Noname manuscript No.
(will be inserted by the editor)

Bayesian Optimization of Functional Output in Inverse Problems

Chaofan Huang · Yi Ren · Emily K. McGuinness · Mark D. Losego · Ryan P. Lively · V. Roshan Joseph

Received: date / Accepted: date

Abstract Motivated by the parameter identification problem of a reaction-diffusion transport model in a vapor phase infiltration processes, we propose a Bayesian optimization procedure for solving the inverse problem that aims to find an input setting that achieves a desired functional output. The proposed algorithm improves over the standard single-objective Bayesian optimization by (i) utilizing the generalized chi-square distribution as a more appropriate predictive distribution for the squared distance objective function in the inverse problems, and (ii) applying functional principal component analysis to reduce the dimensionality of the functional response data, which allows for efficient approximation of the predictive distribution and the subsequent computation of the expected improvement acquisition function.

Keywords Expected Improvement · Gaussian Process · Generalized Chi-Square Distribution · Model Calibration · Functional Principal Component Analysis

Mathematics Subject Classification (2020) 62L05

Chaofan Huang · V. Roshan Joseph*
H. Milton Stewart School of Industrial and Systems Engineering,
Georgia Institute of Technology, Atlanta, GA, 30332
E-mail: chaofan.huang@gatech.edu, roshan@gatech.edu
*To whom the correspondence should be addressed.

Yi Ren · Ryan P. Lively
School of Chemical and Biomolecular Engineering,
Georgia Institute of Technology, Atlanta, GA, 30332
E-mail: yren48@gatech.edu, ryan.lively@chbe.gatech.edu

Emily K. McGuinness · Mark D. Losego
School of Materials Science and Engineering,
Georgia Institute of Technology, Atlanta, GA, 30332
E-mail: emcguinness6@gatech.edu, losego@gatech.edu

1 Introduction

Simulator-based models, such as differential equations, are important tools for modeling the complex physical systems in many scientific and engineering applications. However, the model parameters of interest are generally not known a priori, and thus require estimation from noisy observational data. This is known as the inverse problem [Kirsch, 2011], and it has great implication in chemical engineering [Ren et al., 2021], ecology [Wood, 2010], genetics [Marttinen et al., 2015], and many other fields. The inverse problem is also known as *model calibration* in the statistical literature [Box and Hunter, 1962, Kennedy and O'Hagan, 2001, Joseph and Yan, 2015].

More formally, let $y^* \in \mathcal{Y}$ denote the noisy structured observational data that could be a scalar, a function, or a tensor. Let $H : \Theta \to \mathcal{Y}$ denotes the simulator-based model with unknown parameters $\theta \in \Theta$. Suppose that the model H is correctly specified, the objective of the inverse problem is to identify $\theta^* \in \Theta$ such that

$$y^* = H(\theta^*) + \epsilon, \tag{1}$$

where $\epsilon \in \mathcal{Y}$ represents the unknown observational noise. However, the inverse of H is generally intractable in practice. In order to find θ^*, we formulate the inverse problem as a least squares problem. In other words, we aim to find the model parameters $\theta^* \in \Theta$ such that the discrepancy between $H(\theta^*)$ and y^* is minimized,

$$\theta^* = \arg\min_{\theta \in \Theta} f(\theta) = g(H(\theta), y^*) = \|H(\theta) - y^*\|_2^2, \tag{2}$$

where $g : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$ is the squared distance function, a popular discrepancy measure for the structured data, and the objective $f := g \circ h$ is a composite function with non-negative range. Unfortunately, the optimization problem in (2) is difficult to solve since (i) the model H is so complicated or even black-box such that the analytical gradient/Hessian of f cannot be derived, (ii) the model H could be computationally expensive to evaluate in many real world applications, so we cannot afford the computational cost of getting precise numerical approximation for the gradient/Hessian, and (iii) multiple local minima could exist due to the noise-contaminated observations and possible information loss in the forward process of evaluating the model H. Thus, the objective of this paper is to investigate an efficient optimization procedure to solve (2) using as few evaluations of the model H as possible.

Bayesian optimization (BO) [Kushner, 1964, Mockus, 1994, Jones, 2001, Frazier, 2018] is the state-of-the-art method for solving optimization problem involving an expensive objective function that has multiple local optima, making it a perfect tool for solving the inverse problem in (2). BO builds a probabilistic surrogate model to infer the predictive distribution of any input, and then utilizes the predictive information to select the next sample that maximizes the expected improvement [Jones et al., 1998]. The naive approach is to apply BO with respect to the composite objective function f, but this

ignores the useful information from the model H. To address this shortcoming, Uhrenholt and Jensen [2019] and Matsui et al. [2019] proposed surrogate modeling on the model H and derived the generalized chi-square predictive distribution for f, when g is the squared distance function as in (2). Moreover, Astudillo and Frazier [2019] further generalized this composite objective function BO framework to any well behaved scalar-output function g with analytical first-order derivative.

Motivated by the parameter identification problem of a reaction-diffusion transport model in the vapor phase infiltration (VPI) process where the observational data $y^*(t)$ is a function over time [Ren et al., 2021], we focus on the inverse problem with an expensive model $H(t;\theta)$ that returns a functional response. Prior literature have considered the inverse problem for vector-output model H with both independent components [Uhrenholt and Jensen, 2019] and correlated components [Matsui et al., 2019]. The latter is more related to our interest since functional data are usually represented by some high-dimensional vector with correlated entries for the temporal dependency. However, the approach in Matsui et al. [2019] is computationally expensive for the high dimensional data since (i) vector-valued Gaussian process [Alvarez et al., 2012] is utilized for modeling H, and (ii) spectral decomposition of the predictive covariance is required for deriving the generalized chi-square distribution. Thus, we propose using the functional principal component analysis (FPCA) for dimension reduction of the functional data [Ramsay, 2004], resulting in efficient derivation of the generalized chi-square predictive distribution for f using only the principal component scores. The FPCA has also been used to define probability density for random functions [Delaigle and Hall, 2010] and construct a Gaussian process model with a functional output [Tan, 2018].

This paper is organized as follows. Section 2 presents the physical process, vapor phase infiltration, and the differential equations that motivate our interest on the inverse problem with functional output. Section 3 reviews the standard Bayesian optimization procedure. Section 4 first discusses the modification of Bayesian optimization for a principal approach of solving the inverse problem using the generalized chi-square distribution, and then proposes the BOFO, Bayesian Optimization of Functional Output, to extend the framework to a functional target. Section 5 presents the numerical result of BOFO on the motivating problem and demonstrates how BOFO improves over the standard Bayesian optimization procedure. The article then concludes with some remarks in Section 6.

2 Application: VPI Process

Vapor phase infiltration (VPI) is an emerging chemical modification process for transforming polymers into hybrid organic-inorganic materials with industrially relevant properties. In VPI, the bulk of a polymer is infused with a vapor-phase precursor (often a metal-organic) which becomes entrapped either via a chemical interaction with functional groups on the polymer or through

