

SENSITIVITY ANALYSIS ON INTERACTION EFFECTS OF POLICY-AUGMENTED BAYESIAN NETWORKS

Junkai Zhao¹, Jun Luo¹, Wei Xie², and Zixuan Bai¹

¹Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai, China

² Department of Mechanical and Industrial Engineering, Northeastern University, Boston, MA, USA

ABSTRACT

Biomanufacturing plays an important role in supporting public health and the growth of the bioeconomy. Modeling and studying the interaction effects among various input variables is very critical for obtaining a scientific understanding and process specification in biomanufacturing. In this paper, we use the Shapley-Owen indices to measure the interaction effects for the policy-augmented Bayesian network (PABN) model, which characterizes the risk- and science-based understanding of production bioprocess mechanisms. In order to facilitate efficient interaction effect quantification, we propose a sampling-based simulation estimation framework. In addition, to further improve the computational efficiency, we develop a non-nested simulation algorithm with sequential sampling, which can dynamically allocate the simulation budget to the interactions with high uncertainty and therefore estimate the interaction effects more accurately under a total fixed budget setting.

1 INTRODUCTION

The biomanufacturing industry has made substantial progress in producing biotherapeutics to treat serious illnesses such as cancer, autoimmune disorders, and infectious diseases. Despite these advancements, modern biomanufacturing has encountered several critical challenges. The challenges primarily arise from the complex nature of biomanufacturing processes, which are characterized by high levels of complexity, significant variability, and a lack of comprehensive process data. Biotherapeutics are usually produced in living organisms, particularly cells, where the underlying biological processes are highly complicated. The manufacturing process involves a sequence of intricately connected unit operations, which adds complexity to the production procedures. Furthermore, the collection of process data is significantly limited in biomanufacturing settings due to the lengthy experiment time and the customized nature of new biotherapeutics, which hinders the development of a sophisticated biomanufacturing process.

Due to the high variability, measurement error and lack of enough data, sensitivity analysis is employed to guide the process specification and risk control for biomanufacturing (Xie et al. 2023). Sensitivity analysis studies the fluctuation in a model output caused by changing the model inputs, which can be classified into two classes: local sensitivity analysis and global sensitivity analysis. Local sensitivity analysis focuses on the sensitivity of the output to perturbing the input around a particular value. However, when the input is random, it is difficult to determine local sensitivity without knowing which value the input can actually take. On the other hand, global sensitivity analysis quantifies the variability in the output resulting from the uncertainty in the inputs across the whole range of possible input values. It can help identify the most sensitive random factors in the context of biomanufacturing. Sobol' (1993) first introduced the Sobol' indices to quantify how an input influences an output's variance. Under the assumption of independence between the input variables, the Sobol' indices compare the conditional variance of the output knowing some of the input variables to the total variance of the output. Owen (2014) proposed new sensitivity indices called Shapley effects and showed that these new indices have many advantages over the Sobol' indices for dependent inputs. They are now widely used in global sensitivity analysis (Rabitti and Borgonovo

2020). To improve the computational efficiency of Shapley effects, Song et al. (2016) provided an efficient estimate algorithm with Monte Carlo simulation.

The aforementioned methods focus on measuring the importance of individual model inputs, without considering the interaction effects of different input variables. However, many interactions exist in the biomanufacturing process, including gene-gene interactions (Cordell 2009), cell-environment interactions (Aijaz et al. 2018) and so on. Some interactions may come from common underlying random factors and the systematic observation error of the sensors. Understanding the interactions can benefit the scientific understanding of the biomanufacturing process. For example, it allows one to discriminate whether an interaction is synergistic (i.e., two model inputs are related by positive cooperation) or antagonistic (i.e., two model inputs are related by negative cooperation) (Diouf et al. 2021). The quantity of the interaction effects can also be used to classify parameters and determine which type of studies (i.e., DoE, univariate, or none) should be performed during process characterization (Łacki et al. 2018). To measure the interaction effects of input variables, Rabitti and Borgonovo (2019) proposed the Shapley-Owen index, which generalizes the Shapley effects to the context of interaction quantification. They showed that the Shapley-Owen index can provide information about synergistic and antagonistic effects. However, they only test the performance of the Shapley-Owen index for the stochastic model with an analytical variance function and up to three inputs. As the number of inputs increases, the computational cost would become unaffordable.

In this paper, we consider how to conduct global sensitivity analysis for the interactions among different inputs in biomanufacturing based on the Shapley-Owen index. Our analysis is based on the policy augmented Bayesian network (PABN) proposed by Zheng et al. (2023), which can leverage the existing mechanisms and learn from real-world process data (Zhao et al. 2023). Building on PABN, we introduce the Shapley-Owen based sensitivity analysis to assess the interaction effect. Firstly, as the calculation of the Shapley-Owen index requires massive computational resources, we extend the method in Song et al. (2016) and propose a nested Monte Carlo simulation algorithm for Shapley interaction effect estimation in PABN. Compared to the brute force implementation of the Shapley-Owen index formula, our algorithm can benefit from approximation by sampling and reuse of calculation. Secondly, the sample size for each level of simulation can be hard to be determined without expert knowledge and the error analysis cannot be conducted analytically. To overcome these issues, we propose a non-nested estimator to estimate the Shapley interaction effect, which can facilitate error estimation and can be implemented without the determination of the multilevel sample sizes. In addition, the total number of interactions that need to be quantified grows in the square order of the number of factors, which can be extremely large as the number of factors increases. However, within those interactions, there may be some interactions that exhibit a low level of uncertainty or variability. In order to enhance the accuracy of estimation within a specified computational budget, we propose a sequential budget allocation algorithm that adaptively identifies which interaction has the most uncertainty and then allocates that interaction for one additional simulation replication. By allocating the computation budget sequentially, our algorithm can achieve higher accuracy using the same budget as the previous Shapley value estimation method. We summarize our contributions as follows:

- We propose a sampling-based global sensitivity analysis framework to quantify the impact of the interactions with the Shapley-Owen index in PABN model.
- We develop a non-nested simulation algorithm, which adopts the non-nested estimators and sequential budget allocation strategy to improve the efficiency.
- We demonstrate the performance of our methods in terms of both accuracy and efficiency with empirical study.

The paper is organized as follows. In Section 2, we present the general PABN model. We introduce the Shapley interaction effect and the nested estimation procedure for PABN model in Section 3. In Section 4, we develop the sequential budget allocation algorithm. Then, we study the empirical performance of two proposed algorithms in Section 5 and conclude this paper in Section 6.

2 GENERAL POLICY AUGMENTED KNOWLEDGE GRAPH

In this section, we introduce the general PABN model in Zheng et al. (2023) for the biomanufacturing process. A typical biomanufacturing process comprises multiple unit operations. Within these unit operations, two types of interacting factors play pivotal roles: one is critical quality attributes (CQAs), which can be considered as the states of the bioprocess (e.g., the concentrations of bioproduct); the other is critical process parameters (CPPs), which can be viewed as the actions of the bioprocess (e.g., the feeding rate). Both of them can significantly influence the system's outputs (e.g., drug quality and productivity). In this paper, we represent the CQAs at period t as $\mathbf{s}_t = (s_t^1, \dots, s_t^{d_s})$ and the CPPs at period t as $\mathbf{a}_t = (a_t^1, \dots, a_t^{d_a})$, where H is the total planning horizon, $1 \leq t \leq H$. The n th component of \mathbf{s}_t and \mathbf{a}_t are denoted by s_t^n and a_t^n , respectively. For simplicity, the dimensions of \mathbf{s}_t and \mathbf{a}_t , namely d_s and d_a respectively, are kept constant throughout this study, although they may be time-dependent in reality.

