Published in final edited form as:

Sens Actuators B Chem. 2015 September ; 216: 321–331. doi:10.1016/j.snb.2015.03.071.

Gaussian process based modeling and experimental design for sensor calibration in drifting environments

Zongyu Geng^a, Feng Yang^{a,*}, Xi Chen^b, and Nianqiang Wu^c

^aIndustrial and Management Systems Engineering Department, West Virginia University, Morgantown, WV 26506, USA

^bDepartment of Industrial and Systems Engineering, Virginia Tech, Blacksburg, VA 24061, USA

^cDepartment of Mechanical and Aerospace Engineering, West Virginia University, Morgantown, WV 26506, USA

Abstract

It remains a challenge to accurately calibrate a sensor subject to environmental drift. The calibration task for such a sensor is to quantify the relationship between the sensor's response and its exposure condition, which is specified by not only the analyte concentration but also the environmental factors such as temperature and humidity. This work developed a Gaussian Process (GP)-based procedure for the efficient calibration of sensors in drifting environments. Adopted as the calibration model, GP is not only able to capture the possibly nonlinear relationship between the sensor responses and the various exposure-condition factors, but also able to provide valid statistical inference for uncertainty quantification of the target estimates (e.g., the estimated analyte concentration of an unknown environment). Built on GP's inference ability, an experimental design method was developed to achieve efficient sampling of calibration data in a batch sequential manner. The resulting calibration procedure, which integrates the GP-based modeling and experimental design, was applied on a simulated chemiresistor sensor to demonstrate its effectiveness and its efficiency over the traditional method.

Keywords

Sensor drift; Sensor calibration; Gaussian process model; Design of experiments; Bootstrapping

1. Introduction

Chemical sensors have been widely used in indoor and outdoor environment monitoring, vehicle exhaust measurement, human breath detection, etc [1–3]. It has been long recognized that the responses of chemical sensors, especially chemiresistors, are affected by the drift of environmental factors such as temperature and humidity [4–7]. To reduce detection errors and false alarm, it is important to accurately calibrate a sensor in a drifting environment, which primarily motivated this work. The environmental factors are denoted

*Corresponding author. Tel.: +1 304 293 9477; fax: +1 304 293 4970. feng.yang@mail.wvu.edu. .

as the vector $\mathbf{x}$, and the task of sensor calibration is to establish the functional dependence of the sensor response r upon the analyte concentration c as well as $\mathbf{x}$.

Quantifying the $c - \mathbf{x} - r$ relationship is challenging due to two main reasons: First, the variables (c , $\mathbf{x}$) may affect the response r in a nonlinear fashion and also interact nonlinearly with each other. The underlying mechanism is complicated [8–11] and difficult to be adequately captured by traditional regression analysis [6]. Second, to estimate a calibration model of high dimension, an extremely large sample size is typically required by the classic design of experiments (DOE) [6,12]. Thus, there is a need to develop new modeling and DOE methods for the efficient calibration of sensors subject to environmental drift.

While focusing on calibrating sensors with environmental drift, this work falls into the research efforts to calibrate sensors with general drifting behaviors, which can be classified into two categories [13,14]: external (i.e. environmental) and internal drifts. The latter is caused by the physical and/or chemical changes of the sensor itself, and examples of such changes include re-organization of the sensing materials and irreversible interaction with analytes. When calibrating drifting sensors, most of the literature used a reference-based linear compensation or linear regression to quantify the drifting effects [15–18]. Recognizing the possible nonlinear nature of sensor drifts, powerful nonlinear models have also been employed, such as neural network [6,19], kernel ridge regression [20] and nonlinear supporting vector machine [21]. However, in this stream of nonlinear modeling work, no effort was ever made to quantify the uncertainty of the target estimates (e.g., the analyte concentration estimated by the calibration model from an observed sensor response). This is at least partly due to the difficulties in deriving valid statistical inference (i.e., quantifying model uncertainty) based on those models [22,23]. It is known that statistical inference lays the basis for optimum DOE: Experiments are designed to minimize the uncertainty on the model estimates of interest [24–26]. Thus, optimum DOE is a research issue that has barely been touched in the nonlinear model-based sensor calibration.

In light of the discussions above, our objective is to develop a statistical procedure, which leads to a calibration model of the highest quality by using the least experimental effort. In this work, the calibration model assumes the form of a Gaussian process (GP), which is highly flexible and able to capture practically any continuous functional relationships. GP is chosen over other powerful nonlinear models because of its statistical inference capability [27], which allows for uncertainty quantification and provides the necessary basis for optimum DOE. For sensor calibration, the inference issues are further complicated by the coexistence of forward modeling and inverse estimation (as will become clearer in Section 2.1), and hence a GP-based bootstrap resampling method is developed in this work. The DOE is performed in a batch sequential manner to circumvent the dilemma that the optimum DOE depends on the true $c - \mathbf{x} - r$ relationship, which however, is unknown at the stage of designing experiments [28,25,24]. A learning process is allowed in such a sequential procedure: For the design of a new batch of experiments to be performed, all the information derived from the experimental data already collected is utilized to search for the optimum DOE of that new batch; and the DOE optimization seeks to minimize the calibration model uncertainty with a given batch size.

The remainder of the paper is organized as follows: Section 2 presents the formulation of the calibration model, which takes the form of a GP. The GP-based model fitting and statistical inference issues are discussed in Section 3. The batch sequential procedure for sensor calibration is described in Section 4. Section 5 is devoted to an empirical study to evaluate the effectiveness and efficiency of the calibration procedure. A brief summary is given in Section 6.