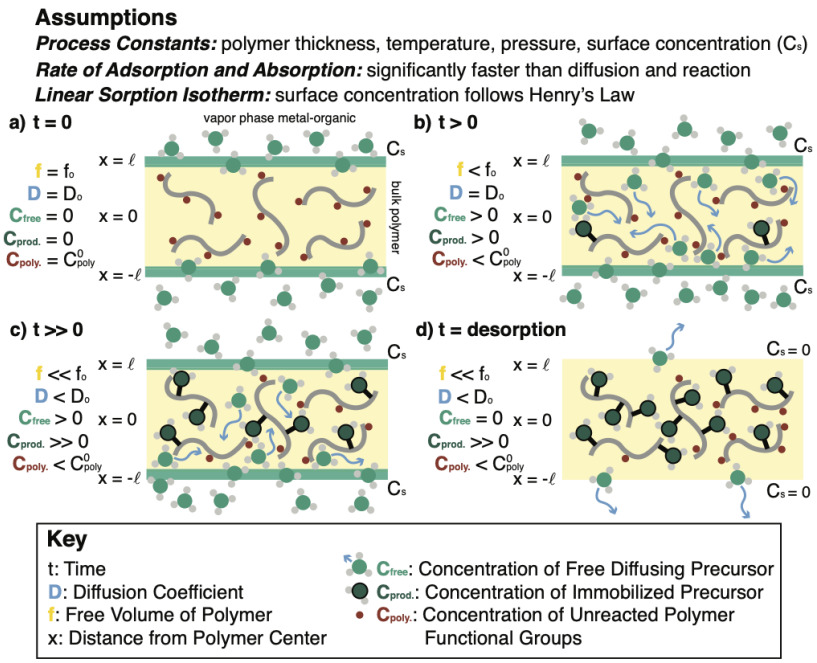


Fig. 1: Detailed steps of a typical VPI process (metal-organic cycle)[1]. a) At $t = 0$, a polymer is exposed to vapor phase metal-organic precursors, equilibrium is reached between polymer surface and vapor pressure. b) As $t > 0$, metal-organic vapor diffuses throughout the polymer, some reacting with polymer functional groups and losing mobility c) As t continues to increase, there will be both free diffusing vapors and immobilized products in the polymer, both contributing to the total mass uptake d) As desorption occurs, surface concentration of precursor drops to 0, immobilized product remains in the polymer while free diffusing precursors leave the polymer.

the introduction of a co-reactant that forms a non-volatile product. VPI is unique in that it modifies polymers post-fabrication, thus maintaining the material's macroscale form (thin films, bulk plastics, fibers, fabrics, etc.). As a result, the application space for VPI spans a number of industries [Leng and Losego, 2017, Waldman et al., 2019, Ingram and Jur, 2019, Azpitarte and Knez, 2018, Subramanian et al., 2019, Losego and Peng, 2019].

While the utility of VPI was the focus of initial research efforts, a recent shift has occurred to understand the fundamental thermodynamic and kinetic principles of the infiltration process. To account for how reactions in VPI

[1] Reprinted with permission from Ren, Yi, Emily K. McGuinness, Chaofan Huang, V. Roshan Joseph, Ryan P. Lively, and Mark D.Losego (2021). "Reaction–Diffusion Transport Model to Predict Precursor Uptake and Spatial Distribution in Vapor-Phase Infiltration Processes". In: *Chemistry of Materials* 33.13, pp. 5210–5222. Copyright 2021 American Chemical Society.

systems influence the transport behavior of vapor phase precursors in the infiltration process, Ren et al. [2021] recently proposed a reaction-diffusion model (see Figure 1) to understand the mass uptake behavior observed from the physical experiment, leading to the following system of partial differential equations (PDEs),

$$\begin{cases} \frac{\partial C_{free}}{\partial t} = D\frac{\partial^2 C_{free}}{\partial x^2} - kC_{free}C_{polymer} \\ \frac{\partial C_{product}}{\partial t} = kC_{free}C_{polymer} \\ D = D_0 \exp(-K'C_{product}) \\ \frac{\partial C_{polymer}}{\partial t} = -kC_{free}C_{polymer} \end{cases} \tag{3}$$

with the following initial and boundary conditions,

$$\begin{cases} C_{free} = 0, & 0 < x < l,\ t = 0 \\ C_{product} = 0, & 0 < x < l,\ t = 0 \\ C_{polymer} = C^0_{polymer}, & 0 < x < l,\ t = 0 \\ \frac{\partial C_{free}}{\partial x} = 0, & x = 0, \quad t > 0 \\ C_{free} = C_s, & x = l, \quad t > 0 \end{cases} \tag{4}$$

where $C_{free}(\text{mol/cm}^3)$ is the concentration of the free diffusing vapor-phase precursor, $C_{polymer}(\text{mol/cm}^3)$ is the concentration of the accessible reactive polymeric functional groups, $C_{product}(\text{mol/cm}^3)$ is the concentration of immobilized product from the reaction between the free diffusing vapor-phase precursor and the polymeric functional groups. The first differential equation describes how the free diffusing vapor-phase precursor diffuses into the polymer while being consumed via reaction; the second equation describes the formation of immobilized product; the third equation captures how diffusivity decreases exponentially with the formation of immobilized product; and the last equation describes the consumption of the polymer reactive groups. Based on the governing equations and boundary conditions, there are five unknown parameters $\theta = \{D_0, C_s, C^0_{polymer}, K', k\}$ that will directly impact the transport process, where $D_0(\text{cm}^2/\text{s})$ is initial diffusivity of the free diffusing vapor-phase precursor, $C_s(\text{mol/cm}^3)$ is the surface concentration of the free diffusing vapor-phase precursor, $C^0_{polymer}(\text{mol/cm}^3)$ is the initial concentration of accessible reactive polymeric functional groups, $K'(\text{cm}^3/\text{mol})$ is the hindering factor describing how immobilized product $C_{product}$ slows down the diffusivity of free diffusing vapor, and $k(\text{cm}^3/\text{mol}\cdot\text{s})$ is the associated reaction rate. There are three more operational parameters, polymer thickness l, temperature, and vapor pressure, that can be experimentally controlled. See supplementary information for more details of the VPI process.

Figure 2 shows experimentally collected data for the mass uptake of the polymer during the VPI process. This data is collected using a quartz crystal microbalance; more information about the experimental data collection can be found in Ren et al. [2021]. The solid black curve shows the total mass uptake of the free diffusing vapor-phase precursor and the immobilized product, i.e. $C_{free}+C_{product}$ after unit adjustment, over time recorded from the experiment

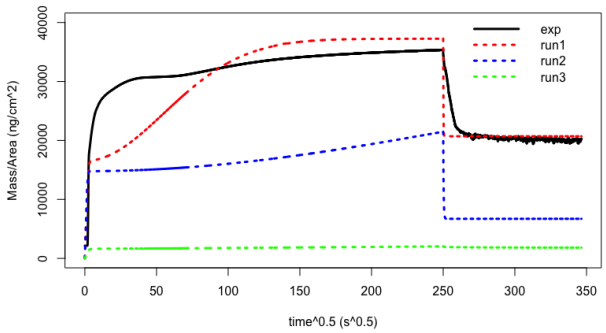


Fig. 2: Black solid curve is the area normalized total mass uptake of the free diffusing vapor-phase precursor and the immobilized product over time recorded from the experiment performed at 130°C, 8.7 Torr vapor pressure, and 483 nm thick polymer film [Ren et al., 2021]. The dashed curves are the corresponding mass uptakes over time from numerically solving the PDEs (3) under three different settings of the five unknown parameters θ.

performed at 130°C, 8.7 Torr vapor pressure, and 483 nm thick polymer film. Thus, our objective is to identify the five unknown parameters θ such that the total mass uptake obtained by numerically solving the PDEs (3) is aligned with the experimental observations. The dashed curves in Figure 2 show the results from solving the PDEs at three different sets of θ. They exhibit different patterns over time and are incompatible with the experimental observation, showing that identifying the correct θ is difficult, not to mention that solving the PDEs via numerical integration is computationally expensive: it takes about one minute to numerically solve the PDEs in (3) in a 2.3 GHz laptop.

When both spatial and temporal experimental observations are available, several gradient matching types of methods have been proposed in the literature to bypass the costly numerical integration step [Xun et al., 2013, Zhao et al., 2021]. However, for the VPI process, the spatial information is aggregated, yielding only a function over time as the experimental observations that we aim to match. For this case of aggregated spatial information, the numerical integration free methods cannot be applied. In fact, many common chemical engineering phenomena, such as the transport processes [Chapter4; Hines and Maddox, 1985], modeled with PDEs will have experimentally collected data that is not easily collectable concurrently in both the space and time domains, but rather will have one of these domains aggregated. This loss of spatial information makes the inverse problem ill-posed as many local minima could exist, but we will see in Section 5 that our proposed Bayesian optimization procedure could circumvent the issue and identify a good optimum using only a few evaluations of the PDEs.