In bioprocesses, kinetic models based on the ordinary or partial differential equations (ODE/PDE) are commonly applied to characterize the dynamics of the system, which means \mathbf{s}_t evolves according to $\frac{d\mathbf{s}_t}{dt} = f_t(\mathbf{s}_t, \mathbf{a}_t)$. Specifically, the functional form of $f_t(\mathbf{s}_t, \mathbf{a}_t; \mathbf{w}_t)$ is expected to be known, while the parameters of the function \mathbf{w}_t , which can be estimated from data, are subject to uncertainty. Drawing upon prior knowledge of the existing ODE/PDE-based kinetics, Zheng et al. (2023) introduced the state transition hybrid model, referred to as, $\mathbf{s}_{t+1} = f_t(\mathbf{s}_t, \mathbf{a}_t; \boldsymbol{\beta}_t) + \mathbf{e}_{t+1}$, where the residual \mathbf{e}_{t+1} is a random vector representing the uncontrollable factors. Suppose that the residuals follow a multivariate Gaussian distribution, i.e., $\mathbf{e} = (\mathbf{e}_1^\top, \mathbf{e}_2^\top, \dots, \mathbf{e}_H^\top) \sim \mathcal{N}(\mathbf{0}, \mathbf{V})$, where v_{ij} in i th row and j th column of \mathbf{V} denotes covariance between the i th component and the j th component of \mathbf{e} . The model parameters can be denoted by $\mathbf{w} = \{\{\boldsymbol{\beta}_t\}_{t=1}^{H-1}, \mathbf{V}\}$ and the distribution of the entire trajectory $\boldsymbol{\tau} = (\mathbf{s}_1, \mathbf{a}_1, \mathbf{s}_2, \mathbf{a}_2, \dots, \mathbf{s}_H)$ of the stochastic decision process (SDP) can be written as $p(\boldsymbol{\tau}) = p(\mathbf{s}_1) \prod_{t=1}^{H-1} p(\mathbf{s}_{t+1} | \mathbf{s}_t, \mathbf{a}_t) p(\mathbf{a}_t | \mathbf{s}_t)$. Given the prior distribution of the model parameters and the historical data $\mathcal{D} = \{\boldsymbol{\tau}^{(l)}\}_{l=1}^L$, where $\boldsymbol{\tau}^{(l)}$ denotes l th observed trajectory in experiments. We can derive the posterior distribution of model parameters $p(\mathbf{w} | \mathcal{D})$, which quantifies the uncertainty in parameter estimation due to the lack of information.

At each period t , we employ the policy π_t to determine the action. Given the policy parameters $\boldsymbol{\theta}_t$, the state vectors \mathbf{s}_t are mapped into the space of all potential action values with $\mathbf{a}_t = \pi_t(\mathbf{s}_t; \boldsymbol{\theta}_t)$. The parametric policy $\pi_{\boldsymbol{\theta}} = \{\pi_t\}_{t=1}^{H-1}$ is the collection of these mappings over the entire planning period and is fully characterized by $\boldsymbol{\theta} = \{\boldsymbol{\theta}_t\}_{t=1}^{H-1}$. Let $r_t(\mathbf{s}_t, \mathbf{a}_t)$ denote the reward function at each period t , the output of interest is the random cumulative reward accrued by adhering to the policy specified by $\boldsymbol{\theta}$ over the planning period, expressed as $r(\boldsymbol{\theta}) = \sum_{t=1}^H r_t(\mathbf{s}_t, \mathbf{a}_t)$. Some key symbols used in the paper are shown in Table 1.

In this study, the focused input variables of sensitivity analysis are the random factors associated with the states, i.e., the residuals. The output of interest is the cumulative reward $Y = r(\boldsymbol{\theta})$ or future states, i.e.,

Table 1: List of key symbols.

Symbol	Definition	Symbol	Definition
\mathbf{s}_t	CQA at period t	\mathbf{a}_t	CPP at period t
H	total horizon	d_s	the dimensions of \mathbf{s}_t
d_a	the dimensions of \mathbf{a}_t	\mathbf{e}_t	uncontrollable factors at period t
\mathbf{w}	model parameters	$\boldsymbol{\theta}_t$	policy parameters at period t
$r_t(\mathbf{s}_t, \mathbf{a}_t)$	the reward function at each period t	$\boldsymbol{\tau}^{(l)}$	l th observed trajectory in experiments
π_t	the policy at period t	Y	the cumulative reward
\mathcal{D}	historical data	\mathcal{I}	the set of all inputs
$g(\cdot \mathbf{w})$	value function given \mathbf{w}	$Sh_{i,j}$	Shapley interaction effect between i and j
$\Pi(\mathcal{I})$	the set of all permutations of the input set \mathcal{I}		
$P_{i,j}(\pi)$	the set of inputs except for j that precede the element i in the permutation π		

$Y \in \{s_t^n, n = 1, \dots, d_s, t = 1, \dots, H\}$. We study how the variance of the Y is affected by the interactions of the random inputs with a predetermined policy so as to better understand system performance, quantify risk, or indicate where interaction intervention or management may be desirable. Our methods can be extended to analyze how the interactions of the model parameters affect the variance of the output.

3 SHAPLEY INTERACTION EFFECT ESTIMATION

In Section 2, we introduced the PABN model that is used to characterize the biomanufacturing process and the focused random factors in the PABN model. In this section, we introduce the framework for how to use the Shapley-Owen index to measure the impact of each interaction among the random factors in PABN on the variance of the final reward. Due to the high dimension of random factors and the existence of model uncertainty in the PABN model described in Section 2, we propose a sampling-based simulation algorithm for the estimation of the Shapley-Owen index in the PABN model, taking into account both computational tractability and model uncertainty.

Let \mathcal{I} denote the set of all inputs, where \mathcal{I} can be the set of random factors $\{e_t^n, n = 1, \dots, d_s, t = 1, \dots, H\}$. Suppose that the process model parameters \mathbf{w} are given, to quantify the contribution of each input $i \in \mathcal{I}$ to the variance of interested output Y , the Shapley value of any input i is defined as

$$\text{Sh}_i(Y|\mathbf{w}) = \sum_{\mathcal{U} \subset \mathcal{I}/\{i\}} \frac{(|\mathcal{I}| - |\mathcal{U}| - 1)!|\mathcal{U}|!}{|\mathcal{I}|!} [g(\mathcal{U} \cup \{i}|\mathbf{w}) - g(\mathcal{U}|\mathbf{w})],$$

where $|\mathcal{U}|$ is the cardinality of subset $\mathcal{U} \subset \mathcal{I}/\{i\}$ and $g(\cdot|\mathbf{w})$ is called the value function associated with a set of inputs, which denotes the variance of the output generated by a subset of inputs \mathcal{U} . The detailed form of $g(\cdot)$ will be discussed later. We can see that $g(\mathcal{U} \cup \{i}|\mathbf{w}) - g(\mathcal{U}|\mathbf{w})$ is the marginal contribution of input i to the subsets of inputs \mathcal{U} in terms of the expectation or variance of the output. Therefore, the Shapley value is the weighted average of the marginal contributions of i to all subsets \mathcal{U} . The interaction effect between two factors i and j can be defined following Rabitti and Borgonovo (2019) as

$$\begin{aligned} \text{Sh}_{i,j}(Y|\mathbf{w}) = & \sum_{\mathcal{U} \in \mathcal{I}/\{i,j\}} \frac{(|\mathcal{I}| - |\mathcal{U}| - 2)!|\mathcal{U}|!}{(|\mathcal{I}| - 1)!} [g(\mathcal{U} \cup \{i,j\}|\mathbf{w}) \\ & - g(\mathcal{U} \cup \{i\}|\mathbf{w}) - g(\mathcal{U} \cup \{j\}|\mathbf{w}) + g(\mathcal{U}|\mathbf{w})], \end{aligned} \quad (1)$$

which is called the Shapley-Owen index and measures the weighted average marginal contribution in the presence of two factors $\{i, j\}$.

