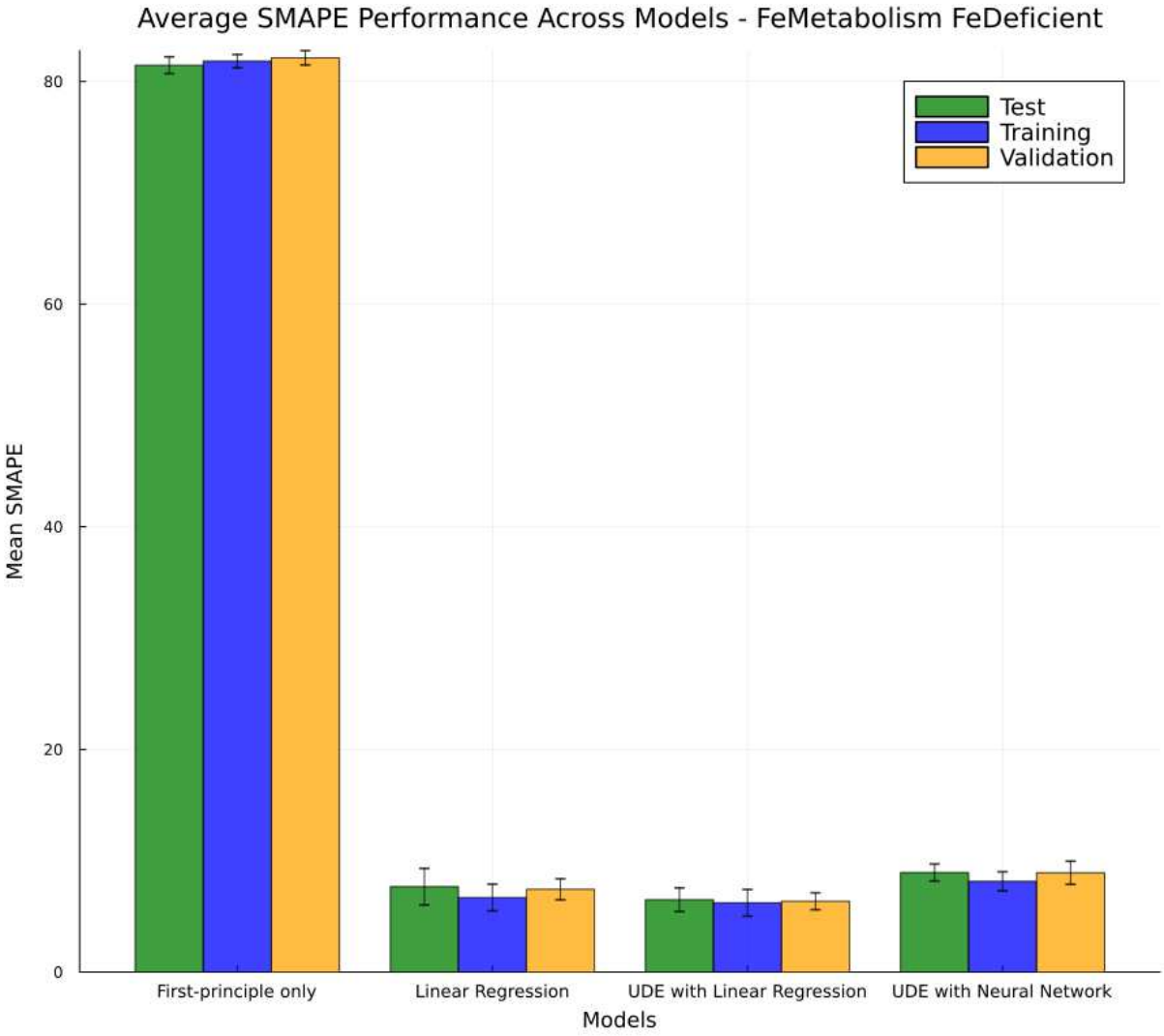


Figure 5.26: Model performance for the FeMetabolism FeDeficient Model, where modeling approaches are compared based on the SMAPE for training, validation, and test sets (each set containing 10 observations). The y-axis represents the mean SMAPE, while the x-axis identifies the model type. The error bars represent the standard deviation.

training sessions were conducted with increased dataset sizes. Training the UDE integrated with linear regression on 100 observations resulted in mean SMAPE scores of 15.51, 15.40, and 15.14, suggesting a deterioration in model generalization as the dataset expanded. Subsequent training with 1000 observations showed an improvement in generalization, with SMAPE scores reducing to 7.28, 7.48, and 7.33 for the training, validation, and test sets, respectively. This indicates an enhanced performance relative to the model trained with 100 observations but still not reaching the efficacy of the initial 10-observation training set (Figure 5.27).

The best prediction of the UDE integrate with linear regression, and trained with 10 observations, achieved a SMAPE score of 3.53, while the poorest prediction recorded a score of 15.53. Figure 5.28 showcases a comparative analysis between the true dynamics and both the best and worst model predictions. Notably, even the least accurate predictions of the model closely followed the true dynamics, although minor discrepancies were

