ADJOINT SENSITIVITY ANALYSIS ON MULTI-SCALE BIOPROCESS STOCHASTIC REACTION NETWORK

Keilung Choy[1], and Wei Xie[1]

[1]Dept. of Mechanical and Industrial Eng., Northeastern University, Boston, MA, USA

ABSTRACT

Motivated by the pressing challenges in the digital twin development for biomanufacturing processes, we introduce an adjoint sensitivity analysis (SA) approach to expedite the learning of mechanistic model parameters. In this paper, we consider enzymatic stochastic reaction networks representing a multi-scale bioprocess mechanistic model that allows us to integrate disparate data from diverse production processes and leverage the information from existing macro-kinetic and genome-scale models. To support forward prediction and backward reasoning, we develop a convergent adjoint SA algorithm studying how the perturbations of model parameters and inputs (e.g., initial state) propagate through enzymatic reaction networks and impact on output trajectory predictions. This SA can provide a sample efficient and interpretable way to assess the sensitivities between inputs and outputs accounting for their causal dependencies. Our empirical study underscores the resilience of these sensitivities and illuminates a deeper comprehension of the regulatory mechanisms behind bioprocess through sensitivities.

1 INTRODUCTION

With prior knowledge of the reaction network structure and regulatory mechanisms, the trajectory output of bioprocess hinges on three key inputs: (1) initial states; (2) actions; and (3) parameters of the mechanistic model. Suppose the effect of actions, such as feeding strategies, on state change is immediate and known. *Therefore, our objective is to develop an interpretable and sample-efficient sensitivity analysis (SA) approach for the multi-scale bioprocess mechanistic model that focuses on investigating the sensitivities between inputs (i.e., initial states and model parameters) and outputs.*

A multi-scale bioprocess mechanistic model can facilitate the development of digital twins and Bio-Foundries for biomanufacturing processes. Within this paradigm, the model's foundation is constructed upon fundamental building blocks, i.e., molecular reaction networks. There exist various challenges to conduct SA on the mechanistic model of enzymatic stochastic reaction networks. One significant feature is its double-stochasticity, which means at any time molecular reaction rates are contingent upon random states, such as species concentrations and environmental variables. Upon formulating the multi-scale bioprocess mechanistic model in the form of stochastic differential equations (SDEs), the drift and diffusion terms could be built based on Michaelis–Menten kinetics, which is the most frequently used kinetic model of enzymatic reactions found in existing literature (Kyriakopoulos et al. 2018). It leverages our existing understanding of the reaction rate structure and encapsulates the double stochasticity inherent in stochastic reaction networks, i.e., both the drift and diffusion terms are contingent upon the current states. Consequently, deriving analytical solutions for such SDEs, characterizing stochastic molecular reaction network dynamics and variations, could be difficult. Moreover, the interactions between input factors and intermediate states further introduce high complexity in SA.

Existing sensitivity analysis approaches could be divided into two categories: local and global sensitivity analysis. Global sensitivity analysis focuses on the impact of significant variations in model inputs. For example, the Sobol method (Sobol 2001) is a global sensitivity analysis approach based on variance

decomposition, which decomposes the variance of the model output into the contributions of different model inputs and parameters. In contrast, the objective of local sensitivity analysis is to quantify the effects of minor perturbations in model inputs and parameters on model predictions. There are a variety of local sensitivity analysis approaches including finite direct differential method (Kramer et al. 1984), Nominal Range Sensitivity Method (Cullen and Frey 2002), and automatic differentiation (Kedem 1980). Unfortunately, most of these existing approaches are not applicable to our study since they scale poorly, either in terms of computational time or memory usage, as the number of parameters and states within the model increases. Given that the multi-scale bioprocess model typically represents a complex system with many inputs and parameters, these limitations pose a significant barrier to the learning process of underlying mechanisms.

In this paper, we formulate the multi-scale bioprocess mechanistic model in SDEs form, accounting for underlying causal interdependencies of an enzymatic stochastic reaction network, and then develop an adjoint SA approach studying the sensitivities between inputs and outputs. It can correctly and efficiently quantify the contribution and criticality of each input and model parameter impacting on the prediction errors of multivariate output trajectories, such as productivity and product critical quality attributes (CQAs). To support forward prediction and backward reasoning, this adjoint sensitivity analysis over the operator of SDEs, characterizing bioprocess mechanisms, exhibits robust scalability even when the complexity of the mechanistic model increases. In addition, since it leverages the structural information of the regulatory reaction network, the adjoint SA can provide sample efficient and interpretable guidance to search inputs and model parameters, accounting for their interactions, to speed up the learning process.

This paper addresses two key challenges in bioprocess modeling through the development of the adjoint SA approach. Firstly, it addresses the issue that the dimensionality of model parameters can be so high that searching for the optimal parameter estimation becomes challenging. The local SA approach assesses the impact of model parameters on prediction accuracy and expedites process mechanism learning. This facilitates digital twin development and more efficient experimental design, which is crucial in the costly, highly regulated biomanufacturing industry. Secondly, it handles the intricate interdependencies in bioprocesses. By leveraging structural information within enzymatic molecular reaction networks, the paper mitigates complexity and streamlines the calculation of input-output sensitivities, alleviating computational bottlenecks encountered in previous local SA studies.

The structure of the paper is organized as follows. In Section 2, we initiate by delineating the multi-scale bioprocess mechanistic model and summarizing its key characteristics. Subsequently, in Section 3, we proceed to construct a metamodel to the multi-scale bioprocess, leveraging the insights from the preceding section. Section 4 is dedicated to local sensitivity analysis algorithm development, which will help us investigate the barriers to the reduction of model prediction errors and provide a guidance to speed up the search for optimal inputs. Building upon this part, Section 4.2 introduces an adjoint SA algorithm on SDEs designed to enhance computational efficiency. Improvements brought by the algorithm will be validated in Section 5 through the empirical study of its finite-sample performance. Finally, in Section 6, we synthesize the findings and insights gathered throughout this study, leading to the paper's conclusion.

2 PROBLEM DESCRIPTION

2.1 Multi-Scale Bioprocess Stochastic Reaction Network

A multi-scale bioprocessing mechanistic model characterizes the causal dependence from molecular to macroscopic kinetics and it is built on the fundamental building block, i.e., enzymatic molecular reaction networks. Suppose the system is composed of I species, denoted by $\boldsymbol{X} = (X_1, X_2, \ldots, X_I)^\top$, interacting with each other through J reactions. At any time t, let $\boldsymbol{s}_t = (s_t^1, s_t^2, \ldots, s_t^I)^\top \in \mathbb{R}_+^I$ be the bioprocess state, where s_t^i denotes the number of molecules of species i. Each j-th reaction with $j = 1, 2, \ldots, J$ is characterized by a reaction vector $\boldsymbol{N}_j \in \mathbb{R}^I$, describing the change in the numbers of I species' molecules when a j-th molecular reaction occurs. The associated reaction rate denoted by v_j, depending on state, such as the current number

of molecules of each species, describes the rate at which the j-th reaction occurs. Specifically, for the j-th reaction equation given by
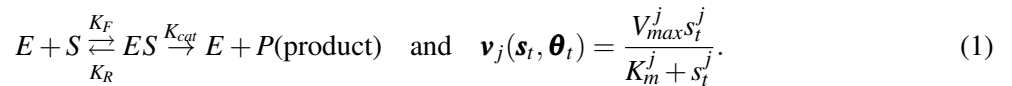
$$p_{j1}X_1 + p_{j2}X_2 + \cdots + p_{jI}X_I \xrightarrow{v_j} q_{j1}X_1 + q_{j2}X_2 + \cdots + q_{jI}X_I,$$
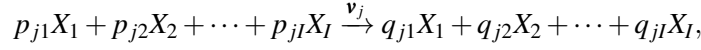
the reaction relational structure, specified by the vector $\boldsymbol{N}_j = (q_{j1} - p_{j1}, q_{j2} - p_{j2}, \ldots, q_{jI} - p_{iI})^\top$, is known for $j = 1, 2, \ldots, J$. Thus, the stoichiometry matrix $\boldsymbol{N} = (\boldsymbol{N}_1, \boldsymbol{N}_2, \ldots, \boldsymbol{N}_J) \in \mathbb{R}^{I \times J}$ characterizes the structure information of the reaction network composed of J reactions. The (i, j)-th element of $\boldsymbol{N}$, denoted by N_{ij}, represents the number of molecules of the i-th species that are either consumed (indicated by a negative value) or produced (indicated by a positive value) in each random occurrence of the j-th reaction.

