European Journal of Operational Research xxx (xxxx) xxx

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Stochastics and Statistics

Stochastic simulation uncertainty analysis to accelerate flexible biomanufacturing process development

Wei Xie[a,*], Russell R. Barton[b], Barry L. Nelson[c], Keqi Wang[a]

[a] Mechanical and Industrial Engineering, Northeastern University, 360 Huntington Avenue, 334 SN, Boston, MA 02115, USA
[b] Supply Chain and Information Systems, Smeal College of Business, 413 Business Building, The Pennsylvania State University, University Park, PA, 16802, USA
[c] Department of Industrial Engineering and Management Sciences, Northwestern University, 2145 Sheridan Road, Room M326, Evanston, IL, 60208-3119, USA

ARTICLE INFO

Article history:
Received 16 March 2022
Accepted 25 January 2023
Available online xxx

Keywords:
Hybrid simulation model
Biomanufacturing systems
Uncertainty quantification (UQ)
Sensitivity analysis (SA)
Gaussian process (GP)

ABSTRACT

Motivated by critical challenges and needs from biopharmaceuticals manufacturing, we propose a general metamodel-assisted stochastic simulation uncertainty analysis framework to accelerate the development of a simulation model with modular design for flexible production processes. There are often very limited process observations. Thus, there exist both simulation and model uncertainties in the system performance estimates. In biopharmaceutical manufacturing, model uncertainty often dominates. The proposed framework can produce a confidence interval that accounts for simulation and model uncertainties by using a metamodel-assisted bootstrapping approach. Furthermore, a variance decomposition is utilized to estimate the relative contributions from each source of model uncertainty, as well as simulation uncertainty. This information can be used to improve the system mean performance estimation. Asymptotic analysis provides theoretical support for our approach, while the empirical study demonstrates that it has good finite-sample performance.

1. Introduction

While the biopharmaceutical industry has developed various innovative bio-drugs for severe diseases, such as cancers, autoimmune disorders, and infectious diseases, the current manufacturing systems are unable to rapidly produce new and existing drugs when needed, largely due to critical challenges, including high complexity, high variability, and very limited process data. Biotherapeutics are manufactured in living organisms (e.g., cells) whose biological processes are very complex. Manufacturing process typically consists of multiple integrated unit operations. There is often very limited data, i.e., having 3–20 process observations is typical in biomanufacturing (O'Brien, Zhang, Daoutidis, & Hu, 2021), reflecting the high cost and long time needed to run lab experiments. Also, the more personalized nature of emerging bio-drugs (e.g., cell and gene therapies) makes it difficult to collect extensive data on every possible variety of drugs and every protein therapy can be unique, which often forces R&D efforts to work with just 3–5 batches.

Simulation can facilitate the development of flexible production systems with modular design. *Hybrid ("mechanistic+statistical") simulation models* can support interpretable and robust decision making, while requiring much less data than purely data-based models. The mechanistic model parameters (such as cell growth rate, oxygen and nutrient uptake rates) can facilitate the learning of underlying biological/physical/chemical (a.k.a. *biophysicochemical*) mechanisms. Thus, in this paper, we suppose that the model family or structure, built on mechanism prior knowledge, is given. The model parameters are estimated from very limited real-world data, which introduces *model uncertainty*. When we create a simulation model to predict the performance of a real system, there exist the errors induced by both simulation estimation uncertainty and process model uncertainty.

In the biomanufacturing literature, modeling of bioprocess dynamics while considering different sources of uncertainty (e.g., batch-to-batch variations, measurement errors, and model uncertainty) is critical (Rodríguez & Frahm, 2021). Model uncertainty quantification can be divided into frequentist and Bayesian approaches. In frequentist inference, model parameter estimation uncertainty is typically quantified via a confidence interval or standard deviation (Möller et al., 2020; Wang, Xie, Martagan, Akcay, & Corlu, 2019). In Bayesian inference, posterior distributions are used to quantify and update model uncertainty (Hernández Rodríguez et al., 2019; Xie, Wang, Li, Xie, & Auclair, 2022).

* Corresponding author.
E-mail addresses: w.xie@northeastern.edu (W. Xie), rbarton@psu.edu (R.R. Barton), nelsonb@northwestern.edu (B.L. Nelson), wang.keq@northeastern.edu (K. Wang).

https://doi.org/10.1016/j.ejor.2023.01.055
0377-2217/

This study is directly related to the existing frequentist and Bayesian approaches on uncertainty quantification and sensitivity analysis; see recent reviews in Corlu, Akcay, & Xie (2020), Borgonovo & Plischke (2016). The Bayesian approaches typically use the posterior distributions of inputs given the real-world data to quantify the input distribution uncertainty; see for example Zouaoui & Wilson (2003), Zouaoui & Wilson (2004), Biller & Corlu (2011). Direct bootstrapping, as frequentist approach, quantifies the impact of input uncertainty using bootstrap resampling of the input data and runs simulations at each bootstrap resample point to estimate the impact on the system mean (Barton & Schruben, 2001; Barton et al., 2007). Compared with the Bayesian approaches, the direct bootstrap can be adapted to any input process without additional analysis (e.g., posterior distribution derivation). The metamodel-assisted bootstrapping approach is further introduced by Barton, Nelson, & Xie (2014). In this framework, the uncertainty is propagated to the output mean by a metamodel, which can be constructed using simulation results from a small number of runs. Thus, this method does not need substantial computational effort.

Built on Barton et al. (2014), *we propose a metamodel-assisted uncertainty quantification and sensitivity analysis (UQ&SA) framework to accelerate the development of flexible manufacturing process with modular design.* As a result we can form a confidence interval (CI) quantifying the overall estimation uncertainty of the system's mean performance. Specifically, bootstrap resampling of the real-world data is used to approximate the model uncertainty. Then, a Gaussian process (GP) metamodel is used to propagate the heterogeneous process model uncertainty to the output mean response. Since model uncertainty typically dominates in the biopharmaceutical manufacturing processes, we further develop sensitivity analysis to quantify the contribution from each source of model uncertainty.

The key contributions of this study are threefold.

- First, we introduce a metamodel-assisted uncertainty quantification (UQ) and sensitivity analysis (SA) framework for hybrid model based simulations. The proposed algorithm can delivery a percentile CI of system mean response, accounting for both model and simulation uncertainties. A further sensitivity analysis can provide the relative contribution from each source of uncertainty. Differing with existing simulation studies in the literature that typically consider the simulation model as a black-box (see for example the review paper Corlu et al., 2020), hybrid model based simulation can leverage existing mechanistic models, facilitate mechanism learning, and support interpretable decision making.
- Second, under the assumption that the unknown mean response surface is a realization of GP, which is a useful representation in many problems, we provide a systematic asymptotic analysis on the proposed GP metamodel assisted UQ and SA framework, including (1) the asymptotic consistency of the proposed CI; and (2) the asymptotic consistency of variance estimators quantifying each source of model uncertainty and simulation uncertainty.
- Third, we provide a comprehensive empirical study to show that the proposed framework has promising finite sample performance, especially under situations with very limited real-world data.

Some existing simulation methodologies can be integrated into the proposed framework to support extensions for computational saving and system risk performance assessment, such as measured by quantiles. Considering the total simulation cost required to achieve consistent estimation of model uncertainty when using the conventional bootstrap resampling techniques, Lam & Qian (2018, 2022) proposed the subsampling techniques as a computational saver to promote the computational efficiency. In addition, the proposed UQ and SA framework can be extended to system quantile performance measure through GP based percentile regression; see for example Zhang, Wang, & Xie (2022), Xie, Wang, & Zhang (2018), Xie, Li, & Zhang (2017).

The remainder of the paper is organized as follows. We present the problem description in Section 2 and give a brief review of the metamodel-assisted bootstrapping approach in Section 3. In Section 4, we provide an algorithm to build an interval quantifying the overall estimation uncertainty of system mean performance, accounting for both model and simulation uncertainties. Then, we provide a variance decomposition approach to estimate the relative contribution from each source of model uncertainty, as well as simulation uncertainty. We provide an empirical study in Section 5 and conclude the paper in Section 6. All proofs are provided in the online appendix.

2. Problem description and proposed framework

A typical biomanufacturing system consists of multiple unit operations, including upstream fermentation for drug substance production and downstream purification to meet quality requirements (Doran, 2012). It can consist of numerous unit operations; see an example illustrated in Fig. 1. Operations typically include (1) fermentation, (2) centrifugation, (3) chromatography, (4) filtration, and (5) quality control. Operation unit (1) belongs to upstream cell culture and target drug substance production process, and (2)–(5) belong to downstream purification process.

To guide reliable and interpretable decision making, a simulation model can be developed based on hybrid models of modules defined according to bioprocess biophysicochemical mechanisms, dynamics, and interdependence of mechanistic parameters. Given very limited real-world data, we take existing mechanistic models as prior knowledge on the structure of mechanism relationships and create *parametric hybrid models*. It can leverage the advantages of mechanistic and statistical models to facilitate mechanism learning and improve sample efficiency and decision interpretability.

The fermentation is the most critical operation unit in the production process and it determines the generation of target drug substance (such as protein monoclonal antibodies or mAbs) and impurities. Here we use a simple example of fermentation on protein production to illustrate bioprocess hybrid modeling. Specifically, the target protein and biomass generation in the exponential-growth phase of fermentation process can be modeled with the cell-growth kinetics mechanism (Doran, 2012). Built on it, we construct a hybrid model capturing bioprocess dynamics and variation, i.e.,

$$X_t = X_0 \cdot e^{\gamma t} + \epsilon_P, \tag{1}$$

where X_t represents the biomass concentration at time t and the growth rate, denoted by γ, depends on biological properties of working cells and culture environments. We model batch-to-batch variation on: (1) the specific growth rate as $\gamma \sim N(\mu_\gamma, \sigma_\gamma^2)$; and (2) raw materials or initial concentration of seed cells as $X_0 \sim N(\mu_0, \sigma_0^2)$. In addition, we model the measurement error or residual as $\epsilon_P \sim N(0, \sigma_P^2)$ to capture the integrated impact from ignored factors. Larger variance from the residual indicates less understanding on underlying bioprocessing mechanisms obtained from the existing exponential growth mechanistic model. Thus, the distributions of residual ϵ_P, protein growth rate γ_P, and raw materials X_0 uniquely characterize the hybrid model of target protein accumulation during the fermentation process.