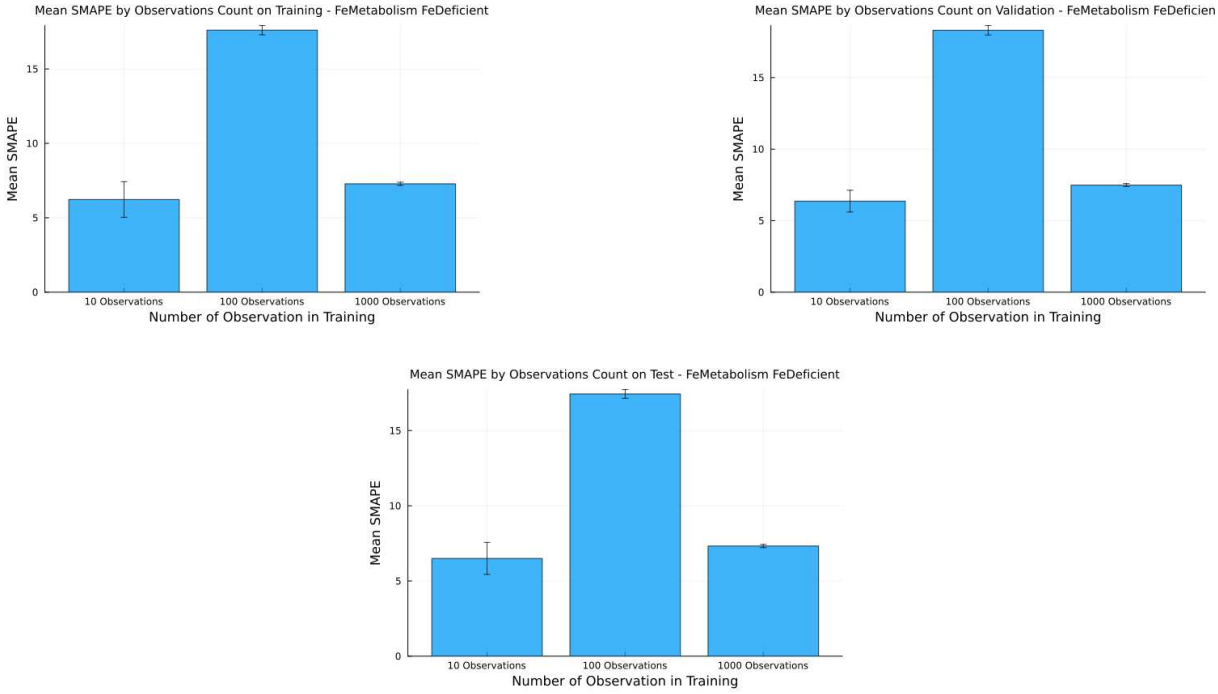


Figure 5.27: Data expansion impact on UDE with linear regression for the FeMetabolism FeDeficient Model. Increasing the number of training observations affects the mean SMAPE for the UDE model integrated with a neural network. The x-axis indicates the number of observations, and the y-axis shows the mean SMAPE across corresponding datasets. The error bars represent the standard deviation.

observed in accurately capturing the dynamics of chemical species x_2.

Upon increasing the training dataset to 100 observations, the range of prediction accuracy widened. The most accurate prediction yielded a SMAPE of 14.98, whereas the least accurate was 27.95. Figure 5.29 displays how the model predictions align with

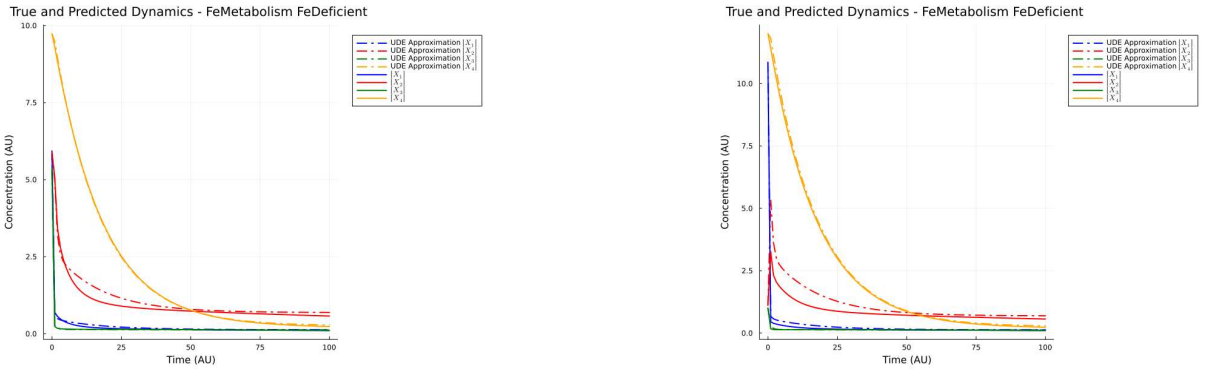


Figure 5.28: Predictive accuracy of the UDE with linear regression for the FeMetabolism FeDeficient model (10 observations): The best (left) and worst (right) predictions from the UDE model trained with 10 observations against the true dynamics of the FeMetabolism FeDeficient model. Solid lines denote true dynamics, while dashed lines represent model predictions.

the true dynamics of each chemical species. Despite a decrease in SMAPE performance compared to the smaller dataset, the model maintained a credible approximation of the true dynamics, with only minor difficulties in approximating the dynamics of species x_2.

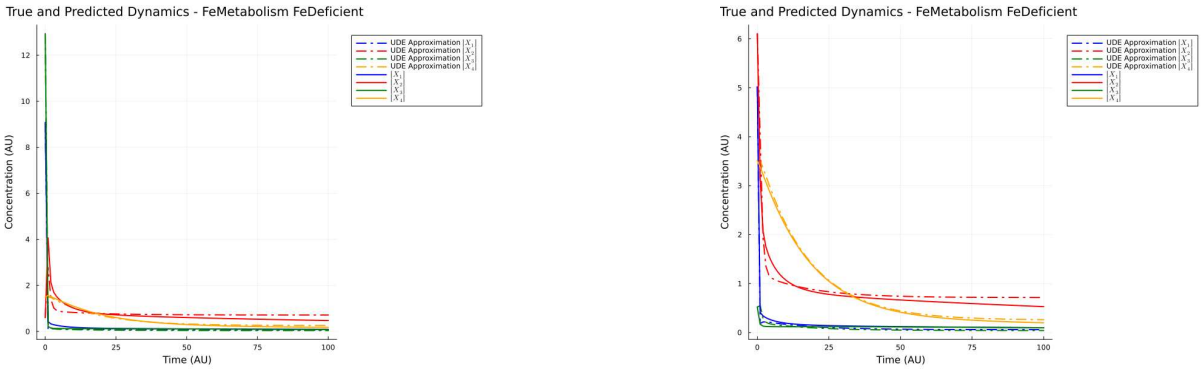


Figure 5.29: Predictive accuracy of the UDE with linear regression for the FeMetabolism FeDeficient model (100 observations): The best (left) and worst (right) predictions from the UDE model trained with 100 observations against the true dynamics of the FeMetabolism FeDeficient model. Solid lines denote true dynamics, while dashed lines represent model predictions.

Expanding the dataset further to 1000 observations, the performance of the model improved significantly for the best predictions but deteriorated for the worst-case scenarios. The best prediction recorded a SMAPE of 3.25, while the worst soared to 35.86. This dichotomy highlights an intriguing aspect of model training scalability; while the average SMAPE across all test observations improved, as visualized in Figure 5.27, the worst-case SMAPE of the predictions worsened compared to that of the model trained with 100 observations. Figure 5.30 contrasts these best and worst predictions, demonstrating that while the best prediction accurately mirrors the dynamics across all chemical species, the worst prediction struggles particularly with species x_2 and x_4.