3 Bayesian Optimization

3.1 Gaussian Process

Let us first review the Gaussian process (GP), the powerful and popularly used probabilistic surrogate model in Bayesian optimization. Following the definition in Rusmassen and Williams [2006], a Gaussian process is a collection of random variables such that any finite number of which have a joint multivariate Gaussian distribution, denoted by

$$f(x) \sim \mathcal{GP}(\mu, \mathcal{K}),\ x \in \mathbb{R}^d, \tag{5}$$

where $\mu : \mathbb{R}^d \to \mathbb{R}$ is the mean function and $\mathcal{K} : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a positive definite covariance function parameterized by some hyperparameters which we obtain using the empirical Bayes method. For any finite set of inputs $X = (x_1, \ldots, x_n)$, we have $f(X) \sim \mathcal{N}(\mu(X), \mathcal{K}(X, X))$. It is typical that we only have noisy observations, i.e., we observe $Y_i = f(x_i) + \epsilon_i$ where we assume additive i.i.d. Gaussian random noise $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$. Let $Y = (Y_1, \ldots, Y_n)$, and it follows that

$$Y \sim \mathcal{N}(\mu(X), \mathcal{K}(X, X) + \eta^2 I_n) \tag{6}$$

Conditional on observing $\mathcal{D}_n = \{(x_i, y_i)\}_{i=1}^n$ and let $\mathbf{y} = (y_1, \ldots, y_n)$, the predictive distribution of f at any unseen test point $x \in \mathbb{R}^d$ is

$$f(x)|\mathcal{D}_n \sim \mathcal{N}(\tilde{\mu}(x), \tilde{\sigma}^2(x)), \tag{7}$$

where

$$\begin{aligned} \tilde{\mu}(x) &= \mu(x) + \mathcal{K}(x, X)(\mathcal{K}(X, X) + \eta^2 I_n)^{-1}(\mathbf{y} - \mu(X)), \\ \tilde{\sigma}^2(x) &= \mathcal{K}(x, x) - \mathcal{K}(x, X)(\mathcal{K}(X, X) + \eta^2 I_n)^{-1}\mathcal{K}(X, x), \end{aligned} \tag{8}$$

can be derived using the property of conditional multivariate Gaussian distribution. For the case where the observations are exact, i.e., $y_i = f(x_i)\ \forall i$, we simply set $\eta^2 = 0$ in (8) to derive the predictive distribution. See Rusmassen and Williams [2006] and Santner et al. [2018] for more details of the GP.

3.2 Standard Bayesian Optimization

Now consider the task of minimizing some black-box function $f : \mathbb{R}^d \to \mathbb{R}$ which could be expensive to evaluate. Under the limited computational budget, we want to allocate the resources smartly to balance between learning the response surface (exploration) and locating the optimal solution (exploitation). Bayesian optimization (BO) offers a principle solution.

For simplicity, consider the case when the evaluation of f is noise free. By using a Gaussian process surrogate model to approximate f, (7) shows that the predictive distribution for any input $x \in \mathbb{R}^d$ conditional on available samples $\mathcal{D}_n = \{(x_i, y_i = f(x_i))\}_{i=1}^n$ is Gaussian with mean $\tilde{\mu}(x)$ and variance $\tilde{\sigma}^2(x)$ defined in (8). Bayesian optimization selects the next sample via an acquisition

function $\alpha : \mathbb{R}^d \to \mathbb{R}$ that assigns utility to any unseen input x based on the inferred predictive distribution from the GP model. In this paper, we focus on the Expected Improvement (EI) acquisition function [Jones et al., 1998]. In EI, for each input x, we compute how much we can expect to improve over the best value $f_{\min} = \min_i f(x_i)$ we have obtained so far, that is to compute

$$\alpha_{\text{EI}}(x; \mathcal{D}_n) = \mathbb{E}[\{f_{\min} - f(x)\}^+ | \mathcal{D}_n], \tag{9}$$

where the notation $\{u\}^+$ stands for $\max(u, 0)$. Given that the predictive distribution of $f(x)$ conditional on the observations $\mathcal{D}$ is Gaussian, the EI can be computed in closed-form,

$$\alpha_{\text{EI}}(x; \mathcal{D}_n) = \tilde{\sigma}(x)\left\{\left(\frac{f_{\min} - \tilde{\mu}(x)}{\tilde{\sigma}(x)}\right)\Phi\left(\frac{f_{\min} - \tilde{\mu}(x)}{\tilde{\sigma}(x)}\right) + \phi\left(\frac{f_{\min} - \tilde{\mu}(x)}{\tilde{\sigma}(x)}\right)\right\} \tag{10}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the c.d.f. and p.d.f. of the standard normal distribution, respectively. We then select the next sample that maximizes the EI,

$$x_{n+1} = \arg\max_{x \in \mathbb{R}^d} \alpha_{\text{EI}}(x; \mathcal{D}_n). \tag{11}$$

By maximizing the EI in (10), we can see that it favors selecting a setting that (i) has large predictive standard deviation $\tilde{\sigma}(x)$ for exploring the response surface where the model has high uncertainty (exploration), or (ii) has small predictive mean $\tilde{\mu}(x)$ that try to locate a better optimum (exploitation). The EI acquisition function naturally balances between exploration and exploitation for the next sample selection, showing why BO is robust against problems with many local minima.

4 Bayesian Optimization for Inverse Problems

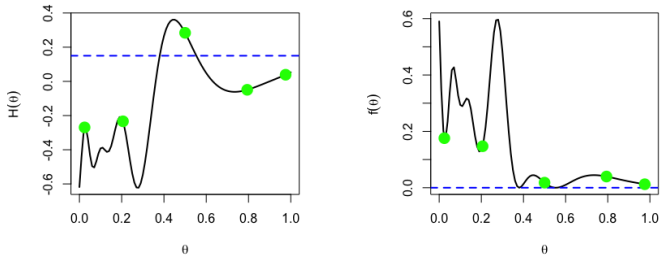
In this section, we discuss the modification to the standard Bayesian optimization procedure for solving the inverse problem in (2). We first discuss the case of scalar-output model H to show that the generalized chi-square distribution is a more appropriate predictive distribution for the the inverse problems. We next extend the framework for the functional-output model H, leading to our proposed method BOFO, Bayesian Optimization of Functional Output.

4.1 Inverse Problems with Scalar Output

Consider a scalar-output simulator-based model $H : \Theta \to \mathbb{R}$ with the target observed data $y^* \in \mathbb{R}$, and the objective of the inverse problem is to solve for

$$\theta^* = \arg\min_{\theta \in \Theta} f(\theta) = g(H(\theta), y^*) = \|H(\theta) - y^*\|_2^2 = (H(\theta) - y^*)^2 \tag{12}$$

where $g : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ is the squared distance function. Following the standard BO procedure, we apply GP model on f. This approach suffers from two deficiencies: (i) the information from the model H is masked, which could



(a) $H(\theta)$ (b) $f(\theta)$

Fig. 3: Left panel shows the model H in (13) with blue dashed line indicating the target $y^* = 0.15$. Right panel shows the corresponding squared loss function $f(\theta) = \|H(\theta) - y^*\|_2^2$ that we want to minimize. The green circles are the initial five observations.

contain useful insight for the optimization, and (ii) the predictive distribution from GP is symmetric and has support in the negative domain, while the squared distance type of objective in (12) only yields non-negative value and the error distribution is skewed.