The normality assumption is often used in the biopharmaceutical literature to model batch-to-batch variations and measurement errors since they are often induced by many underlying factors; see for example Mockus, Peterson, Lainez, & Reklaitis (2015). In

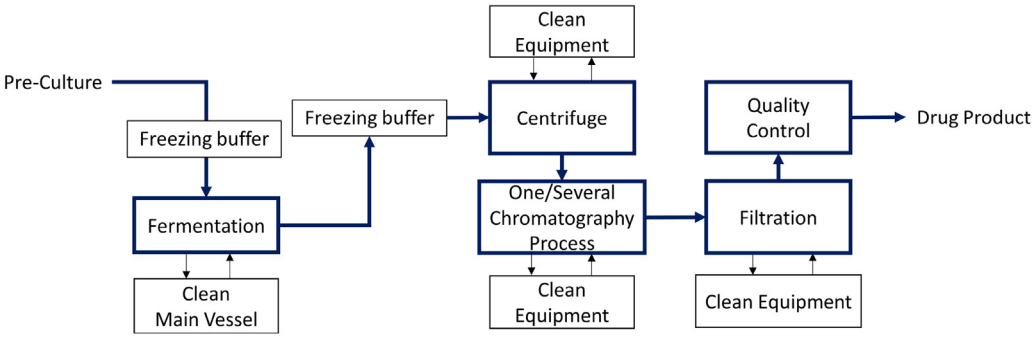

Fig. 1. An illustrative example of integrated biomanufacturing process.

addition, there is often very limited data. In our previous study, we used real-world fermentation process data with the size of 8 batches to conduct the hypothesis test which validates the normality assumption (Xie et al., 2022).

An integrated biomanufacturing system is often composed of multiple interconnected modules. Suppose that the simulation model is a function of L parametric multivariate and univariate models $F \equiv \{F_1, F_2, \ldots, F_L\}$ characterizing the underlying bioprocess dynamics and variations. Each ℓth model F_ℓ can be uniquely characterized by h_ℓ unknown parameters. In the simple fermentation example mentioned above in (1), the variation of residual ϵ_P is characterized by model F_1 specified by parameter σ_P^2; the batch-to-batch variation on the growth rate γ is characterized by model F_2 specified by parameters $\{\mu_\gamma, \sigma_\gamma^2\}$ and the raw material uncertainty is characterized by model F_3 specified by parameters $\{\mu_0, \sigma_0^2\}$.

Each h_ℓ-parameter distribution is uniquely specified by its first (finite) h_ℓ moments, which is true for the distributions that are most often used in stochastic simulation. The moments are chosen as the input variables for the metamodel of the system response surface because when they are close, the corresponding distributions will be similar and therefore generate similar outputs. Let $\mathbf{x}_{[\ell]}$ denote an $h_\ell \times 1$ vector of the first h_ℓ moments for the ℓth model and $d = \sum_{\ell=1}^{L} h_\ell$. Then, by stacking $\mathbf{x}_{[\ell]}$ with $\ell = 1, 2, \ldots, L$ together, we have a $d \times 1$ dimensional input vector, denoted by $\mathbf{x}$. Notice that $F = \{F_1, F_2, \ldots, F_L\}$ is completely characterized by the collection of model moments $\mathbf{x} = (\mathbf{x}_{[1]}, \mathbf{x}_{[2]}, \ldots, \mathbf{x}_{[L]})^\top$.

The output from the jth replication of a simulation with model moments $\mathbf{x}$ can be written as

$$Y_j(\mathbf{x}) = \mu(\mathbf{x}) + \epsilon_j(\mathbf{x}) \tag{2}$$

where $\mu(\mathbf{x}) = \mathrm{E}[Y_j(\mathbf{x})]$ denotes the unknown expected performance (e.g., productivity of protein drug substance) and $\epsilon_j(\mathbf{x})$ represents the simulation error with mean zero. The simulation output depends on the choice of process models. Let $\Psi \equiv \{\mathbf{x} \in \Re^d$: the random variable $Y(\mathbf{x})$ is defined and$\mu(\mathbf{x})$is finite$\}$ denote the region of interest. We assume $\mu(\mathbf{x})$ is continuous for $\mathbf{x} \in \Psi$.

The underlying "correct" process models, denoted by $F^c \equiv \{F_1^c, F_2^c, \ldots, F_L^c\}$, specified by the moments, $\mathbf{x}_c = (\mathbf{x}_{[1],c}, \mathbf{x}_{[2],c}, \ldots, \mathbf{x}_{[L],c})$, are unknown and are estimated from a finite sample of real-world data. Suppose that the set of true parameters $\mathbf{x}_c$ is in the interior of Ψ. Our goal is to find a $(1-\alpha)$100% CI, denoted by $[Q_L, Q_U]$, such that

$$\Pr\{\mu(\mathbf{x}_c) \in [Q_L, Q_U]\} = 1 - \alpha, \tag{3}$$

which quantifies the overall estimation uncertainty of system mean performance, accounting for simulation and model uncertainties. Then, if this interval is too wide, we further develop a variance decomposition to quantify the contribution from each source of model uncertainty, which can guide more data collection and improve the system mean response estimation.

The true moments $\mathbf{x}_c$ are unknown and estimated based on a finite sample $\mathbf{Z_m}$ from F^c. Let m_ℓ denote the number of i.i.d. real-world observations available from the ℓth model, i.e., $\mathbf{Z}_{\ell,m_\ell} \equiv \{Z_{\ell,1}, Z_{\ell,2}, \ldots, Z_{\ell,m_\ell}\}$ with $Z_{\ell,i} \overset{i.i.d}{\sim} F_\ell^c$, $i = 1, 2, \ldots, m_\ell$. Let $\mathbf{Z_m} = \{\mathbf{Z}_{\ell,m_\ell}, \ell = 1, 2, \ldots, L\}$ be the collection of samples from all L model distributions in F^c, where $\mathbf{m} = (m_1, m_2, \ldots, m_L)$. Let $\mathbf{X_m}$ be a $d \times 1$ dimensional moment estimator that is a function of $\mathbf{Z_m}$ written as $\mathbf{X_m} = \mathbf{X}(\mathbf{Z_m})$. Specifically, $\mathbf{X}_{\ell,m_\ell} = \mathbf{X}_\ell(\mathbf{Z}_{\ell,m_\ell})$ and $\mathbf{X}_\mathbf{m}^T = (\mathbf{X}_{1,m_1}^T, \mathbf{X}_{2,m_2}^T, \ldots, \mathbf{X}_{L,m_L}^T)$. Let $F_{\mathbf{X_m}}^c$ represent the true, unknown distribution of $\mathbf{X_m}$. Therefore, the impact of model uncertainty is captured by the sampling distribution of $\mu(\mathbf{X_m})$ with $\mathbf{X_m} \sim F_{\mathbf{X_m}}^c$. The real-world data are a particular realization of $\mathbf{Z_m}$, say $\mathbf{z}_\mathbf{m}^{(0)}$. Given a finite sample of real-world data $\mathbf{z}_\mathbf{m}^{(0)}$, we use bootstrap resampling to approximate $F_{\mathbf{X_m}}^c$ and a metamodel to represent $\mu(\mathbf{x})$. Notice that the components of the moment estimator $\mathbf{X_m}$ can be statistically dependent.

Suppose each experiment is expensive. The proposed metamodel-assisted bootstrapping uncertainty analysis framework can accelerate the development of a simulation model for a flexible and integrated real manufacturing system with modular design. Since the underlying response surface $\mu(\cdot)$ is unknown, we model our prior belief about $\mu(\cdot)$ by a Gaussian Process (GP). Given a set of stochastic simulation outputs, the GP-based belief is updated by a posterior distribution, denoted by $M_p(\cdot)$. When we use this metamodel to propagate the sampling distribution of $\mathbf{X_m}$ to the output mean, it introduces the simulation uncertainty induced by finite simulation runs (i.e., finite design points and finite run length in each simulation run). *Thus, the estimation uncertainty of underlying system mean performance $\mu(\mathbf{x}_c)$ is characterized by the compound random variable, $M_p(\mathbf{X_m})$, accounting for both model and simulation uncertainties.* Based on the variability of $M_p(\mathbf{X_m})$, we can construct an interval estimator $[Q_L, Q_U]$ in (3) to quantify the overall estimation uncertainty of real system mean response $\mu(\mathbf{x}_c)$.

We further develop a variance decomposition measuring the contributions to $\mathrm{Var}[M_p(\mathbf{X_m})]$ from simulation uncertainty quantified by GP $M_p(\cdot)$ and model uncertainty quantified by the sampling distribution of $\mathbf{X}_\mathbf{m}^T = (\mathbf{X}_{1,m_1}^T, \mathbf{X}_{2,m_2}^T, \ldots, \mathbf{X}_{L,m_L}^T)$. Therefore, if this interval is too wide, our study can guide further data collection to efficiently update the simulation model to faithfully represent the real system and improve the estimation accuracy of $\mu(\mathbf{x}_c)$.

If the simulation uncertainty dominates, we will allocate more computational resource to improve our knowledge on the mean response surface $\mu(\cdot)$. However, in biopharmaceutical manufacturing with high stochasticity and very limited process observations, model uncertainty often dominates. The distribution of

model uncertainty depends on heterogeneous process observations, as well as the complexity of the underlying mechanisms and inherent stochasticity at each part of the integrated biomanufacturing system. Thus, if certain model uncertainty, say $\mathbf{X}_{\ell,m_\ell}$ with $\ell = 1, 2, \ldots, L$, dominates the system performance estimation uncertainty, it will guide us collecting the additional real-world data there to improve the simulation model.

3. Metamodel-Assisted bootstrapping for UQ

We introduce the metamodel-assisted bootstrapping and provide the algorithm for uncertainty analysis. Basically, we first find the space-filling design points covering the most likely bootstrap samples of model moments, denoted by $\widehat{\mathbf{X}}_{\mathbf{m}}^{(b)}$ with $b = 1, 2, \ldots, B$, quantifying the model uncertainty. Then, we run simulations and construct the GP or stochastic kriging (SK) metamodel for the mean response surface $\mu(\cdot)$ quantifying simulation uncertainty in Section 3.1. This metamodel is used to propagate the model uncertainty to output mean. We introduce the metamodel-assisted bootstrapping in Section 3.2 to construct an interval of $\mu(\mathbf{x}_c)$ accounting for both simulation and model uncertainties, and show its asymptotic consistency in Section 3.3.