Algorithm 1: Shapley Interaction Effect Estimation on Random Factors of PABN

Input: the number of samples for model parameters K , the number of permutations M , the outer and inner sample size for the simulation of value function N_O and N_I , the policy parameters $\boldsymbol{\theta}$, the posterior distribution of model parameters $p(\mathbf{w}|\mathcal{D})$, the set of inputs \mathcal{I} , $\hat{S}h_{i,j} = 0$ for $\{i, j\} \in \mathcal{I}$.

1. Generate the samples of the marginal contribution and calculate the sum of samples:

for $k = 1, 2, \dots, K$ **do**

 Sample model parameters $\mathbf{w}^{(k)}$ from the posterior distribution $p(\mathbf{w}|\mathcal{D})$:

for $m = 1, 2, \dots, M$ **do**

 Sample the random permutation $\boldsymbol{\pi}^{(m)}$ from $\Pi(\mathcal{I})$;

for all set $\{i, j\} \subseteq \mathcal{I}$ **do**

$\hat{S}h_{i,j} = \hat{S}h_{i,j} + \hat{g}_{nested}(P_{i,j}(\boldsymbol{\pi}^{(m)}) \cup \{i, j\}|\mathbf{w}^{(k)}) - \hat{g}_{nested}(P_{i,j}(\boldsymbol{\pi}^{(m)}) \cup \{i\}|\mathbf{w}^{(k)}) - \hat{g}_{nested}(P_{i,j}(\boldsymbol{\pi}^{(m)}) \cup \{j\}|\mathbf{w}^{(k)}) + \hat{g}_{nested}(P_{i,j}(\boldsymbol{\pi}^{(m)})|\mathbf{w}^{(k)})$.

2. Calculate the estimation of the Shapley interaction effect: $\hat{S}h_{i,j} = \hat{S}h_{i,j}/(KM)$ for $\{i, j\} \in \mathcal{I}$.

$$\pi = \{1,3,2,4\} \left\{ \begin{array}{l}
 \Delta_{1,3}(\pi) = \hat{g}(\{1,3\}|\mathbf{w}) - \hat{g}(\{1\}|\mathbf{w}) - \hat{g}(\{3\}|\mathbf{w}) + \hat{g}(\emptyset|\mathbf{w}) \\
 \Delta_{1,2}(\pi) = \hat{g}(\{1,2\}|\mathbf{w}) - \hat{g}(\{1\}|\mathbf{w}) - \hat{g}(\{2\}|\mathbf{w}) + \hat{g}(\emptyset|\mathbf{w}) \\
 \Delta_{1,4}(\pi) = \hat{g}(\{1,4\}|\mathbf{w}) - \hat{g}(\{1\}|\mathbf{w}) - \hat{g}(\{4\}|\mathbf{w}) + \hat{g}(\emptyset|\mathbf{w}) \\
 \Delta_{2,3}(\pi) = \hat{g}(\{1,2,3\}|\mathbf{w}) - \hat{g}(\{1,3\}|\mathbf{w}) - \hat{g}(\{1,2\}|\mathbf{w}) + \hat{g}(\{1\}|\mathbf{w}) \\
 \Delta_{3,4}(\pi) = \hat{g}(\{1,3,4\}|\mathbf{w}) - \hat{g}(\{1,3\}|\mathbf{w}) - \hat{g}(\{1,4\}|\mathbf{w}) + \hat{g}(\{1\}|\mathbf{w}) \\
 \Delta_{2,4}(\pi) = \hat{g}(\{1,2,3,4\}|\mathbf{w}) - \hat{g}(\{1,2,3\}|\mathbf{w}) - \hat{g}(\{1,3,4\}|\mathbf{w}) + \hat{g}(\{1,3\}|\mathbf{w})
 \end{array} \right.$$

Figure 1: An illustration of the reused calculation in Shapley-Owen index estimation for a sampled permutation. Duplicate terms are framed and displayed in the same color.

It is important to notice that as $|\mathcal{S}|$ increases, the number of subset \mathcal{U} increases exponentially. To address this issue, we can rewrite Equation (1) as

$$\text{Sh}_{i,j}(Y|\mathbf{w}) = \sum_{\pi \in \Pi(\mathcal{S})} \frac{1}{|\mathcal{S}|!} [g(P_{i,j}(\pi) \cup \{i, j\}|\mathbf{w}) - g(P_{i,j}(\pi) \cup \{i\}|\mathbf{w}) - g(P_{i,j}(\pi) \cup \{j\}|\mathbf{w}) + g(P_{i,j}(\pi)|\mathbf{w})],$$

where $\Pi(\mathcal{S})$ is the set of all permutations of the input set \mathcal{S} , $P_{i,j}(\pi)$ is the set of inputs except for j that precede the element i in the permutation π . For instance, if the set of inputs is denoted as $\mathcal{S} = \{1, 2, 3, 4, 5\}$, one permutation of the \mathcal{S} can be $\pi = \{3, 1, 2, 5, 4\}$, then $P_{2,4}(\pi) = \{3, 1\}$. We can randomly sample permutations from $\Pi(\mathcal{S})$ and average the incremental marginal contributions $g(P_{i,j}(\pi) \cup \{i, j\}|\mathbf{w}) - g(P_{i,j}(\pi) \cup \{i\}|\mathbf{w}) - g(P_{i,j}(\pi) \cup \{j\}|\mathbf{w}) + g(P_{i,j}(\pi)|\mathbf{w})$ to approximate the real Shapley-Owen index.

Notice that there are some redundant calculations in the Shapley interaction effect estimation. We can record the previous calculation results to reduce computational efforts. For instance, if there are 4 inputs and the sampled permutation $\pi = \{1, 3, 2, 4\}$, the computation process is shown in Figure 1, where $\Delta_{i,j}(\pi) = g(P_{i,j}(\pi) \cup \{i, j\}|\mathbf{w}) - g(P_{i,j}(\pi) \cup \{i\}|\mathbf{w}) - g(P_{i,j}(\pi) \cup \{j\}|\mathbf{w}) + g(P_{i,j}(\pi)|\mathbf{w})$. We frame the terms that are calculated more than once and the duplicate terms are marked with the same color. In this example, $\hat{g}(\emptyset|\mathbf{w})$, $\hat{g}(\{1\}|\mathbf{w})$, $\hat{g}(\{1, 3\}|\mathbf{w})$, $\hat{g}(\{1, 2\}|\mathbf{w})$, $\hat{g}(\{1, 4\}|\mathbf{w})$, $\hat{g}(\{1, 3, 4\}|\mathbf{w})$, and $\hat{g}(\{1, 2, 3\}|\mathbf{w})$ are calculated more than once. We can record them for subsequent calculations and therefore reduce the computational load by approximately half.