In this paper, we suppose that the structure of the reaction network represented by matrix $\boldsymbol{N}$ is known. The regulation mechanism of each j-th reaction is characterized by the reaction rate function $\boldsymbol{v}_j$, which is associated with the current system state $\boldsymbol{s}_t$ and the mechanistic model parameters denoted by $\boldsymbol{\theta}_t$. Let $\boldsymbol{R}_t$ be a vector of the occurrences of each molecular reaction in a given short time interval $(t, t + \Delta t]$ and the system state is updated from $\boldsymbol{s}_t$ to $\boldsymbol{s}_{t+1}$. Since a molecular reaction will occur when one molecule collides, binds, and reacts with another one while molecules move around randomly, driven by stochastic thermodynamics of Brownian motion (Golightly and Wilkinson 2005), the occurrences of molecular reactions are modeled by non-homogeneous Poisson process. Therefore, the state transition model becomes,

$$\boldsymbol{s}_{t+1} = \boldsymbol{s}_t + \boldsymbol{N} \cdot \boldsymbol{R}_t \quad \text{with} \quad \boldsymbol{R}_t \sim \text{Poisson}(\boldsymbol{v}(\boldsymbol{s}_t, \boldsymbol{\theta}_t)),$$

where $\boldsymbol{N} \cdot \boldsymbol{R}_t$ represents the net amount of reaction outputs during time interval $(t, t + \Delta t]$.

Michaelis–Menten (MM) kinetics is commonly used to model the regulation mechanisms of enzymatic reaction networks and the flux rates $\boldsymbol{v}(\boldsymbol{s}_t, \boldsymbol{\theta}_t)$ depend on the state (Michaelis and Menten 2007). In an enzymatic molecular reaction as shown in the first equation in (1), the substrate (S) initially forms a reversible complex (ES) with the enzyme (E), i.e., the enzyme and substrate have to interact for the enzyme to be able to perform its catalytic function to produce the product (P), with K_F, K_R, and K_{cat} representing kinetic rates. For MM kinetics as shown in the second equation in (1), we assume the enzyme is either present as the free enzyme or as the ES complex, i.e., $[E]_{total} = [E] + [ES]$ with $[E]$ denoting the concentration of the enzyme. Suppose the rate of formation of the ES complex is equal to the rate of dissociation plus the breakdown, i.e., $K_F[E][S] = [ES](K_R + K_{cat})$. Thus, *the parameters in MM kinetics characterize the regulation mechanisms of enzymatic reaction network*: (1) $V_{max}^j = K_{cat}[E]_{total}$, is the maximum possible velocity of the j-th molecular reaction that can occur when all the enzyme molecules are bound with the substrate, i.e., $[E]_{total} = [ES]$; and (2) $K_m^j = \frac{K_R + K_{cat}}{K_F} = \frac{[E][S]}{[ES]}$ is a dissociation constant for the ES complex,

$$E + S \underset{K_R}{\overset{K_F}{\rightleftarrows}} ES \overset{K_{cat}}{\to} E + P(\text{product}) \quad \text{and} \quad v_j(\boldsymbol{s}_t, \boldsymbol{\theta}_t) = \frac{V_{max}^j s_t^j}{K_m^j + s_t^j}. \tag{1}$$

By applying diffusion approximation, the state transition model becomes,

$$p(\boldsymbol{s}_{t+1}|\boldsymbol{s}_t, \boldsymbol{\theta}_t) \sim \begin{cases} \mathcal{N}(\boldsymbol{s}_t + \boldsymbol{N}\boldsymbol{v}(\boldsymbol{s}_t, \boldsymbol{\theta}_t)\Delta t, \boldsymbol{N} diag(\boldsymbol{v}(\boldsymbol{s}_t, \boldsymbol{\theta}_t))\boldsymbol{N}^\top \Delta t), \boldsymbol{s}_{t+1} > \boldsymbol{0}, \\ 0, \boldsymbol{s}_{t+1} \le \boldsymbol{0}. \end{cases} \tag{2}$$

Thus, the multi-scale bioprocess mechanistic model is further represented as a system of Stochastic Differential Equations (SDEs), intricately reliant on parameters $\boldsymbol{\theta}_t$ and current states $\boldsymbol{s}_t$. The diffusion term is necessary due to the inherent stochasticity of molecular reactions in the bioprocess:

$$(d\boldsymbol{s}_t, d\boldsymbol{\theta}_t)^\top = (\mu(\boldsymbol{s}_t, \boldsymbol{\theta}_t), f(\boldsymbol{s}_t, \boldsymbol{\theta}_t))^\top dt + (\sigma(\boldsymbol{s}_t, \boldsymbol{\theta}_t), g(\boldsymbol{s}_t, \boldsymbol{\theta}_t))^\top dW_t, \tag{3}$$

where μ, f are the drift terms and σ, g are the diffusion terms defined as:

$$\begin{aligned} (\mu(\boldsymbol{s}_t, \boldsymbol{\theta}_t), f(\boldsymbol{s}_t, \boldsymbol{\theta}_t))^\top &= \text{E}[(\boldsymbol{s}_{t+\Delta t}, \boldsymbol{\theta}_{t+\Delta t})^\top - (\boldsymbol{s}_t, \boldsymbol{\theta}_t)^\top | \boldsymbol{s}_t, \boldsymbol{\theta}_t], \\ (\sigma^2(\boldsymbol{s}_t, \boldsymbol{\theta}_t), g^2(\boldsymbol{s}_t, \boldsymbol{\theta}_t))^\top &= \text{Var}[(\boldsymbol{s}_{t+\Delta t}, \boldsymbol{\theta}_{t+\Delta t})^\top | \boldsymbol{s}_t, \boldsymbol{\theta}_t]. \end{aligned}$$

The specific form of infinitesimal mean and variance on the right side will be derived in Section 3. This SDE system is the foundational groundwork for the subsequent local sensitivity analysis we will conduct. In this paper, $\boldsymbol{\theta}_t$ remains constant, resulting in both f and g becoming 0. However, given that we model the dynamics of $\boldsymbol{\theta}_t$ in SDE form as shown in Equation (3), our approach can be readily extended to accommodate the scenarios where $\boldsymbol{\theta}_t$ varies over time.

Given any feasible policy with action, i.e., $\boldsymbol{a}_t = \pi(\boldsymbol{s}_t)$ at any decision time t, we consider the state transition model characterizing the process mechanistic dynamics and inherent stochasticity, i.e.,

$$\boldsymbol{s}_{t+1} \sim p(\boldsymbol{s}_{t+1}|\boldsymbol{s}_t,\boldsymbol{a}_t;\boldsymbol{\theta}).$$

For the biomanufacturing process, suppose the impact of decision $\boldsymbol{a}_t$ (e.g., feeding strategy) on the state is known and happens immediately; that means we get the post-decision state denoted by $\boldsymbol{s}_t' = \boldsymbol{f}(\boldsymbol{s}_t,\boldsymbol{a}_t)$ with a known function $\boldsymbol{f}$. For notation simplification, we ignore the impact of action. The proposed sensitivity analysis over inputs (i.e., states and model parameters) is extendable to account for the policy effect.

