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

Cristiano G.S. Campos[4], **Ronaldo N. Sousa**[1,2,3], **Hugo A. Armelin**[2,3], **Marcelo S. Reis**[2,3,4]

[1]Instituto de Matemática e Estatística, Universidade de São Paulo (USP), São Paulo, Brazil

[2]Instituto de Química, USP, São Paulo, Brazil

[3]CeTICS center, Laboratório de Ciclo Celular, Instituto Butantan, São Paulo, Brazil

[4]Recod lab, Instituto de Computação, Universidade Estadual de Campinas (UNICAMP), Campinas, Brazil

msreis@unicamp.br, cristiano.campos@students.ic.unicamp.br

Abstract. *The regulation of cellular processes is governed by chains of chemical reactions, known as cell signaling pathways. A key challenge in modeling these pathways is the "lack of isolation problem", where reactions within the model fail to interact with those in the broader cellular context, reducing prediction accuracy in first-principle models. Moreover, often some first-principle parameters are missing and must be inferred from data. To address this, we propose a hybrid modeling approach combining ordinary differential equation (ODE)-based first-principle models with neural network-based data-driven models, which jointly infers both neural network weights and missing first-principle parameters. Computational experiments using an iron metabolism model and a model implementation based on universal differential equations (UDEs) demonstrated significant improvements in prediction accuracy compared to first-principle models. These results support UDE-based hybrid models as effective tools for studying the complex dynamics of biological systems.*

1. Introduction

Cell signaling pathways are essential to understanding the complex processes that regulate cellular behavior. These pathways consist of chemical species, primarily proteins, that interact through chains of reactions. Information is transmitted via concentration changes of these species over time, often involving protein-protein interactions or post-translational modifications such as phosphorylation. Modeling these pathways is a significant challenge in systems biology due to their intricate interactions and regulatory mechanisms, and also the lack of measurements of many kinetic constants of the reactions in the pathway. Ordinary differential equations (ODEs) have traditionally been used to model cell signaling pathways by describing the rates of reactions [Reis et al. 2017]. Kinetic constants are either sourced from literature or repositories like SABIO-RK [Wittig et al. 2011], while unknown constants are inferred from time-series data through optimization methods. However, a persistent issue with ODE models

is the ”lack of isolation” problem, where the exclusion of reactions involving unmodeled species results in models that fail to capture the complete system dynamics, reducing predictive accuracy [Sousa et al. 2023]. Expanding the model or introducing hypothetical species are common but flawed approaches, often leading to overfitting or unrealistic models. In recent years, data-driven approaches, especially machine learning, have offered alternatives for modeling biological systems. While these methods excel at handling complex datasets, they lack interpretability, which is crucial for understanding underlying biological mechanisms. To address this, hybrid models combining mechanistic ODEs with neural networks have gained attention, offering both interpretability and predictive power. Notably, Universal Differential Equations (UDEs) [Rackauckas et al. 2021] provide a framework to integrate these models, enabling more accurate representations of biological systems. Despite their promise, UDE applications to cell signaling pathways still face the challenge of inferring both neural network weights and unknown first-principle parameters, such as rate constants, simultaneously—an issue that remains unresolved in the current literature. This work aims to address this gap by proposing a UDE-based approach for modeling cell signaling pathways that jointly infers unknown parameters and neural network weights from data.

The remainder of this paper is organized as follows: in Section 2, we present and discuss some papers from the literature that are relevant to this work; in Section 3, we describe the proposed methodology, including the UDE-based hybrid model, the training process, data acquisition, model assessment, experimental setup and case study; in Section 4, we show the results obtained in the experiments and discuss them; finally, in Section 5, we make final remarks about this work and point out possible future paths for this research line.

2. Related Works

Hybrid models have become an essential tool in chemical kinetics and cell signaling pathway studies, as they combine the interpretability of traditional first-principle models with the adaptability of data-driven approaches. These models address the limitations of ordinary differential equations (ODEs), which often struggle to capture the full complexity of biological systems due to omitted interactions or species. In response, hybrid models integrate machine learning components, such as neural networks, into first-principle models to enhance prediction accuracy and system representation.

For instance, Zander and colleagues introduced a hybrid model combining neural networks with real-world data to predict dynamic behavior in biochemical systems [Zander et al. 1999]. Wouver and colleagues further explored hybrid modeling with radial basis function (RBF) networks to handle real datasets in kinetic systems [V. Wouwer et al. 2004]. These studies demonstrated the potential of integrating data-driven methods with traditional modeling approaches, though challenges remain regarding the generalization of these models across diverse scenarios.