Next, we consider the estimation of the value function. Let $X_{\mathcal{U}}$ be the vector of random inputs included in \mathcal{U} . In this study, we use $g(\mathcal{U}) = \text{Var}[E[Y|X_{\mathcal{U}}]]$, which can be interpreted as the component of the output variance that can be explained by a functional dependence on the input in \mathcal{U} . Using the variance decomposition formula, it can be rewritten as $\text{Var}[Y] - E[\text{Var}[Y|X_{\mathcal{U}}]]$, which can be interpreted as the average decrease in output variance when one receives perfect information about $X_{\mathcal{U}}$. To estimate $g(\mathcal{U})$, Sun et al. (2011) introduced the following unbiased nested Monte Carlo estimator: Given the number of outer samples N_O and the number of inner samples N_I , let $y_{n_1, n_2}, n_1 = 1, \dots, N_O, n_2 = 1, \dots, N_I$ be the overall samples of Y , which is obtained by firstly sampling $x_{\mathcal{U}}$ from $F_{X_{\mathcal{U}}}$, the joint distribution of random factors in \mathcal{U} , then sampling Y from $F_{Y|X_{\mathcal{U}}=x_{\mathcal{U}}}$, the conditional distribution of Y given $X_{\mathcal{U}} = x_{\mathcal{U}}$. Let $\bar{y}_{n_1} = \frac{1}{N_I} \sum_{n_2=1}^{N_I} y_{n_1, n_2}$, $\bar{y} = \frac{1}{N_O} \sum_{n_1=1}^{N_O} \bar{y}_{n_1}$, the quantity of interest is estimated by

$$\hat{g}_{\text{nested}}(\mathcal{U}) = \frac{1}{N_O - 1} \sum_{n_1=1}^{N_O} (\bar{y}_{n_1} - \bar{y})^2 - \frac{1}{N_O N_I (N_I - 1)} \sum_{n_1=1}^{N_O} \sum_{n_2=1}^{N_I} (y_{n_1, n_2} - \bar{y}_{n_1})^2,$$

The aforementioned calculation is based on a fixed \mathbf{w} . To handle the model parameters uncertainty of PABN, we can draw samples $\mathbf{w}^{(k)}$, $k = 1, \dots, K$, from the posterior distribution $p(\mathbf{w}|\mathcal{D})$, then the final

estimator of the Shapley-Owen index is

$$\hat{Sh}_{i,j} = \frac{1}{KM} \sum_{k=1}^K \sum_{m=1}^M \left[\hat{g}_{nested}(P_{i,j}(\boldsymbol{\pi}^{(m)}) \cup \{i, j\} | \mathbf{w}^{(k)}) - \hat{g}_{nested}(P_{i,j}(\boldsymbol{\pi}^{(m)}) \cup \{i\} | \mathbf{w}^{(k)}) - \hat{g}_{nested}(P_{i,j}(\boldsymbol{\pi}^{(m)}) \cup \{j\} | \mathbf{w}^{(k)}) + \hat{g}_{nested}(P_{i,j}(\boldsymbol{\pi}^{(m)}) | \mathbf{w}^{(k)}) \right].$$

The complete statement of the interaction effect estimation procedure is given in Algorithm 1.

4 COMPUTATIONAL BUDGET ALLOCATION

In practice, the calculation of the Shapley interaction effect using Algorithm 1 faces several challenges. Firstly, the implementation of Algorithm 1 requires the proper choice of the sample size in each level, i.e., the number of K , M , N_O and N_I , which could be difficult to properly predetermine. Secondly, the complex nested simulation procedure hinders the analysis of the performance guarantee and estimation error of the algorithm. Thirdly, for each sampled permutation, Algorithm 1 accumulates the marginal contribution for all subset $\{i, j\} \subseteq \mathcal{S}$, which allocates the total computational effort to each subset $\{i, j\}$ equally. However, if one alternative has very low variance, then it may only require very few samples to accurately estimate its performance. Therefore, equal budget allocation can be inefficient. In this section, we provide a non-nested unbiased estimator for the Shapley interaction effect $Sh_{i,j}$, based on which we then propose an adaptive budget allocation procedure for the estimation of $Sh_{i,j}$.

Let $X^{(0)}$ and $X^{(1)}$ be two independent and identically distributed (i.i.d.) random variables generated from F_X , $Y^{(1)}$ and $Y^{(2)}$ are i.i.d. random variables following distribution $F_{Y|X^{(0)}, \mathbf{w}}$, the conditional distribution of Y given $X = X^{(0)}$ and model parameter \mathbf{w} , and $Y^{(3)}$ follows the distribution $F_{Y|X^{(1)}, \mathbf{w}}$, the conditional distribution of Y given $X = X^{(1)}$ and model parameter \mathbf{w} . Then according to Cheng and Zhang (2021), $\text{Var}[E(Y|X)] = E[Y^{(1)}(Y^{(2)} - Y^{(3)})]$. Therefore, let $x_{\mathcal{U}}^{(0)}$ and $x_{\mathcal{U}}^{(1)}$ be two random samples generated from $F_{X_{\mathcal{U}}|\mathbf{w}}$, a non-nested unbiased estimator of the value function can be $\hat{g}_{nn}(\mathcal{U} | \mathbf{w}) = y^{(1)}(y^{(2)} - y^{(3)})$, where $y^{(1)}$ and $y^{(2)}$ are sampled from $F_{Y|x_{\mathcal{U}}^{(0)}, \mathbf{w}}$, $y^{(3)}$ is sampled from $F_{Y|x_{\mathcal{U}}^{(1)}, \mathbf{w}}$. Let $\text{Unif}(\Pi(\mathcal{S}))$ denote the uniform distribution over the permutation space $\Pi(\mathcal{S})$. Now we obtain a new estimator of the Shapley interaction effect, which can be written as

$$\hat{Sh}_{ij} = \frac{1}{N_{i,j}} \sum_{n=1}^{N_{i,j}} \hat{g}_{nn}(P_{i,j}(\boldsymbol{\pi}^{(n)}) \cup \{i, j\} | \mathbf{w}^{(n)}) - \hat{g}_{nn}(P_{i,j}(\boldsymbol{\pi}^{(n)}) \cup \{i\} | \mathbf{w}^{(n)}) - \hat{g}_{nn}(P_{i,j}(\boldsymbol{\pi}^{(n)}) \cup \{j\} | \mathbf{w}^{(n)}) + \hat{g}_{nn}(P_{i,j}(\boldsymbol{\pi}^{(n)}) | \mathbf{w}^{(n)})$$

where $\boldsymbol{\pi}^{(n)}$ and $\mathbf{w}^{(n)}$ are sampled from $\text{Unif}(\Pi(\mathcal{S}))$ and $p(\mathbf{w} | \mathcal{D})$, respectively, and $N_{i,j}$ is the sample size used to estimate Sh_{ij} .

In Algorithm 1, the computational resources are allocated to each subset $\{i, j\} \subseteq \mathcal{S}$ equally. However, subset $\{i, j\}$ with greater uncertainty in $\hat{Sh}_{i,j}$ may require more samples to reduce the estimation error. As a result, the equal allocation approach could not efficiently use the simulation resource. To avoid such issue, we develop a sequential procedure that can efficiently employ the simulation budget to estimate $\hat{Sh}_{i,j}$ for each $\{i, j\} \subseteq \mathcal{S}$. As the focus of this paper is the estimation of the true Shapley-Owen indices rather than the identification of the most influential interactions, we use the confidence interval to quantify the estimation uncertainty and to guide the allocation of total simulation budget.