2.2 Adjoint Sensitivity Analysis on Model Parameters and State Variables

Local sensitivity analysis studies the changes in the model prediction outputs with respect to initial input values, i.e., $(\boldsymbol{s}_0,\boldsymbol{\theta}_0)$. The variations around this local point are quantified by the sensitivity coefficients (Zi 2011) that are intimately connected to the solution of the SDE system (3). In Section 3, we will derive $\mu(\boldsymbol{s}_t,\boldsymbol{\theta}_t)$ and $\sigma(\boldsymbol{s}_t,\boldsymbol{\theta}_t)$, showing that both functions are characterized by infinite differentiability, with their first-order derivatives being bounded, i.e., $\mu, \sigma \in C_b^{\infty,1}$. Consequently, once the initial values are given, a unique solution to the system (3) is guaranteed to exist. We use $\Phi_{0,t_2}(\boldsymbol{s}_0,\boldsymbol{\theta}_0)$ to represent the solution of the SDEs in (3) at time t_2 and it is called forward flow satisfying the property:

$$\Phi_{0,t_2}(\boldsymbol{s}_0,\boldsymbol{\theta}_0) = \Phi_{0,t}(\Phi_{t,t_2}(\boldsymbol{s}_0,\boldsymbol{\theta}_0)) \quad \text{for} \quad 0 \le t \le t_2.$$

To simulate $\Phi_{0,t_2}(\boldsymbol{s}_0,\boldsymbol{\theta}_0)$, we consider its calculus form based on Equation (3):

$$
\begin{aligned}
\Phi_{0,t_2}(\boldsymbol{s}_0,\boldsymbol{\theta}_0) &= (\boldsymbol{s}_0,\boldsymbol{\theta}_0)^\top + \int_0^{t_2} (\mu(\Phi_{0,t}(\boldsymbol{s}_0,\boldsymbol{\theta}_0)), f(\Phi_{0,t}(\boldsymbol{s}_0,\boldsymbol{\theta}_0)))^\top dt \\
&+ \int_0^{t_2} (\sigma(\Phi_{0,t}(\boldsymbol{s}_0,\boldsymbol{\theta}_0)), g(\Phi_{0,t}(\boldsymbol{s}_0,\boldsymbol{\theta}_0)))^\top \circ dW_t,
\end{aligned}
\tag{4}
$$

where $\circ dW_t$ represents the Stratonovich stochastic integral. For a continuous semimartingale $\{f_t\}_{t<T}$ adapted to the forward filtration $\{\mathscr{F}_{0,t}\}_{t<T}$, the Stratonovich stochastic integral is:

$$\int_0^T f_t \circ dW_t = \lim_{|\Pi| \to 0} \sum_{i=1}^N \frac{(f_{t_{i-1}} + f_{t_i})}{2}(W_{t_i} - W_{t_{i-1}}),$$

where $\Pi = \{0 = t_0 < 0 < \ldots < t_N = T\}$ is a partition of time interval $[0,T]$ and $|\Pi| = \max_n(t_n - t_{n-1})$. The reason to introduce Stratonovich stochastic integral is that we could generate the inverse flow $\psi_{0,t_2} \equiv \Phi_{0,t_2}^{-1}$ from SDE system (3) based on Equation (4) (Kunita 2019):

$$
\begin{aligned}
\psi_{0,t_2}(\boldsymbol{s}_{t_2},\boldsymbol{\theta}_{t_2}) &= (\boldsymbol{s}_{t_2},\boldsymbol{\theta}_{t_2})^\top - \int_0^{t_2} (\mu(\psi_{t,t_2}(\boldsymbol{s}_{t_2},\boldsymbol{\theta}_{t_2})), f(\psi_{t,t_2}(\boldsymbol{s}_{t_2},\boldsymbol{\theta}_{t_2})))^\top dt \\
&- \int_0^{t_2} (\sigma(\psi_{t,t_2}(\boldsymbol{s}_{t_2},\boldsymbol{\theta}_{t_2})), g(\psi_{t,t_2}(\boldsymbol{s}_{t_2},\boldsymbol{\theta}_{t_2})))^\top \circ d\widetilde{W}_t,
\end{aligned}
\tag{5}
$$

where $\widetilde{W}_t$ is the backward Wiener process defined as $\widetilde{W}_t = W_t - W_T$ for any $t < T$. It is adapted to the backward filtration $\{\mathscr{F}_{t,T}\}_{t<T}$. The difference between Equations (4) and (5) is only the negative sign, and such symmetry is attributed to the use of Stratonovich stochastic integral.

We could further define a scalar loss function $\mathscr{L}$ of $\Phi_{0,T}(s_0, \theta_0)$. Then, the loss for system (3) becomes $\mathscr{L}(\Phi_{0,T}(s_0, \theta_0))$ and the sensitivity coefficient for s_0 and θ_0 becomes,

$$A_{0,T}(s_0, \theta_0) \equiv \left(\frac{\partial \mathscr{L}(\Phi_{0,T}(s_0, \theta_0))}{\partial s_0}, \frac{\partial \mathscr{L}(\Phi_{0,T}(s_0, \theta_0))}{\partial \theta_0} \right)^\top. \tag{6}$$

Then, based on the chain rule, we have

$$A_{0,T}(s_0, \theta_0) = \nabla_\Phi \mathscr{L}(\Phi_{0,T}(s_0, \theta_0)) \nabla \Phi_{0,T}(s_0, \theta_0). \tag{7}$$

For $\nabla \Phi_{0,T}(s_0, \theta_0)$ in (6), we first derive $\left(\frac{\partial s_T}{\partial s_0}, \frac{\partial \theta_T}{\partial \theta_0} \right)^\top$ under the assumption that $(s_T, \theta_T)^\top$ is deterministic, i.e., not dependent on the Wiener process W_t. We then extend to the case that $(s_T, \theta_T)^\top$ is stochastically obtained from forward flow $\Phi_{0,T}(s_0, \theta_0)$ from SDE system (3) and derive $\left(\frac{\partial \Phi_{0,T}(s_0, \theta_0)}{\partial s_0}, \frac{\partial \Phi_{0,T}(s_0, \theta_0)}{\partial \theta_0} \right)^\top$, which implements the structural information in the stochastic reaction network through a dual process of forward and backward propagation.

Then, for $\nabla_\Phi \mathscr{L}(\Phi_{0,T}(s_0, \theta_0))$ in (7), it is determined by the loss function $\mathscr{L}$. Denote the observed bioprocess states as s^c, the underlying true parameters of the real reaction network as θ^c and the output as $O(s^c)$ which is a scalar function of states. Here the superscript "c" indicates data and parameters from the real system. Then the simulated output is denoted by $O(s)$, where s is the simulated state with initial state $s_0 = s_0^c$ and θ_0 as the input of simulation model. $\mathscr{L}$ is a scalar function measuring the difference of observed and simulated outputs, such as mean squared error (MSE) , *i.e.*, $\mathrm{E}[(O(s^c) - O(s))^2 | s_0, \theta_0]$. Then we could generate $\nabla_\Phi \mathscr{L}(\Phi_{0,T}(s_0))$. Combining the above results, we can derive the sensitivities $A_{0,T}(s_0, \theta_0)$, which can identify the key contributor to the model prediction MSE.