Narayanan and colleagues extended the application of hybrid models by introducing functional transformations in simulations, addressing the limitations of real-time data availability [Narayanan et al. 2022]. Li and colleagues applied long short-term memory (LSTM) networks in real-world datasets, offering a neural network solution for sequential data prediction in biological systems [Li et al. 2022]. Similarly, Dong and colleagues

utilized LSTM-based hybrid models to improve predictions of cell signaling pathway dynamics, showing that machine learning approaches can augment ODE-based models in complex systems [Dong et al. 2023].

The introduction of universal differential equations (UDEs) has significantly advanced hybrid modeling [Rackauckas et al. 2021]. Bangi and colleagues were among the first to apply UDEs in conjunction with neural networks to model beta-carotene production in yeast; this approach demonstrated that UDEs could effectively integrate neural networks into ODE models, yielding superior results compared to traditional methods [Bangi et al. 2022]. Santana and colleagues further demonstrated the utility of UDEs by applying them to sorption kinetics, showcasing their ability to handle complex biological systems with enhanced accuracy [Santana et al. 2023].

Lima and colleagues employed UDEs in both real and simulated datasets, highlighting their ability to balance interpretability and flexibility in hybrid modeling [Lima et al. 2023]. Despite these advances, a major research gap persists: the simultaneous estimation of unknown first-principle parameters, such as reaction rate constants, and neural network weights. Current methods often assume that these parameters are fully known, which is impractical in many real-world scenarios.

In summary, while UDE-based hybrid models hold great promise in addressing the challenges of traditional ODE approaches, the simultaneous estimation of mechanistic parameters and neural network components remains an open problem. This gap presents an opportunity for further research to develop robust methodologies that can fully leverage the power of hybrid models in biological and chemical systems.

3. Methodology

3.1. The inference of hybrid model

Our proposed methodology leverages the strengths of first-principle models and neural networks through the universal differential equation (UDE) framework to create a hybrid model [Rackauckas et al. 2021]. This model integrates the structured, mechanistic understanding of biological processes with the flexibility of machine learning, allowing us to better capture complex dynamics. The hybrid model is represented as:

$$\dot{\mathbf{x}} = f(\mathbf{x}(t), \mathbf{u}(t), U(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\omega}); \boldsymbol{\theta}), \tag{1}$$

where U denotes the neural network, $\boldsymbol{\omega}$ is a vector with the neural network weights, and $\boldsymbol{\theta}$ represents the parameters of the first-principle model. Training this hybrid model involves calculating the gradients of the solution with respect to $\boldsymbol{\omega}$, using the ADAM optimizer for rapid convergence and BFGS for fine-tuning the solution. ADAM is effective for quickly reaching a good approximation due to its adaptive learning rate, making it ideal for the early stages of optimization when large adjustments are beneficial. Once close to a minimum, BFGS is used to refine the solution, leveraging curvature information to take smaller, more accurate steps, thereby enhancing the precision and stability of convergence.

We incorporate time series data from various initial conditions into the training process, optimizing a loss function designed to handle multiple time series simultaneously. To infer the parameters of the first-principle model and the neural network weights

concurrently, we treat the first-principle parameters as additional optimization variables. By concatenating these parameters with the vector of the neural network weights, the optimization algorithm can update all parameters simultaneously.

To prevent the neural network from dominating the hybrid model, which could result in zeroing out first-principle parameters, we apply regularization to the neural network weights and penalize negative values for first-principle parameters using a rectified linear unit (ReLU). This ensures biologically plausible estimates and helps maintain the balance between both components of the hybrid model.

3.2. Training process

The dataset, which consists of time series of concentration levels of the species in the first-principle model, is split equally into training, validation, and testing sets. Early stopping, based on the validation set's mean absolute error (MAE), halts training if no improvement occurs after 100 iterations, preventing overfitting. To further reduce overfitting, we apply L2 regularization to the loss function, defined as:

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

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}.

We begin optimization with ADAM using a 0.1 learning rate for $2,000$ iterations, then switch to BFGS for $1,000$ iterations with a backtracking line search to avoid instability. Neural network weights are initialized near zero for stability.

In joint inference, a penalty of $100,000$ is applied to prevent negative first-principle parameter values, ensuring biological plausibility.

3.3. Data acquisition

The data for this study was generated through *in silico* simulations of a complete cell signaling pathway, denoted as $\mathcal{E}$, retrieved from Odebase [Lüders et al. 2022], and supplemented with a toy model. Random initial values were assigned, and the resulting initial value problems (IVPs) were solved over the time interval $[0, 100]$, sampled at 101 equally spaced points. Therefore, we obtained different time series, simulating real-world experiments where a cell line is stimulated with different compounds (*i.e.,*, different initial conditions).