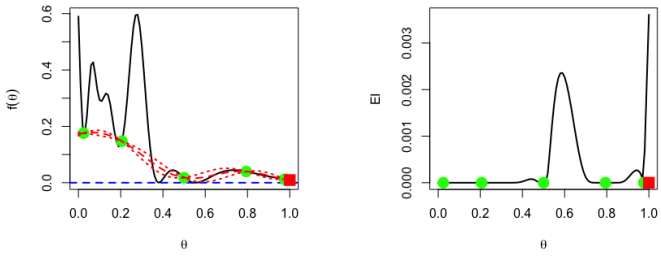
To illustrate the shortcoming, let us consider the following model [Xiong et al., 2007],

$$H(\theta) = \sin(30(\theta - 0.9)^4)\cos(2(\theta - 0.9)) + (\theta - 0.9)/2, \ \forall \theta \in [0, 1], \tag{13}$$

with the target $y^* = 0.15$. Figure 3 shows the model H and the corresponding squared loss function f that we want to minimize. We can see that this problem has two minima in the domain $[0, 1]$. Suppose that we have five exact observations $\mathcal{D}_5 = \{(\theta_i, H(\theta_i), f(\theta_i) = (H(\theta_i) - y^*)^2)\}_{i=1}^5$ (green circles in Figure 3) on the 5 Chebyshev nodes of $[0, 1]$. Left panel of Figure 4 shows the predictive distribution of f at any unseen input $\theta \in [0, 1]$ after GP modeling on f. The predictive distribution yields some negative values at around $\theta = 0.6$, reflecting a flawed expectation about the behavior of f in those unexplored regions. Moreover, the predictive distribution completely misses one of the minima at around 0.4, which can also be seen from the EI value in the right panel of Figure 4.

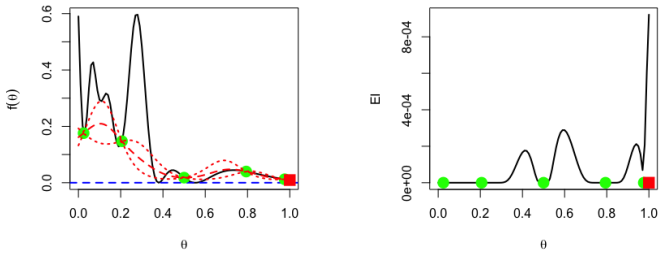
One simple approach to take care of the non-negative range of f is to perform GP modeling on $\log f$ [Gutmann and Corander, 2016], then the predictive distribution for $\log f$ at any input x is Gaussian with some mean $\tilde{\mu}(\theta)$ and variance $\tilde{\sigma}^2(\theta)$. Thus, it follows that the predictive distribution for f is log-Gaussian, and we can compute the corresponding EI analytically,

$$\begin{aligned} \alpha_{\mathrm{EI}}(\theta; \mathcal{D}_n) = & \Phi\left(\frac{\log(f_{\min}) - \tilde{\mu}(\theta)}{\tilde{\sigma}(\theta)}\right) f_{\min} - \\ & \Phi\left(\frac{\log(f_{\min}) - \tilde{\mu}(\theta) - \tilde{\sigma}^2(\theta)}{\tilde{\sigma}(\theta)}\right) \exp\left\{\tilde{\mu}(\theta) + \frac{\tilde{\sigma}^2(\theta)}{2}\right\}, \end{aligned} \tag{14}$$



(a) Predictive Distribution (b) Expected Improvement

Fig. 4: Left panel shows the predictive mean in red dashed line and 90% Gaussian predictive interval in red dotted lines. Right panel shows the corresponding EI acquisition value. The green circles are the initial five observations, and the red square is the sample selected by EI.

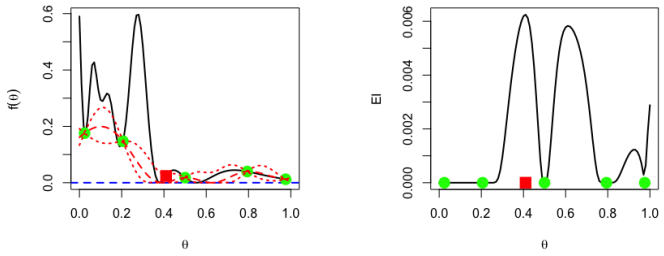


(a) Predictive Distribution (b) Expected Improvement

Fig. 5: Left panel shows the predictive mean in red dashed line and 90% log-Gaussian predictive interval in red dotted lines. Right panel shows the corresponding EI acquisition value. The green circles are the initial five observations, and the red square is the sample selected by EI.

where $f_{\min} = \min_i f(\theta_i)$ is the best value we have from the samples. We can see that the predictive distribution no longer has support on the negative domain (left panel of Figure 5). Although it is able to yield positive EI for the two local minima, it again fails to identify them as the most important regions for the next step exploration (right panel of Figure 5), suffering the same drawback of the standard BO approach on f that ignores the information from the model H (Figure 4).

Recently Uhrenholt and Jensen [2019] and Matsui et al. [2019] proposed a principal remedy to the forgoing problem by using the generalized chi-square distribution as the predictive distribution for the squared distance type of ob-



(a) Predictive Distribution (b) Expected Improvement

Fig. 6: Left panel shows the predictive mean in red dashed line and 90% generalized chi-square predictive interval in red dotted lines. Right panel shows the corresponding EI acquisition value. The green circles are the initial five observations, and the red square is the sample selected by EI.

jective function. They apply GP modeling on H, then for any unseen input θ, $H(\theta)|\mathcal{D}_n$ follows a Gaussian distribution. Hence, $f(\theta)|\mathcal{D}_n$ is the square of a Gaussian random variable, which follows the generalized chi-square distribution [Imhof, 1961]. See supplementary information for details. Next, we need to compute the EI under the generalized chi-square predictive distribution. Let G_{χ^2} and g_{χ^2} denotes the c.d.f. and p.d.f. of the generalized chi-square distribution respectively. Let $f_{\min} = \min_i f(\theta_i)$ be the best value we have from the samples, then the EI of any input θ can be computed by

$$\begin{aligned}
\alpha_{\mathrm{EI}}(\theta; \mathcal{D}_n) &= \mathbb{E}[\{f_{\min} - f(\theta)\}^+ | \mathcal{D}_n] \\
&= \int_0^{f_{\min}} (f_{\min} - t) g_{\chi^2}(t) dt \\
&= f_{\min} G_{\chi^2}(f_{\min}) - \int_0^{f_{\min}} t g_{\chi^2}(t) dt \\
&= f_{\min} G_{\chi^2}(f_{\min}) - \left(\{t G_{\chi^2}(t)\}|_0^{f_{\min}} - \int_0^{f_{\min}} G_{\chi^2}(t) dt \right) \\
&= \int_0^{f_{\min}} G_{\chi^2}(t) dt,
\end{aligned} \tag{15}$$

which can be efficiently estimated using quadrature [Golub and Welsch, 1969]. The EI in (15) requires computing the left tail probability of the generalized chi-square distribution, but the existing methods such as Imhof's [Imhof, 1961] and Liu's [Liu et al., 2009] methods implemented in the R package `CompQuadForm` [Duchesne and de Micheaux, 2010] yield poor estimation when the tail probability is very small. To address this issue, we use importance sampling on support points [Mak and Joseph, 2018], a recently developed Quasi-Monte Carlo point set, for more robust approximation of the small tail

probability that often occurs in the EI computation. See supplementary information for details.

Left panel of Figure 6 shows the predictive generalized chi-square distribution of f at any unseen input θ. We can see that the generalized chi-square predictive distribution has support on the non-negative domain only, reflecting the true behavior of the function f. Moreover, this approach incorporates the information available from the model H, which helps identify the two regions around the minima as the key regions for the next step exploration, while the two aforementioned standard BO approaches all fail to recognize. This demonstrates the advantage of using generalized chi-square as the predictive distribution when applying BO in the inverse problem.