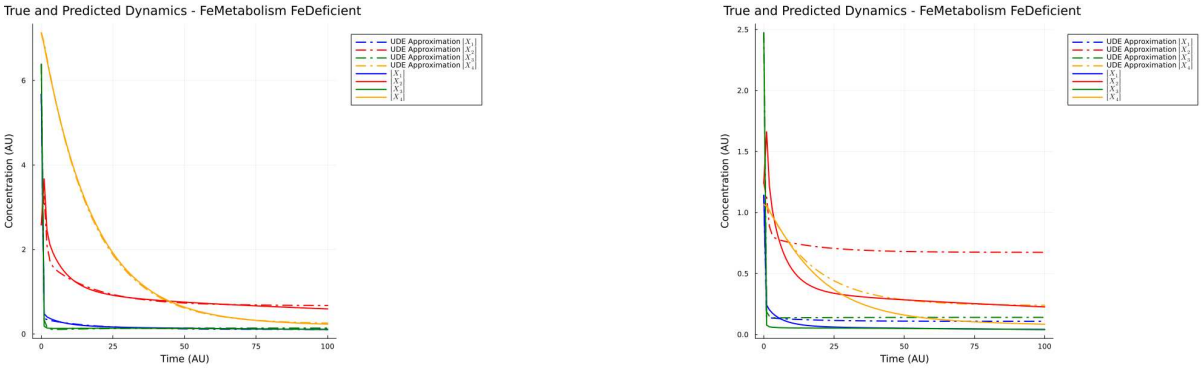


Figure 5.30: Predictive accuracy of the UDE with linear regression for the FeMetabolism FeDeficient model (1000 observations): The best (left) and worst (right) predictions from the UDE model trained with 1000 observations against the true dynamics of the FeMetabolism FeDeficient model. Solid lines denote true dynamics, while dashed lines represent model predictions.

Figure 5.31 demonstrates the effects of removing and then estimating first-principle parameters while concurrently adjusting the weights of the linear regression model, using datasets comprising 10 observations for training, validation, and testing. The analysis reveals that the sensitivity of the model to parameter removal varies; specifically, removing the following sets of first-principle parameters: $\{k_4\}, \{k_{17}\}$, and $\{k_{17}, k_6\}$. When these parameters are excluded and their values are estimated alongside the regression weights, the resulting mean SMAPE deteriorates compared to scenarios where all first-principle parameters are held constant. However, despite this decrease in performance, the SMAPE remains markedly better than that of the model relying solely on first-principle model, with the notable exception occurring when all first-principle parameters are removed.

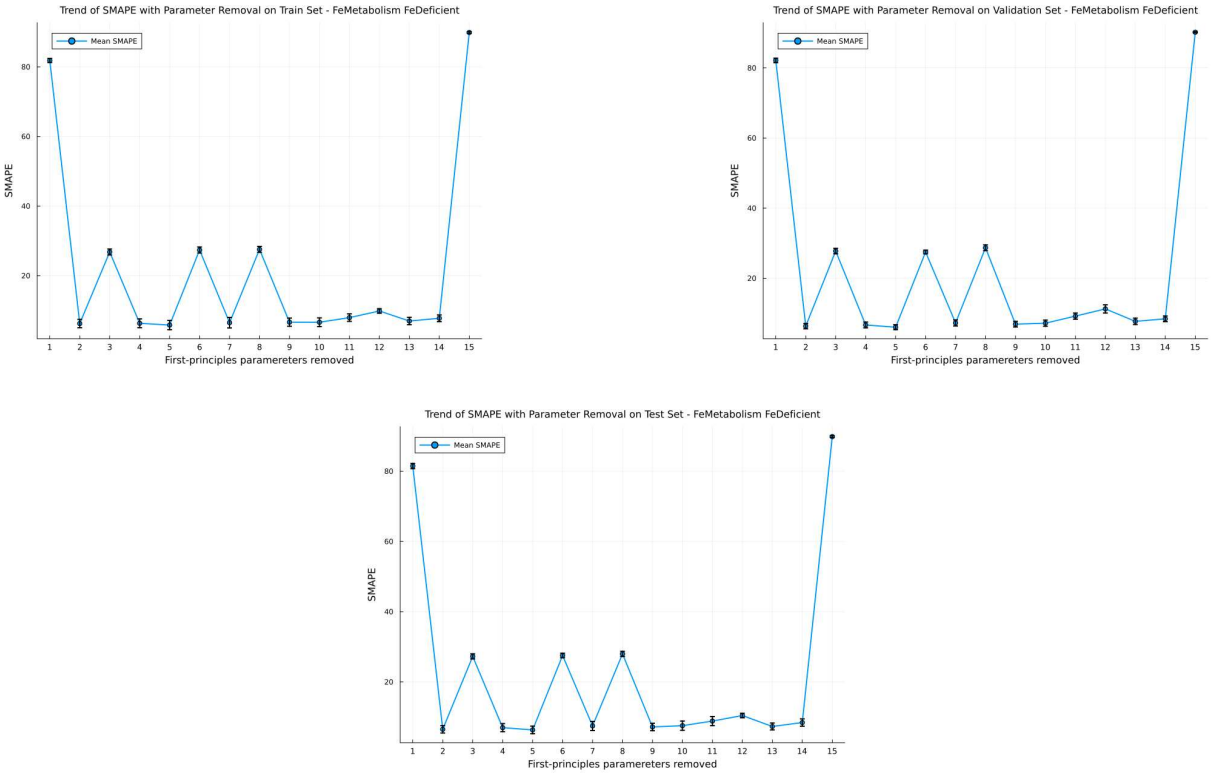


Figure 5.31: Impact of first-principle parameter removal on mean SMAPE for the FeDeficient FeMetabolism model. It is depicted here the effect on the mean SMAPE of iteratively removing sets of first-principle parameters, using 10 noisy observations for training set, 10 for validation set, and 10 for test set. The y-axis denotes the mean SMAPE of the respective dataset, and the x-axis lists the sets of parameters removed (the full list of the sequence of removed rate constants is available in Table A.3). The error bars represent the standard deviation.

5.4.1 Experimentation with noisy data for the FeDeficient and FeMetabolism pathway model

The first-principle model, under noisy conditions, recorded a mean SMAPE of 81.84, 80.97, and 79.63 for the training, validation, and test sets respectively. In comparison,

the purely linear regression model demonstrated substantially improved performance with mean SMAPE of 6.14, 7.43, and 7.09, significantly surpassing the first-principle model.

The UDE-based models, while not matching the performance of the purely linear regression, still showed considerable resilience to noise. The UDE integrated with linear regression achieved SMAPE scores of 6.91, 8.06, and 8.02 across the training, validation, and test sets, respectively. Similarly, the UDE model using a neural network architecture recorded a mean SMAPE of 8.03, 9.46, and 8.35. These comparisons are visually represented in Figure 5.32.

Further analysis focused on the UDE model integrated with linear regression to evaluate its predictive capabilities under noisy conditions. The best prediction from this model attained a SMAPE of 4.43, while the poorest was at 19.50. Figure 5.33 shows that even in the worst-case scenario, this model maintained a reasonable approximation of the true dynamics across all chemical species, illustrating its robustness.