To create a subset $\mathcal{F}$, we selected species common to both $\mathcal{E}$ and $\mathcal{F}$, forming the dataset for parameter inference. Noise was added to the data using a Gaussian distribution with standard deviation:

$$\sigma = 0.05\overline{x_i}, \tag{3}$$

where $\overline{x_i}$ is the mean of state variable x_i. This approach ensured that model training, validation, and testing reflected realistic biological variability.

3.4. Model assessment

To evaluate the effectiveness of the hybrid model in capturing cell signaling dynamics, particularly under uncertainty, we used the symmetric mean absolute percentage error (SMAPE) metric. SMAPE provides a normalized measure of prediction accuracy, making it well-suited for comparing model performance across different scales. It is defined as:

$$SMAPE = \frac{100}{NIJ} \sum_{n=1}^{N} \sum_{i=1}^{I} \sum_{j=1}^{J} \frac{|\hat{x}_{ij}^{(n)} - x_{ij}^{(n)}|}{\hat{x}_{ij}^{(n)} + x_{ij}^{(n)}}. \tag{4}$$

This metric allows for a percentage-based evaluation, ensuring a consistent comparison of the model's predictive accuracy across varying magnitudes of target variables, providing a robust assessment of the model's performance under diverse experimental conditions.

3.5. Description of the experiments

The experimental design consists of two phases to evaluate the resilience and accuracy of the hybrid model, particularly under the lack of isolation problem and unknown first-principle parameters.

First phase: This phase focuses on inferring neural network weights, assuming the first-principle parameters are known. We initially assess the ability of the model to compensate for missing inputs using 30 time series, divided equally for training, validation, and testing.

Second phase: In this phase, we jointly infer both first-principle parameters and neural network weights. Starting with one missing parameter at a time, we iteratively remove more parameters to test the adaptability and performance of the model, as measured by Symmetric Mean Absolute Percentage Error (SMAPE). Due to computational limits, this phase uses 30 time series, ensuring efficiency while rigorously testing parameter inference.

3.6. Case study: iron metabolism pathway

As a case study, we collected a model from the literature called "FeMetabolism FeDeficient" [Lopes et al. 2010]. This model describes the dynamics of iron metabolism through 17 chemical species and 33 first-principle parameters, covering 29 irreversible reactions [Lopes et al. 2010]. The model is based on studies of C57BL6 wild-type mice under different dietary conditions: iron-deficient, iron-adequate, and iron-loaded diets [Lopes et al. 2010]. The pathway cutout, highlighted in red in Figure 1, includes 4 chemical species and 4 first-principle parameters, supporting 3 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 built with the species of the cutout pathway are described