3.1. Stochastic kriging metamodel

Since the outputs from simulations include simulation variability that often changes significantly across the design space of process models specified by moments $\mathbf{x}$, SK is introduced to distinguish the uncertainty about the response surface from the simulation uncertainty (Ankenman, Nelson, & Staum, 2010; Kleijnen, 2017). Suppose that the underlying unknown response surface can be thought of as a realization of a stationary GP. The simulation output Y is modeled as,

$$Y_j(\mathbf{x}) = \beta_0 + W(\mathbf{x}) + \epsilon_j(\mathbf{x}) \tag{4}$$

where $\mathbf{x}$ denotes a $d \times 1$ vector of model moments. SK uses a mean-zero, second-order stationary GP $W(\mathbf{x})$ to account for the spatial dependence of the response surface. Thus, the uncertainty about the true response surface $\mu(\mathbf{x})$ is represented by a GP $M(\mathbf{x}) \equiv \beta_0 + W(\mathbf{x})$ (note that β_0 can be replaced by a more general trend term $\mathbf{f}(\mathbf{x})^\top \beta$). For many, but not all, simulation settings the output is an average of a large number of more basic outputs, so a normal approximation can be applied: $\epsilon(\mathbf{x}) \sim \mathrm{N}(0, \sigma_\epsilon^2(\mathbf{x}))$.

In SK, the covariance between $W(\mathbf{x})$ and $W(\mathbf{x}')$ quantifies how knowledge of the surface at some design points affects the prediction of the surface. A parametric form of the spatial covariance, denoted by $\Sigma(\mathbf{x}, \mathbf{x}') = \mathrm{Cov}[W(\mathbf{x}), W(\mathbf{x}')] = \tau^2 r(\mathbf{x} - \mathbf{x}')$, is typically assumed where τ^2 denotes the variance and $r(\cdot)$ is a correlation function that depends only on the distance $\mathbf{x} - \mathbf{x}'$. Based on our previous study (Xie, Nelson, & Staum, 2010), we use the product-form Gaussian correlation function $r(\mathbf{x} - \mathbf{x}') = \exp(-\sum_{j=1}^{d} \theta_j (x_j - x_j')^2)$ for the empirical evaluation in Section 5. Let $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_d)$ represent the correlation parameters. Thus, the prior knowledge of the response surface $\mu(\mathbf{x})$ is represented by a Gaussian process, i.e., $M(\mathbf{x}) \sim \mathrm{GP}(\beta_0, \tau^2 r(\mathbf{x} - \mathbf{x}'))$.

To reduce the uncertainty about $\mu(\mathbf{x})$, we choose an experiment design consisting of pairs $\mathcal{D} \equiv \{(\mathbf{x}_i, n_i), i = 1, 2, \ldots, k\}$ at which to run simulations and collect observations, where $(\mathbf{x}_i, n_i)$ denotes the location and the number of replications, respectively, at the ith design point. The design that we recommend is described in more detail in Appendix D. The simulation outputs at $\mathcal{D}$ are $\mathbf{Y}_\mathcal{D} \equiv \{(Y_1(\mathbf{x}_i), Y_2(\mathbf{x}_i), \ldots, Y_{n_i}(\mathbf{x}_i)); i = 1, 2, \ldots, k\}$ and the sample mean at design point $\mathbf{x}_i$ is $\bar{Y}(\mathbf{x}_i) = \sum_{j=1}^{n_i} Y_j(\mathbf{x}_i)/n_i$. Let the sample means at all k design points be $\bar{\mathbf{Y}}_\mathcal{D} = (\bar{Y}(\mathbf{x}_1), \bar{Y}(\mathbf{x}_2), \ldots, \bar{Y}(\mathbf{x}_k))^T$. Set the simulations at different design points independent. Then, the variance of $\bar{\mathbf{Y}}_\mathcal{D}$ is represented by a $k \times k$ diagonal matrix $C = \mathrm{diag}\{\sigma_\epsilon^2(\mathbf{x}_1)/n_1, \sigma_\epsilon^2(\mathbf{x}_2)/n_2, \ldots, \sigma_\epsilon^2(\mathbf{x}_k)/n_k\}$.

Let Σ be the $k \times k$ spatial covariance matrix of the design points and let $\Sigma(\mathbf{x}, \cdot)$ be the $k \times 1$ spatial covariance vector between the design points and a fixed prediction point $\mathbf{x}$. If the parameters $(\tau^2, \boldsymbol{\theta}, C)$ are known, then the metamodel or simulation uncertainty can be characterized by a refined GP $M_p(\mathbf{x})$ that denotes the conditional distribution of $M(\mathbf{x})$ given simulation outputs $\bar{\mathbf{Y}}_\mathcal{D}$,

$$M_p(\mathbf{x}) \sim \mathrm{GP}\big(m_p(\mathbf{x}), \sigma_p^2(\mathbf{x})\big) \tag{5}$$

where the minimum mean squared error (MSE) linear unbiased predictor is

$$m_p(\mathbf{x}) = \widehat{\beta}_0 + \Sigma(\mathbf{x}, \cdot)^\top (\Sigma + C)^{-1} \big(\bar{\mathbf{Y}}_\mathcal{D} - \widehat{\beta}_0 \cdot 1_{k\times 1}\big), \tag{6}$$

and the corresponding variance is

$$\begin{aligned} \sigma_p^2(\mathbf{x}) = {} & \tau^2 - \Sigma(\mathbf{x}, \cdot)^\top (\Sigma + C)^{-1} \Sigma(\mathbf{x}, \cdot) \\ & + \eta^\top \big[1_{k\times 1}^\top (\Sigma + C)^{-1} 1_{k\times 1}\big]^{-1} \eta \end{aligned} \tag{7}$$

where $\widehat{\beta}_0 = [1_{k\times 1}^\top (\Sigma + C)^{-1} 1_{k\times 1}]^{-1} 1_{k\times 1}^\top (\Sigma + C)^{-1} \bar{\mathbf{Y}}_\mathcal{D}$ and $\eta = 1 - 1_{k\times 1}^\top (\Sigma + C)^{-1} \Sigma(\mathbf{x}, \cdot)$ (Ankenman et al., 2010). The spatial correlation parameters τ^2 and $\boldsymbol{\theta}$ are estimated by using MLEs. The sample variance is used as an estimate for the simulation variance at design points C. By plugging $(\widehat{\beta}_0, \widehat{\tau}^2, \widehat{\boldsymbol{\theta}}, \widehat{C})$ into Eqs. (6) and (7), we can obtain the estimated mean $\widehat{m}_p(\mathbf{x})$ and variance $\widehat{\sigma}_p^2(\mathbf{x})$. Thus, the metamodel we use is $\widehat{\mu}(\mathbf{x}) = \widehat{m}_p(\mathbf{x})$ with marginal variance estimated by $\widehat{\sigma}_p^2(\mathbf{x})$.

Ankenman et al. (2010) demonstrate that $\widehat{m}_p(\mathbf{x})$ is still an unbiased predictor even with the plug-in estimator $\widehat{C}$, and the variance inflation of $\sigma_p^2(\mathbf{x})$ caused by using $\widehat{C}$ is typically small. In the asymptotic analysis, we assume that the parameters $(\tau^2, \boldsymbol{\theta}, C)$ are known. This is necessary (and common in the kriging literature) because including the effect of parameter estimation is mathematically intractable. Further, there is both theoretical and empirical evidence that in many cases prediction accuracy is minimally affected by using estimated hyperparameters; see Wang et al. (2021).

3.2. Metamodel-Assisted bootstrapping for uncertainty quantification

The proposed metamodel-assisted bootstrapping can provide a CI for the true mean performance, which accounts for both model and simulation uncertainties. Since $m_p(\mathbf{x})$ is an unbiased predictor under the GP assumption, $\sigma_p^2(\mathbf{x}) = 0$ for all $\mathbf{x}$ would imply that there is no simulation uncertainty due either to a finite number of design points $\mathbf{x}_i$ or finite number of replications n_i; that is, $m_p(\mathbf{x}) = \mu(\mathbf{x})$. Unfortunately, if the budget is tight relative to the complexity of the true response surface, then the effect of simulation uncertainty could be substantial, resulting in significant undercoverage of the confidence interval of Barton et al. (2014) as we show in Section 5. The new interval introduced here does not suffer this degradation, and therefore is robust to the amount of simulation effort that can be expended.

The kriging literature is the foundation for our work; see for instance Santner, Williams, & Notz (2003). Kriging models uncertainty about the function as a GP $M(\cdot)$ by assuming $\mu(\cdot)$ is a realization of $M(\cdot)$. An interval constructed to cover the conditional distribution of $M(\mathbf{x}_0)$ given the values at the design points is often interpreted as a CI for $\mu(\mathbf{x}_0)$; see for example Picheny, Ginsbourger, Roustant, Haftka, & Kim (2010). The success of this paradigm is not because the function of interest is actually random—it is not—but because in many problems the conditional GP appears to be a robust characterization of the remaining response-surface uncertainty.

We adopt the kriging paradigm but with two key differences: our prediction point $\mathbf{x}_c$ is unknown and must be estimated from real-world data, and our function $\mu(\cdot)$ can only be evaluated in the presence of stochastic simulation noise. Given the simulation outputs $\bar{\mathbf{Y}}_{\mathcal{D}}$, the remaining uncertainty about $\mu(\cdot)$ is characterized by the conditional GP $M_p(\cdot)$. To account for the impact from both model and simulation uncertainties, we construct an interval $[C_L, C_U]$ covering $M_p(\mathbf{x}_c)$ with probability $(1-\alpha)100\%$, i.e.,

$$\Pr\{M_p(\mathbf{x}_c) \in [C_L, C_U]\} = 1-\alpha. \tag{8}$$

Since the conditional coverage is $1-\alpha$, the unconditional coverage of $M(\mathbf{x}_c)$ is $1-\alpha$ as well. *The revised objective (8) is connected to our objective (3) through the assumption that the function $\mu(\cdot)$ is a realization of the GP $M(\cdot)$.* A procedure that delivers an interval satisfying (8) will be a good approximation for a CI procedure satisfying (3) if $M_p(\cdot)$ faithfully represents the remaining uncertainty about $\mu(\cdot)$. This is clearly an approximation because in any real problem $\mu(\cdot)$ is a fixed function, therefore we refer to $[C_L, C_U]$ as an approximation for the CI (ACI).

Based on a hierarchical approach, we propose Algorithm 1 to build $(1-\alpha)100\%$ bootstrap percentile ACIs to achieve (8). In this procedure, Step 1 provides an experiment design to build a SK metamodel, which is central to the metamodel-assisted bootstrapping approach. Since the system model uncertainty is quantified with bootstrapped samples, we want the metamodel to correctly predict the responses at these sample points $\widehat{\mathbf{X}}_{\mathbf{m}} \sim \widehat{F}_{\widehat{\mathbf{X}}_{\mathbf{m}}}(\cdot|\mathbf{z}_{\mathbf{m}}^{(0)})$. Thus, the metamodel needs to be accurate and precise in a design space that covers the "most likely" bootstrap moment estimates, which can be achieved by the experiment design proposed by Barton et al. (2014). Specifically, they find the smallest ellipsoid denoted by E that covers the most likely bootstrap resampled moments and then generate a space-filling design that covers E; see the details in online Appendix D.