Lastly, Figure 5.34 details the impact of selectively removing and then estimating sets of first-principle parameters alongside the model weights under noisy data conditions. Interestingly, our approach yielded better performance with noisy data than with clean data. Although the removal of parameters $\{k_{17}, k_{18}, k_3\}$ resulted in poorer outcomes, the overall mean SMAPE was still markedly improved compared to using the first-principle model alone. The stability of other parameter sets remained consistent even when compared to scenarios where all first-principle parameters were assumed known.

5.5 Discussion

This section provides a discussion of the results presented in the previous sections, focusing on the performance of UDE-based models compared to traditional first-principle models. The data clearly indicate that all UDE-based models significantly outperform the first-principle models alone, suggesting that this hybrid approach is promising for modeling complex biological systems. Across various scenarios, the UDE-based models effectively captured the true dynamics of several observations. However, there were challenges in accurately modeling certain chemical species, particularly in the worst-case scenarios. An exception to this was observed in the FeMetabolism FeDeficient pathway model, which consistently performed well across all tests.

A critical observation from our experiments is the dependency of model performance on data quality. Since the synthetic data was generated randomly, regions of the observation domain of each case study were adequately represented in the training set, leading to discrepancies in model output, especially in the worst-case scenarios.

Expanding the dataset did not uniformly enhance the performance of the UDE-based models. This might be attributed to the random nature of the data expansion, which sometimes failed to introduce necessary diversity in the curve types. In certain instances, even though the expanded dataset generally improved the mean SMAPE on the test set, the worst-case predictions were poorer than those obtained from the first-principle model alone.

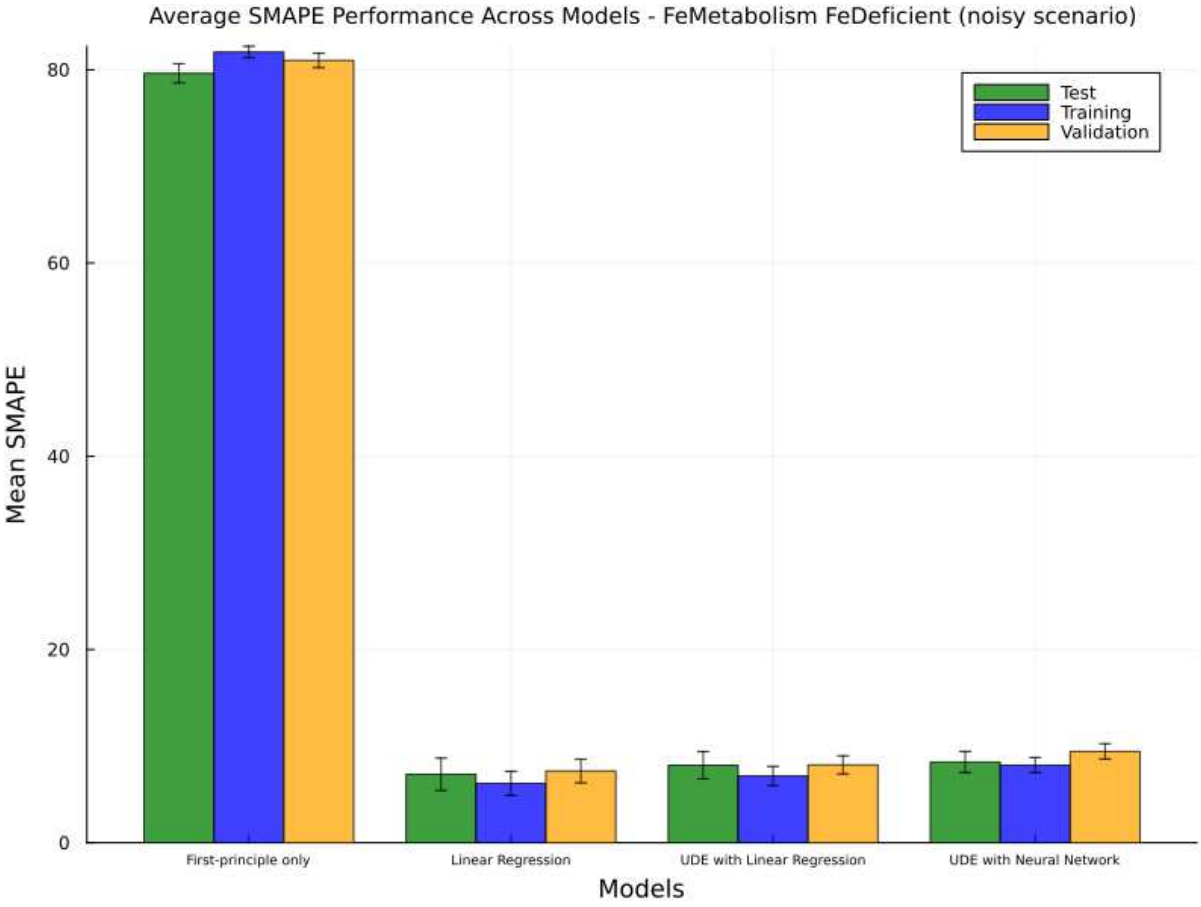


Figure 5.32: Model performance across noisy datasets for the FeMetabolism FeDeficient Model. This graph compares different modeling approaches under noisy conditions by showcasing the SMAPE values across training, validation, and test datasets, each with 10 observations. The y-axis displays the mean SMAPE for noisy data, while the x-axis categorizes the models. The error bars represent the standard deviation.

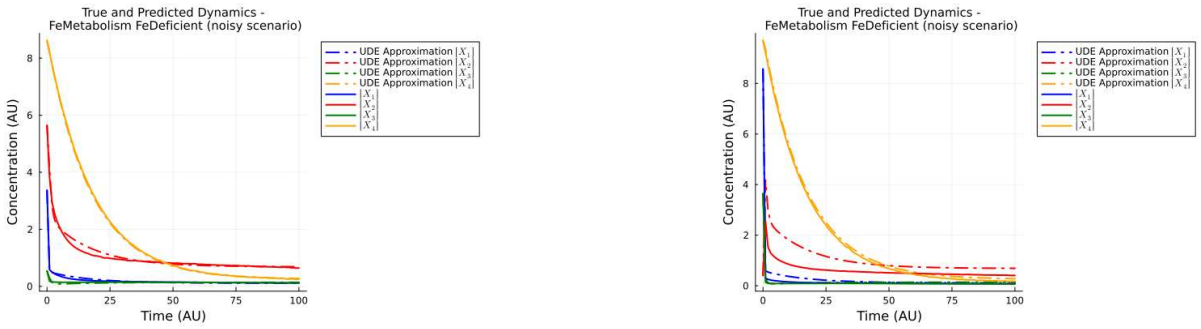


Figure 5.33: Predictive accuracy of the UDE with linear regression in noisy conditions for the JFeMetabolism FeDeficient model (10 observations). The best (left) and worst (right) predictions from the UDE model trained with 10 observations under noisy conditions, this figure contrasts the model predictions (dashed lines) against the true dynamics (solid lines) of the JFeMetabolism FeDeficient model.