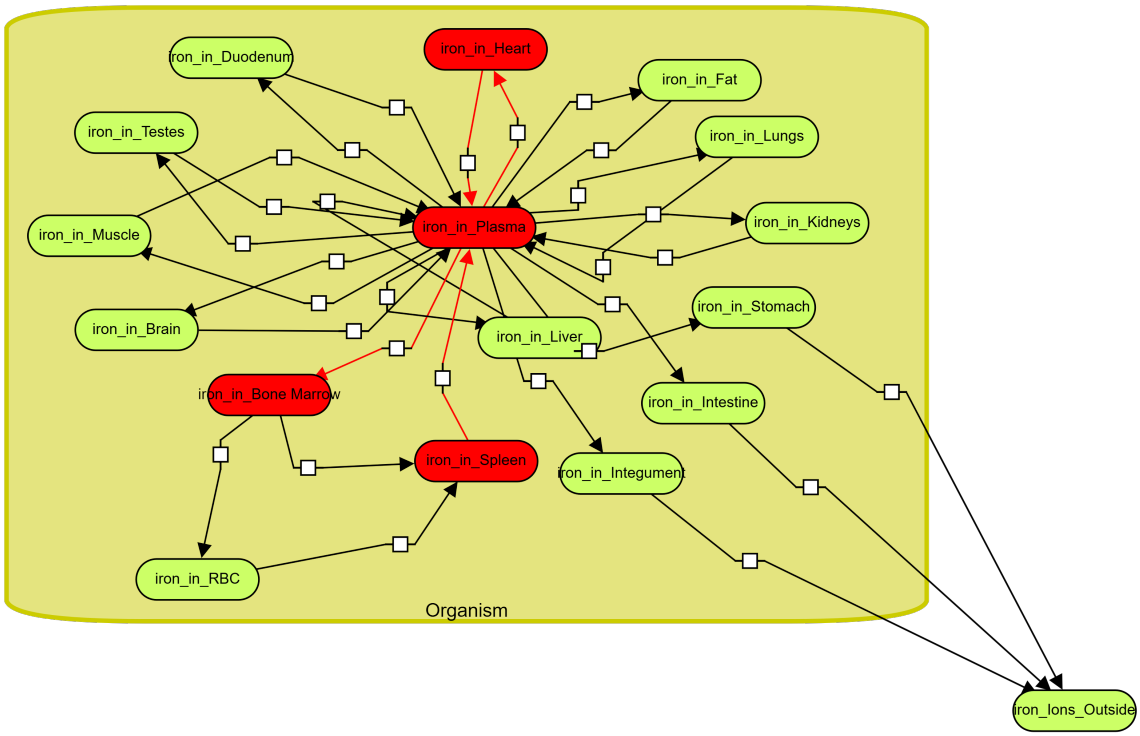

Figure 1. SBGN diagram of the iron metabolism cell signaling pathway. Nodes symbolize chemical species (such as proteins) and edges denote reactions. The pathway cutout is highlighted by red nodes and red edges, emphasizing the specific segments and interactions under study.

by the following equations:

$$\frac{d[x_1]}{dt} = -[x_1]k_3 - [x_1]k_{17} + [x_4]k_{18} \tag{5a}$$

$$\frac{[dx_2]}{dt} = [x_1]k_3 \tag{5b}$$

$$\frac{d[x_3]}{dt} = -[x_3]k_6 \tag{5c}$$

$$\frac{d[x_4]}{dt} = [x_1]k_{17}. \tag{5d}$$

The corresponding first-principle parameters are listed in Table 1.

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

Table 1. First-principle parameters of the iron metabolism pathway cutout.

4. Results

This section presents the results of our investigation into two cell signaling pathways using the methodology from Section 3. We compared the performance of traditional

first-principle models against hybrid models, including: a linear regression baseline, an ODE-based model, a UDE-based model with linear regression, and a UDE-based model with neural networks. The neural network consists of 4 layers, with 4 input neurons and 4 output neurons. The hidden layer contains 5 neurons, using the ReLU activation function, while the output layer applies a linear activation function.

The first-principle model under noisy conditions showed mean SMAPE values of 81.84, 80.97, and 79.63 for the training, validation, and test sets, respectively. In contrast, the linear regression model performed significantly better, with SMAPE scores of, respectively, 6.14, 7.43, and 7.09.

The UDE-based models, though not outperforming linear regression, exhibited strong noise resilience. The UDE with linear regression had SMAPE values of 6.91, 8.06, and 8.02, in training, validation and test sets, respectively, while the UDE with neural networks scored, respectively, 8.03, 9.46, and 8.35. These results are shown in Figure 2.