Based on the central limit theorems, we first construct the asymptotic confidence interval for $\hat{Sh}_{i,j}$ as in Cheng and Zhang (2021):

Proposition 1 Given the sample size $N_{i,j}$ and the significance level of α ,

$$\Pr \left\{ \hat{Sh}_{i,j} \in \left(\hat{Sh}_{i,j} + z_{\alpha/2} \hat{\sigma}_{i,j} / \sqrt{N_{i,j}}, \hat{Sh}_{i,j} - z_{\alpha/2} \hat{\sigma}_{i,j} / \sqrt{N_{i,j}} \right) \right\} \simeq 1 - \alpha,$$

Algorithm 2: Shapley Interaction Effect Estimation with Sequential Budget Allocation

Input: the number of total iterations N , the number of iterations for Stage 1 N_0 , the policy parameters $\boldsymbol{\theta}$, the posterior distribution of model parameters $p(\mathbf{w}|\mathcal{D})$, the set of inputs \mathcal{I} , $\hat{S}h_{i,j} = 0$, $\widehat{Sh}_{i,j}^2 = 0$ and $N_{i,j} = 0$ for $\{i, j\} \in \mathcal{I}$.

Stage 1. Pilot simulation to estimate the sample standard deviation:

for $n = 1, 2, \dots, N_0$ **do**

Sample model parameters $\mathbf{w}^{(n)}$ and permutation $\pi^{(n)}$ from $p(\mathbf{w}|\mathcal{D})$ and $\Pi(\mathcal{I})$;

for all set $\{i, j\} \subseteq \mathcal{I}$ **do**

$N_{i,j} = N_{i,j} + 1$;

$\Delta_{i,j}^{(n)} = \hat{g}_{nn}(P_{i,j}(\pi^{(n)}) \cup \{i, j\} | \mathbf{w}^{(n)}) - \hat{g}_{nn}(P_{i,j}(\pi^{(n)}) \cup \{i\} | \mathbf{w}^{(n)}) - \hat{g}_{nn}(P_{i,j}(\pi^{(n)}) \cup \{j\} | \mathbf{w}^{(n)}) + \hat{g}_{nn}(P_{i,j}(\pi^{(n)}) | \mathbf{w}^{(n)})$;

$\hat{S}h_{i,j} = \hat{S}h_{i,j} + \Delta_{i,j}^{(n)}$, $\widehat{Sh}_{i,j}^2 = \widehat{Sh}_{i,j}^2 + (\Delta_{i,j}^{(n)})^2$;

$$\hat{\sigma}_{i,j} = \sqrt{\frac{\widehat{Sh}_{i,j}^2 + (\hat{S}h_{i,j})^2 / N_0}{N_0 - 1}}$$

Stage 2. Sequential allocation of the sample budget:

for $n = N_0 + 1, N_0 + 2, \dots, N$ **do**

Calculate the gradient with respect to the half length $\text{gHL}_{i,j}$ for all $\{i, j\} \subseteq \mathcal{I}$;

Sample model parameters $\mathbf{w}^{(n)}$ and permutation $\pi^{(n)}$ from $p(\mathbf{w}|\mathcal{D})$ and $\Pi(\mathcal{I})$;

Find $\{i^*, j^*\} = \text{argmax}_{i,j} \text{gHL}_{i,j}$ and i^* that is the leftmost element of $\{i^*, j^*\}$ in $\pi^{(n)}$;

Construct the group $G_m(i^*, \pi^{(n)})$;

for all set $\{i, j\} \subseteq G_m(i^*, \pi^{(n)})$ **do**

$N_{i,j} = N_{i,j} + 1$;

$\hat{S}h_{i,j} = \hat{S}h_{i,j} + \hat{g}_{nn}(P_{i,j}(\pi^{(n)}) \cup \{i, j\} | \mathbf{w}^{(n)}) - \hat{g}_{nn}(P_{i,j}(\pi^{(n)}) \cup \{i\} | \mathbf{w}^{(n)}) - \hat{g}_{nn}(P_{i,j}(\pi^{(n)}) \cup \{j\} | \mathbf{w}^{(n)}) + \hat{g}_{nn}(P_{i,j}(\pi^{(n)}) | \mathbf{w}^{(n)})$.

Stage 3. Calculate the Shapley interaction effect: $\hat{S}h_{i,j} = \hat{S}h_{i,j} / N_{i,j}$ for $\{i, j\} \in \mathcal{I}$.

where z_α is the $\alpha/2$ quantile of the standard normal distribution, $\hat{\sigma}_{i,j}$ is the sample standard deviation, and \simeq indicates the asymptotic equality as $N_{i,j} \rightarrow \infty$.

In Proposition 1, the half length of the asymptotic confidence interval $\text{HL}_{i,j} = z_{\alpha/2} \hat{\sigma}_{i,j} / \sqrt{N_{i,j}}$ quantifies the estimation uncertainty due to finite sample size. The corresponding absolute value for the gradient of $\text{HL}_{i,j}$ with respect to the number of samples $N_{i,j}$ is $\text{gHL}_{i,j} = \frac{z_{\alpha/2} \hat{\sigma}_{i,j}}{2N_{i,j}^{3/2}}$, which reflects the benefit of one

additional sample in estimating $\hat{S}h_{i,j}$. Therefore, instead of allocating the simulation budget to each $\{i, j\}$ uniformly, one can sequentially allocate the simulation budget to $\{i^*, j^*\}$ with the greatest $\text{gHL}_{i,j}$ at each iteration until the simulation budget is exhausted.

As we described in Section 3, Algorithm 1 can benefit from the reused calculation of the value function. To take advantage of the reused calculation, we can modify the sequential procedure as follows: Without loss of generality, let $G(i^*, \pi^{(n)}) \subseteq \mathcal{I}$ denote the group of subsets with cardinality of 2 whose first element is i^* . At each iteration, we allocate the simulation efforts to $\{i^*, j\}$ s that belong to $G(i^*, \pi^{(n)})$ and have top m gHL among $G(i^*, \pi^{(n)})$, which is denoted by $G_m(i^*, \pi^{(n)})$. These $\{i^*, j\}$ s have a common item $\hat{g}(P_{i^*,j}(\pi^{(n)}) | \mathbf{w}^{(n)}) - \hat{g}(P_{i^*,j}(\pi^{(n)}) \cup \{i^*\} | \mathbf{w}^{(n)})$ when estimating the value function, which can be calculated once and then recorded to reduce the computation load. Algorithm 2 presents the pseudo-code of the sequential budget allocation procedure for Shapley interaction effect estimation.

To further reduce the estimation error of the interaction effect estimation, we apply the quasi-Monte Carlo method in Algorithm 2, which achieves a lower estimation error of the Shapley-Owen index compared

to the standard Monte Carlo method by careful selection of samples. We first generate low discrepancy point sets (e.g., Halton sequence), then map n th low discrepancy point $\mathbf{x}^{(n)}$ to an actual sample $\{\mathbf{w}^{(n)}, \pi^{(n)}\}$. More specifically, let d_w denote the dimension of $\mathbf{w}^{(n)}$ and d_π denote the dimension of $\pi^{(n)}$, the size and dimension of the generated low discrepancy point sets be N and $d_w + d_\pi - 2$, respectively. The first d_w dimensions of $\mathbf{x}^{(n)}$ are mapped to $\mathbf{w}^{(n)}$ according to Lemieux (2009) and the remaining $(d_\pi - 2)$ dimensions are mapped to $\pi^{(n)}$ using the transformation method in Mitchell et al. (2022).