Based on the experiment design provided in Step 1, we run simulations and construct a metamodel in Step 2 by fitting $(\beta_0, \tau^2, \boldsymbol{\theta}, C)$. Given the metamodel, we predict the simulation's mean responses at different model settings corresponding to bootstrap resampled moments. The bootstrap resampled moments are drawn from the bootstrap distribution denoted by $\widehat{F}_{\widehat{\mathbf{X}}_{\mathbf{m}}}(\cdot|\mathbf{z}_{\mathbf{m}}^{(0)})$. In Step 3(a), we generate bootstrapped model moments. Then, we return a $(1-\alpha)100\%$ interval estimators as shown in Algorithm 1. Notice that Step 3(b) accounts for the model uncertainty and Step 3(c) accounts for the simulation uncertainty. Thus, this procedure provides two types of intervals: **(a)** CI_0, proposed in Barton et al. (2014), returns an estimate of $[Q_L, Q_U]$ in Eq. (3) by assuming $\widehat{m}_p(\mathbf{x}) = \mu(\mathbf{x})$; that is, it only accounts for model uncertainty and will be in error if there is substantial simulation uncertainty. **(b)** CI_+ returns an estimate of $[C_L, C_U]$ in Eq. (8). This ACI accounts for both model and simulation uncertainty. As the simulation uncertainty decreases, CI_0 and CI_+ become closer and closer to each other. Before evaluating the finite-sample performance of CI_+ in Section 5, we establish its asymptotic consistency for objective (8) in Section 3.3. Then, in Steps 4 and 5, variance decomposition is developed to quantify the contribution from each source of uncertainty, which will be studied in Section 4.

3.3. Asymptotic consistency study on interval CI_+

In this section, we show that the ACI CI_+ provided in Algorithm 1 satisfies Eq. (8) asymptotically. *The asymptotic consistency of this interval is proved under the assumption that the true response surface $\mu(\mathbf{x})$ is a realization of a GP with all parameters known except β_0.* Under this assumption, $M_p(\mathbf{x})$ characterizes the remaining simulation uncertainty after observing $\bar{\mathbf{Y}}_{\mathcal{D}}$. Since the model uncertainty is asymptotically correctly quantified by the bootstrap moment estimator $\widehat{\mathbf{X}}_{\mathbf{m}}$, the distribution of $M_p(\widehat{\mathbf{X}}_{\mathbf{m}})$ accounts for both model and simulation uncertainties. Theorem 3.1 shows that this interval satisfies objective (8) asymptotically. The detailed proof is provided in online Appendix B.

Algorithm 1: Metamodel-Assisted Bootstrap for UQ and SA.

Input: Given real-world data $\mathbf{z}_{\mathbf{m}}^{(0)} = \{\mathbf{z}_{\ell,m_\ell}^{(0)}, \ell = 1,2,\ldots,L\}$
Output: Estimated CI and ACI quantifying the overall estimation uncertainty of $\mu(\mathbf{x}_c)$; Estimated model variance $\widehat{\sigma}_I^2$, simulation variance $\widehat{\sigma}_M^2$ and uncertainty contribution $\widehat{s}_\ell$ from ℓth model.

Function of Uncertainty Quantification (UQ):

Step 1: Build the design space covering the most likely bootstrap moment estimates of process models, and choose a space-filling experiment design $\mathcal{D} = \{(\mathbf{x}_i, n_i), i = 1,2,\ldots,k\}$ as described in online Appendix D.

Step 2: Run simulations at design points to obtain outputs $\mathbf{Y}_{\mathcal{D}}$. Compute the sample average $\bar{Y}(\mathbf{x}_i)$ and sample variance $S^2(\mathbf{x}_i)$ of the simulation outputs, $i = 1,2,\ldots,k$. Fit the SK metamodel parameters $(\beta_0, \tau^2, \boldsymbol{\theta}, C)$ to obtain $\widehat{m}_p(\mathbf{x})$ and $\widehat{\sigma}_p^2(\mathbf{x})$ using $\left(\bar{Y}(\mathbf{x}_i), S^2(\mathbf{x}_i), \mathbf{x}_i\right)$, $i = 1,2,\ldots,k$.

Step 3: for $b = 1$ *to* B **do**

Step 3(a): Draw m_ℓ samples with replacement from $\mathbf{z}_{\ell,m_\ell}^{(0)}$, denoted by $\mathbf{Z}_{\ell,m_\ell}^{(b)}$, and calculate the corresponding $h_\ell \times 1$ vector of bootstrap moment estimates denoted by $\widehat{\mathbf{X}}_{\ell,m_\ell}^{(b)} = \mathbf{X}_\ell(\mathbf{Z}_{\ell,m_\ell}^{(b)})$ for $\ell = 1,2,\ldots,L$. Then stack the results for all L processes to obtain a $d \times 1$ vector $\widehat{\mathbf{X}}_{\mathbf{m}}^{(b)}$.

Step 3(b): Let $\widehat{\mu}_b \equiv \widehat{m}_p(\widehat{\mathbf{X}}_{\mathbf{m}}^{(b)})$.

Step 3(c): Draw $\widehat{M}_b \sim \mathrm{N}\left(\widehat{m}_p(\widehat{\mathbf{X}}_{\mathbf{m}}^{(b)}), \widehat{\sigma}_p^2(\widehat{\mathbf{X}}_{\mathbf{m}}^{(b)})\right)$.

Return (1) Estimated $(1-\alpha)100\%$ bootstrap percentile CI and ACI; **(2)** Estimated model variance and simulation variance,

$$\mathrm{CI}_0 = \left[\widehat{\mu}_{(\lceil B\frac{\alpha}{2}\rceil)}, \widehat{\mu}_{(\lceil B(1-\frac{\alpha}{2})\rceil)}\right], \quad \mathrm{CI}_+ = \left[\widehat{M}_{(\lceil B\frac{\alpha}{2}\rceil)}, \widehat{M}_{(\lceil B(1-\frac{\alpha}{2})\rceil)}\right],$$

$$\widehat{\sigma}_I^2 = \sum_{b=1}^{B}(\widehat{\mu}_b - \widehat{\mu})^2/(B-1), \quad \widehat{\sigma}_M^2 = \sum_{b=1}^{B}\widehat{\sigma}_p^2(\widehat{\mathbf{X}}_{\mathbf{m}}^{(b)})/B,$$

where $\widehat{\mu}_{(1)} \le \widehat{\mu}_{(2)} \le \cdots \le \widehat{\mu}_{(B)}$ and $\widehat{M}_{(1)} \le \widehat{M}_{(2)} \le \cdots \le \widehat{M}_{(B)}$ are the sorted values, and $\widehat{\mu} = \sum_{b=1}^{B}\widehat{\mu}_b/B$.

Function of Sensitivity Analysis (SA):

for *each* $\mathcal{J} \subseteq \mathcal{L}$ *with* $\mathcal{L} = \{1,2,\ldots,L\}$ **do**

Step 4: Generate bootstrap samples $\widehat{\mathbf{X}}_{\mathcal{J}}^{(b)}$ for $b = 1,2,\ldots,B'$ and obtain simulation output prediction $\widehat{m}_p(\mathbf{x}_{-\mathcal{J}}^{(0)}, \widehat{\mathbf{X}}_{\mathcal{J}}^{(b)})$.

Step 5: Estimate the cost function $c(\mathcal{J})$ by (13).

Return Estimated ℓth model uncertainty contribution $\widehat{s}_\ell$ through (12) with $\ell = 1,2,\ldots,L$.

Theorem 3.1. *Suppose that Assumptions ($\star$) in online Appendix A hold. Then the interval $\mathrm{CI}_+ = [M_{(\lceil B\frac{\alpha}{2}\rceil)}, M_{(\lceil B(1-\frac{\alpha}{2})\rceil)}]$ is asymptotically consistent,*

$$\lim_{m\to\infty}\lim_{B\to\infty}\Pr\left\{M_{(\lceil B\alpha/2\rceil)} \le M_p(\mathbf{x}_c) \le M_{(\lceil B(1-\alpha/2)\rceil)}\right\} = 1-\alpha. \tag{9}$$

4. Variance decomposition for uncertainty analysis

In a practical setting, what is the next step if the interval CI_+ is so wide that we are uncomfortable making decisions based on estimates with that level of error? We suggest gaining some sense of the relative contribution from each source of uncertainty as a

guide toward either collecting more real-world process data to reduce the model uncertainty or running more simulations to improve the system mean response estimation at any given models F. The overall estimation variance of system true performance $\mu(\mathbf{x}_c)$ is quantified by $\mathrm{Var}[M_p(\mathbf{X}_m)]$. In Section 4.1, we propose a variance decomposition approach to quantify the contribution from simulation and model uncertainties. Compared with the existing studies on estimating the relative contributions, such as Song & Nelson (2013), our variance decomposition does not require the homogeneity assumption, i.e., the simulation noise has a constant variance. Since the effect of model uncertainty is induced by the complex interactions of estimation uncertainties from L models $(F_1, F_2, \ldots, F_L)$, we further decompose it by using Shapley value (SV) based global sensitivity analysis to correctly quantify the contribution from each source of model uncertainty in Section 4.2. This information can provide a guide on which model F_ℓ to collect more real-world data and improve the system mean performance estimation. Then, we provide the asymptotic consistency study over the variance component estimation for each source of uncertainty in Section 4.3.

4.1. Simulation and model uncertainty contribution quantification

Suppose that the parameters $(\tau^2, \boldsymbol{\theta}, C)$ are known, the simulation uncertainty can be characterized by a GP, and the simulation error follows a normal distribution. Then given the simulation outputs $\bar{\mathbf{Y}}_{\mathcal{D}}$, the simulation uncertainty is characterized by a GP, i.e., $M_p(\mathbf{x}) \sim \mathrm{N}(m_p(\mathbf{x}), \sigma_p^2(\mathbf{x}))$. Conditional on $\bar{\mathbf{Y}}_{\mathcal{D}}$, both $m_p(\mathbf{x})$ and $\sigma_p^2(\mathbf{x})$ are fixed functions. For notation simplification, all of following derivations are conditional on the simulation outputs $\bar{\mathbf{Y}}_{\mathcal{D}}$, but we will suppress the "$|\bar{\mathbf{Y}}_{\mathcal{D}}$".