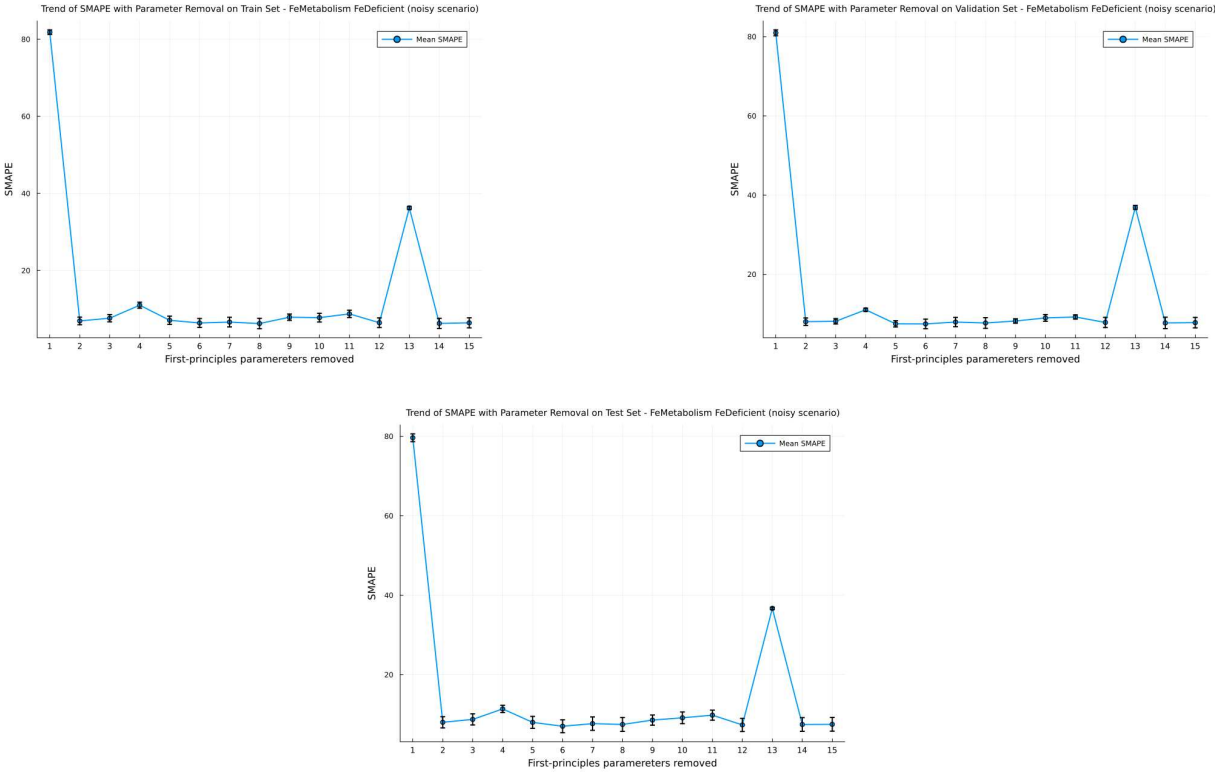


Figure 5.34: Impact of first-principle parameter removal on mean SMAPE for the FeMetabolism FeDeficient model, in noisy conditions. It is depicted here the effect on the mean SMAPE of iteratively removing sets of first-principle parameters, using 10 noisy observations for training set, 10 for validation set, and 10 for test set. The y-axis denotes the mean SMAPE of the respective dataset, and the x-axis lists the sets of parameters removed (the full list of the sequence of removed rate constants is available in Table A.3). The error bars represent the standard deviation.

The approach of jointly inferring first-principle parameters alongside neural network weights holds significant promise. However, one limitation is the identifiability issue with first-principle parameters. In many cases, the results from jointly estimating a removed set of first-principle parameters with neural network weights were comparable to, or even better than, those where all first-principle parameters were assumed known. Nevertheless, it is important to note that in models sensitive to specific first-principle parameters, such as the Toy Model, our approach faced challenges in maintaining performance.

Moreover, the robustness of our approach was tested with noisy data (a necessity to evaluate the feasibility of application of this approach in real-world data), which appeared to affect the model outcomes minimally, suggesting that the methodology is resilient to data quality issues, preserving performance even under less ideal conditions.

Chapter 6

Conclusion

This work provided a way of integrating hybrid modeling techniques in the study of cell signaling pathways, particularly when one or more first-principle parameters are unknown. Our findings indicate that our hybrid model significantly outperforms traditional first-principle models, showcasing enhanced predictive accuracy. However, the process of jointly inferring first-principle parameters and neural network weights, while promising, has revealed complexities that warrant further investigation. Specifically, challenges related to numerical instability and suboptimal estimations of first-principle parameters, despite achieving satisfactory SMAPE values, need to be addressed in future studies.

6.1 Answers to the research questions

After carefully analyzing the results yielded by our methodology, we can finally answer the research questions made in Section 1.1:

RQ 1 *How well can we identify hybrid models of cell signaling pathways by employing time series data of species present in the first-principle part of the model?*

Answer: The performance of our UDE-based model in identifying cell signaling pathways varies significantly across different chemical species. For certain species, the model accurately captures the interactions with external proteins, exhibiting a high degree of generalization. However, for other species, the model faces challenges in accurately modeling these interactions, indicating variability in model performance across the spectrum of species within the system.

RQ 2 *How can we efficiently infer the parameters of the hybrid model, encompassing both the mechanistic (first-principle) and data-driven components?*

Answer: The efficacy of parameter inference in our hybrid model is notably influenced by the specific set of first-principle parameters being removed and estimated. Some parameter sets when removed lead to inferior model performance, even poorer than using the first-principle model alone. Conversely, other parameter removals result in stable performance, and some even enhance model accuracy compared to scenarios where all first-principle parameters are presumed known. This variability underscores the sensitivity of our model to the specific parameters being adjusted.

In conclusion, our investigation into the use of UDE-based hybrid models for cell signaling pathways has yielded nuanced findings. The effectiveness of the hybrid model in capturing the dynamics of cellular pathways varies significantly across different chemical species. Specifically, the model excels in accurately modeling interactions for certain species, showing a high degree of generalization and predictive accuracy. This success depends on the model's ability to integrate interactions with external proteins effectively. However, it encounters difficulties with other species, where the performance inconsistency is largely attributable to the quality and diversity of the training data available. The variations in data quality, especially the representativeness and diversity of curve types for different species within the training set, significantly impact the model's ability to generalize effectively across all species.

