Neural Stochastic Differential Equations for Bioprocesses

Author:	Claus Wirnsperger
Affiliation:	DataHow AG, R&D Team
Scope:	Internship
Skills:	Coding in python, Machine Learning, Critical thinking (Bioprocess knowledge not crucial)
Date:	September 26, 2025
Contact:	h.narayanan@datahow.ch

Executive Summary

The goal of this project to implement a neural stochastic differential equation based modeling framework for bioprocesses. The main motivation of this project is to explore the possibility of implementing a theoretically improved approach for our existing modeling framework, particularly for the estimation of uncertainty across variables and times.

Background

Brief introduction / Context

One of the main features of DataHowLab is the decision making under uncertainty and the analysis of tradeoffs in designing new experiments - under consideration of the probability of staying within certain quality constraints and staying above or below certain yield or other productivity requirements. To enable algorithms that allow us to explore these tradeoffs, we require a theoretically sound implementation of the estimation of uncertainty of each predicted value along time for the different process variables. The current modeling strategy uses the concept called hybrid modeling-that combines mechanistic model based on systems of ordinary differential equations (ODEs) with machine learning (ML) model. Stochastic differential equations (SDEs) are a natural extension to ODEs that evolve in continuous time subject to uncertainty. Multiple approaches exist to unify neural networks and SDEs to create neural SDEs. The goal of this project is to explore the possibility of applying neural SDEs to bioprocess data.

Current approaches / State of the art

A basic hybrid for ODEs would be:

$$\frac{dX_i}{dt} = f(X, Z) + \text{In}_i - \text{Out}_i$$

$$f(X, Z) = NN(X, Z)$$

SDEs instead consist of a deterministic and stochastic term.

$$dX_{i,t} = f(t, X_{i,t}, Z)dt + g(t, X_{i,t}, Z)dW_{i,t} \tag{1}$$

The straightforward augmentation into the ML framework can be realized by discretizing the differential equations and learning a mapping between the current time to a future time point.

$$X_{i,t+1} = X_{i,t} + \Delta t \left(f(X_{i,t}, Z) + \text{In}_i + \text{Out}_i + g(X_{i,t}, Z)\frac{\Delta W_{i,t}}{\Delta t} \right) \tag{2}$$

Kidger et al. (2021) argues that fitting neural SDEs can be approached as a special case of Wasserstein GANs. Tzen and Raginsky (2019) consider the diffusion limit of deep latent Gaussian models for time-inhomogeneous Markov chains (specifically here the noise process in the SDE is assumed to be Gaussian). General treatment of DLGMs is found in Rezende et al. (2014). Another application of SDEs for irregularly sampled timeseries data is described in Oh et al. (2025). These papers can act as an initial starting point into the literature.

Challenges

Multiple challenges exist at first glance:

- How to define a unified approach to developing SDEs that can be applied across multiple process formats and unit operation encountered in bioprocesses (e.g., mammalian cell cultures, microbial cultivation etc)

- How to parameterize the noise process $dW_{i,t}$.

- How to train such a model (typically posteriors are not analytically tractable).

Furthermore, evaluating the quality of the estimated uncertainties is not a straight forward task. While heuristics and some metrics exist, they can be biased or show an incomplete picture of the overall accuracy of the estimation.

Project rationale / Approach

The extension of ODEs to SDEs is a natural way to introduce uncertainty into our hybrid models. A unified implementation of the SDEs across different process formats and process units, coupling with the neural networks (i.e., Neural SDEs) and defining an effective training procedure are key aspects to be addressed in this direction.

Objectives

1. **Developing the SDEs system:** Formulate the SDEs for the different process formats and process units that currently exist in our software. These are mammalian cell cultures, microbial cultivation, in-vitro transcription of mRNA and post-transcriptional capping of mRNA.

2. **Implement the model and the training procedure:** Implement the different Neural SDEs and develop a sound training procedures.

3. **Capability assessment:** Inspect not just the performance metrics (i.e. relative RMSE) of the individual predictions across time but also the estimated uncertainties given the quality metrics provided and compare with our existing approaches.

4. **Stress testing and limitations:** Identify the potential shortcomings/pitfall of the neural SDEs under various combinations of data frequencies, missing values etc.

Work Plan

Data and Resources

Some implementations of SDEs already exist in Python, however depending on the difficulty of the task it could make sense to implement our own solvers. Otherwise branching out to other programming languages like Julia might be worthwhile.

Timeline

Timeline subject to the scope. Preference: Longer the better.

Phase	Tasks	Target Dates
Exploration	Literature review.	Wk 1–2
Prototype	Build initial Neural SDE; sanity checks.	Wk 4–8
Benchmark	Compare with existing models; profiling.	Wk 8–10
Stress Tests	Noise/frequency/duration studies; limitations.	Wk 10–11
Wrap-up	Documentation; presentation; next steps.	Wk 12

Expected Outcome

At the end of this project, we aim to have a clear assessment of neural SDEs specifically as an extension of neural ODEs to reliably estimate the uncertainty in the predictions of process variables over time.

References

Kidger, P., Foster, J., Li, X., Oberhauser, H., and Lyons, T. (2021). Neural sdes as infinite-dimensional gans.

Oh, Y., Lim, D.-Y., and Kim, S. (2025). Stable neural stochastic differential equations in analyzing irregular time series data.

Rezende, D. J., Mohamed, S., and Wierstra, D. (2014). Stochastic backpropagation and approximate inference in deep generative models.

Tzen, B. and Raginsky, M. (2019). Neural stochastic differential equations: Deep latent gaussian models in the diffusion limit.