3 METAMODELING AND PREDICTION

In this section, we will derive the expression of μ and σ, we start with two-step transition from initial states,

$$
\begin{aligned}
\mathrm{E}(s_2 | s_0, \theta_0) &= \int p(s_1, \theta_1 | s_0, \theta_0) \mathrm{E}(s_2 | s_1, \theta_1) d(s_1, \theta_1) = \int p(s_1, \theta_1 | s_0, \theta_0)[s_1 + N\nu(s_1, \theta_1)\Delta t] d(s_1, \theta_1) \\
&= \mathrm{E}(s_1 | s_0, \theta_0) + \int p(s_1, \theta_1 | s_0, \theta_0) N\nu(s_1, \theta_1)\Delta t d(s_1, \theta_1).
\end{aligned}
\tag{8}
$$

By combining Equation (8) and $\mathrm{E}(s_1 | s_0, \theta_0) = s_0 + N\nu(s_0, \theta_0)\Delta t$, we have

$$\mathrm{E}(s_2 | s_0, \theta_0) = s_0 + N\nu(s_0, \theta_0)\Delta t + N \int p(s_1, \theta_1 | s_0, \theta_0)\nu(s_1, \theta_1)\Delta t d(s_1, \theta_1). \tag{9}$$

We need to calculate $\int p(s_1, \theta_1 | s_0, \theta_0)\nu(s_1, \theta_1)\Delta t d(s_1, \theta_1)$ in Equation (9), i.e., the expectation of flux rate after one step transition. For each j-th reaction, its reaction rate $v_j(s_1, \theta_1)$ is related to s_1^j, and by applying the MM kinetics in (1), we have

$$
\begin{aligned}
\int p(s_1, \theta_1 | s_0, \theta_0) v_j(s_1, \theta_1)\Delta t d(s_1, \theta_1) &= \int p(s_1^j, \theta_1 | s_0, \theta_0) \frac{V_{max}^j s_1^j}{K_m^j + s_1^j} \Delta t d(s_1, \theta_1) \\
&= \int p(s_1^j, \theta_1 | s_0, \theta_0) \left(V_{max}^j - \frac{V_{max}^j K_m^j}{K_m^j + s_1^j} \right) \Delta t d(s_1, \theta_1) \\
&= V_{max}^j \Delta t - V_{max}^j K_m^j \Delta t \int p(s_1^j, \theta_1 | s_0, \theta_0) \frac{1}{K_m^j + s_1^j} d(s_1, \theta_1).
\end{aligned}
\tag{10}
$$

To compute the second term in Equation (10), we first consider the probability distribution function $p(s_1^j, \boldsymbol{\theta}_1 | \boldsymbol{s}_0, \boldsymbol{\theta}_0)$. Since $\boldsymbol{\theta}_t$ is constant over time, we will put our focus on $p(s_1^j | \boldsymbol{s}_0, \boldsymbol{\theta}_0)$. As shown in Section 2.1, we've already derived the state transition model in Equation (2). This model adopts a truncated normal distribution, which is motivated by the fact that the concentration of molecules, represented by s_1^j, cannot be negative in reality. Furthermore, the introduction of the truncated normal distribution is due to the absence of an inverse moment for the standard normal distribution. To address this issue, we will first review Theorem 1 and then apply it to approximate the second term in Equation (10).

Theorem 1 (Hall (1979)) Suppose $\mu > 0$ and X is a random variable with density

$$\phi(X) = \begin{cases} \frac{k}{\sqrt{2\pi}\sigma} e^{\frac{-(X-\mu)^2}{2\sigma^2}}, X \geq a > 0, \\ 0, X < a, \end{cases}$$

where k is the normalization constant. Then for each $\hat{\sigma} = \frac{\sigma}{\mu} \leq \frac{1}{5}$ and any value of a satisfying the inequality $\hat{\sigma}^2 \leq \frac{a}{\mu} \leq \frac{1}{25}$, we have: $\mu E[X^{-1}] = I(\hat{\sigma}) + e_1, |e_1| < 8000\hat{\sigma}^{12} < 3.3 \times 10^{-5}$, where $I(\hat{\sigma})$ is given in terms of Dawson's Integral by the following equations:

$$I(\hat{\sigma}) = \frac{\sqrt{2}}{\hat{\sigma}} D\left(\frac{1}{\sqrt{2}\hat{\sigma}}\right), \quad D(x) = e^{-x^2} \int_0^x e^{t^2} dt.$$

Before applying Theorem 1, we rewrite Equation (10) by changing the variable,

$$\int p(s_1^j | \boldsymbol{s}_0, \boldsymbol{\theta}_0) \frac{1}{K_m^j + s_1^j} ds_1^j \overset{\hat{s}_1^j = s_1^j + K_m^j}{=\!=\!=\!=} \int p(\hat{s}_1^j | \boldsymbol{s}_0, \boldsymbol{\theta}_0) \frac{1}{\hat{s}_1^j} d\hat{s}_1^j.$$

When $\boldsymbol{s}_0 \gg \Delta t$, both conditions $\hat{\sigma} \leq \frac{1}{5}$ and $\hat{\sigma}^2 \leq \frac{a}{\mu} \leq \frac{1}{25}$ could be satisfied. Thus, Theorem 1 is applicable to the approximation of $\int p(s_1^j | \boldsymbol{s}_0, \boldsymbol{\theta}_0) \frac{1}{K_m^j + s_1^j} ds_1^j$, which means:

$$\int p(s_1^j | \boldsymbol{s}_0, \boldsymbol{\theta}_0) \frac{1}{K_m^j + s_1^j} ds_1^j \approx \frac{I(\hat{\sigma}_0^j)}{\mu_0^j}, \tag{11}$$

$$\mu_0^j = s_0^j + \sum_{k=1}^{R} \boldsymbol{N}_{j,k} \boldsymbol{v}_k(\boldsymbol{s}_0, \boldsymbol{\theta}_0) \Delta t + K_m^j, \quad \sigma_0^j = \sqrt{\sum_{k=1}^{R} \boldsymbol{N}_{j,k}^2 \boldsymbol{v}_k(\boldsymbol{s}_0, \boldsymbol{\theta}_0) \Delta t}, \quad \hat{\sigma}_0^j = \frac{\sigma_0^j}{\mu_0^j}.$$

Since Dawson's Integral does not have an analytic expression, we introduce an analytic approximation of Dawson's Integral (Filobello-Nino et al. 2019): $D(x) \approx \frac{1}{2x} + \frac{1}{4x^3}$ for $x > 2.68$. Based on Filobello-Nino's results, the approximation above can control the relative error below 2.5% (Filobello-Nino et al. 2019). Since we follow Theorem 1's assumption that $\hat{\sigma}_0 \leq \frac{1}{5}$, we could derive $\frac{1}{\sqrt{2}\hat{\sigma}_0} \geq \frac{1}{\frac{\sqrt{2}}{5}} > 2.68$. Thus:

$$\frac{I(\hat{\sigma}_0^j)}{\mu_0^j} = \frac{\frac{\sqrt{2}}{\hat{\sigma}_0^j} D\left(\frac{1}{\sqrt{2}\hat{\sigma}_0^j}\right)}{\mu_0^j} \approx \frac{\frac{\sqrt{2}}{\hat{\sigma}_0^j} \frac{\sqrt{2}\hat{\sigma}_0^j}{2}\left(1 + \frac{(\sqrt{2}\hat{\sigma}_0^j)^2}{2}\right)}{\mu_0^j} = \frac{1 + (\hat{\sigma}_0^j)^2}{\mu_0^j} = \frac{(\sigma_0^j)^2 + (\mu_0^j)^2}{(\mu_0^j)^3}. \tag{12}$$

Combining results from Equations (9), (10), (11), and (12), we can get,

$$E(\boldsymbol{s}_2 | \boldsymbol{s}_0, \boldsymbol{\theta}_0) \approx f(\boldsymbol{v}(\boldsymbol{s}_0, \boldsymbol{\theta}_0)) \triangleq \boldsymbol{s}_0 + \boldsymbol{N}\boldsymbol{v}(\boldsymbol{s}_0, \boldsymbol{\theta}_0)\Delta t + \boldsymbol{N}V_{max}\Delta t - \boldsymbol{N}\Delta t\left(V_{max} \odot K_m \odot \frac{I(\hat{\sigma}_0)}{\mu_0}\right),$$

where $\frac{I(\hat{\sigma}_0)}{\mu_0}, V_{max}$ and K_m are R-dimensional vectors with jth component equal to $\frac{I(\hat{\sigma}_0^j)}{\mu_0^j}, V_{max}^j$ and K_m^j.

