Leveraging Jacobian Analysis for Interpretability and Control in Hybrid Bioprocess Models

Author: Karolina Biniek, Lucas Cosier
Affiliation: DataHow AG, R&D Team
Scope: Internship
Skills: Curiosity and willingness to tackle challenging research problems
Strong background and experience in Machine Learning
Good programming skills in Python
Familiarity with autodifferentiation
Date: September 26, 2025
Contact: h.narayanan@datahow.ch

Executive Summary

This project proposes developing a Jacobian-based interpretability framework for hybrid bioprocess models that combine neural networks with mechanistic mass balances. While these models achieve strong predictive performance through learned residual dynamics, they remain fundamentally black boxes. By extracting and analyzing the Jacobian, we aim to reveal metabolic control patterns and enable early detection of process deviations. The framework will provide both interpretability tools for understanding model decisions and predictive capabilities for proactive process control, transforming black-box models into actionable grey-box systems that maintain performance while offering mechanistic insights.

Background

Brief introduction / Context

Bioprocesses are dynamic systems with complex interactions, making them difficult to capture with either purely mechanistic or purely data-driven models. Hybrid models address this limitation by combining machine learning (ML) with a mechanistic backbone, thereby improving predictive performance. However, interpretability remains a key concern, as these models often function as black boxes. To support deeper process insight and effective control, it is essential not only to accurately forecast system trajectories but also to uncover the underlying factors that drive these outcomes.

Current approaches / State of the art

Among the models used for hybrid modeling, recurrent neural networks (RNNs) are particularly well suited. At the same time, their complexity has motivated the development of interpretability approaches in deep learning, where techniques such as feature attribution or attention mechanisms aim to indicate which inputs most strongly influence predictions.

Challenges

Despite strong predictive performance, hybrid RNNs remain difficult to interpret as their predictions arise from complex, recurrent transformations across many timesteps. In addition, bioprocess data often contain discontinuities introduced by bolus feeding, for instance, which create non-smooth dynamics and complicate sensitivity analysis. The large number of features further results in large sensitivity matrices at every timestep, which are difficult to analyze without dimensionality reduction.

Project rationale / Approach

The Jacobian offers a principled interpretability tool, providing a mechanistically grounded alternative to attention-based methods. The Euler-step integration in RNNs yields direct access to

$$\frac{\partial X_{t+1}}{\partial X_t}$$

As the Jacobian reflects only localized dynamics at a given timestep, the key challenge for interpretability lies in studying how these matrices evolve over time. Eigendecomposition could reveal shifts in metabolic control (e.g., glucose to lactate limitation) and anticipate process deviations before they are observable. To handle bolus-induced discontinuities, continuous and discrete Jacobians can be analyzed separately, while their high dimensionality ($n_{features} \times n_{features}$ at every timestep) can be addressed through reduction techniques such as singular value decomposition (SVD).

Objectives

1. **Jacobian extraction and analysis:** Implement routines to compute $\partial X_{t+1}/\partial X_t$ at every timestep of the RNN.
2. **Eigendecomposition and sensitivity modes:** Characterize dominant eigenvalues/vectors to identify metabolic control shifts.
3. **Handling discontinuities:** Devise strategies to separate continuous Jacobians from discrete bolus-induced jumps.
4. **Application to interpretability:** Develop tools that highlight metabolite influences, compare against other methods.
5. **Exploratory predictive applications:** Use Jacobian-based sensitivities to forecast depletion events or rank initial conditions for optimization of product yield.

Methods and Work Plan

Data and Resources

- Datasets: In-silico datasets with bolus feeding
- Frameworks: Python, PyTorch

Timeline

Timeline subject to the scope. Preference: The longer, the better.

Timeline

Phase	**Tasks**	**Target Dates**
Exploration	Literature review on Jacobian analysis; baseline hybrid RNN setup.	Wk 1–3
Implementation	Jacobian extraction; feeding event handling.	Wk 4–6
Analysis	Eigendecomposition; sensitivity backprop to endpoints.	Wk 7–9
Applications	Interpretability demos; predictive depletion forecasting.	Wk 10–11
Wrap-up	Documentation, evaluation vs alternatives, final report.	Wk 12

Expected Outcome

At the end of this project, we will have developed a Jacobian-based interpretability toolkit for hybrid RNN bioprocess models.