4.2 Inverse Problem with Functional Output

Now consider a functional-output model $H(t;\theta)$ with the target functional $y^*(t)$ defined in a compact interval $\mathcal{T}$. The inverse problem aims to solve for

$$\begin{aligned}\theta^* = \arg\min_{\theta\in\Theta} f(\theta) &= g(H(t;\theta), y^*(t)) \\ &= \|H(t;\theta) - y^*(t)\|_2^2 \\ &= \int_{t\in\mathcal{T}} \{H(t;\theta) - y^*(t)\}^2 dt,\end{aligned} \tag{16}$$

where g computes the squared distance between any two functions. We first introduce the functional principal component analysis (FPCA) that allows for simplification of the integral in (16) to a summation of the squared errors in principal component scores. Next, we present BOFO, Bayesian Optimization of Functional Output, that can efficiently solve for θ^* in (16).

4.2.1 Functional Principal Component Analysis

Similar to the principal component analysis (PCA), Ramsay [2004] derives the similar decomposition for the functional data, which is known as the functional principal component analysis (FPCA). It is also termed as Karhunen-Loève expansion or generalized Fourier expansion. Let $Y(t)$ be a random function supported on a compact interval $\mathcal{T}$ with mean function $\mu(t)$ and positive definite covariance function

$$\kappa(s,t) \equiv \mathrm{Cov}\{Y(s), Y(t)\} = \sum_{k=1}^{\infty} \tau_k \psi_k(s)\psi_k(t), \tag{17}$$

where equality follows from the spectral decomposition. $\tau_1 \geq \tau_2 \geq \cdots$ are the eigenvalues with the respective orthonormal eigenvectors $\psi_1, \psi_2, \ldots$. The functions $\{\psi_k\}_{k=1}^{\infty}$ form a basis for the space of square-integrable functions on $\mathcal{T}$, i.e., for any square-integrable function $y_i(t)$ on $\mathcal{T}$,

$$y_i(t) = \mu(t) + \sum_{k=1}^{\infty} \tau_k^{1/2} \beta_k^i \psi_j(t), \tag{18}$$

where $\beta_k^i = \tau_k^{1/2} \int_{\mathcal{I}} y_i(t)\psi_k(t)dt$ is the k-th principal component score. Given that the eigenvectors ψ_k's are orthonormal, the squared distance between any two square-integrable function $y_i(t)$ and $y_j(t)$ can be simplified to

$$\|y_i(t) - y_j(t)\|_2^2 = \sum_{k=1}^{\infty} \tau_k(\beta_k^i - \beta_k^j)^2 \approx \sum_{k=1}^{K} \tau_k(\beta_k^i - \beta_k^j)^2, \tag{19}$$

where we only retain the top K eigenvalues with their respective eigenvectors for the approximation. The rule-of-thumb is to select K such that 99% of the total variance can be explained, i.e., we find the smallest K that $\sum_{k=1}^{K} \tau_k / \sum_{k=1}^{\infty} \tau_k > 0.99$. We use the R package `fda` [Ramsay et al., 2020] to estimate the mean function $\mu(t)$, eigenvalue τ_k's, and the eigenvector ψ_k's from data.

4.2.2 Bayesian Optimization for Inverse Problem with Functional Output

Assume that both the model output $H(t;\theta)$ and the target $y^*(t)$ are square-integrable in the compact interval $\mathcal{T}$. Following the conclusion in (19), the objective function in (16) can be simplified to

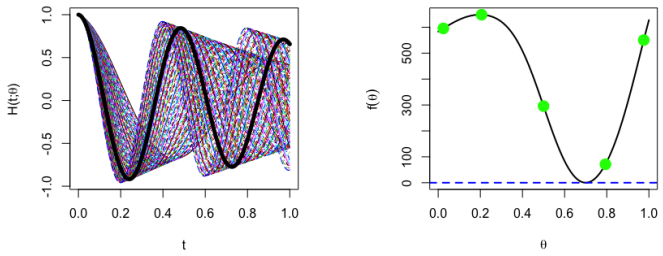
$$f(\theta) = \|H(t;\theta) - y^*(t)\|_2^2 \approx \sum_{k=1}^{K} \tau_k(\beta_k(\theta) - \beta_k^*)^2, \tag{20}$$

where $\beta_k(\theta)$ is the k-th principal component score for $H(t;\theta)$ and β_k^* is the k-th principal component score for the target $y^*(t)$. Thus, we only require GP modeling on the principal component score $\beta_k(\theta)$'s. Moreover, by the orthonormality of the eigenvectors, the principal component scores are independent, so we can fit independent GP model for each $\beta_k(\theta)$. We use the `mlegp` package [Dancik and Dorman, 2008] available in R for the GP fitting. Conditional on observing $\mathcal{D}_n = \{(\theta_i, y_i(t) = H(t;\theta_i))\}_{i=1}^n$ and computing all corresponding principal component scores $\{\beta_k^i\}_{i=1}^n{}_{k=1}^K$, the predictive distribution of $\beta_k(\theta)$ at any unseen data θ is again Gaussian for all $k = 1, \ldots, K$. It follows that $f(\theta)|\mathcal{D}_n$ is approximated by a weighted sum of independent squared Gaussian random variables, which is again the generalized chi-square distribution [Imhof, 1961] with specific set of parameters (see supplementary information). Thus, the EI acquisition function we derived for the generalized chi-square distribution in (15) in the scalar-output case can also be applied.

Consider the following functional-output model with parameter $\theta \in [0, 1]$,

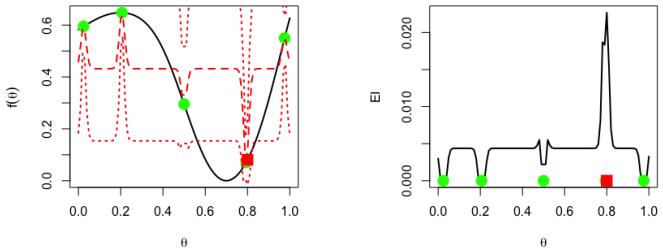
$$H(t;\theta) = \exp\{-5t(1-\theta/2)^2/6\} \cos\left(5t\sqrt{1 + (1+2\theta)^2 - (1-\theta/2)^3/4}\right), \ \forall t \in [0,1], \tag{21}$$

and the target functional $y^*(t) = H(t; 0.7)$. Figure 7 shows the functional output of $H(t;\theta)$ at different value of $\theta \in [0, 1]$ and the corresponding squared loss function f that we want to minimize over. Again, assume that the observations are exact, and we have five initial observations $\mathcal{D}_5 = \{(\theta_i, y_i(t) = H(t;\theta_i), f(\theta_i) = \|y_i(t) - y^*(t)\|_2^2)\}_{i=1}^5$. Let us first look at the performance of



(a) $h(x)$ (b) $f(x)$

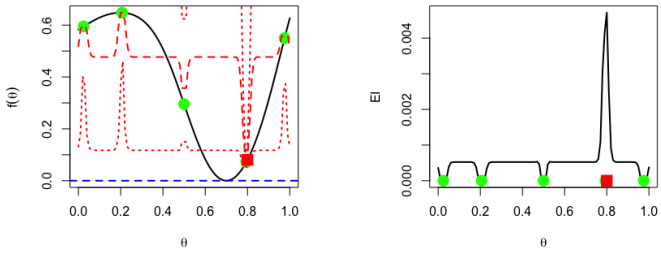
Fig. 7: Left panel shows the functional-output of $H(t;\theta)$ (21) at different $\theta \in [0,1]$. The black solid line indicating the target $y^* = H(t;0.7)$. Right panel shows the corresponding squared loss function $f(\theta) = \|H(t;\theta) - y^*\|_2^2$. The green circles are the initial five observations.