In addition, we can derive $\text{Var}(s_2|s_0, \boldsymbol{\theta}_0)$ following above process:

$$
\begin{aligned}
\text{Var}(s_2|s_0, \boldsymbol{\theta}_0) &= \int p(s_1, \boldsymbol{\theta}_1|s_0, \boldsymbol{\theta}_0) \text{Var}(s_2|s_1, \boldsymbol{\theta}_1) d(s_1, \boldsymbol{\theta}_1) \\
&= \int p(s_1, \boldsymbol{\theta}_1|s_0, \boldsymbol{\theta}_0) \left[\boldsymbol{N} diag(\boldsymbol{v}(s_1, \boldsymbol{\theta}_1)) \boldsymbol{N}^\top \Delta t \right] d(s_1, \boldsymbol{\theta}_1) \\
&= \boldsymbol{N} diag \left(V_{max} - \left(V_{max} \odot K_m \odot \frac{I(\hat{\sigma}_0)}{\mu_0} \right) \right) \boldsymbol{N}^\top \Delta t.
\end{aligned}
$$

Since we know the distribution of s_2 conditioned on s_0, we could extend it to time t and write the form of μ and σ:

$$ ds_t = \mu(s_t, \boldsymbol{\theta}_t) dt + \sigma(s_t, \boldsymbol{\theta}_t) dW_t, $$

$$ \mu(s_t, \boldsymbol{\theta}_t) = \text{E}[s_{t+2} - s_t|s_t, \boldsymbol{\theta}_t] \approx \boldsymbol{N}\boldsymbol{v}(s_t, \boldsymbol{\theta}_t) \Delta t + \boldsymbol{N}V_{max}\Delta t - \boldsymbol{N}\Delta t \left(V_{max} \odot K_m \odot \frac{I(\hat{\sigma}_t)}{\mu_t} \right), $$

$$ \sigma^2(s_t, \boldsymbol{\theta}_t) = \text{Var}[s_{t+2}|s_t, \boldsymbol{\theta}_t] \approx \boldsymbol{N} diag \left(V_{max} - \left(V_{max} \odot K_m \odot \frac{I(\hat{\sigma}_t)}{\mu_t} \right) \right) \boldsymbol{N}^\top \Delta t. $$

We could also calculate $\frac{\partial \mu(s_t, \boldsymbol{\theta}_t)}{\partial s_t}, \frac{\partial \mu(s_t, \boldsymbol{\theta}_t)}{\partial \boldsymbol{\theta}_t}, \frac{\partial \sigma(s_t, \boldsymbol{\theta}_t)}{\partial s_t}$ and $\frac{\partial \sigma(s_t, \boldsymbol{\theta}_t)}{\partial \boldsymbol{\theta}_t}$ which will be the input of the algorithm in Section 4.

4 LOCAL SENSITIVITY ANALYSIS

4.1 Convergence Analysis

We start from the gradient of the backward flow represented in Equation (5),

$$
\begin{aligned}
\triangledown \psi_{0,t_2}(s_{t_2}, \boldsymbol{\theta}_{t_2}) &= \triangledown s_{t_2} - \triangledown \int_0^{t_2} (\mu(\psi_{t,t_2}(s_{t_2}, \boldsymbol{\theta}_{t_2})), f(\psi_{t,t_2}(s_{t_2}, \boldsymbol{\theta}_{t_2})))^\top dt \\
&\quad - \triangledown \int_0^{t_2} (\sigma(\psi_{t,t_2}(s_{t_2}, \boldsymbol{\theta}_{t_2})), g(\psi_{t,t_2}(s_{t_2}, \boldsymbol{\theta}_{t_2})))^\top \circ d\widetilde{W}_t.
\end{aligned}
$$

The gradient $\triangledown s_0$ is the identity matrix $\mathbb{I}$ with d dimension, where d is the sum of the dimension of bioprocess states s_t and the parameter set $\boldsymbol{\theta}_t$. For the last two terms on the right side of equation, based on Proposition 2.4.3 and Theorem 3.4.3 from Kunita (2019), we can switch the order of derivative and integral, i.e.,

$$
\begin{aligned}
\triangledown \psi_{0,t_2}(s_{t_2}, \boldsymbol{\theta}_{t_2}) &= \mathbb{I}_d - \int_0^{t_2} (\triangledown_\psi \mu(\psi_{t,t_2}(s_{t_2}, \boldsymbol{\theta}_{t_2})), \triangledown_\psi f(\psi_{t,t_2}(s_{t_2}, \boldsymbol{\theta}_{t_2})))^\top \triangledown \psi_{t,t_2}(s_{t_2}, \boldsymbol{\theta}_{t_2}) dt \\
&\quad - \int_0^{t_2} (\triangledown_\psi \sigma(\psi_{t,t_2}(s_{t_2}, \boldsymbol{\theta}_{t_2})), \triangledown_\psi g(\psi_{t,t_2}(s_{t_2}, \boldsymbol{\theta}_{t_2})))^\top \triangledown \psi_{t,t_2}(s_{t_2}, \boldsymbol{\theta}_{t_2}) \circ d\widetilde{W}_t. \quad (13)
\end{aligned}
$$

Since ψ_{0,t_2} is the inverse function of Φ_{0,t_2}, we have $\Phi_{0,t_2}(\psi_{0,t_2}(s_{t_2}, \boldsymbol{\theta}_{t_2})) = (s_{t_2}, \boldsymbol{\theta}_{t_2})$. Therefore, by applying the chain rule, $\triangledown \Phi_{0,t_2}(\psi_{0,t_2}(s_{t_2}, \boldsymbol{\theta}_{t_2})) \triangledown \psi_{0,t_2}(s_{t_2}, \boldsymbol{\theta}_{t_2}) = \mathbb{I}_d$. Then by applying Stratonovich version of Itô's formula (Theorem 2.4.1 (Kunita 2019)) on Equation (13), we have,

$$ \frac{\partial \triangledown \Phi_{0,t_2}(\psi_{0,t_2}(s_{t_2}, \boldsymbol{\theta}_{t_2}))}{\partial \triangledown \psi_{0,t_2}(s_{t_2}, \boldsymbol{\theta}_{t_2})} = \frac{\partial (\triangledown \psi_{0,t_2}(s_{t_2}, \boldsymbol{\theta}_{t_2}))^{-1}}{\partial \triangledown \psi_{0,t_2}(s_{t_2}, \boldsymbol{\theta}_{t_2})} = -\frac{1}{\triangledown \psi_{0,t_2}(s_{t_2}, \boldsymbol{\theta}_{t_2})^2}, $$