To quantify the relative contribution of model and simulation uncertainties, we decompose the total variance of $M_p(\mathbf{X_m})$ into two parts:

$$\begin{aligned}\sigma_T^2 &\equiv \mathrm{Var}[M_p(\mathbf{X_m})] \\ &= \mathrm{E}\{\mathrm{Var}[M_p(\mathbf{X_m})|\mathbf{X_m}]\} + \mathrm{Var}\{\mathrm{E}[M_p(\mathbf{X_m})|\mathbf{X_m}]\} \\ &= \mathrm{E}[\sigma_p^2(\mathbf{X_m})] + \mathrm{Var}[m_p(\mathbf{X_m})]. \end{aligned} \tag{10}$$

The term $\sigma_M^2 \equiv \mathrm{E}[\sigma_p^2(\mathbf{X_m})]$ is a measure of the simulation uncertainty: the expected metamodel variance weighted by the density of moment estimator $\mathbf{X_m}$. This weighting makes sense because the accuracy of the metamodel in regions with higher density is more important for the estimation of system mean performance. The term $\sigma_I^2 \equiv \mathrm{Var}[m_p(\mathbf{X_m})]$ is a measure of model uncertainty when we replace the unknown true response surface $\mu(\cdot)$ with its best linear unbiased estimate $m_p(\cdot)$.

If the simulation uncertainty disappears (i.e., $\sigma_p^2(\cdot) = 0$), then $\sigma_M^2 = 0$, CI_0 and CI_+ coincide. On the other hand, as $m \to \infty$ (more and more real-world data), $\mathbf{X_m} \overset{a.s.}{\to} \mathbf{x}_c$ and since $m_p(\mathbf{x})$ is continuous we have $\sigma_I^2 = 0$; therefore, the width of CI_0 shrinks to zero as does coverage since there is remaining simulation uncertainty in general. However, because CI_+ accounts for simulation uncertainty it still provides asymptotically consistent coverage. This effect is demonstrated by the empirical study in Section 5.

Our decomposition allows us to express the total variance in Eq. (10) as the sum of two variances measuring model and simulation uncertainties: $\sigma_T^2 = \sigma_I^2 + \sigma_M^2$. In the metamodel-assisted bootstrapping framework, we can estimate each variance component as follows:

- Total variance: $\widehat{\sigma}_T^2 = \sum_{b=1}^B (M_b - \bar{M})^2/(B-1)$, where $\bar{M} = \sum_{b=1}^B M_b/B$.
- Model variance: $\widehat{\sigma}_I^2 = \sum_{b=1}^B (\mu_b - \bar{\mu})^2/(B-1)$, where $\bar{\mu} = \sum_{b=1}^B \mu_b/B$.
- Simulation variance: $\widehat{\sigma}_M^2 = \sum_{b=1}^B \sigma_p^2(\widehat{\mathbf{X}}_{\mathbf{m}}^{(b)})/B$.

The ratio $\widehat{\sigma}_I/\widehat{\sigma}_T$ provides an estimate of the relative contribution from model uncertainty on CI_+. If it is close to 1, the contribution from simulation uncertainty can be ignored. Thus, this ratio can help a decision maker determine where to put more effort: If the model variance dominates, then get more real-world data (if possible). If the simulation variance dominates, then it can be reduced by more simulations, which can be a combination of additional design points and additional replications at existing design points. If neither dominates, then both activities are necessary to reduce CI_+ to a practically useful size.

4.2. Variance decomposition for model uncertainty analysis

The overall model variance $\sigma_I^2 = \mathrm{Var}[m_p(\mathbf{X_m})]$ is induced by the estimation uncertainty of correct moments $\mathbf{x}_c = (\mathbf{x}_{[1],c}, \mathbf{x}_{[2],c}, \ldots, \mathbf{x}_{[L],c})$ for process models $F^c = \{F_1^c, F_2^c, \ldots, F_L^c\}$. To efficiently identify the bottlenecks and reduce the impact of model uncertainty, we are interested in quantifying the contribution of moment estimation uncertainty of $\mathbf{X}_{\ell,m_\ell} = \mathbf{X}_\ell(\mathbf{Z}_{\ell,m_\ell})$ for each ℓth model F_ℓ. To approximate the estimation uncertainty of $\mathbf{X}_{\ell,m_\ell}$ with $\ell = 1, 2, \ldots, L$, the bootstrap resampled moments are drawn from the bootstrap distribution, $\widehat{\mathbf{X}}_{\ell,m_\ell} \sim F_{\mathbf{X}_{\ell,m_\ell}}(\cdot|\mathbf{z}_{\ell,m_\ell}^{(0)})$.

Motivated by the SV based sensitivity analysis (see for example Song, Nelson, & Staum, 2016), the overall model variance σ_I^2 in (10) can be decomposed as the sum of contributions from each source of model uncertainty,

$$\begin{aligned}\sigma_I^2 &= \mathrm{Var}[m_p(\mathbf{X_m})] \\ &= \mathrm{Var}[m_p(\mathbf{X}_{1,m_1}, \mathbf{X}_{2,m_2}, \ldots, \mathbf{X}_{L,m_L})] = \sum_{\ell=1}^L s_\ell, \end{aligned} \tag{11}$$

with s_ℓ quantifying the contribution from the ℓth model uncertainty,

$$s_\ell = \sum_{\mathcal{J} \subseteq \mathcal{L}\setminus\{\ell\}} \frac{(L - |\mathcal{J}| - 1)!|\mathcal{J}|!}{L!}[c(\mathcal{J} \cup \{\ell\}) - c(\mathcal{J})], \tag{12}$$

where $\mathcal{L} = \{1, 2, \ldots, L\}$ denotes the index set of L sources of model uncertainty and $|\cdot|$ indicates the set size. Here, for any subset $\mathcal{J} \subseteq \mathcal{L}$, we use the total effect based cost function $c(\mathcal{J}) = \mathrm{E}[\mathrm{Var}[m_p(\mathbf{X_m})|\mathbf{X}_{-\mathcal{J}}]]$ measuring the expected remaining variance when all other model moments, denoted by $\mathbf{X}_{-\mathcal{J}}$, are conditionally fixed, where $-\mathcal{J}$ denotes the remaining subset $\mathcal{L} \setminus \mathcal{J}$.

The metamodel-assisted bootstrap resampling is used to estimate the contribution from each source of model uncertainty (Algorithm 1). Basically, for any model with the index $i \notin \mathcal{J}$ or $i \in \mathcal{L} \setminus \mathcal{J}$, we take the sample moment $\mathbf{x}_{i,m_i}^{(0)}$ as true one. Denote these model moments by $\mathbf{x}_{-\mathcal{J}}^{(0)}$. Then, for the model with index $j \in \mathcal{J}$, we draw with replacement to generate the bootstrap sample moments quantifying the corresponding model uncertainty, $\widehat{\mathbf{X}}_{j,m_j}^{(b)} \sim F_{\mathbf{X}_{j,m_j}}(\cdot|\mathbf{z}_{j,m_j}^{(0)})$ with $b = 1, 2, \ldots, B'$. We represent the combination of bootstrap moment samples for all model moments with index $j \in \mathcal{J}$ by $\widehat{\mathbf{X}}_{\mathcal{J}}^{(b)}$. Thus, we estimate $c(\mathcal{J})$ by a Monte Carlo sampling approach,

$$\widehat{c}(\mathcal{J}) = \frac{1}{B'-1}\sum_{b=1}^{B'}\left[\widehat{m}_p\left(\mathbf{x}_{-\mathcal{J}}^{(0)}, \widehat{\mathbf{X}}_{\mathcal{J}}^{(b)}\right) - \bar{m}_{\mathcal{J}}\right]^2 \tag{13}$$

where $\bar{m}_{\mathcal{J}} = \sum_{b=1}^{B'} \widehat{m}_p(\mathbf{x}_{-\mathcal{J}}^{(0)}, \widehat{\mathbf{X}}_{\mathcal{J}}^{(b)})/B'$. By plugging $\widehat{c}(\mathcal{J})$ into Eq. (12), we can get the estimator $\widehat{s}_\ell$ quantifying the contribution from the ℓth model uncertainty to $\mathrm{Var}[m_p(\mathbf{X_m})]$. An efficient approximation algorithm, using the randomly selected subset instead of all possible index sets permutations, can be used to reduce the computational burden; see Song et al. (2016).

4.3. Asymptotic consistency study of variance contribution estimation

We provide the asymptotic consistency study of variance contribution estimation from each source of uncertainty; see Theorems 4.1, 4.2, and 4.3.

Theorem 4.1. *Suppose that Assumptions 1–4 in online Appendix A hold. Then conditional on* $\bar{\mathbf{Y}}_{\mathcal{D}}$, *the variance component estimators* $\widehat{\sigma}_M^2, \widehat{\sigma}_I^2, \widehat{\sigma}_T^2$ *are consistent as* $m, B \to \infty$, *where as* $m \to \infty$ *we have* $m_\ell/m \to c_\ell$, $\ell = 1, 2, \ldots, L$, *for a constant* $c_\ell > 0$. *Specifically,*

- *As* $m \to \infty$, *the model uncertainty disappears:*

$$\lim_{m\to\infty} \sigma_M^2 = \sigma_p^2(\mathbf{x}_c),\ \lim_{m\to\infty} \sigma_I^2 = 0 \text{ and } \lim_{m\to\infty} \sigma_T^2 = \sigma_p^2(\mathbf{x}_c).$$

- *As* $m \to \infty$ *and* $B \to \infty$ *in an iterated limit, the variance component estimators are consistent:*

$$\lim_{m\to\infty}\lim_{B\to\infty} \widehat{\sigma}_M^2 = \lim_{m\to\infty} \sigma_M^2 = \sigma_p^2(\mathbf{x}_c),$$
$$\lim_{m\to\infty}\lim_{B\to\infty} \widehat{\sigma}_I^2 = \lim_{m\to\infty} \sigma_I^2 = 0,$$
$$\lim_{m\to\infty}\lim_{B\to\infty} \widehat{\sigma}_T^2 = \lim_{m\to\infty} \sigma_T^2 = \sigma_p^2(\mathbf{x}_c),$$
$$\lim_{m\to\infty}\lim_{B\to\infty} \widehat{s}_\ell = \lim_{m\to\infty} s_\ell = 0 \text{ for } \ell = 1, 2, \ldots, L.$$

Theorem 4.1 demonstrates that the variance components estimators $\widehat{\sigma}_I^2$, $\widehat{\sigma}_M^2$, $\widehat{\sigma}_T^2$, and $\widehat{s}_\ell$ for $\ell = 1, 2, \ldots, L$ are consistent. However, we can see that the model uncertainty disappears as $m \to \infty$. In addition, we study the consistency of scaled versions of σ_I^2 and $\widehat{\sigma}_I^2$ in Theorem 4.2, showing that $m\sigma_I^2$ and $m\widehat{\sigma}_I^2$ converge to the same non-zero constant.