(a) Predictive Distribution (b) Expected Improvement

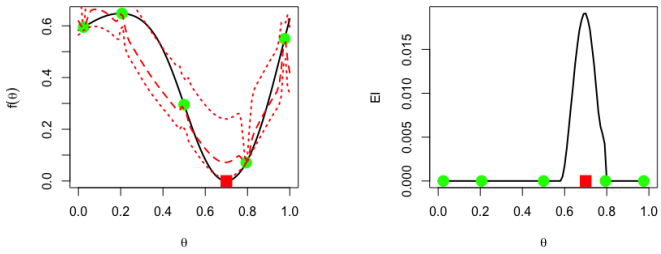
Fig. 8: Left panel shows the predictive mean in red dashed line and 90% Gaussian predictive interval in red dotted lines. Right panel shows the corresponding EI acquisition value. The green circles are the initial five observations, and the red square is the sample selected by EI.

applying standard Bayesian optimization with respect to the composite objective function f, which ignores the information from the model H and the behavior of non-negative range. Figure 8 shows both the predictive distribution of f at any input θ and the corresponding EI value. We can see that the performance is poor, as the GP model fails to learn the behavior of f, which is possibly due to too few observations. Even after applying the GP model on $\log f$ to account for the non-negative range behavior, the performance is still disastrous by looking at Figure 9. Last, let us incorporate the available information from the model H by fitting GP on the principal component scores. Left panel of Figure 10 shows the generalized chi-square predictive distribution



(a) Predictive Distribution (b) Expected Improvement

Fig. 9: Left panel shows the predictive mean in red dashed line and 90% log-Gaussian predictive interval in red dotted lines. Right panel shows the corresponding EI acquisition value. The green circles are the initial five observations, and the red square is the sample selected by EI.



(a) Predictive Distribution (b) Expected Improvement

Fig. 10: Left panel shows the predictive mean in red dashed line and 90% generalized chi-square predictive interval in red dotted lines. Right panel shows the corresponding EI acquisition value. The green circles are the initial five observations, and the red square is the sample selected by EI.

of f at any input θ, we can see that it can better capture the true behavior of f, e.g. the trend and the support on non-negative domain. Moreover, it is able to assign higher utility to regions around the optimum by looking at the EI plot, showing its advantage against the standard BO approach of ignoring available information from the model H. Algorithm 1 details the full Bayesian optimization procedure for efficiently solving the inverse problem involving a functional-output model H.

Algorithm 1: BOFO: Bayesian Optimization of Functional Output.

Objective: Solve (16).

Initial Observations:

$\mathcal{D}_n = \{(\theta_i, y_i(t) = H(t;\theta_i), f(\theta_i) = \|y_i(t) - y^*(t)\|_2^2)\}_{i=1}^n$.

while *not converges* **do**

- Estimate the mean function $\mu(t)$, eigenvalues τ_k's, and eigenvectors ψ_k's from the available observations $\{y_i(t) = H(t;\theta_i)\}_{i=1}^n$.
- Compute the principal component scores $\{\beta_k^i\}_{i=1}^n{}_{k=1}^K$ for the available observations, and the principal component scores $\{\beta_k^*\}_{k=1}^K$ for the target $y^*(t)$.
- Fit independent GP model for each principal component score $\beta_k(\theta)$ using $\{\beta_k^i\}_{i=1}^n$.
- Compute the predictive generalized chi-square distribution for $f(\theta)$ on any unseen test point θ and find the next sample by solving

$$\theta_{n+1} = \arg\max_{\theta\in\Theta} \alpha_{\mathrm{EI}}(\theta, \mathcal{D}_n)$$

  via the Nelder-Mead method [Nelder and Mead, 1965].
- Evaluate $y_{n+1}(t) = H(t;\theta_{n+1})$ and add to the observations

$$\mathcal{D}_{n+1} = \mathcal{D}_n \cup \{(\theta_{n+1}, y_{n+1}(t) = H(t;\theta_{n+1}), f(\theta_{n+1}) = \|y_{n+1}(t) - y^*(t)\|_2^2)\}.$$

- Set $n = n + 1$.

end

Return: $\theta^* = \arg\min_{\theta\in\{\theta_i\}_{i=1}^n} f(\theta)$, the sample with the best result.

D_0	C_s	$C^0_{polymer}$	K'	k
[1.0e-12, 1.0e-9]	[4.0e-3, 5.0e-3]	[5.0e-3, 6.0e-3]	[5.0e2, 2.5e3]	[1.0e-3, 1.0e1]

Table 1: Feasible domain space Θ of the five unknown parameters.

5 Numerical Results

We now demonstrate the efficiency of our proposed method BOFO (Algorithm 1) on the motivating model calibration problem of the VPI process. Recall that the goal is to identify the set of unknown parameters $\theta = \{D_0, C_s, C^0_{polymer}, K', k\}$ such that the output from the PDEs $H(t;\theta)$ is aligned with the experimental data $y^*(t)$ (black solid curve in Figure 2).

Recommended by the rule-of-thumb computer experiment sample size of $n = 10p$ where p is the number of unknown parameters [Loeppky et al., 2009], we start with 50 initial observations. To ensure good space-filling properties in all subset of the factors, we use the 50-run maximum projection (MaxPro) design [Joseph et al., 2015] in the feasible domain space defined by Table 1 with rationale provided in the supplementary information. Since the range for

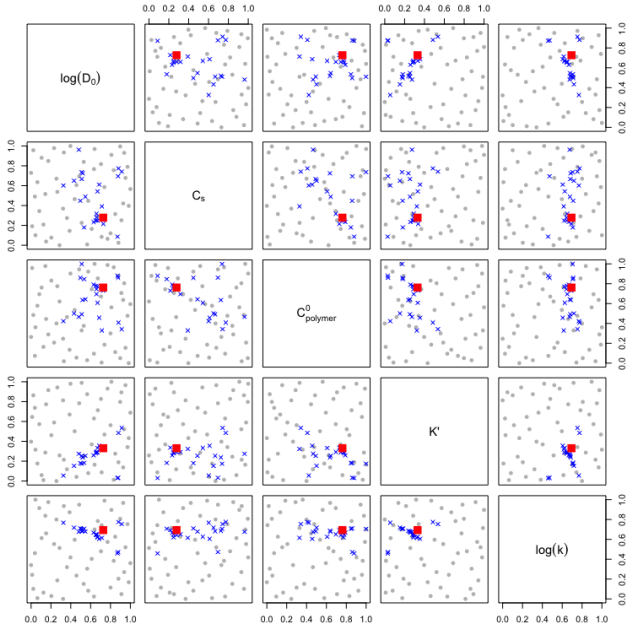
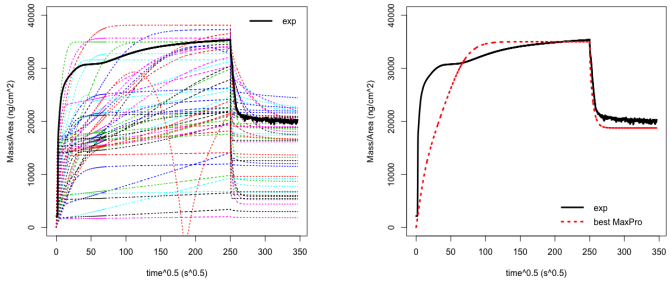


Fig. 11: All samples from applying BOFO on the motivation problem after scaling to $[0, 1]$. Gray circles are the initial 50-point MaxPro design. Blue crosses are the 25 adaptive samples selected by EI. Red square is the best sample θ^* with the smallest square error to the experimental data $y^*(t)$.

diffusivity and reactivity is huge, spanning orders of magnitude, the values for both parameters can be considered useful and accurate when the correct order of magnitude is determined in chemical engineering. In this work, to ensure a comprehensive parameter range while maintaining high calibration efficiency, the diffusivity D_0 and the reaction rate k are calibrated based on log scale. See gray circles in Figure 11 for the 50-point MaxPro design. The left panel of Figure 12 shows the outputs of evaluating the PDEs at the 50-run MaxPro design. We can see that the magnitudes are very different, with few values in the 1,000's while many are in 10,000's. Also, inconsistent patterns are observed from run to run. Furthermore, even for the MaxPro run with the smallest square error to the experimental data, the discrepancy is still large (right panel of Figure 12), showing the difficulty of our problem! Note that from the left panel of Figure 12, there is one MaxPro run that yields negative values, which are due to the numerical approximation errors from solving the PDEs. We remove this sample in the subsequent Bayesian optimization step.