where

$$\nabla \Phi_{0,t_2}(\psi_{0,t_2}(s_{t_2}, \theta_{t_2}))$$
$$= \mathbb{I}_d - \int_0^{t_2} (\nabla_\psi \mu(\psi_{t,t_2}(s_{t_2}, \theta_{t_2})), \nabla_\psi f(\psi_{t,t_2}(s_{t_2}, \theta_{t_2})))^\top \nabla \psi_{t,t_2}(s_{t_2}, \theta_{t_2}) \frac{\partial \nabla \Phi_{t,t_2}(\psi_{t,t_2}(s_{t_2}, \theta_{t_2}))}{\partial \nabla \psi_{t,t_2}(s_{t_2}, \theta_{t_2})} dt$$
$$- \int_0^{t_2} (\nabla_\psi \sigma(\psi_{t,t_2}(s_{t_2}, \theta_{t_2})), \nabla_\psi g(\psi_{t,t_2}(s_{t_2}, \theta_{t_2})))^\top \nabla \psi_{t,t_2}(s_{t_2}, \theta_{t_2}) \frac{\partial \nabla \Phi_{t,t_2}(\psi_{t,t_2}(s_{t_2}, \theta_{t_2}))}{\partial \nabla \psi_{t,t_2}(s_{t_2}, \theta_{t_2})} \circ d\widetilde{W}_t$$
$$= \mathbb{I}_d + \int_0^{t_2} (\nabla_\psi \mu(\psi_{t,t_2}(s_{t_2}, \theta_{t_2})), \nabla_\psi f(\psi_{t,t_2}(s_{t_2}, \theta_{t_2})))^\top \nabla \Phi_{t,t_2}(\psi_{t,t_2}(s_{t_2}, \theta_{t_2})) dt$$
$$+ \int_0^{t_2} (\nabla_\psi \sigma(\psi_{t,t_2}(s_{t_2}, \theta_{t_2})), \nabla_\psi g(\psi_{t,t_2}(s_{t_2}, \theta_{t_2})))^\top \nabla \Phi_{t,t_2}(\psi_{t,t_2}(s_{t_2}, \theta_{t_2})) \circ d\widetilde{W}_t. \tag{14}$$

Let $\widetilde{A}_{0,t_2}(s_{t_2}, \theta_{t_2}) = A_{0,t_2}(\psi_{0,t_2}(s_{t_2}, \theta_{t_2}))$. Then, by combining Equation (14) and

$$A_{0,t_2}(s_0, \theta_0) = \nabla \mathscr{L}(\Phi_{0,t_2}(s_0, \theta_0)) \nabla \Phi_{0,t_2}(s_0, \theta_0),$$

we can derive

$$\widetilde{A}_{0,t_2}(s_{t_2}, \theta_{t_2}) = A_{0,t_2}(\psi_{0,t_2}(s_{t_2}, \theta_{t_2})) = \nabla \mathscr{L}(\Phi_{0,t_2}(\psi_{0,t_2}(s_{t_2}, \theta_{t_2}))) \nabla \Phi_{0,t_2}(\psi_{0,t_2}(s_{t_2}, \theta_{t_2}))$$
$$= \nabla \mathscr{L}(s_{t_2}) + \int_0^{t_2} (\nabla_\psi \mu(\psi_{t,t_2}(s_{t_2}, \theta_{t_2})), \nabla_\psi f(\psi_{t,t_2}(s_{t_2}, \theta_{t_2})))^\top \widetilde{A}_{t,t_2}(s_{t_2}, \theta_{t_2}) dt$$
$$+ \int_0^{t_2} (\nabla_\psi \sigma(\psi_{t,t_2}(s_{t_2}, \theta_{t_2})), \nabla_\psi g(\psi_{t,t_2}(s_{t_2}, \theta_{t_2})))^\top \widetilde{A}_{t,t_2}(s_{t_2}, \theta_{t_2}) \circ d\widetilde{W}_t. \tag{15}$$

From Equations (5) and (15), we can see that the drift and diffusion terms for $\psi_{0,t_2}(s_{t_2}, \theta_{t_2})$ and $\widetilde{A}_{0,t_2}(s_{t_2}, \theta_{t_2})$ are $C_b^{\infty,1}$, which means the system $(\widetilde{A}_{0,t_2}(s_{t_2}, \theta_{t_2}), \psi_{0,t_2}(s_{t_2}, \theta_{t_2}))$ has a unique strong solution. Thus, we could define $\widetilde{A}_{0,T}(s_{t_2}, \theta_{t_2}) = F(s_{t_2}, \theta_{t_2}, W)$ where $W = \{W_t\}_{0 \le t \le T}$ is a path of Wiener process and $F : \mathbb{R}^d \times C([0,1], \mathbb{R}^J) \to \mathbb{R}^d$ is a deterministic measurable function, which is also called Itô map. In our case, the number of reactions denoted by J defines the dimension of the Wiener process deployed to simulate the stochastic reaction network. Note that $A_{0,t_2}(s_0, \theta_0) = \widetilde{A}_{0,t_2}(\Phi_{0,t_2}(s_0, \theta_0))$. In Section 2.2, we've already shown that $\Phi_{0,t_2}(s_0, \theta_0)$ has a unique strong solution. Similar to F, we define $G : \mathbb{R}^d \times C([0,1], \mathbb{R}^J) \to \mathbb{R}^d$ as the solution map for forward flow, i.e., $G(s_0, \theta_0, W) = \Phi_{0,T}(s_0, \theta_0)$. Then, apparently:

$$A_{0,T}(s_0, \theta_0) = \widetilde{A}_{0,T}(G(s_0, \theta_0, W)) = F(G(s_0, \theta_0, W), W).$$

4.2 Algorithm Development

Based on Equations (4), (5), and (15), with $\frac{\partial \mu(s_t, \theta_t)}{\partial s_t}$, $\frac{\partial \mu(s_t, \theta_t)}{\partial \theta_t}$, $\frac{\partial \sigma(s_t, \theta_t)}{\partial s_t}$ and $\frac{\partial \sigma(s_t, \theta_t)}{\partial \theta_t}$ as inputs, we could approximate $A_{0,T}(s_0, \theta_0)$. Since the system is described in Stratonovich integral, the most commonly used Euler-Maruyama Scheme is not applicable, which requires the system to be represented in Itô integral. Thus, we instead deployed the Euler-Heun method below:

$$(s_{t+1}, \theta_{t+1})^\top = (s_t, \theta_t)^\top + (\mu(s_t, \theta_t), f(s_t, \theta_t))^\top \Delta t$$
$$+ \frac{1}{2} \left[(\sigma(s_t, \theta_t), g(s_t, \theta_t))^\top + (\sigma(\bar{s}_t, \overline{\theta}_t), g(\bar{s}_t, \overline{\theta}_t))^\top \right] (W_t - W_{t-1}),$$
$$(\bar{s}_t, \overline{\theta}_t)^\top = (s_t, \theta_t)^\top + (\sigma(s_t, \theta_t), g(s_t, \theta_t))^\top (W_t - W_{t-1}).$$

Based on the schemes, the procedure of adjoint sensitivity analysis on the SDEs-based mechanistic model is described in Algorithm 1. The algorithm is built on the Euler-Heun Scheme, using the gradients of

the drift and diffusion terms as inputs, and produces the expected gradient of the simulation prediction MSE as its output. The algorithm begins with the generation of sample paths of the Wiener process W_t^n in Step 1. Subsequently, as depicted in Steps 2 and 3, we conduct simulations under s_0 and θ_0 to obtain the simulated output denoted by $O(s)$. In the empirical study, the observed states s^c are generated from the model with a predefined set of true parameters θ^c, using the same path W_t^n to control randomness. W_t^n is queried again in the backward pass $\widetilde{W}_t^n$ to generate the backward flow $\psi_{(t-1)\Delta t,T}^n(\Phi_{0,T}^n(s_0,\theta_0))$. By combining observed output, simulated output, and the backward flow, we derive sensitivity $\widetilde{A}_{(t-1)\Delta t,T}^n(\Phi_{0,T}^n(s_0,\theta_0))$ corresponding to path W_t^n in Step 6. By repeating these steps, we generate N sample paths, and the mean value of $\widetilde{A}_{(t-1)\Delta t,T}^n(\Phi_{0,T}^n(s_0,\theta_0))$ for each sample path serves as the final sensitivity, as shown in Step 7.

Algorithm 1: Adjoint Sensitivity Analysis on SDEs based on MSE of Output Prediction.