5 NUMERICAL EXPERIMENT

In this section, we present the numerical results for two types of PABN where we estimate the Shapley-Owen indices. The first one is the linear Gaussian PABN in Zheng et al. (2023), which is valid for bioprocesses with online monitoring, meaning that the monitoring frequency is on a faster time scale than the evolution of bioprocess dynamics. As the value function $g(\cdot)$ for linear PABN can be analytically calculated, we can obtain the Shapley-Owen indices precisely. Therefore, we use linear PABN to test the performance of the Algorithm 1 and 2 on estimating the Shapley-Owen indices. In addition, we conduct the ablation study on the choice of sample size at each level of Algorithm 1 and the impact of the sampling methods (Monte Carlo and quasi-Monte Carlo) in Algorithm 2. The second one is a PABN application with nonlinear transition functions that describe the general mature cell-progenitor negative feedback process. We analyze the estimated value of the Shapley-Owen indices for this real-world PABN and demonstrate the effectiveness of these indices in capturing interaction effects and providing guidance for decision-making.

5.1 Linear Gaussian PABN Model

For the linear Gaussian PABN model, we consider the case that the transition function has a linear form and the random factors follow a Gaussian distribution. More accurately, let β_t^s denote the $d_s \times d_s$ matrix whose (n, q) -th element is the linear coefficient β_t^{nq} corresponding to the effect of state s_t^n on the next state s_{t+1}^q . Similarly, let β_t^a be the $d_a \times d_s$ matrix representing the effects of each component of \mathbf{a}_t on each component of \mathbf{s}_{t+1} . Then the dynamic function becomes $\mathbf{s}_{t+1} = \mu_{t+1}^s + (\beta_t^s)^\top (\mathbf{s}_t - \mu_t^s) + (\beta_t^a)^\top (\mathbf{a}_t - \mu_t^a) + \mathbf{e}_{t+1}$, where $\mu_t^s = (\mu_t^1, \dots, \mu_t^{d_s})$, $\mu_t^a = (\lambda_t^1, \dots, \lambda_t^{d_a})$ and $\mathbf{e} = (\mathbf{e}_1^\top, \mathbf{e}_2^\top, \dots, \mathbf{e}_H^\top) \sim \mathcal{N}(\mathbf{0}, \mathbf{V})$. The model parameters $\mathbf{w} = (\mu^s, \mu^a, \beta, \mathbf{V})$, where $\mu^s = \{\mu_t^s\}_{t=1}^H$, $\mu^a = \{\mu_t^a\}_{t=1}^H$, $\beta = \{(\beta_t^a, \beta_t^s)\}_{t=1}^H$. The policy and reward functions are $\mathbf{a}_t = \mu_t^a + \theta_t^\top (\mathbf{s}_t - \mu_t^s)$ and $r_t(\mathbf{s}_t, \mathbf{a}_t) = m_t + \mathbf{b}_t^\top \mathbf{a}_t + \mathbf{c}_t^\top \mathbf{s}_t$, where θ_t is an $d_s \times d_a$ matrix.

Given model parameters \mathbf{w} , the cumulative reward $r(\theta|\mathbf{w})$ can be written as $r(\theta|\mathbf{w}) = \sum_{t=1}^H r_t(\mathbf{s}_t, \mathbf{a}_t|\mathbf{w}) = \gamma + \sum_{t=1}^H \alpha_t (\sum_{t'=1}^t \mathbf{R}_{t',t-1} \mathbf{e}_{t'})$, where γ denotes the term that excludes random factors, $\alpha_t = \mathbf{b}_t^\top \theta_t^\top + \mathbf{c}_t^\top$,

$\mathbf{R}_{t',t} = \prod_{t''=t'}^t \left[(\beta_{t''}^s)^\top + (\beta_{t''}^a)^\top \theta_{t''}^\top \right]$ represents the product of pathway coefficients from time step t' to

t and $\mathbf{R}_{t',t'-1} = \mathbb{I}_{d_s \times d_s}$, the $d_s \times d_s$ identity matrix. Let $\mathbf{R}_{t'} = \sum_{t=t'}^H \alpha_t \mathbf{R}_{t',t-1}$, the variance of $r(\theta|\mathbf{w})$ is $\text{Var}[r(\theta|\mathbf{w})] = \text{Var}[\sum_{t=1}^H \alpha_t (\sum_{t'=1}^t \mathbf{R}_{t',t-1} \mathbf{e}_{t'})] = \sum_{t_1=1}^H \sum_{t_2=1}^H [\mathbf{R}_{t_1} \text{Cov}(\mathbf{e}_{t_1}, \mathbf{e}_{t_2}) \mathbf{R}_{t_2}^\top]$.

Now we calculate the value function for the linear Gaussian PABN. Let $\mathbf{R} = (\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_H)$, the cumulative reward can be written as $r(\theta|\mathbf{w}) = \gamma + \mathbf{R}\mathbf{e}$. If we denote $\mathbf{R}_{\mathcal{U}}$ as the corresponding coefficient vector in \mathbf{R} for input set \mathcal{U} , and similarly, let $\Sigma_{-\mathcal{U}, \mathcal{U}}$ be the the covariance matrix between random inputs not in \mathcal{U} and random inputs in \mathcal{U} , $\Sigma_{\mathcal{U}}$ denote the covariance matrix between random inputs in \mathcal{U} , then we can derive the value function $g(\mathcal{U}|\mathbf{w})$ explicitly as follows:

$$g(\mathcal{U}|\mathbf{w}) = (\mathbf{R}_{-\mathcal{U}} \Sigma_{-\mathcal{U}, \mathcal{U}} \Sigma_{\mathcal{U}}^{-1} + \mathbf{R}_{\mathcal{U}}) \Sigma_{\mathcal{U}} (\mathbf{R}_{-\mathcal{U}} \Sigma_{-\mathcal{U}, \mathcal{U}} \Sigma_{\mathcal{U}}^{-1} + \mathbf{R}_{\mathcal{U}})^\top. \quad (2)$$

In our experiments, we set the number of states $d_s = 3$, the number of actions $d_a = 1$, and the time horizon $T = 2$. This configuration results in 6 factors and 15 interactions. The posterior distribution of the model parameters $p(\mathbf{w}|\mathcal{D})$ is set to be multivariate Gaussian. We first test the performance of Algorithm 1 with different budget allocation strategies. Following Hart et al. (2019), we calculate the mean squared error (MSE) to measure the performance of the algorithm, which is defined as $\text{MSE} = \frac{1}{15} \sum_{\{i,j\} \subseteq \mathcal{I}} (\text{Sh}_{i,j} - \hat{\text{Sh}}_{i,j})^2$,

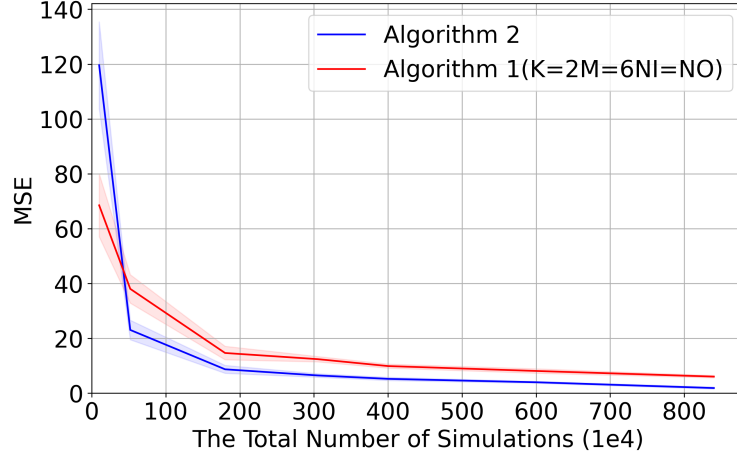


Figure 2: MSE of Algorithm 1 and Algorithm 2

where $Sh_{i,j}$ is the real Shapley-Owen value estimated by using a large number of model parameter samples and calculating the value function with Equation (2), $\hat{Sh}_{i,j}$ is obtained from Algorithm 1.