Theorem 4.2. *Suppose Assumptions 1–6 in online Appendix A hold. Then we have* $\lim_{m\to\infty} m\sigma_I^2 = \lim_{m\to\infty}\lim_{B\to\infty} m\widehat{\sigma}_I^2 = \sigma_\mu^2$ *almost surely, where* σ_μ^2 *is a positive constant.*

Theorem 4.3. *Suppose Assumptions 1–6 in online Appendix A hold. Then we have* $\lim_{m\to\infty} ms_\ell = \lim_{m\to\infty}\lim_{B\to\infty} m\widehat{s}_\ell = \sigma_s^2$ *almost surely, where* σ_s^2 *is a positive constant.*

Theorems 4.1 –4.3 give the asymptotic properties of the variance component estimators, guaranteeing: (1) $\widehat{\sigma}_I/\widehat{\sigma}_T$ is a consistent estimator for the relative contribution of model uncertainty to the overall estimation uncertainty; and (2) $\widehat{s}_\ell$ is a consistent estimator of the contribution from the ℓth model uncertainty. The detailed proof is provided in online Appendix C. We will empirically evaluate its finite-sample performance in Section 5 where we form the variance component estimators by inserting $(\widehat{\tau}^2, \widehat{\boldsymbol{\theta}}, \widehat{C})$ for the unknown parameters $(\tau^2, \boldsymbol{\theta}, C)$.

5. Empirical study

We study the finite sample performance of the proposed metamodel-assisted uncertainty analysis framework and compare it with the direct bootstrap approach. We consider a biopharmaceutical manufacturing example in Sections 5.1. A cell culture process hybrid model for cell therapy manufacturing is studied in Section 5.2. Additionally, a queueing network example is provided in online Appendix E. The proposed framework demonstrates good and robust performance under different experiment settings in terms of (1) the amount of real-world data m which controls the level of model uncertainty; (2) the simulation budget N which controls the simulation uncertainty; and (3) the number of design points k for GP metamodel construction, with N, is used to control the metamodel uncertainty.

The empirical results show that the proposed framework can provide better performance than the direct bootstrap approach. The new ACI CI_+ is robust to different levels of real-world data m, number of design points k, and simulation budget N in terms of replications. When simulation uncertainty is significant, CI_0 tends to have undercoverage that becomes more serious as m increases. Since CI_+ accounts for both simulation and model uncertainties, it does not exhibit this degradation. The ratio $\widehat{\sigma}_I/\widehat{\sigma}_T$ is a useful measure of the relative contribution of model uncertainty to overall statistical uncertainty and the SV-based sensitivity analysis further quantifies the contribution from each source of model uncertainty.

Table 1
The underlying true process model parameters.

	Protein Concentration	Impurity Concentration
Initial Biomass	$X_0 \sim \mathcal{N}(15.98, 4.17^2)$	N.A.
Growth Rate	$\gamma \sim \mathcal{N}(0.0475, 0.008^2)$	
Residual	$\epsilon_P \sim \mathcal{N}(0, 0.4918^2)$	$\epsilon_I \sim \mathcal{N}(0, 0.4918^2)$
Centrifuge	N.A.	$Q \sim \text{Unif}(0.4, 0.5)$
Chromatography	$Q_p \sim \text{Unif}(0.4833, 0.5907)$	$Q_I \sim \text{Unif}(0.1458, 0.1782)$
Filtration	N.A.	$Q_{fr} \sim \text{Unif}(0.99, 1)$

5.1. A biopharmaceutical manufacturing example

We consider the biomanufacturing example illustrated in Fig. 1; see the details in Wang et al. (2019). We are interested in estimating the expected productivity of an antigen protein drug, i.e., $\mu(\mathbf{x}_c)$. The protein and impurity accumulations in the exponential-growth phase of fermentation process are modeled with the hybrid models, i.e., $X_t = X_0 \cdot e^{\gamma t} + \epsilon_P$ and $I_t = I_0 \cdot e^{\gamma t} + \epsilon_I$ with $0 \le t \le T$, where γ is the growth rate, X_0 and I_0 are the starting amounts of biomass and impurity. We consider the fixed harvest time $T = 54$ and the fixed initial impurity amount $I_0 = 14.64$.

The downstream purification process includes centrifuge, chromatography, filtration, and quality control. Random proportions of protein and impurity are removed at each operation unit, except at the quality control step. **(1) Centrifuge Step.** The protein and impurity levels before and after centrifuge are denoted by (X_F, I_F) and (X_C, I_C). We assume that this step does not change the protein level, i.e., $X_C \equiv X_F$ (Delahaye, Lawrence, Ward, & Hoare, 2015), and it removes a random proportion of impurity, i.e., $I_C = Q \cdot I_F$. **(2) Chromatography Step.** For chromatography, random removal proportions of protein and impurity, denoted by Q_P and Q_I, follow uniform distributions (Martagan, Krishnamurthy, Leland, & Maravelias, 2017). The target protein and impurity levels before and after chromatography are denoted by (X_C, I_C) and (X_P, I_P), and we have $X_P = Q_P \cdot X_C$ and $I_P = Q_I \cdot I_C$. **(3) Filtration Step.** Filtration works as a polishing procedure and it slightly reduces the impurity. Denote the protein and impurity levels before and after filtration with (X_P, I_P) and (X_{fr}, I_{fr}). Thus, $I_{fr} = Q_{fr} \cdot I_P$. and $X_{fr} = X_P$. **(4) Quality Control Step.** During the quality control step, if the impurity percentage $\frac{I_{fr}}{X_{fr}+I_{fr}}$ is greater than the requirement, say $\omega = 25\%$, the corresponding batch is discarded. Therefore, the expected productivity of each batch is defined as:

$$\mu(\mathbf{x}_c) = \text{E}\left[X_{fr} \cdot 1\left(\frac{I_{fr}}{X_{fr}+I_{fr}} \le \omega\right)\right].$$

Thus, this biopharmaceutical manufacturing example has $L = 8$ process models: (1) F_1 modeling the residual or measurement error ϵ_P; (2) F_2 modeling the batch-to-batch variation of the growth rate γ; (3) F_3 modeling the variation of the initial biomass X_0; (4) F_4 modeling the residual ϵ_I of impurity and metabolic waste accumulation; (5) F_5 modeling the random impurity removal ratio Q at centrifuge step; (6) F_6 and F_7 modeling the random removal ratios, Q_p and Q_I, of protein and impurity at chromatography step; and (7) F_8 modeling the random impurity removal ratio Q_{fr} at filtration step. All the underlying true model parameters are summarized in Table 1. In the empirical study, we assume that these parameters are unknown and they are estimated with finite observations with

Table 2
The CIs results (SD) of the expected productivity and $\widehat{\sigma}_I^2/\widehat{\sigma}_T^2$ when $N = 2000$.

$m = 10$	Metamodel-Assisted Uncertainty Analysis			Direct Bootstrap
	$k = 20$, n=100	$k = 40$, n=50	$k = 80$, n=25	
Coverage of CI_0	84.80%	88.20%	89.40%	99.60%
Coverage of CI_+	88.60%	90.40%	92.00%	
CI_0 Width	89.60 (32.99)	99.19 (37.08)	98.54 (33.62)	224.21 (81.25)
CI_+ Width	103.21 (35.81)	109.60 (38.66)	102.81 (35.09)	
$\widehat{\sigma}_I^2/\widehat{\sigma}_T^2$	80.26%	88.99%	89.51%	61.10%
$m = 20$	Metamodel-Assisted Uncertainty Analysis			Direct Bootstrap
	$k = 20$, n=100	$k = 40$, n=50	$k = 80$, n=25	
Coverage of CI_0	85.80%	89.60%	89.60%	100.00%
Coverage of CI_+	92.60%	93.00%	92.60%	
CI_0 Width	64.09 (19.68)	66.98 (17.18)	70.96 (17.84)	205.85 (48.12)
CI_+ Width	75.46 (20.93)	74.69 (17.29)	79.01 (19.37)	
$\widehat{\sigma}_I^2/\widehat{\sigma}_T^2$	76.55%	84.15%	83.62%	63.69%
$m = 40$	Metamodel-Assisted Uncertainty Analysis			Direct Bootstrap
	$k = 20$, n=100	$k = 40$, n=50	$k = 80$, n=25	
Coverage of CI_0	84.20%	91.00%	86.40%	100.00%
Coverage of CI_+	93.80%	95.40%	92.40%	
CI_0 Width	43.62 (11.38)	47.58 (10.26)	48.09 (10.42)	196.75 (32.74)
CI_+ Width	55.23 (12.59)	56.79 (10.86)	57.74 (11.73)	
$\widehat{\sigma}_I^2/\widehat{\sigma}_T^2$	68.46%	74.73%	73.96%	65.03%

size m. Since we often have very limited biopharmaceutical manufacturing process data available in the real world, we focus on the cases with $m = 10, 20, 40$ and let $m_\ell = m$ for $\ell = 1, 2, \ldots, L$.

We assess the performances of CI_+ and CI_0 especially under the situation when the system has large simulation uncertainty. Therefore, the run length for each replication is set as 2 after the warm up equal to 25 in terms of the number of batches. For the proposed metamodel-assisted uncertainty analysis framework, when we build the GP metamodel, we set the number of design points $k = 20, 40, 80$. The same number of replications is assigned to each design point, i.e., $n_j = n = N/k$ for $j = 1, 2, \ldots, k$. To precisely estimate the percentile interval quantifying the system mean performance estimation uncertainty, we set the number of bootstrap resampled moments $B = 1000$ (Barton et al., 2014). We compare the performance of our proposed framework with *direct bootstrap* under the same computational budget. In the direct bootstrap approach, we run simulations at each bootstrapped moments to estimate the system mean response and equally allocate the simulation budget. It means that the number of replications at each bootstrapped moment sample is $n^d = N/B$. To assess the coverage of CIs, we conduct a side experiment with 10^6 run length and 40 replications to estimate the true mean response and obtain $\mu(\mathbf{x}_c) = 116.759 \pm 0.006$.