Input:

- Observed states s^c from the real system.
- Initial parameters θ_0.
- Start time t_0, end time T.
- Gradient of drift and diffusion term $\frac{\partial \mu(s_t,\theta_t)}{\partial s_t}$, $\frac{\partial \mu(s_t,\theta_t)}{\partial \theta_t}$, $\frac{\partial \sigma(s_t,\theta_t)}{\partial s_t}$ and $\frac{\partial \sigma(s_t,\theta_t)}{\partial \theta_t}$.
- Number of iterations N.
- Grid size Δt, number of steps $S = \frac{T}{\Delta t}$.

Output: $\mathrm{E}\left[\left(\frac{\partial (O(s)-O(s^c))^2}{\partial s_0}, \frac{\partial (O(s)-O(s^c))^2}{\partial \theta_0} \right)^{\top} \middle| (s_0,\theta_0) \right]$.

1. Calculate the real system output $O(s^c)$ and define $s_0 = s_0^c$;

for $n = 1,\ldots,N$ **do**

    for $t \leftarrow 1$ **to** S **do**

        2. Generate a sample path of Wiener process W_t^n with grid size Δt;

        3. Based on W_t^n, generate $\Phi_{0,t\Delta t}^n(s_0,\theta_0)$ from Equation (4) with Euler-Heun Scheme;

    4. Generate backward pass $\widetilde{W}_t^n$ based on forward pass W_t^n;

    5. Calculate the model prediction output $O(\Phi_{0,T}^n(s_0,\theta_0))$;

    for $t \leftarrow S$ **to** 1 **do**

        6. Based on $\widetilde{W}_t^n$, generate $\psi_{(t-1)\Delta t,T}^n(\Phi_{0,T}^n(s_0,\theta_0))$ from Equation (5) with Euler-Heun Scheme;

        7. Based on $\psi_{(t-1)\Delta t,T}^n(\Phi_{0,T}^n(s_0,\theta_0))$, generate $\widetilde{A}_{(t-1)\Delta t,T}^n(\Phi_{0,T}^n(s_0,\theta_0))$ from Equation (15) with Euler-Heun Scheme, $\nabla \mathscr{L}(s_T) = 2|O(s)-O(s^c)|$;

8. Calculate $\mathrm{E}\left[\left(\frac{\partial (O(s)-O(s^c))^2}{\partial s_0}, \frac{\partial (O(s)-O(s^c))^2}{\partial \theta_0} \right)^{\top} \middle| (s_0,\theta_0) \right] = \frac{1}{N} \sum_{n=1}^{N} \widetilde{A}_{0,T}^n(\Phi_{0,T}^n(s_0,\theta_0))$.

From Theorem 2, as the grid or time size $\Delta t \to 0$, the output of Algorithm 1 will converge pathwise (i.e., almost surely) from any fixed starting point to true local sensitivity. This sensitivity analysis, accounting for the interdependency of multi-scale mechanistic model, accelerates the search for the optimal parameters θ that minimizes the model prediction's MSE.

Theorem 2 (Li et al. (2020)) Suppose the schemes F_h and G_h, with $h = \frac{T}{L}$ denoting the size of the grid for solvers, satisfy the following conditions:

- $F_h(s_t,\theta_t,W) \to F(s_t,\theta_t,W)$, $G_h(s_t,\theta_t,W) \to G(s_t,\theta_t,W)$ as $h \to 0$;
- For $\forall M > 0$, $\sup_{s_t;\theta_t \leq M} |F_h(s_t,\theta_t,W) - F(s_t,\theta_t,W)| \to 0$ as $h \to 0$.
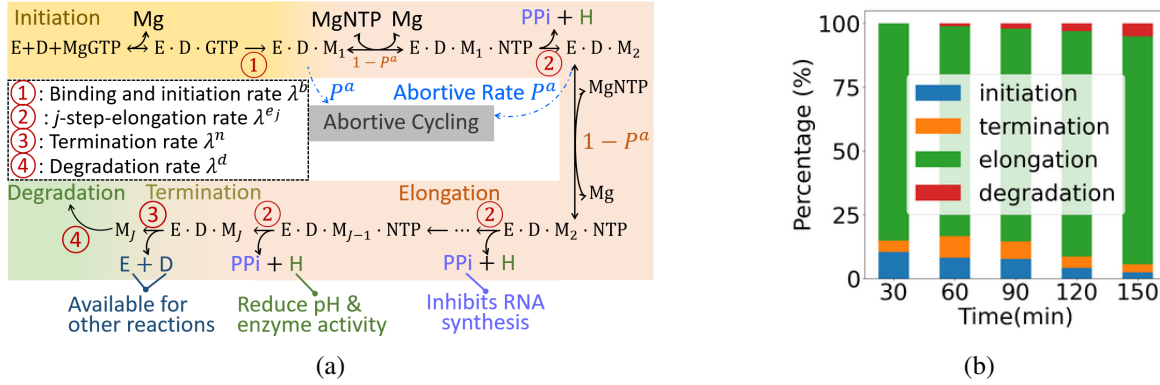
(a)

(b)

Figure 1: Interaction between subprocesses of the IVT system and their relative importance. (a) The reaction network for the IVT process (Wang et al. 2023). (b) Relative importance of the IVT system subprocesses parameter estimation on the output predictions at different times based on SA results.

Then for any starting point s_0 and $\boldsymbol{\theta}_0$, we have,

$$F_h(G_h(s_0, \boldsymbol{\theta}_0, W), W) \to F(G(s_0, \boldsymbol{\theta}_0, W), W) = A_{0,T}(s_0, \boldsymbol{\theta}_0), \forall s_0 \in \mathbb{R}^d.$$

5 EMPIRICAL STUDY

In this section, we conduct the empirical study by using the In-Vitro Transcription (IVT) system example to validate our SA algorithm. The bioprocess model from Wang et al. (2023) with the true model parameters denoted by $\boldsymbol{\theta}^c$ is used to assess the performance of proposed adjoint SA algorithm. We start the section by providing an in-depth interpretation of the sensitivity analysis results obtained from Algorithm 1. We then assess the efficacy of our approach by showing the decrease rate in the simulation output prediction MSE over multiple search iterations of model parameters.

The IVT process is mainly composed of four sub-processes: (1) initiation; (2) elongation; (3) termination; and (4) degradation of generated RNA molecule product. The molecular reaction network is illustrated in Figure 1a. One insightful interpretation of the sensitivity analysis results is to investigate the relative importance of each subprocess, which is inferred from the sensitivity of their corresponding parameters. Aligned with this reasoning, Figure 1b presents the relative importance of each subprocess within the IVT system. Notably, the result underscores the pivotal role of elongation reactions in the IVT process regardless of termination time T. Furthermore, it is observed that the impact of degradation increases over time, which fits our expectation since the synthesis rate of mRNA transcripts decreases as the IVT process approaches completion. Concurrently, the accumulation of generated RNA transcripts leads to an increase in the degradation rate at the system level. Therefore, the relevance of parameters associated with the degradation model becomes increasingly significant over time. Conversely, the influence of the initiation process appears to diminish over time as the raw materials (i.e., NTPs) used to synthesize RNA products are consumed.

To further assess the performance of the proposed adjoint SA, we benchmark our approach with a state-of-art gradient estimation approach (Fu 2015) that estimates the gradient by using finite difference (FD) simulation estimator, i.e.,

$$\frac{O(\Phi_{0,T}^n(s_0, \hat{\boldsymbol{\theta}} + c_k e_k)) - O(\Phi_{0,T}^n(s_0, \hat{\boldsymbol{\theta}} - c_k e_k))}{2c_k}$$

for the n-th iteration of parameter search, where c_k is the perturbation and e_k denotes a unit vector with the element in k-th dimension equal to 1 and all remaining elements equal to 0. In our case, we define $c_k = p \times \hat{\boldsymbol{\theta}}_k$ where p is a percentage selected to be 5%, 10%, and 20%.