Additionally, the process of inferring parameters for these hybrid models is highly sensitive to the selection of first-principle parameters removed for estimation. This aspect of the methodology can either degrade or enhance the model's performance, demonstrating a critical dependency on the nature and role of the specific parameters adjusted during the model training process. These findings underscore the complexity and nuanced challenges faced in the effective development of hybrid modeling in systems biology, guiding future research towards improving model robustness and parameter estimation methods.

6.2 List of publications and other contributions

Besides the main scientific contribution of this dissertation, additional contributions have been made, which are listed below:

1. Our research on the inference of UDE-based hybrid models with fixed first-principle parameters was presented at the Brazilian Symposium on Bioinformatics (BSB) in 2023. This work was subsequently published in the proceedings of the event [32];

2. The methodology for joint inference of first-principle parameters and neural network weights was showcased through a poster presentation at BSB 2023;

3. The source code for all experiments conducted as part of this study is publicly available at the following GitHub repository: `cristiano-campos/UDE-based_hybrid_modeling.com`.

6.3 Suggestions for future research

Based on the current challenges in jointly infer first-principle parameter and neural network weights, future research should explore the potential of multiple shooting techniques as a means to enhance the predictive accuracy of the model and address the issue of numerical instability observed in some scenarios. Additionally, the investigation of the LDA loss function as a novel optimization strategy offers a promising avenue [36]. This approach, focusing on comparing angles and lengths of vectors within the state space independently, could provide a more nuanced understanding of model performance and accuracy. These

suggestions aim to build upon the foundational work presented in this dissertation, encouraging further exploration and development in the field of hybrid modeling.

Bibliography

[1] Faizan M. S. Bangi, Katy Kao, and Joseph S. Kwon. Physics-informed neural networks for hybrid modeling of lab-scale batch fermentation for beta-carotene production using saccharomyces cerevisiae. *Chemical Engineering Research and Design*, 179:415–423, 2022. DOI: 10.1016/j.cherd.2022.01.041.

[2] Dexter Barrows and Silvana Ilie. Parameter estimation for the reaction–diffusion master equation. *AIP Advances*, 13(6):065318, 06 2023. DOI: 10.1063/5.0150292.

[3] Vijaysekhar Chellaboina, Sanjay P. Bhat, Wassim M. Haddad, and Dennis S. Bernstein. Modeling and analysis of mass-action kinetics. *IEEE Control Systems Magazine*, 29(4):60–78, 2009. DOI: 10.1109/MCS.2009.932926.

[4] Shijian Dong, Yuzhu Zhang, and Xingxing Zhou. Intelligent hybrid modeling of complex leaching system based on lstm neural network. *Systems*, 11(2), 2023. DOI: 10.3390/systems11020078.

[5] Attila Gabor, Marco Tognetti, Alice Driessen, Jovan Tanevski, Baosen Guo, Wencai Cao, He Shen, Thomas Yu, Verena Chung, Single Cell Signaling in Breast Cancer DREAM Consortium members, et al. Cell-to-cell and type-to-type heterogeneity of signaling networks: insights from the crowd. *Molecular Systems Biology*, 17(10):e10402, 2021. DOI: 10.15252/msb.202110402.

[6] Manvel Gasparyan and Shodhan Rao. Parameter estimation for kinetic models of chemical reaction networks from partial experimental data of species' concentrations. *Bioengineering*, 10(9), 2023. DOI: 10.3390/bioengineering10091056.

[7] Amanda S. Guimarães. An algorithm to simplify systems of differential equations that describe the kinetics of chemical reactions. Master's thesis, University of São Paulo, 2016. DOI: 10.11606/D.45.2016.tde-23082016-170051.

[8] Albin Heimerson and Johan Ruuskanen. Extending microservice model validity using universal differential equations. *IFAC-PapersOnLine*, 56(2):2401–2406, 2023. DOI: 10.1016/j.ifacol.2023.10.1214.

[9] Marta R. Hidalgo, Alicia Amadoz, Cankut Çubuk, José Carbonell-Caballero, and Joaquín Dopazo. Models of cell signaling uncover molecular mechanisms of high-risk neuroblastoma and predict disease outcome. *Biology Direct*, 13(1):16, 2018. DOI: 10.1186/s13062-018-0219-4.

[10] Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, and Jonathan Taylor. *An introduction to statistical learning with applications in python*. Springer Nature, 2023.

[11] Jin Deok Joo. The use of intra-cellular signaling pathways in anesthesiology and pain medicine field. *Korean Journal of Anesthesiology*, 57(3):277–283, 2009. DOI: 10.4097/kjae.2009.57.3.277.

[12] James Koch, Zhao Chen, Aaron Tuor, Ján Drgona, and Vrabie Vrabie. Structural inference of networked dynamical systems with universal differential equations. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 33(2):023103, 02 2023. DOI: 10.1063/5.0109093.

[13] Bruce Kuwahara and Chris T. Bauch. Predicting covid-19 pandemic waves with biologically and behaviorally informed universal differential equations. *Heliyon*, 10(4):e25363, 2024. DOI: 10.1016/j.heliyon.2024.e25363.

[14] Leo Lagunes. W10: Mathematical modeling of cell signaling. `https://www.youtube.com/watch?v=GuKZF2sMK7U`, 2022. Accessed: 2024-02-05.

[15] Dongheon Lee, Arul Jayaraman, and Joseph S. Kwon. Development of a hybrid model for a partially known intracellular signaling pathway through correction term estimation and neural network modeling. *PLOS Computational Biology*, 16(12):1–31, 2020. DOI: 10.1371/journal.pcbi.1008472.

[16] Kaili Li, Haoran Duan, Linfeng Liu, Ruihong Qiu, Ben van den Akker, Bing-Jie Ni, Tong Chen, Hongzhi Yin, Zhiguo Yuan, and Liu Ye. An integrated first principal and deep learning approach for modeling nitrous oxide emissions from wastewater treatment plants. *Environmental Science and Technology*, 56(4):2816–2826, 2022. DOI: 10.1021/acs.est.1c05020.

[17] Fernando R.D. Lima, Carine M. Rebello, Erbet A. Costa, Vinícius V. Santana, Marcellus G.F. de Moares, Amaro G. Barreto, Argimiro R. Secchi, Maurício B. de Souza, and Idelfonso B.R. Nogueira. Improved modeling of crystallization processes by universal differential equations. *Chemical Engineering Research and Design*, 200:538–549, 2023. DOI: 10.1016/j.cherd.2023.11.032.

[18] Tiago J. Lopes, Tatyana Luganskaja, Maja S. Vujić, Matthias W. Hentze, Martina U. Muckenthaler, Klaus Schümann, and Jens G. Reich. Systems analysis of iron metabolism: the network of iron pools and fluxes. *BMC Systems Biology*, 2010. DOI: 10.1186/1752-0509-4-112.