For a comparison to our proposed method BOFO, we also consider (i) random sampling from the feasible domain space, (ii) standard BO procedure on f, (iii) standard BO procedure on $\log f$ to account for the non-negative



(a) 50 MaxPro Runs (b) Best MaxPro Run

Fig. 12: Left panel shows the PDEs' outputs from the initial 50-point MaxPro runs. Right panel shows the best MaxPro run to the experimental data (black solid curve).

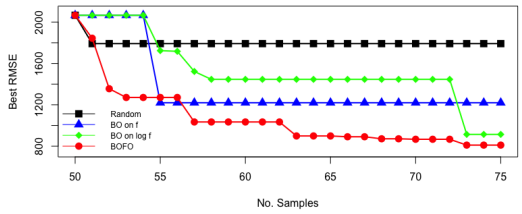
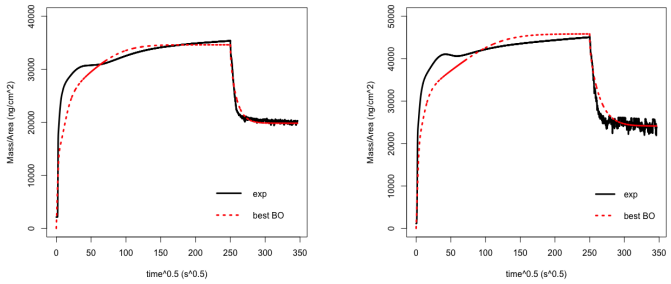


Fig. 13: Best RMSE (root mean square error) versus the number of BO samples (PDEs evaluation) using the same initial 50-point MaxPro design on the motivation example.

range. We run all four approaches by another 25 evaluations of the PDEs for a fair comparison. Figure 13 shows that BOFO outperforms the other three approaches. Though the standard BO procedure on $\log f$ is also able to locate a good design in terms of the squared error, we can see that from Figure 13, this approach spends more time exploring the less important regions, in other words, we do not see a consistent drop in the best squared error as we evaluate the PDEs at more design points. This is expected since the response surface of the objective function f is too complicated to learn without using the output information available from the PDEs $H(t;\theta)$, demonstrating the robustness of our BOFO algorithm. Next, we discuss the additional runtime that is incurred by BOFO over the standard BO procedure. For the problem considered, the top three principal components explain 99% of the variance, drastically reducing the problem dimension from 6,601 to 3, where the system of PDEs is evaluated at 6,601 time points. Thus, we only require fitting three GP models,

	Parameter	Train	Test
Experiment	Temperature (°C)	130	130
	Thickness (nm)	483	607
	Pressure (Torr)	8.7	**10.5**
PDEs	D_0 (cm^2/s)	1.510e-10	1.510e-10
	C_s (mol/cm^3)	4.274e-3	**5.158e-3**
	$C^0_{polymer}$ (mol/cm^3)	5.749e-3	5.749e-3
	K' (cm^3/mol)	1160	1160
	k ($cm^3/mol \cdot s$)	0.592	0.592

Table 2: The best set of the five unknown parameters θ^* identified by our method (Algorithm 1) on the 8.7 Torr experimental run (train set), and then estimated the parameters for the 10.5 Torr run (test set) by multiplying C_s with the ratio 10.5/8.7 to account for the increase in pressure.



(a) Train (8.7 Torr) (b) Test (10.5 Torr)

Fig. 14: Left panel shows the PDEs' output at the θ^* identified by our method (Algorithm 1) on the 8.7 Torr run (see Table 2). Right panel shows the PDEs' output at the adjusted θ^* on the 10.5 Torr run (see Table 2). Both 8.7 and 10.5 Torr experimental data are in solid black curves.

yielding additional computational burden that is negligible compared to the cost of numerically solving the PDEs.

The blue crosses in Figure 11 shows the 25 adaptive samples selected by our proposed BOFO. We can see that these 25 samples are also spread out in the feasible domain space, indicating the consideration of both exploration and exploitation of our method to avoid getting stuck at some local optimum. Table 2 lists the value for the best set of parameters θ^* identified by BOFO using the 8.7 Torr experimental data, and the evaluation of PDEs at θ^* is presented in the left panel of Figure 14. We can see that the PDEs' output at θ^* aligns much better with the experimental data than the output from the initial 50 MaxPro runs (Figure 12). To further assess the performance of θ^*, we use the parameters θ^* learned from the 8.7 Torr run to predict the mass change behavior of the VPI process at 10.5 Torr pressure with the same

temperature setting of 130°C. As mentioned in Ren et al. [2021], due to the increase in pressure from 8.7 Torr to 10.5 Torr, we should consider the same ratio (10.5/8.7) increment for the surface concentration of the free diffusing vapor C_s with the adjusted value listed in Table 2. From the right panel of Figure 14, we can see that the output from the PDEs at the estimated θ^* is also similar to the 10.5 Torr experimental data, showing the generalizability of the PDEs with θ^* identified by our proposed method to predict the mass change behavior of the VPI process at different experimental settings. Having this generalizability is important since we no longer need to run a physical experiment to study the mass uptake behavior at new polymer thickness or vapor pressure setting. We could save time (the physical experiment of the VPI process takes more than 30 hours) and expense from purchasing the material and constructing the systems capable of performing in situ characterization of the process. More importantly, only 75 evaluations of the expensive PDEs numerical solver is used to identify the good set of the unknown parameters θ^* from the large feasible domain space Θ in Table 1, showing the efficiency of our proposed algorithm BOFO on the inverse problems with functional output. The source codes for the comparisons are available at `https://github.com/BillHuang01/BOFO`.

6 Conclusion

In this paper we present a Bayesian optimization algorithm, BOFO (Algorithm 1), for efficiently solving inverse problems with a functional-output model, i.e., that is to identify the unknown input θ^* such that the expensive functional-output model $H(t;\theta^*)$ can achieve a desired target functional output $y^*(t)$. An important use of BOFO is in the model calibration problem of the PDEs with spatially aggregated observational data, such as the VPI process presented in Section 2. To tackle this difficult inverse problem, there are two improvements proposed in this paper over the standard Bayesian optimization procedure: (i) we suggest the use of generalized chi-square distribution as a more appropriate predictive distribution since it principally capture the correct probabilistic assumption, (ii) we propose the use of functional principal component analysis for functional data dimensional reduction, and moreover, we further show that this approach yields an efficient procedure to compute the predictive distribution and the corresponding expected improvement acquisition function. The proposed BOFO algorithm shows successful result in solving the motivated model calibration problem from the VPI process. Moreover, BOFO could be adapted for solving the inverse problems with the target observed data being a tensor by replacing the functional principal component analysis with tensor decomposition. Last, the discussion of the chi-square distribution as a more appropriate prediction distribution also sheds light on the approximate Bayesian computation literature where Gaussian process is used to model the discrepancy measure for speeding up the inference [Gutmann and Corander, 2016, Järvenpää et al., 2018].

Supplementary Information

codes are available at `https://github.com/BillHuang01/BOFO`.
supplementary.pdf contains the supplementary information of this paper, including details of VPI process, generalized chi-square distribution, and etc.

Acknowledgements This material is based upon work supported by the National Science Foundation (DMREF-1921873). EKM was also supported by the Department of Defense (DoD) through the National Defense Science & Engineering Graduate Fellowship (NDSEG) Program.

References

M. A. Alvarez, L. Rosasco, and N. D. Lawrence. Kernels for vector-valued functions: A review. *Foundations and Trends® in Machine Learning*, 4(3): 195–266, 2012.