5.1.1. Biomanufacturing system uncertainty quantification

Tables 2 and 3 show the mean and standard deviation (SD) results of width and coverage of 95% CIs, quantifying the overall estimation uncertainty of the expected productivity, obtained by the proposed metamodel-assisted uncertainty analysis framework and the direct bootstrap approach, when the simulation computational budget is $N = 2000, 4000$. We also record the ratio of model uncertainty to total variance $\widehat{\sigma}_I^2/\widehat{\sigma}_T^2$. All results are based on 500 macro-replications. As m increases, the contribution of model uncertainty, measured by $\widehat{\sigma}_I^2/\widehat{\sigma}_T^2$, decreases. The coverage of CI_+ is constantly better and closer to the nominal value of 95% compared with CI_0. The direct bootstrap approach has substantial over coverage issue, which was described and explained in Barton et al. (2007). Since each experiment can be expensive and the average value of each batch of bio-drugs excesses one million, this over coverage issue can lead to overly conservative decision making and dramatically impact the profit. Given the fixed computational budget, as the number of real-world data m increases, the mean and SD of the interval widths decrease, and the coverage becomes closer to the nominal value. *Overall, the proposed metamodel-assisted uncertainty analysis will provide better performance, especially under the situation with very limited amount of real-world data and high model uncertainty,* which often happens in the biopharmaceutical manufacturing industry.

5.1.2. Biomanufacturing system variance decomposition

When the model uncertainty plays a dominate impact on the system performance estimation uncertainty, it is critical to identify the key source, which can be used to efficiently improve the simulation model. Based on the analytical study in Section 4.2, the means with 95% CI of the relative contribution from each ℓth model uncertainty, i.e., $(\widehat{s}_\ell/\widehat{\sigma}_I^2 \times 100\%)$, are recorded in Table 4. The results are estimated based on 100 macro-replications. We set the number of bootstrapped moments used for the variance estimation $B' = 2000$. Since the model uncertainty of protein generation process characterized by models for $\{\epsilon_P, \gamma, X_0\}$ dominates, we gradually increase m_ℓ with $\ell = 1, 2, 3$ as $m' = 10, 20, 40$, while fixing the number of real-world data for remaining models $m_\ell = 10$ for $\ell = 4, 5, \ldots, L$. The order of importance, $\epsilon_P > \gamma > X_0$ is consistent across all sample sizes. Of the remaining variables, the removal proportion of protein at chromatography, Q_P, provides the largest proportion of contribution across all sample sizes and it increases dramatically as the sample size increases. As the sample size m' increases, the relative contribution from model uncertainty of $\{X_0, \gamma, \epsilon_P\}$ reduces. The overall model uncertainty, measured by σ_I, also decreases with increasing sample size.

This case study is motivated by a real animal bio-drug production. The quality requirement, i.e., $\frac{I_{fr}}{X_{fr}+I_{fr}} \leq \omega$ with $\omega = 25\%$, is relatively easy to meet through downstream purification. Thus, the results in Table 4 indicate that the influence of the impurity pathway parameters is negligible. This observation does not hold in general, especially for antigen proteins for human beings that typically have much more restrictive quality requirements (say $\omega = 1\%$).

Table 3
The CIs results (SD) of the expected productivity and $\widehat{\sigma}_I^2/\widehat{\sigma}_T^2$ when $N = 4000$.

$m = 10$	Metamodel-Assisted Uncertainty Analysis			Direct Bootstrap
	$k = 20$, n=200	$k = 40$, n=100	$k = 80$, n=50	
Coverage of CI_0	86.80%	89.40%	91.20%	99.40%
Coverage of CI_+	91.20%	90.60%	92.80%	
CI_0 Width	91.96 (33.84)	103.24 (37.66)	99.45 (35.73)	178.19 (64.18)
CI_+ Width	102.84 (34.90)	108.23 (38.96)	103.46 (36.72)	
$\widehat{\sigma}_I^2/\widehat{\sigma}_T^2$	83.52%	92.78%	93.38%	72.20%
$m = 20$	Metamodel-Assisted Uncertainty Analysis			Direct Bootstrap
	$k = 20$, n=200	$k = 40$, n=100	$k = 80$, n=50	
Coverage of CI_0	88.80%	91.20%	91.60%	100.00%
Coverage of CI_+	92.40%	93.00%	93.40%	
CI_0 Width	65.58 (21.10)	69.49 (16.92)	73.37 (17.45)	156.94 (39.69)
CI_+ Width	75.87 (21.68)	74.64 (17.59)	78.50 (18.43)	
$\widehat{\sigma}_I^2/\widehat{\sigma}_T^2$	79.08%	88.86%	89.13%	75.56%
$m = 40$	Metamodel-Assisted Uncertainty Analysis			Direct Bootstrap
	$k = 20$, n=200	$k = 40$, n=100	$k = 80$, n=50	
Coverage of CI_0	87.00%	93.60%	91.60%	100.00%
Coverage of CI_+	94.80%	95.00%	94.80%	
CI_0 Width	45.69 (11.59)	49.66 (9.71)	51.19 (10.27)	145.75 (30.15)
CI_+ Width	54.52 (12.42)	55.87 (10.04)	57.18 (10.79)	
$\widehat{\sigma}_I^2/\widehat{\sigma}_T^2$	74.93%	82.37%	82.99%	77.71%

Table 4
The relative contributions from each model uncertainty when $m' = 10, 20, 40$.

Process Model	m'=10	m'=20	m'=40
ϵ_P	44.51% ± 5.05%	40.32% ± 4.74%	37.73% ± 4.53%
γ	35.18% ± 4.93%	31.89% ± 4.18%	28.44% ± 3.69%
X_0	15.04% ± 4.55%	14.34% ± 4.28%	11.88% ± 3.89%
Q_P	3.87% ± 0.88%	10.44% ± 2.16%	18.06% ± 3.32%
ϵ_I	0.83% ± 1.11%	1.90% ± 1.81%	2.28% ± 2.02%
Q	0.18% ± 0.35%	0.33% ± 0.53%	0.70% ± 0.89%
Q_I	0.29% ± 0.23%	0.35% ± 0.37%	0.45% ± 0.54%
Q_{fr}	0.04% ± 0.12%	0.23% ± 0.19%	0.64% ± 0.94%
$\widehat{\sigma}_I$	25.43 ± 2.01	18.10 ± 1.21	13.51 ± 0.65

5.2. Cell culture expansion scheduling for cell therapy manufacturing

Here we use the erythroblast cell therapy manufacturing example presented in Glen, Cheeseman, Stacey, & Thomas (2018) to assess the performance of proposed framework. The cell culture of erythroblast exhibits two phases: a relatively uninhibited growth phase followed by an inhibited phase. The hybrid model cell growth and inhibitor accumulation is

$$\rho_{t+1} = \rho_t + \Delta t \cdot r^g \rho_t \left(1 - \left(1 + e^{(k^s(k^c - I_t))}\right)^{-1}\right) + e_t^\rho,$$

$$I_{t+1} = I_t + \Delta t \cdot \left(\frac{\rho_{t+1} - \rho_t}{\Delta t} - r^d I_t\right) + e_t^I,$$

where Δt represents the time interval, ρ_t and I_t represent the cell density and the unobservable inhibitor concentration at the tth time step. The kinetic coefficients r^g, k^s, k^c and r^d denote the cell growth rate, inhibitor sensitivity, inhibitor threshold, and inhibitor decay. The residuals follow the normal distributions, i.e., $e_t^\rho \sim N(0, (v^\rho)^2)$ and $e_t^I \sim N(0, (v^I)^2)$. There is raw material uncertainty for seed cell density, i.e., $\rho_0 \sim N(\mu_\rho, \sigma_\rho^2)$. The initial inhibitor concentration equals to 0 due to the fresh medium, i.e., $I_0 = 0$. Additionally, the investigation from Glen et al. (2018) shows that the growth rate has significant variability cross different donors. Therefore, we incorporate batch-to-batch variation by considering the random effect on the growth rate, i.e., $r^g \sim \mathcal{N}(\mu^g, (\sigma^g)^2)$.

Thus, this erythroblast cell therapy manufacturing example has $L = 7$ process models: (1) F_1 for ρ_0; (2) F_2 for e^ρ; (3) F_3 for e^I; (4) F_4 for r^g; and (5–7) the degenerate distributions F_5, F_6, F_7 for bioprocess kinetic parameters k^s, k^c, r^d. Set the underlying true parameters as $\{\mu_\rho, \sigma_\rho, v^\rho, v^I, \mu^g, \sigma^g\} = \{3, 0.03, 0.01, 0.01, 0.037, 0.008\}$ and $\{k^s, k^c, r^d\} = \{3.4, 2.6, 0.005\}$, which are validated by using the real-world data presented in Glen et al. (2018). In this empirical study, we assume that all these parameters are unknown and estimated with a finite amount of real-world data with size m. The cell density data are collected every 4 hours, i.e., $\Delta t = 4$ hours. Thus, we have m trajectory observations, i.e., $\boldsymbol{\tau}^{(i)} \equiv (\rho_0^{(i)}, \rho_1^{(i)}, \ldots, \rho_T^{(i)})$ with $i = 1, 2, \ldots, m$.

At any time t, if the batch-extension is performed, the original batch is scaled up to a λ times larger cell culture vessel filling with fresh medium. That means the cell density ρ and the concentration of inhibitor I decrease to $1/\lambda$ of original values. In this example, suppose that the batch-extension is scheduled at the 24th hour (corresponding to time step $t = \frac{24}{\Delta t} + 1 = 7$). Then, the original batch is scaled up to $\lambda = 4$ fold. The cell culture process ends at $T = 40$ hours (corresponding to time step $t = \frac{T}{\Delta t} + 1 = 11$). Our goal is to estimate the expected productivity in terms of total biomass of target cells, i.e., $\mu(\mathbf{x}_c) = \mathrm{E}[\rho_T \cdot \lambda]$.

We focus on the cases with $m = 3, 6, 20$ and let $m_\ell = m$ for $\ell = 1, 2, \ldots, L$. The total simulation budget is set to be $N = 4000$ replications. We compare the performance of our proposed framework with direct bootstrap approach under the same computational budget. For the proposed metamodel-assisted uncertainty analysis framework, we set the number of design points $k = 20, 40, 80$. The same number of replications is assigned to each design point, i.e., $n_j = n = N/k$ for $j = 1, 2, \ldots, k$. The number of bootstrap resampled moments is set as $B = 1000$. In the direct bootstrap approach, the number of replications allocated at each bootstrapped moment sample is $n^d = N/B = 4$. To assess the coverage of CIs, we conduct a side experiment with 10^6 batches and 20 replications to estimate the true mean response and obtain $\mu(\mathbf{x}_c) = 17.32 \pm 0.004$.