[19] Zhihao Lou and John Reinitz. Parallel simulated annealing using an adaptive resampling interval. *Parallel Computing*, 53:23–31, 2016. DOI: 10.1016/j.parco.2016.02.001.

[20] Christoph Lüders, Thomas Sturm, and Ovidiu Radulescu. ODEbase: A repository of ODE systems for systems biology. *Bioinformatics Advances*, 2(1), 2022. DOI: 10.1093/bioadv/vbac027.

[21] Arnout V. Messem Manvel Gasparyan and Shodhan Rao. Parameter estimation for models of chemical reaction networks from experimental data of reaction rates. *International Journal of Control*, 96(2):392–407, 2023. DOI: 10.1080/00207179.2021.1998636.

[22] Nick I. Markevich, Jan B. Hoek, and Boris N. Kholodenko. Signaling switches and bistability arising from multisite phosphorylation in protein kinase cascades . *Journal of Cell Biology*, 164(3):353–359, 2004. DOI: 10.1083/jcb.200308060.

[23] André T. G. Mello. A proposal of new methods for parameter estimation in ordinary differential equations. Master's thesis, University of São Paulo, 2020. DOI: 10.11606/D.45.2020.tde-10112020-091717.

[24] Paul J. Myers, Sung H. Lee, and Matthew J. Lazzara. Mechanistic and data-driven models of cell signaling: Tools for fundamental discovery and rational design of therapy. *Current Opinion in Systems Biology*, 28:100349, 2021. DOI: 10.1016/j.coisb.2021.05.010.

[25] Harini Narayanan, Mariano N. C. Bournazou, Gonzalo G. Gosálbez, and Alessandro Butté. Functional-hybrid modeling through automated adaptive symbolic regression for interpretable mathematical expressions. *Chemical Engineering Journal*, 430:133032, 2022. DOI: 10.1016/j.cej.2021.133032.

[26] Matthew J. Page, Joanne E. McKenzie, Patrick M. Bossuyt, Isabelle Boutron, Tammy C. Hoffmann, Cynthia D. Mulrow, Larissa Shamseer, Jennifer M. Tetzlaff, Elie A. Akl, Sue E. Brennan, Roger Chou, Julie Glanville, Jeremy M. Grimshaw, Asbjørn Hróbjartsson, Manoj M. Lalu, Tianjing Li, Elizabeth W. Loder, Evan Mayo-Wilson, Steve McDonald, Luke A. McGuinness, Lesley A. Stewart, James Thomas, Andrea C. Tricco, Vivian A. Welch, Penny Whiting, and David Moher. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372, 2021. DOI: 10.1136/bmj.n71.

[27] Christopher Rackauckas, Yingbo Ma, Julius Martensen, Collin Warner, Kirill Zubov, Rohit Supekar, Dominic Skinner, Ali Ramadhan, and Alan Edelman. Universal differential equations for scientific machine learning, 2021. DOI: 10.48550/arXiv.2001.04385.

[28] Padmini Rangamani and Ravi Iyengar. Modelling cellular signalling systems. *Essays in Biochemestry*, 45:83–94, 2008. DOI: 10.1042/bse0450083.

[29] Marcelo S. Reis and Juliane Liepe. An approximate Bayesian computation-based approach to tackle the lack of isolation problem in signaling pathway modeling. Technical report, Max Planck Institute, 2020.

[30] Marcelo S. Reis, Vincent Noël, Matheus H.S. Dias, Layra L. Albuquerque, Amanda S. Guimarães, Lulu Wu, Junior. Barrera, and Hugo A. Armelin. An interdisciplinary approach for designing kinetic models of the Ras/MAPK signaling pathway. In

Methods in Molecular Biology Special Edition on Kinase Signaling Networks, pages 455–474. Humana Press, New York, 2017. DOI: 10.1007/978-1-4939-7154-1_28.

[31] Vinicius V. Santana, Erbet Costa, Carine M. Rebello, Ana M. Ribeiro, Christopher Rackauckas, and Idelfonso B.R. Nogueira. Efficient hybrid modeling and sorption model discovery for non-linear advection-diffusion-sorption systems: A systematic scientific machine learning approach. *Chemical Engineering Science*, 282:119223, 2023. DOI: 10.1016/j.ces.2023.119223.

[32] Ronaldo N. Sousa, Cristiano G. S. Campos, Willian Wang, Ronaldo F. Hashimoto, Hugo A. Armelin, and Marcelo S. Reis. Exploring identifiability in hybrid models of cell signaling pathways. In Marcelo S. Reis and Raquel C. de Melo-Minardi, editors, *Advances in Bioinformatics and Computational Biology*, pages 148–159, Cham, 2023. Springer Nature Switzerland. DOI: 10.1007/978-3-031-42715-2_14.

[33] Han Tang. Uncertain chemical reaction equation with delay. *Journal of Ambient Intelligence and Humanized Computing*, 14(4):3867–3874, 2023. DOI: 10.1007/s12652-022-04458-9.

[34] Allain V. Wouwer, Christine Renotte, and Philippe Bogaerts. Biological reaction modeling using radial basis function networks. *Computers and Chemical Engineering*, 28(11):2157–2164, 2004. DOI: 10.1016/j.compchemeng.2004.03.003.

[35] Peter J. M. van Laarhoven and Emile H. L. Aarts. *Simulated annealing*, pages 7–15. Springer Netherlands, Dordrecht, 1987. DOI: 978-94-015-7744-1_2.

[36] Pipa G. A Vortmeyer-Kley R, Nieters P. trajectory-based loss function to learn missing terms in bifurcating dynamical systems. *Scientific Reports*, 2021. DOI: 10.1080/00207179.2021.1998636.

[37] Ulrike Wittig, Renate Kania, Martin Golebiewski, Maja Rey, Lei Shi, Lenneke Jong, Enkhjargal Algaa, Andreas Weidemann, Heidrun Sauer-Danzwith, Saqib Mir, Olga Krebs, Meik Bittkowski, Elina Wetsch, Isabel Rojas, and Wolfgang Müller. SABIO-RK—database for biochemical reaction kinetics. *Nucleic Acids Research*, 40(D1):D790–D796, 11 2011. DOI: 10.1093/nar/gkr1046.

[38] Satoshi Yamada, Satoru Shiono, Akiko Joo, and Akihiko Yoshimura. Control mechanism of jak/stat signal transduction pathway. *FEBS Letters*, 534(1-3):190–196, 2003. DOI: 10.1016/S0014-5793(02)03842-5.

[39] Hans-Jörg Zander, Roland Dittmeyer, and Josef Wagenhuber. Dynamic modeling of chemical reaction systems with neural networks and hybrid models. *Chemical Engineering Technology*, 1999. DOI: 10.1002/(SICI)1521-4125(199907)22:7<571::AID-CEAT571>3.0.CO;2-5.