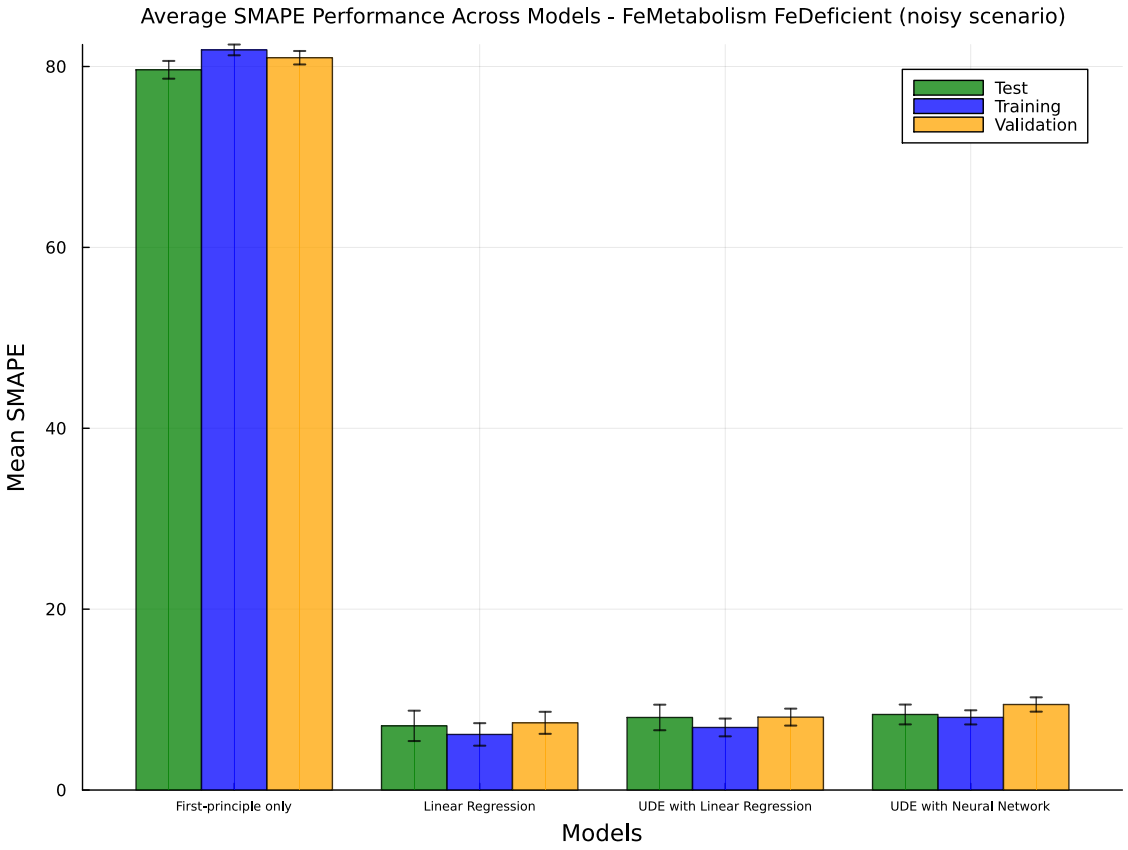

Figure 2. Model performance for the iron metabolism model across noisy datasets, showing SMAPE values for training, validation, and test sets (10 observations each). The y-axis shows mean SMAPE, and the x-axis categorizes the models. Error bars indicate standard deviation.

Further analysis evaluated the predictive performance of the UDE model with linear regression under noisy conditions. The best prediction achieved a SMAPE of 4.43 (see the Figure 3), while the worst reached 19.50. As shown in Figure 4, even in the worst case, the model maintained a reasonable approximation of the true dynamics across all chemical species, demonstrating its robustness.

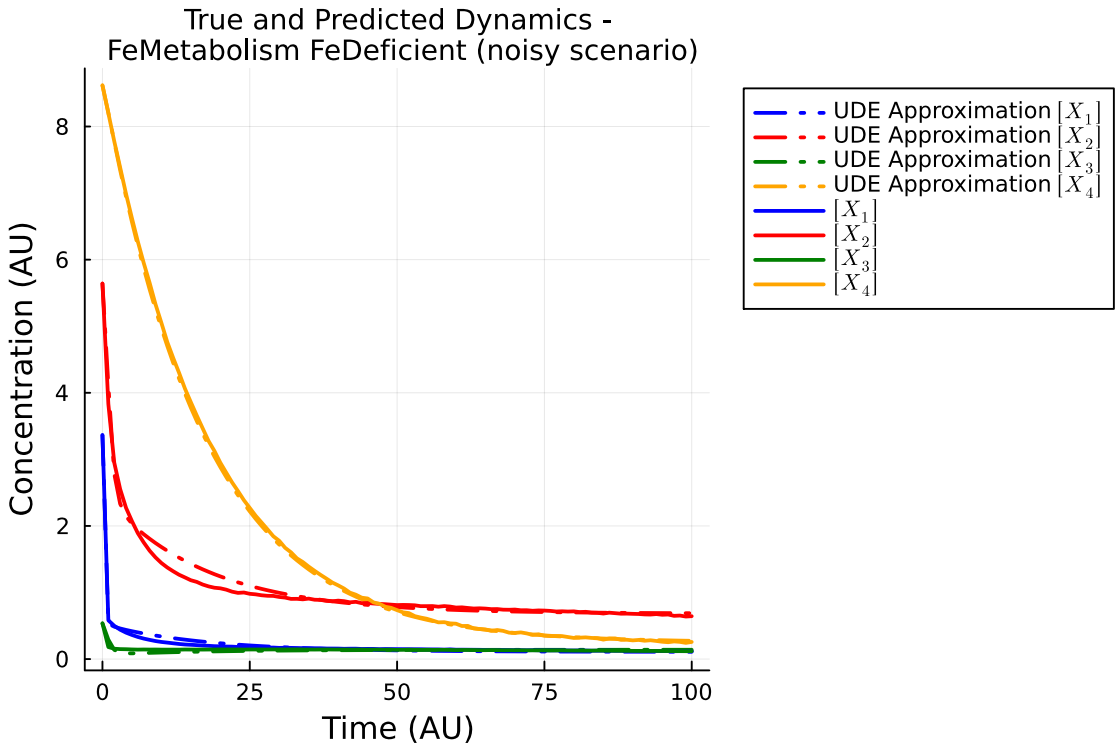

Figure 3. Best predictions (dashed lines) of the UDE with linear regression compared to true dynamics (solid lines) for the iron metabolism model under noisy conditions.