We first conduct ablation study to determine the sample size for each level of simulation given a fixed computational budget, i.e., to decide K, M, N_O and N_I . As we cannot find any literature that discusses the optimal allocation strategy for such four-level nested simulation, we apply the grid search and assign a number in 1, 2, 3, 4, 5, 6 to each level of simulation. Fixing $K * M * N_O * N_I = 30000$, we use the ratio of 4 levels (i.e., $K : M : N_O : N_I$) to allocate the simulation budget. We find that the budget allocation ratio with the smallest MSE is $K : M : N_O : N_I = 6 : 3 : 6 : 1$, which means a good budget allocation policy tends to bound the number of inner level samples N_I . In order to validate this conclusion, we randomly sample 10 different budget allocation ratios for K, M and N_O (i.e., $K : M : N_O$), and for each ratio, we vary the number assigned to N_I and record the corresponding MSE averaged by 10 macro-replications in Table 2. We find that the MSE of Algorithm 1 increases as the proportion of inner sample size N_I increases, which is consistent with the conclusion in Sun et al. (2011) that, as the computational budget increases, the optimal number of inner-level samples remains bounded.

Next, we compare the performance of Algorithm 1 and Algorithm 2. For Algorithm 1, we use the budget allocation policy $K : M : N_O : N_I = 6 : 3 : 6 : 1$ and for Algorithm 2, we set $N_0 = 0.02N$ and $m = 2$. The result is illustrated in Figure 2, which indicates that the performance of Algorithm 2 is superior over

Table 2: MSE of Algorithm 1 using different budget allocation ratios with 90% confidence interval.

$K:M:N_O$	Number assigned to N_I (increasing order)				
	1	2	3	4	5
1:1:1	20.9 (3.9)	26.9 (8.6)	28.3 (7.1)	67.2 (17.0)	81.4 (18.5)
1:3:5	17.2 (2.4)	19.2 (3.6)	27.9 (5.5)	28.8 (7.3)	40.6 (7.8)
2:3:6	17.5 (5.1)	19.1 (4.2)	22.5 (6.1)	28.9 (4.9)	44.9 (8.3)
6:2:4	18.6 (1.9)	20.4 (5.1)	23.9 (4.3)	28.6 (4.0)	68.2 (13.3)
4:3:3	16.9 (2.9)	18.3 (3.4)	22.9 (4.2)	25.8 (8.5)	49.1 (7.2)
1:6:1	16.9 (1.3)	19.0 (5.1)	22.1 (7.2)	40.3 (4.0)	44.3 (10.2)
5:4:3	20.7 (3.2)	25.0 (7.2)	26.2 (4.5)	36.6 (8.1)	48.3 (9.5)
2:1:6	18.3 (2.1)	26.0 (8.0)	27.8 (7.8)	69.9 (14.0)	90.0 (7.8)
5:5:1	17.2 (3.7)	21.4 (3.2)	26.0 (6.3)	39.7 (7.9)	94.7 (22.2)
3:1:3	18.8 (1.7)	24.3 (5.2)	40.5 (10.9)	45.2 (7.4)	62.4 (25.7)

Table 3: MSE of algorithm 2 using MC and QMC implementation with 90% confidence interval.

Sample size	180000	360000	540000	720000	900000
MC	110.2 (9.0)	62.4 (4.7)	26.0 (2.4)	22.7 (1.9)	20.5 (1.8)
QMC	87.1 (7.7)	49.9 (3.9)	20.5 (2.1)	17.7 (1.9)	16.2 (1.3)

Algorithm 1 except when the budget is very limited. This is because the estimation of the variance might be inaccurate without a sufficient simulation budget. Finally, we compare the performance of Algorithm 2 using classical Monte Carlo (MC) and quasi Monte Carlo (QMC) sampling methods. As shown in Table 3, we can see the QMC can significantly reduce the estimation error of Algorithm 2.

5.2 Nonlinear PABN Model

For the nonlinear PABN model, we use a general mature cell-progenitor negative feedback model in Stacey et al. (2018) as an example to interpret the results of Shapley-Owen indices. In this system, cytokine-mediated negative feedback suppresses the bulking up of progenitors in the presence of mature cells, and the productivity of a volume of medium is therefore determined by the accumulation of inhibitory cytokines. The transition model that characterizes the evolution of progenitor density (S), product density (P) and inhibitor concentration (I) over time can be described as $\frac{dS}{dt} = \frac{r_g}{(1+\exp(a(I-b)))} - r_c S$, $\frac{dP}{dt} = r_c S - r_d P$, $\frac{dI}{dt} = r_p P$, where r_g is growth rate, r_c is conversion rate from progenitor to product, r_d is product death rate, r_p is production rate of inhibitor, a is inhibitor sensitivity and b is inhibitor threshold. Therefore, the state vector in this system is $\mathbf{s}_t = \{S_t, P_t, I_t\}$. The values of these parameters are set according to Stacey et al. (2018). By reducing the concentration of inhibitor by a media dilution, there is the potential to increase productivity. Therefore, the action in this system is the fraction of the concentration reduction of I at each period. For illustration, we simply adopt the policy that we dilute the concentration to be half of the current concentration at each decision time. The output of interest is the reward of final product density minus the cost of the inhibitor concentration reduction.

Let the random factors associated with S , P and I at period t be e_t^S , e_t^P , and e_t^I . Following McAdams et al. (1997), the culture PH can be negatively related to the progenitor density and positively related to the product density and inhibitor concentration over a specific range, we set $e_t^i = l^i e_t^{PH} + e_t^{i'}$ for $i \in \{S, P, I\}$, where e_t^{PH} denotes the randomness associated with the culture PH, l^i is a coefficient whose sign indicates the correlation relationship, and $e_t^{i'}$ denotes other sources of randomness. We vary the values of $\mathbf{l} = \{l^S, l^P, l^I\}$ as "no dependence" ($\mathbf{l} = \{0, 0, 0\}$), "low dependence" ($\mathbf{l} = \{-0.5, 0.5, 0.5\}$), "strong dependence" ($\mathbf{l} = \{-1, 1, 1\}$) to show the Shapley-Owen indices at different level of correlation.

The estimated Shapley-Owen index is shown in Figure 3. First, it is evident that the interaction effects of (S, P) , (S, I) and (P, I) diminish as time proceeds. This is because the interactions at the early stage can be amplified and accumulated to impact the final output through the nonlinear transition model. Second, we find that as the level of correlations grows, the magnitude of the Shapley-Owen indices also increases to capture the interaction effects on the output variance, which shows the power of Shapley-Owen indices in measuring the interaction effects. Notice that such interactions among the random factors often occur due to the underlying factors. Therefore, a systematic experiment on factors with large interaction effects can help detect the underlying factors that contribute greatly to the system variation. Third, under the scenario of strong dependence, we find that the introduced correlation between the random factors associated with P and I can reduce the output variance. Therefore, a sophisticated control of the product density based on the inhibitor concentration can help improve the process stability. For example, Liu et al. (2015) suggested that one can use activated carbon to control the concentration of inhibitors during butanol production and show that it can yield high butanol production and reduce the output variation.