R. Astudillo and P. Frazier. Bayesian optimization of composite functions. In *International Conference on Machine Learning*, pages 354–363. PMLR, 2019.

I. Azpitarte and M. Knez. Vapor phase infiltration: from a bioinspired process to technologic application, a prospective review. *MRS Communications*, 8 (3):727–741, 2018. doi: 10.1557/mrc.2018.126. URL `https://doi.org/10.1557/mrc.2018.126`.

G. E. Box and W. G. Hunter. A useful method for model-building. *Technometrics*, 4(3):301–318, 1962.

G. M. Dancik and K. S. Dorman. mlegp: statistical analysis for computer models of biological systems using r. *Bioinformatics*, 24(17):1966–1967, 2008.

A. Delaigle and P. Hall. Defining probability density for a distribution of random functions. *The Annals of Statistics*, 38(2):1171–1193, 2010.

P. Duchesne and P. L. de Micheaux. Computing the distribution of quadratic forms: Further comparisons between the liu-tang-zhang approximation and exact methods. *Computational Statistics and Data Analysis*, 54:858–862, 2010.

P. I. Frazier. Bayesian optimization. *INFORMS Tutorials*, pages 255–278, 2018.

G. H. Golub and J. H. Welsch. Calculation of gauss quadrature rules. *Mathematics of computation*, 23(106):221–230, 1969.

M. U. Gutmann and J. Corander. Bayesian optimization for likelihood-free inference of simulator-based statistical models. *Journal of Machine Learning Research*, 2016.

A. L. Hines and R. N. Maddox. *Mass transfer: fundamentals and applications*, volume 434. Prentice-Hall Englewood Cliffs, NJ, 1985.

J.-P. Imhof. Computing the distribution of quadratic forms in normal variables. *Biometrika*, 48(3/4):419–426, 1961.

W. F. Ingram and J. S. Jur. Properties and applications of vapor infiltration into polymeric substrates. *Jom*, 71(1):238–245, 2019. doi: 10.1007/s11837-018-3157-9. URL https://doi.org/10.1007/s11837-018-3157-9.

M. Järvenpää, M. U. Gutmann, A. Vehtari, P. Marttinen, et al. Gaussian process modelling in approximate bayesian computation to estimate horizontal gene transfer in bacteria. *Annals of Applied Statistics*, 12(4):2228–2251, 2018.

D. R. Jones. A taxonomy of global optimization methods based on response surfaces. *Journal of global optimization*, 21(4):345–383, 2001.

D. R. Jones, M. Schonlau, and W. J. Welch. Efficient global optimization of expensive black-box functions. *Journal of Global optimization*, 13(4):455–492, 1998.

V. R. Joseph and H. Yan. Engineering-driven statistical adjustment and calibration. *Technometrics*, 57(2):257–267, 2015.

V. R. Joseph, E. Gul, and S. Ba. Maximum projection designs for computer experiments. *Biometrika*, 102(2):371–380, 2015.

M. C. Kennedy and A. O'Hagan. Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3):425–464, 2001.

A. Kirsch. *An introduction to the mathematical theory of inverse problems*, volume 120. Springer Science & Business Media, 2011.

H. J. Kushner. A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise. 1964.

C. Z. Leng and M. D. Losego. Vapor phase infiltration (vpi) for transforming polymers into organic–inorganic hybrid materials: a critical review of current progress and future challenges. *Mater. Horiz.*, 4:747–771, 2017. doi: 10.1039/C7MH00196G. URL http://dx.doi.org/10.1039/C7MH00196G.

H. Liu, Y. Tang, and H. H. Zhang. A new chi-square approximation to the distribution of non-negative definite quadratic forms in non-central normal variables. *Computational Statistics & Data Analysis*, 53(4):853–856, 2009.

J. L. Loeppky, J. Sacks, and W. J. Welch. Choosing the sample size of a computer experiment: A practical guide. *Technometrics*, 51(4):366–376, 2009.

M. D. Losego and Q. Peng. *Atomic Layer Deposition and Vapor Phase Infiltration*, chapter 5, pages 135–159. John Wiley & Sons, Ltd, 2019. ISBN 9783527819249. doi: https://doi.org/10.1002/9783527819249.ch5. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/9783527819249.ch5.

S. Mak and V. R. Joseph. Support points. *The Annals of Statistics*, 46(6A): 2562–2592, 2018.

P. Marttinen, N. J. Croucher, M. U. Gutmann, J. Corander, and W. P. Hanage. Recombination produces coherent bacterial species clusters in both core and accessory genomes. *Microbial Genomics*, 1(5), 2015.

K. Matsui, S. Kusakawa, K. Ando, K. Kutsukake, T. Ujihara, and I. Takeuchi. Bayesian active learning for structured output design. *arXiv preprint arXiv:1911.03671*, 2019.

J. Mockus. Application of bayesian approach to numerical methods of global and stochastic optimization. *Journal of Global Optimization*, 4(4):347–365, 1994.

J. A. Nelder and R. Mead. A Simplex Method for Function Minimization. *The Computer Journal*, 7(4):308–313, 01 1965. ISSN 0010-4620. doi: 10.1093/comjnl/7.4.308. URL https://doi.org/10.1093/comjnl/7.4.308.

J. O. Ramsay. Functional data analysis. *Encyclopedia of Statistical Sciences*, 4, 2004.

J. O. Ramsay, S. Graves, and G. Hooker. *fda: Functional Data Analysis*, 2020. URL https://CRAN.R-project.org/package=fda. R package version 5.1.9.

Y. Ren, E. K. McGuinness, C. Huang, V. R. Joseph, R. P. Lively, and M. D. Losego. Reaction–diffusion transport model to predict precursor uptake and spatial distribution in vapor-phase infiltration processes. *Chemistry of Materials*, 33(13):5210–5222, 2021. doi: 10.1021/acs.chemmater.1c01283. URL https://doi.org/10.1021/acs.chemmater.1c01283.

C. Rusmassen and C. Williams. *Gaussian process for machine learning*. MIT Press, 2006.

T. J. Santner, B. J. Williams, and W. I. Notz. *The design and analysis of computer experiments*. Springer, 2018.

A. Subramanian, N. Tiwale, and C.-Y. Nam. Review of recent advances in applications of vapor-phase material infiltration based on atomic layer deposition. *JOM*, 71(1):185–196, 2019. doi: 10.1007/s11837-018-3141-4. URL https://doi.org/10.1007/s11837-018-3141-4.

M. H. Tan. Gaussian process modeling of a functional output with information from boundary and initial conditions and analytical approximations. *Technometrics*, 60(2):209–221, 2018.

A. K. Uhrenholt and B. S. Jensen. Efficient bayesian optimization for target vector estimation. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 2661–2670. PMLR, 2019.

R. Z. Waldman, D. J. Mandia, A. Yanguas-Gil, A. B. F. Martinson, J. W. Elam, and S. B. Darling. The chemical physics of sequential infiltration synthesis—a thermodynamic and kinetic perspective. *The Journal of Chemical Physics*, 151(19):190901, 2019. doi: 10.1063/1.5128108. URL https://doi.org/10.1063/1.5128108.

S. N. Wood. Statistical inference for noisy nonlinear ecological dynamic systems. *Nature*, 466(7310):1102–1104, 2010.

Y. Xiong, W. Chen, D. Apley, and X. Ding. A non-stationary covariance-based kriging method for metamodelling in engineering design. *International Journal for Numerical Methods in Engineering*, 71(6):733–756, 2007.

X. Xun, J. Cao, B. Mallick, A. Maity, and R. J. Carroll. Parameter estimation of partial differential equation models. *Journal of the American Statistical Association*, 108(503):1009–1020, 2013.

Y. Zhao, X. Huo, and Y. Mei. Identification of underlying dynamic system from noisy data with splines. *arXiv preprint arXiv:2103.10231*, 2021.