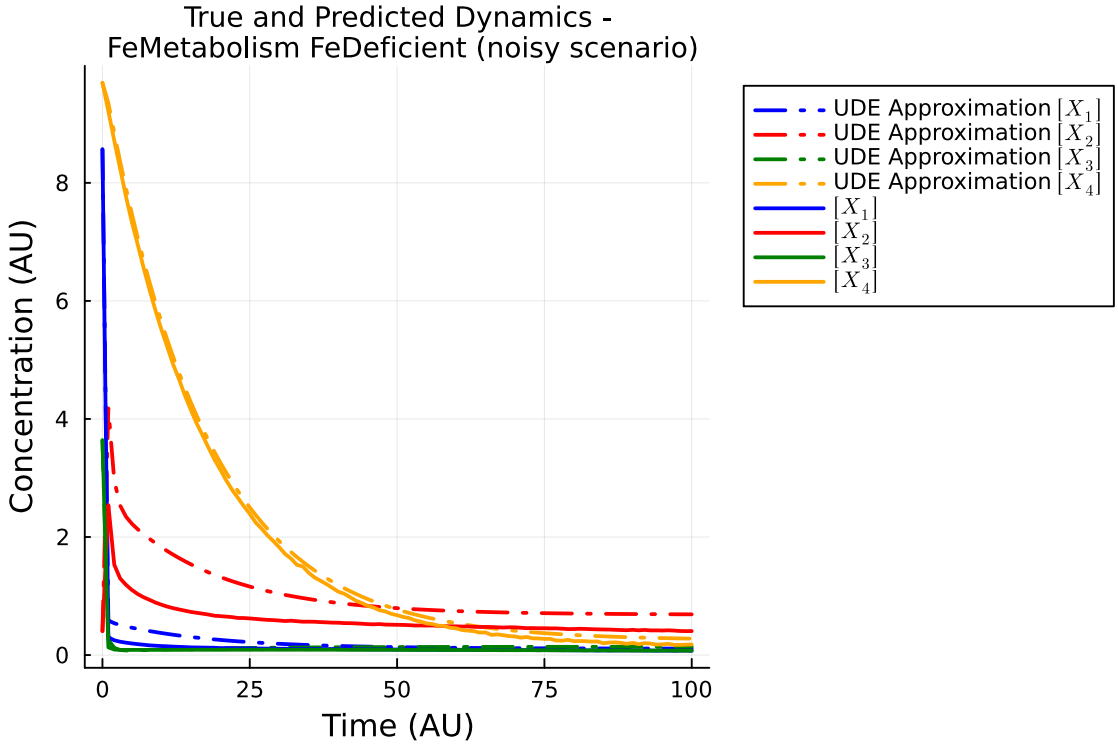

Figure 4. Worst predictions (dashed lines) of the UDE with linear regression compared to true dynamics (solid lines) for the iron metabolism model under noisy conditions.

Figure 5 illustrates the effect of removing and estimating different sets of first-principle parameters along with model weights under noisy conditions. The parameter sets are indexed as follows: index 1 represents the first-principle model without any removal; index 2 corresponds to the UDE-based model without removal; index 3 involves the removal of parameter k_6; index 4 represents the removal of k_3; index 5 involves the removal of k_{18}; and index 6 represents the removal of k_{17}. Continuing with combinations of parameter removals, index 7 corresponds to the removal of k_{18} and k_6; index 8 involves the removal of k_{17} and k_6; index 9 includes the removal of k_3 and k_6; and index 10 represents the removal of k_{17} and k_3. The analysis proceeds with index 11, which includes the removal of k_{17}, k_{18}, and k_6; index 12 involves the removal of k_{17}, k_{18}, and k_3; index 13 corresponds to the removal of k_{18}, k_3, and k_6; index 14 involves the removal of k_{17}, k_3, and k_6; and finally, index 15 represents the removal of k_{17}, k_{18}, k_3, and k_6.

Although the removal of the subset of parameters k_{17}, k_{18}, k_3 (index 12) led to poorer results, the overall mean SMAPE was still significantly improved compared to the first-principle model. Other parameter sets exhibited stability, even when compared to scenarios where all first-principle parameters were assumed to be known.

5. Conclusion

This study presents a robust approach to integrating hybrid modeling techniques for cell signaling pathways when key first-principle parameters are unknown. Our findings highlight the clear advantages of hybrid models, particularly in their ability to capture complex dynamics and deliver higher predictive accuracy than traditional first-principle models (ODE-based models). The integration of data-driven methods, such as neural networks, with mechanistic models offers greater flexibility in modeling biological systems, especially in scenarios involving noise and incomplete information.