Table 1: Final RNA yield and its 95% confidence interval with $\hat{\boldsymbol{\theta}}$ estimated with different approaches.

Approach	Observed	Adjoint sensitivity	Finite difference (p=5%)	Finite difference (p=10%)	Finite difference (p=20%)
Yield (μM)	73.48 ± 0.28	73.05 ± 0.27	72.66 ± 0.32	74.37 ± 0.39	74.51 ± 0.31

(a) MSE

(b) Relative error between $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}^c$
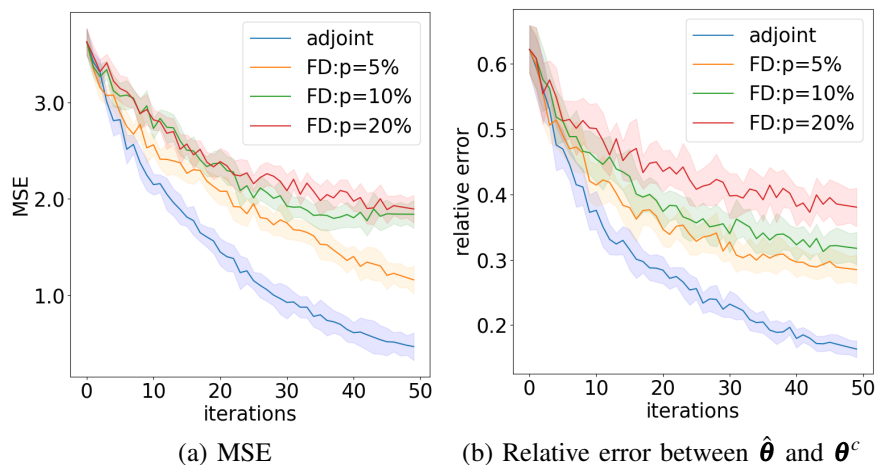
Figure 2: The results of IVT output prediction MSE and model parameter estimation relative error obtained by using adjoint SA and finite difference approaches.

Figure 2a illustrates the convergence of the Mean Squared Error (MSE) of IVT process output prediction over the iterations. It's evident that as the number of iterations increases, the adjoint SA approach surpasses the finite difference gradient estimation approach. Nonetheless, same as the finite difference estimator, the convergence speed gradually decreases, and the lower bound shows no further improvement once it reaches approximately $0.4(g/L)^2$, which could be associated with a fixed step size, i.e., $\alpha = 0.01$, used during the gradient search process. This slowdown could also be attributed to our MSE objective measure that solely relies on the final RNA yield. This discrete data provide very limited information about the interactions within the complex reaction network. We also show the specific RNA yield prediction by using the model with parameters $\hat{\boldsymbol{\theta}}$ estimated with different approaches in Table 1. In contrast, the final RNA yield and its 95% confidence interval (CI) obtained from the system with true parameters $\boldsymbol{\theta}^c$ are 73.48 ± 0.28. There is an overlap between the 95% confidence interval of the adjoint SA approach's predictions and the observed RNA yield, indicating no statistically significant difference between them. In contrast, the finite difference approach shows a significant prediction discrepancy from the observed values, demonstrating that the adjoint SA approach outperforms the finite difference method.

Furthermore, we present the results of the expected relative error between $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}^c$, i.e., $E\left[|\frac{\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^c}{\boldsymbol{\theta}^c}|\right]$, in Figure 2b. Our approach consistently outperforms the finite difference estimator, as the expected relative error diminishes more significantly after 50 iterations compared with the finite difference estimator regardless of the p value chosen. However, similar to the MSE results, the lower bound of the relative difference stagnates at approximately 15%, which could be attributed to the same limitation mentioned above.

6 CONCLUSION

In this paper, we present an adjoint sensitivity analysis approach on the stochastic differential equations (SDEs) designed to expedite the calibration of stochastic reaction network mechanistic model parameters. We first formulate the multi-scale bioprocess stochastic reaction network in the form of SDEs. Then, we develop an adjoint SA algorithm on SDEs for computing local sensitivities of model parameters and validate the convergence of the algorithm. Our empirical study underscores the importance of model parameters of enzymatic stochastic reaction networks and provides empirical evidence of the efficacy of our approach. Moving forward, we aim to leverage this capability by refining the loss function. Rather than solely relying on the Mean Squared Error (MSE) of the final output, we intend to integrate additional bioprocess states and consider the disparities between simulated and real process trajectories, which promises to enhance the effectiveness and accuracy of our method.

REFERENCES

Cullen, A. C. and H. C. Frey. 2002. "Probabilistic Techniques in Exposure Assessment: A Handbook for Dealing with Variability and Uncertainty in Models and Inputs". *International Journal of General Systems* 31(1):97–110.

Filobello-Nino, U. A., H. Vazquez-Leal, A. Herrera-May, R. Ambrosio, R. Castañeda-Sheissa, V. Jimenez-Fernandez *et al*. 2019. "A Handy, Accurate, Invertible and Integrable Expression for Dawson's Function". *Acta Universitaria* 29:18.

Fu, M. C. 2015. "Stochastic Gradient Estimation". In *Handbook of Simulation Optimization*, edited by M. C. Fu, 105–147. New York: Springer.

Golightly, A. and D. J. Wilkinson. 2005. "Bayesian Inference for Stochastic Kinetic Models Using a Diffusion Approximation". *Biometrics* 61(3):781–788.

Hall, R. L. 1979. "Inverse Moments for a Class of Truncated Normal Distributions". *The Indian Journal of Statistics* 41(1-2):66–76.

Kedem, G. 1980. "Automatic Differentiation of Computer Programs". *ACM Transactions on Mathematical Software* 6(2):150–165.

Kramer, M. A., H. Rabitz, J. M. Calo, and R. J. Kee. 1984. "Sensitivity Analysis in Chemical Kinetics: Recent Developments and Computational Comparisons". *International Journal of Chemical Kinetics* 16(5):559–578.

Kunita, H. 2019. *Stochastic Flows and Jump-Diffusions*. New York: Springer.

Kyriakopoulos, S., K. S. Ang, M. Lakshmanan, Z. Huang, S. Yoon, R. Gunawan *et al*. 2018. "Kinetic Modeling of Mammalian Cell Culture Bioprocessing: The Quest to Advance Biomanufacturing". *Biotechnology Journal* 13(3):1700229.

Li, X., T.-K. L. Wong, R. T. Chen, and D. Duvenaud. 2020. "Scalable Gradients for Stochastic Differential Equations". In *International Conference on Artificial Intelligence and Statistics*, August 26th-28th, Virtual Conference, 3870–3882.

Michaelis, L. and M. L. Menten. 2007. "Die Kinetik der Invertinwirkung". *Biochemische Zeitschrift* 49:333 – 369.

Sobol, I. M. 2001. "Global Sensitivity Indices for Nonlinear Mathematical Models and Their Monte Carlo Estimates". *Mathematics and Computers in Simulation* 55(1-3):271–280.

Wang, K., W. Xie, and H. Zheng. 2023. "Stochastic Molecular Reaction Queueing Network Modeling for in Vitro Transcription Process". In *2023 Winter Simulation Conference (WSC)*, 1900–1911 https://doi.org/10.1109/WSC60868.2023.10408139.

Zi, Z. 2011. "Sensitivity Analysis Approaches Applied to Systems Biology Models". *IET Systems Biology* 5(6):336–346.

AUTHOR BIOGRAPHIES

Keilung Choy is Ph.D. candidate of Mechanical and Industrial Engineering (MIE) at Northeastern University (NU) with expertise in machine learning, computer simulation, and digital twin. Email: choy.k@northeastern.edu.

WEI XIE is an assistant professor in Mechanical and Industrial Engineering at Northeastern University. Her research interests include AI/ML, computer simulation, data analytics, and stochastic optimization. Her email address is w.xie@northeastern.edu.