Table 5 records the mean and standard deviation (SD) results of width and coverage of 95% CIs, quantifying the overall estimation uncertainty of the expected productivity, obtained by the proposed metamodel-assisted uncertainty analysis framework and the direct bootstrap approach. We also record the ratio of

Table 5
The CIs results (SD) of the expected productivity and $\widehat{\sigma}_I^2/\widehat{\sigma}_T^2$ when $N = 4000$.

$m = 3$	Metamodel-Assisted Uncertainty Analysis			Direct Bootstrap
	$k = 20$, n=200	$k = 40$, n=100	$k = 80$, n=50	
Coverage of CI_0	83.20%	86.20%	84.40%	99.80%
Coverage of CI_+	90.20%	91.20%	90.80%	
CI_0 Width	4.67 (2.11)	4.23 (2.42)	4.13 (2.45)	7.12 (3.75)
CI_+ Width	5.03 (2.75)	5.36 (2.76)	5.21 (2.52)	
$\widehat{\sigma}_I^2/\widehat{\sigma}_T^2$	86.17%	90.23%	87.32%	87.21%
$m = 6$	Metamodel-Assisted Uncertainty Analysis			Direct Bootstrap
	$k = 20$, n=100	$k = 40$, n=50	$k = 80$, n=25	
Coverage of CI_0	89.60%	89.00%	89.20%	100.00%
Coverage of CI_+	92.80%	93.40%	91.60%	
CI_0 Width Mean	2.99 (1.86)	3.14 (1.76)	3.25 (1.78)	5.35 (2.52)
CI_+ Width Mean	3.34 (1.91)	3.42 (1.82)	3.43 (1.84)	
$\widehat{\sigma}_I^2/\widehat{\sigma}_T^2$	86.34%	90.41%	91.02%	74.83%
$m = 20$	Metamodel-Assisted Uncertainty Analysis			Direct Bootstrap
	$k = 20$, n=100	$k = 40$, n=50	$k = 80$, n=25	
Coverage of CI_0	93.40%	94.00%	93.60%	97.80%
Coverage of CI_+	95.40%	95.00%	95.20%	
CI_0 Width Mean	1.68 (1.05)	1.72 (1.09)	1.74 (1.12)	3.84 (1.72)
CI_+ Width Mean	1.79 (1.10)	1.83 (1.12)	1.86 (1.15)	
$\widehat{\sigma}_I^2/\widehat{\sigma}_T^2$	81.20%	85.16%	84.65%	80.14%

model uncertainty to total variance $\widehat{\sigma}_I^2/\widehat{\sigma}_T^2$. All results are based on 500 macro-replications. The coverage of CI_+ is much closer to the nominal value of 95%, when compare with CI_0. The direct bootstrap again exhibits overcoverage and provides much wider confidence interval width means and standard deviations. Given the fixed computational budget, as the number of real-world data m increases, the mean and SD of the interval widths decrease, and the coverage becomes closer to the nominal value.

6. Conclusions

To efficiently develop a simulation model to improve the assessment of the mean response for flexible and integrated biomanufacturing systems with modular design, we propose a metamodel-assisted bootstrapping uncertainty quantification and sensitivity analysis framework. Process model uncertainty is approximated by the bootstrap and an equation-based stochastic kriging metamodel is used to propagate the model uncertainty to the output mean. The simulation uncertainty is derived using properties of stochastic kriging. This framework delivers an interval quantifying the system mean response estimation accuracy accounting for both simulation and model uncertainties. The asymptotic consistency of this interval is proved under the assumption that the true response surface is a realization of a Gaussian process and certain parameters are known. Given very limited real-world observations and high stochastic uncertainty, the model uncertainty often dominates, especially for personalized bio-drug manufacturing. We provide a variance decomposition quantifying the relative contribution from each source of model uncertainty, as well as simulation uncertainty. While the asymptotic analysis shows correctness for the proposed framework, the empirical study on multiple biomanufacturing and service examples demonstrates that it also has good finite-sample performance.

Acknowledgments

This paper is based upon work supported by the National Science Foundation under Grant No. CMMI-0900354 and CMMI-1068473, National Institute of Standards and Technology (70NANB17H002), Department of Commerce. We also would like to thank the anonymous reviewers for their comments that have helped us improve the manuscript.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ejor.2023.01.055.

References

Ankenman, B. E., Nelson, B. L., & Staum, J. (2010). Stochastic kriging for simulation metamodeling. *Operations Research, 58*, 371–382.
Barton, R. R., Nelson, B. L., & Xie, W. (2014). Quantifying input uncertainty via simulation confidence interval. *Informs Journal on Computing, 26*, 74–87.
Barton, R. R., & Schruben, L. W. (2001). Resampling methods for input modeling. *Proceeding of the 2001 winter simulation conference (Cat. No. 01CH37304)* (pp. 372–378). IEEE.
Barton, R. R., et al. (2007). Presenting a more complete characterization of uncertainty: Can it be done. In *Proceedings of the 2007 informs simulation society research workshop* (pp. 26–60). INFORMS Simulation Society.
Biller, B., & Corlu, C. G. (2011). Accounting for parameter uncertainty in large-scale stochastic simulations with correlated inputs. *Operations Research, 59*(3), 661–673.
Borgonovo, E., & Plischke, E. (2016). Sensitivity analysis: A review of recent advances. *European Journal of Operational Research, 248*(3), 869–887.
Corlu, C. G., Akcay, A., & Xie, W. (2020). Stochastic simulation under input uncertainty: A review. *Operations Research Perspectives*, 100162.
Delahaye, M., Lawrence, K., Ward, S., & Hoare, M. (2015). An ultra scale-down analysis of the recovery by dead-end centrifugation of human cells for therapy. *Biotechnology and Bioengineering, 112*(5), 997–1011.
Doran, P. M. (2012). *Bioprocess engineering principles*. Academic Press.
Glen, K. E., Cheeseman, E. A., Stacey, A. J., & Thomas, R. J. (2018). A mechanistic model of erythroblast growth inhibition providing a framework for optimisation of cell therapy manufacturing. *Biochemical Engineering Journal, 133*, 28–38.
Hernández Rodríguez, T., Posch, C., Schmutzhard, J., Stettner, J., Weihs, C., Pörtner, R., & Frahm, B. (2019). Predicting industrial-scale cell culture seed trains–a Bayesian framework for model fitting and parameter estimation, dealing with uncertainty in measurements and model parameters, applied to a nonlinear kinetic cell culture model, using an MCMC method. *Biotechnology and Bioengineering, 116*(11), 2944–2959.
Kleijnen, J. P. (2017). Regression and kriging metamodels with their experimental designs in simulation: A review. *European Journal of Operational Research, 256*(1), 1–16.
Lam, H., & Qian, H. (2018). Subsampling variance for input uncertainty quantification. In *2018 Winter simulation conference (WSC)* (pp. 1611–1622). IEEE.
Lam, H., & Qian, H. (2022). Subsampling to enhance efficiency in input uncertainty quantification. *Operations Research, 70*(3), 1891–1913.
Martagan, T., Krishnamurthy, A., Leland, P. A., & Maravelias, C. T. (2017). Performance guarantees and optimal purification decisions for engineered proteins. *Operations Research, 66*(1), 18–41.

Mockus, L., Peterson, J. J., Lainez, J. M., & Reklaitis, G. V. (2015). Batch-to-batch variation: A key component for modeling chemical manufacturing processes. *Organic Process Research & Development, 19*(8), 908–914.
Möller, J., Rodríguez, T. H., Müller, J., Arndt, L., Kuchemüller, K. B., Frahm, B., … Pörtner, R. (2020). Model uncertainty-based evaluation of process strategies during scale-up of biopharmaceutical processes. *Computers & Chemical Engineering, 134*, 106693.
O'Brien, C. M., Zhang, Q., Daoutidis, P., & Hu, W.-S. (2021). A hybrid mechanistic-empirical model for in silico mammalian cell bioprocess simulation. *Metabolic Engineering, 66*, 31–40.
Picheny, V., Ginsbourger, D., Roustant, O., Haftka, R. T., & Kim, N. (2010). Adaptive designs of experiments for accurate approximation of a target region. *Journal of Mechanical Design, 132*, 071008.
Rodríguez, T. H., & Frahm, B. (2021). Digital seed train twins and statistical methods. *Advances in Biochemical Engineering and Biotechnology, 176*, 97–131.
Santner, T. J., Williams, B. J., & Notz, W. I. (2003). *The design and analysis of computer experiments*. New York: Springer.
Song, E., & Nelson, B. L. (2013). A quicker assessment of input uncertainty. In *2013 Winter simulations conference (WSC)* (pp. 474–485). IEEE.
Song, E., Nelson, B. L., & Staum, J. (2016). Shapley effects for global sensitivity analysis: Theory and computation. *SIAM/ASA Journal on Uncertainty Quantification, 4*(1), 1060–1083.
Wang, B., Xie, W., Martagan, T., Akcay, A., & Corlu, C. G. (2019). Stochastic simulation model development for biopharmaceutical production process risk analysis and stability control. In *2019 Winter simulation conference (WSC)* (pp. 1989–2000). IEEE.
Wang, P., Mihaylova, L., Chakraborty, R., Munir, S., Mayfield, M., Alam, K., … Fang, H. (2021). A Gaussian process method with uncertainty quantification for air quality monitoring. *Atmosphere, 12*(1344), 18.
Xie, W., Li, C., & Zhang, P. (2017). A factor-based Bayesian framework for risk analysis in stochastic simulations. *ACM Transactions on Modeling and Computer Simulation (TOMACS), 27*(4), 1–31.
Xie, W., Nelson, B. L., & Staum, J. (2010). The influence of correlation functions on stochastic kriging metamodels. In *2010 Winter simulation conference (WSC)* (pp. 1067–1078). IEEE.
Xie, W., Wang, B., Li, C., Xie, D., & Auclair, J. (2022). Interpretable biomanufacturing process risk and sensitivity analyses for quality-by-design and stability control. *Naval Research Logistics (NRL), 69*(3), 461–483.
Xie, W., Wang, B., & Zhang, Q. (2018). Metamodel-assisted risk analysis for stochastic simulation with input uncertainty. In *2018 Winter simulation conference (WSC)* (pp. 1766–1777). IEEE.
Zhang, Q., Wang, B., & Xie, W. (2022). A pooled percentile estimator for parallel simulations. *Journal of Simulation, 16*(1), 73–83.
Zouaoui, F., & Wilson, J. R. (2003). Accounting for parameter uncertainty in simulation input modeling. *IIE Transactions, 35*, 781–792.
Zouaoui, F., & Wilson, J. R. (2004). Accounting for input-model and input-parameter uncertainties in simulation. *IIE Transactions, 36*, 1135–1151.