Despite the superior performance of the hybrid model, particularly in noisy conditions, several challenges emerged during the process of jointly inferring first-principle parameters and neural network weights. Numerical instability and suboptimal parameter estimates occurred in some cases, even when SMAPE values indicated reasonable predictive accuracy. These issues underline the need for further refinement in both the optimization algorithms and the regularization techniques used to maintain the balance between mechanistic accuracy and neural network flexibility. The removal of certain critical parameters simultaneously, such as $\{k_{17}, k_{18}, k_3\}$, led to a noticeable drop in performance, indicating potential dependencies on specific parameters that should be investigated in future studies.

In conclusion, while the hybrid approach shows great promise for advancing the study of cell signaling pathways, particularly in situations where traditional models are limited, there is room for improvement. Future work should focus on addressing the challenges of numerical instability, refining the methods for joint parameter inference, and exploring the model's performance across a broader range of pathways and conditions. This will be essential for ensuring that hybrid models become a reliable and scalable tool for studying complex biological systems.

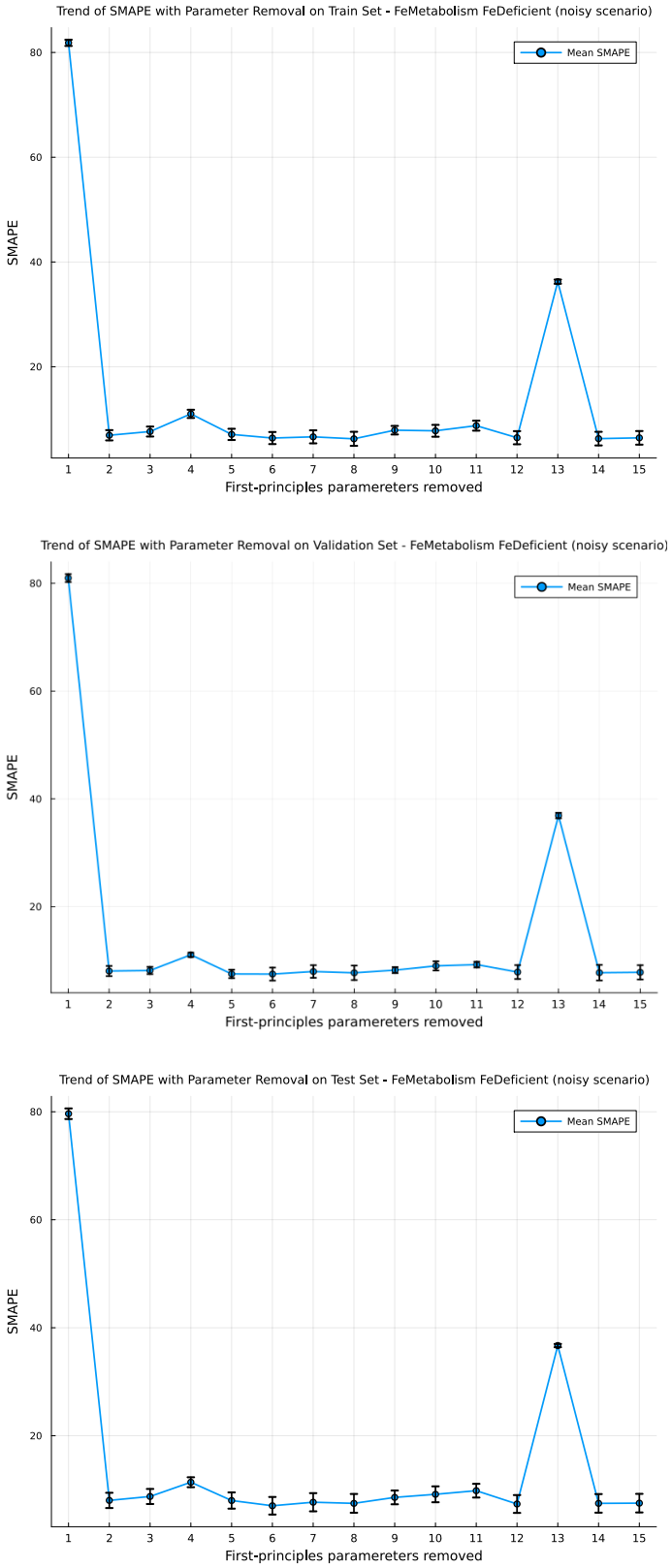

Figure 5. Impact of first-principle parameter removal on mean SMAPE for the iron metabolism model under noisy conditions. The mean SMAPE is shown for different sets of removed parameters using 10 noisy observations each for training, validation, and test sets. The y-axis represents mean SMAPE, and the x-axis lists the removed parameters. Error bars indicate standard deviation.