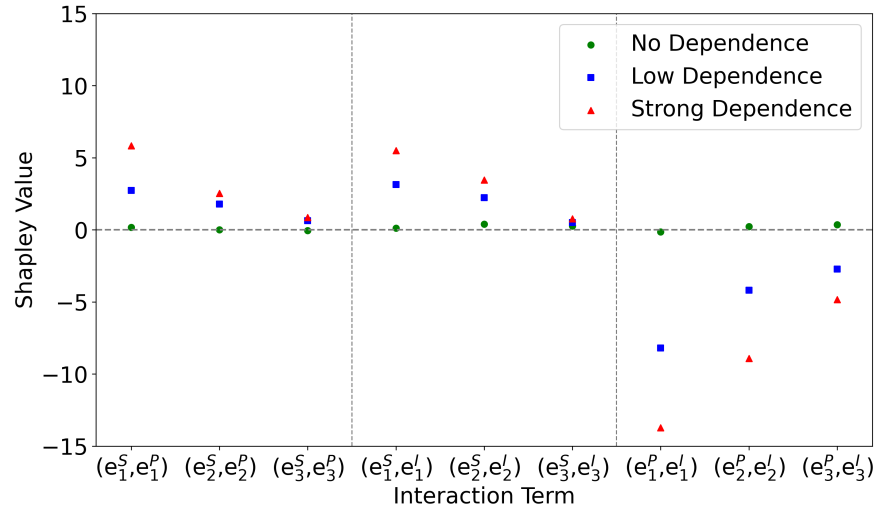


Figure 3: Shapley interaction effects for mature cell-progenitor negative feedback model.

6 CONCLUSION

In this paper, we present a method for conducting a global sensitivity analysis on the interaction effects of the PABN model in biomanufacturing. We use the Shapley-Owen index to quantify the interaction effects and introduce a sampling-based framework to approximate the interaction measures. To enhance computational efficiency, we further incorporate a non-nested simulation method and a sequential budget allocation scheme. Numerical results show that our algorithms have superior empirical performance in terms of accuracy and efficiency. In future research, it will be worth considering combining the variance reduction techniques tailored to the quantification of interaction effects for PABN models that involve high noise. In addition, how to use both the main effects and the interaction effects for factor screening in biomanufacturing is also an interesting problem.

ACKNOWLEDGMENTS

The first, second, and fourth authors were supported in part by the National Natural Science Foundation of China (Grants 72031006, 72394370, 72394375).

REFERENCES

- Aijaz, A., M. Li, D. Smith, D. Khong, C. LeBlon, O. S. Fenton, , , *et al.* 2018. “Biomanufacturing for clinically advanced cell therapies”. *Nature Biomedical Engineering* 2(6):362–376.
- Cheng, H.-F. and K. Zhang. 2021. “Non-nested estimators for the central moments of a conditional expectation and their convergence properties”. *Operations Research Letters* 49(5):625–632.
- Cordell, H. J. 2009. “Detecting gene–gene interactions that underlie human diseases”. *Nature Reviews Genetics* 10(6):392–404.
- Diouf, A., H. Mokrani, E. Afenya, and B. I. Camara. 2021. “Computation of the conditions for anti-angiogenesis and gene therapy synergistic effects: Sensitivity analysis and robustness of target solutions”. *Journal of Theoretical Biology* 528:110850.
- Hart, J. L., J. Bessac, and E. M. Constantinescu. 2019. “Global sensitivity analysis for statistical model parameters”. *SIAM/ASA Journal on Uncertainty Quantification* 7(1):67–92.
- Łacki, K. M., J. Joseph, and K. O. Eriksson. 2018. “Downstream process design, scale-up principles, and process modeling”. In *Biopharmaceutical Processing*, 637–674. Elsevier.
- Lemieux, C. 2009. *Monte Carlo and Quasi-Monte Carlo Sampling*. 1 ed. Springer Series in Statistics., Springer New York.

- Liu, K., H. K. Atiyeh, O. Pardo-Planas, T. C. Ezeji, V. Ujor, J. C. Overton, , *et al.* 2015. “Butanol production from hydrothermolysis-pretreated switchgrass: quantification of inhibitors and detoxification of hydrolyzate”. *Bioresource Technology* 189:292–301.
- McAdams, T. A., W. M. Miller, and E. T. Papoutsakis. 1997. “Variations in culture pH affect the cloning efficiency and differentiation of progenitor cells in ex vivo haemopoiesis”. *British Journal of Haematology* 97(4):889–895.
- Mitchell, R., J. Cooper, E. Frank, and G. Holmes. 2022. “Sampling permutations for shapley value estimation”. *Journal of Machine Learning Research* 23(43):1–46.
- Owen, A. B. 2014. “Sobol’ indices and Shapley value”. *SIAM/ASA Journal on Uncertainty Quantification* 2(1):245–251.
- Rabitti, G. and E. Borgonovo. 2019. “A Shapley–Owen index for interaction quantification”. *SIAM/ASA Journal on Uncertainty Quantification* 7(3):1060–1075.
- Rabitti, G. and E. Borgonovo. 2020. “Is mortality or interest rate the most important risk in annuity models? A comparison of sensitivity analysis methods”. *Insurance: Mathematics and Economics* 95:48–58.
- Sobol’, I. 1993. “Sensitivity estimates for nonlinear mathematical models”. *Mathematical Modeling and Computational Experiment* 1:407.
- Song, E., B. L. Nelson, and J. Staum. 2016. “Shapley effects for global sensitivity analysis: Theory and computation”. *SIAM/ASA Journal on Uncertainty Quantification* 4(1):1060–1083.
- Stacey, A. J., E. A. Cheeseman, K. E. Glen, R. L. Moore and R. J. Thomas. 2018. “Experimentally integrated dynamic modelling for intuitive optimisation of cell based processes and manufacture”. *Biochemical Engineering Journal* 132:130–138.
- Sun, Y., D. W. Apley, and J. Staum. 2011. “Efficient nested simulation for estimating the variance of a conditional expectation”. *Operations Research* 59(4):998–1007.
- Xie, W., R. R. Barton, B. L. Nelson, and K. Wang. 2023. “Stochastic simulation uncertainty analysis to accelerate flexible biomanufacturing process development”. *European Journal of Operational Research* 310(1):238–248.
- Zhao, J., J. Luo, and W. Xie. 2023. “Policy-Augmented Bayesian Network Optimization with Global Convergence”. In *2023 Winter Simulation Conference (WSC)*, 3553–3564. IEEE.
- Zheng, H., W. Xie, I. O. Ryzhov, and D. Xie. 2023. “Policy Optimization in Dynamic Bayesian Network Hybrid Models of Biomanufacturing Processes”. *INFORMS Journal on Computing* 35(1):66–82.

AUTHOR BIOGRAPHIES

JUNKAI ZHAO is a Ph.D. student in the Antai College of Economics and Management at Shanghai Jiao Tong University. His research interest is simulation optimization. His email address is zhaojunkai@sjtu.edu.cn.

JUN LUO is a professor in the Antai College of Economics and Management at Shanghai Jiao Tong University. His primary research interests are stochastic simulation and simulation optimization. His email address is jluo_ms@sjtu.edu.cn.

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.

ZIXUAN BAI is a master student in the Antai College of Economics and Management at Shanghai Jiao Tong University. His email address is zixuanbai@sjtu.edu.cn.