Appendix A

Lists of removed sets of rate constants

In this appendix, we present the lists of the specific sequences of first-principle parameter sets (rate constants) removal used in our experiments. These sequences illustrate the iterative removal of first-principle parameter sets and the subsequent estimation of parameters contained within those sets, alongside adjustments to neural network weights.

Index	Set of first-principles paremeters removed
1	First-principle alone (No removal)
2	UDE-based model (No removal)
3	k3
4	k4
5	k5
6	k3-k4
7	k3-k5
8	k4-k5
9	k3-k4-k5

Table A.1: Complete sequence of first-principle parameter removal sets for the MAPK orderedElementary model.

Index	Set of first-principles parameter removed
1	First-principle alone (No removal)
2	UDE-based model (No removal)
3	k4
4	k3
5	k5
6	k3-k4
7	k4-k5
8	k3-k5
9	k3-k4-k5

Table A.2: Complete sequence of first-principle parameter removal sets for the MAPK orderedElementary model using noisy data.

Index	Set of first-principles parameter removed
1	First-principle alone (No removal)
2	UDE-based model (No removal)
3	k6
4	k3
5	k18
6	k17
7	k18-k6
8	k17-k6
9	k3-k6
10	k17-k3
11	k17-k18-k6
12	k17-k18-k3
13	k18-k3-k6
14	k17-k3-k6
15	k17-k18-k3-k6

Table A.3: Complete sequence of first-principle parameter removal sets for the FeMetabolism FeDeficient model.

Index	Set of first-principles parameter removed
1	First-principle alone (no removal)
2	UDE-based model (no removal)
3	k6
4	k18
5	k17
6	k3
7	k18-k6
8	k3-k6
9	k17-k18
10	k18-k3
11	k17-k18-k6
12	k18-k3-k6
13	k17-k18-k3
14	k17-k3-k6
15	k17-k18-k3-k6

Table A.4: Complete sequence of first-principle parameter removal sets for the FeMetabolism FeDeficient model using noisy data.

index	Set of first-principles parameter removed
1	First-principle alone (no removal)
2	UDE-based model (no removal)
3	k59
4	k58
5	k50
6	k51
7	k53-k61
8	k51-k58
9	k51-k59
10	k58-k61
11	k58-k59
12	k53-k58-k59
13	k50-k51-k58
14	k51-k59-k61
15	k53-k59-k61
16	k58-k59-k61
17	k50-k53-k58-k61
18	k50-k53-k58-k59
19	k53-k58-k59-k61
20	k50-k51-k53-k59
21	k51-k53-k58-k61
22	k50-k51-k53-k58
23	k51-k53-k58-k59-k61
24	k50-k53-k58-k59-k61
25	k50-k51-k53-k58-k59
26	k50-k51-k53-k59-k61
27	k50-k51-k53-k58-k61
28	k50-k51-k53-k58-k59-k61

Table A.5: Complete sequence of first-principle parameter removal sets for the JAK-STAT model.

index	Set of first-principles parameter removed
1	First-principle alone (no removal)
2	UDE-based model (no removal)
3	k51
4	k50
5	k53
6	k61
7	k58
8	k51-k58
9	k59-k61
10	k58-k59
11	k51-k59
12	k53-k61
13	k53-k59
14	k50-k53-k59
15	k53-k58-k59
16	k51-k58-k61
17	k50-k53-k61
18	k51-k53-k59
19	k53-k58-k59-k61
20	k50-k51-k53-k59
21	k50-k51-k53-k58-k61
22	k50-k51-k53-k59-k61
23	k50-k51-k53-k58-k59
24	k50-k51-k53-k58-k59-k61

Table A.6: Complete sequence of first-principle parameter removal sets for the JAK-STAT model using noisy data.

Index	Set of first-principles parameter removed
1	First-principle alone (no removal)
2	UDE-based model (no removal)
3	kcat2
4	kf4
5	kcat4
6	kcat3-kr4
7	kf2-kr1
8	kr3-kr4
9	kcat1-kr2
10	kcat2-kf1-kr2
11	kcat4-kf1-kr1
12	kcat2-kcat3-kf2
13	kf1-kr1-kr2
14	kcat2-kf3-kr2-kr3
15	kcat4-kf2-kf4-kr3
16	kcat1-kf1-kf2-kr1
17	kcat1-kcat4-kf4-kr3
18	kcat2-kcat3-kf1-kf2-kr1
19	kcat3-kf1-kf4-kr2-kr3
20	kcat3-kf2-kf3-kr1-kr2
21	kcat4-kf2-kf3-kr2-kr3
22	kcat2-kcat4-kf1-kf2-kf3-kr1
23	kcat3-kf1-kf2-kf4-kr1-kr3
24	kcat1-kf1-kf2-kf3-kf4-kr3
25	kcat3-kf2-kf3-kf4-kr2-kr3
26	kcat1-kcat2-kcat4-kf4-kr1-kr3-kr4
27	kcat1-kcat2-kcat3-kr1-kr2-kr3-kr4
28	kcat1-kcat3-kcat4-kf1-kf3-kr1-kr3
29	kcat3-kcat4-kf1-kf2-kf4-kr1-kr3
30	kcat1-kcat3-kcat4-kf1-kf2-kf3-kf4-kr2
31	kcat2-kcat3-kcat4-kf2-kf3-kf4-kr1-kr3
32	kcat3-kcat4-kf3-kf4-kr1-kr2-kr3-kr4
33	kcat1-kcat2-kcat3-kcat4-kf1-kf4-kr2-kr3-kr4
34	kcat1-kcat2-kcat3-kcat4-kf2-kf3-kr1-kr2-kr4
35	kcat1-kcat2-kcat3-kcat4-kf1-kf2-kf4-kr2-kr3-kr4
36	kcat2-kcat3-kcat4-kf1-kf2-kf3-kf4-kr1-kr3-kr4
37	kcat2-kcat3-kcat4-kf1-kf2-kf3-kf4-kr2-kr3-kr4
38	kcat1-kcat2-kcat4-kf2-kf3-kf4-kr1-kr2-kr3-kr4
39	kcat1-kcat2-kcat3-kcat4-kf1-kf2-kf3-kf4-kr1-kr3-kr4
40	kcat1-kcat2-kcat3-kcat4-kf1-kf2-kf3-kf4-kr1-kr2-kr4
41	kcat1-kcat2-kcat3-kcat4-kf1-kf3-kf4-kr1-kr2-kr3-kr4
42	kcat1-kcat3-kcat4-kf1-kf2-kf3-kf4-kr1-kr2-kr3-kr4
43	kcat1-kcat2-kcat3-kcat4-kf1-kf2-kf3-kf4-kr1-kr2-kr3-kr4

Table A.7: Full list of the sequence of the first-principle removal for Toy Model.