Acknowledgements

This work was supported by CAPES, BECAS Santander, and by grants 13/07467-1, 19/21619-5, and 19/24580-2, São Paulo Research Foundation (FAPESP).

References

[Bangi et al. 2022] Bangi, F. M. S., Kao, K., and Kwon, J. S. (2022). 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. DOI: 10.1016/j.cherd.2022.01.041.

[Dong et al. 2023] Dong, S., Zhang, Y., and Zhou, X. (2023). Intelligent hybrid modeling of complex leaching system based on lstm neural network. *Systems*, 11(2). DOI: 10.3390/systems11020078.

[Li et al. 2022] Li, K., Duan, H., Liu, L., Qiu, R., van den Akker, B., Ni, B.-J., Chen, T., Yin, H., Yuan, Z., and Ye, L. (2022). 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. DOI: 10.1021/acs.est.1c05020.

[Lima et al. 2023] Lima, F. R., Rebello, C. M., Costa, E. A., Santana, V. V., de Moares, M. G., Barreto, A. G., Secchi, A. R., de Souza, M. B., and Nogueira, I. B. (2023). Improved modeling of crystallization processes by universal differential equations. *Chemical Engineering Research and Design*, 200:538–549. DOI: 10.1016/j.cherd.2023.11.032.

[Lopes et al. 2010] Lopes, T. J., Luganskaja, T., Vujić, M. S., Hentze, M. W., Muckenthaler, M. U., Schümann, K., and Reich, J. G. (2010). Systems analysis of iron metabolism: the network of iron pools and fluxes. *BMC Systems Biology*. DOI: 10.1186/1752-0509-4-112.

[Lüders et al. 2022] Lüders, C., Sturm, T., and Radulescu, O. (2022). ODEbase: A repository of ODE systems for systems biology. *Bioinformatics Advances*, 2(1). DOI: 10.1093/bioadv/vbac027.

[Narayanan et al. 2022] Narayanan, H., C. Bournazou, M. N., G. Gosálbez, G., and Butté, A. (2022). Functional-hybrid modeling through automated adaptive symbolic regression for interpretable mathematical expressions. *Chemical Engineering Journal*, 430:133032. DOI: 10.1016/j.cej.2021.133032.

[Rackauckas et al. 2021] Rackauckas, C., Ma, Y., Martensen, J., Warner, C., Zubov, K., Supekar, R., Skinner, D., Ramadhan, A., and Edelman, A. (2021). Universal differential equations for scientific machine learning. DOI: 10.48550/arXiv.2001.04385.

[Reis et al. 2017] Reis, M. S., Noël, V., Dias, M. H., Albuquerque, L. L., Guimarães, A. S., Wu, L., Barrera, J., and Armelin, H. A. (2017). 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. DOI: 10.1007/978-1-4939-7154-1_28.

[Santana et al. 2023] Santana, V. V., Costa, E., Rebello, C. M., Ribeiro, A. M., Rackauckas, C., and Nogueira, I. B. (2023). 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. DOI: 10.1016/j.ces.2023.119223.

[Sousa et al. 2023] Sousa, R. N., Campos, C. G. S., Wang, W., Hashimoto, R. F., Armelin, H. A., and Reis, M. S. (2023). Exploring identifiability in hybrid models of cell signaling pathways. In Reis, M. S. and de Melo-Minardi, R. C., editors, *Advances in Bioinformatics and Computational Biology*, pages 148–159, Cham. Springer Nature Switzerland. DOI: 10.1007/978-3-031-42715-2_14.

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

[Wittig et al. 2011] Wittig, U., Kania, R., Golebiewski, M., Rey, M., Shi, L., Jong, L., Algaa, E., Weidemann, A., Sauer-Danzwith, H., Mir, S., Krebs, O., Bittkowski, M., Wetsch, E., Rojas, I., and Müller, W. (2011). SABIO-RK—database for biochemical reaction kinetics. *Nucleic Acids Research*, 40(D1):D790–D796. DOI: 10.1093/nar/gkr1046.

[Zander et al. 1999] Zander, H.-J., Dittmeyer, R., and Wagenhuber, J. (1999). Dynamic modeling of chemical reaction systems with neural networks and hybrid models. *Chemical Engineering Technology*. DOI: 10.1002/(SICI)1521-4125(199907)22:7¡571::AID-CEAT571¿3.0.CO;2-5.