2. Formulation of calibration model

2.1. Calibration model

For a sensor exposed to drifting environments, its calibration model needs to functionally relate the sensor response r to the target analyte concentration c as well as the environmental factors $\mathbf{x}$. For notational convenience, all the exposure-condition factors are denoted as the vector $\mathbf{w} = (c, \mathbf{x}^\top)^\top$ of d dimension, with d being a positive integer. The sensor response can be generally written as

$$r(\mathbf{w}) = \mathbb{E}[r(\mathbf{w})] + \epsilon = F(\mathbf{w}) + \epsilon, \quad \mathbf{w} \in \mathcal{W} \quad (1)$$

where $F(\mathbf{w})$ quantifies the expected sensor response $\mathbb{E}[r(\mathbf{w})]$ as a function of $\mathbf{w}$. The feasible region of interest for exposure conditions is represented by $\mathcal{W}$. The sensor response r is subject to random error ϵ , which arises from instrument noise, variation of sensing materials, etc. The random error ϵ is assumed to be independent and identically distributed (i.i.d.). The exposure condition $\mathbf{w}$ is considered as deterministic, which complies with the actual experimental settings: In calibration experiments, $\mathbf{w}$ can be specified with extremely high accuracy and precision.

The calibration model fitted from experimental data is denoted as $\hat{F}(\mathbf{w})$, or equivalently $\hat{F}(c, \mathbf{x})$. In operational use, $\hat{F}(c, \mathbf{x})$ is coupled with the sensor device to quantify the target analyte. The observed sensor response r_0 and the observed values $\mathbf{x}_0$ for the environmental factors (e.g., temperature and humidity) are used to estimate the underlying analyte concentration c_0 as follows

$$\hat{c}_0 = \hat{F}^{-1}(r_0, \mathbf{x}_0). \quad (2)$$

In (2), $\hat{F}^{-1}$ denotes the inverse function of $\hat{F}$ with respect to (w.r.t.) c at given $\mathbf{x}_0$. This inverse function is assumed to exist in the region $\mathcal{W}$ considered, because otherwise the sensor would not be applicable at all.

Model (1) is also referred to as forward model, since it reflects the forward direction of calibration experiments: For a pre-specified $\mathbf{w}$, the response r is observed. Whereas, when the calibration model is coupled with the sensor device for monitoring, the inverse estimation of analyte concentration from observed r needs to be performed, as shown in (2).

2.2. Gaussian process model

As mentioned in Section 1, a GP is chosen in this work to model the functional dependence of r upon $\mathbf{w}$, and the GP calibration model (1) is represented as

$$r = F(\mathbf{w}) + \epsilon = \mu + M(\mathbf{w}) + \epsilon, \quad (3)$$

where μ is the mean parameter, and $M(\mathbf{w})$ is a realization of a mean-zero stationary GP with constant variance σ^2 . The random error ϵ follows a continuous probability distribution (e.g., normal). In this work, it is assumed that the variance of ϵ is a constant σ^2 , which is a typical assumption for sensor calibration [29,30]. However, it is worthy of noting that the GP model can be extended to accommodate general variance structures of the errors, as in Ankenman et al. [31].

The GP $M(\cdot)$ is characterized by its correlation function [32], which is denoted as

$\text{Corr}(M(\mathbf{w}), M(\mathbf{w}'))$, where $\mathbf{w} = (w_1, w_2, \dots, w_d)^\top$ and $\mathbf{w}' = (w'_1, w'_2, \dots, w'_d)^\top$ represent two distinct exposure conditions. In this work, we adopted the widely-used squared exponential form [27] for the correlation function:

$$\text{Corr}(M(\mathbf{w}), M(\mathbf{w}')) = K(\mathbf{w}, \mathbf{w}') = \exp \left\{ \sum_{h=1}^d -\theta_h (w_h - w'_h)^2 \right\}, \quad (4)$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_d)$ is a vector of unknown parameters to be estimated from the experimental data. It is required that $\theta_h > 0$, and a smaller value of θ_h tends to provide a smoother response surface. Other choices of the correlation forms can be found in the literature [24,27].

For a collection of I distinct exposure conditions $\{\mathbf{w}_i; i = 1, 2, \dots, I\}$ with I being a positive integer, an $I \times I$ correlation matrix $\mathbf{R}(\boldsymbol{\theta})$ is defined as:

$$\mathbf{R}(\boldsymbol{\theta}) = \begin{pmatrix} 1 & K(\mathbf{w}_1, \mathbf{w}_2) & \dots & K(\mathbf{w}_1, \mathbf{w}_I) \\ K(\mathbf{w}_2, \mathbf{w}_1) & 1 & \dots & K(\mathbf{w}_2, \mathbf{w}_I) \\ \vdots & \vdots & \ddots & \vdots \\ K(\mathbf{w}_I, \mathbf{w}_1) & K(\mathbf{w}_I, \mathbf{w}_2) & \dots & 1 \end{pmatrix}. \quad (5)$$

Along with an arbitrary condition $\mathbf{w}_0$, we further define the following $I \times 1$ vector

$$\mathbf{v}(\mathbf{w}_0; \boldsymbol{\theta}) = \begin{pmatrix} K(\mathbf{w}_0, \mathbf{w}_1) \\ K(\mathbf{w}_0, \mathbf{w}_2) \\ \vdots \\ K(\mathbf{w}_0, \mathbf{w}_I) \end{pmatrix}. \quad (6)$$

3. Model estimation and inferences

3.1. Experimental data for sensor calibration

To calibrate a sensor, experimental data has to be collected at a range of exposure conditions. The calibration sample data can be represented as

$$\{(\mathbf{w}_i, r_j(\mathbf{w}_i)); i=1, 2, \dots, I, j=1, 2, \dots, n\}. \quad (7)$$

In (7), $\mathbf{w}_i$ denotes the i th design point (exposure condition at which experiments are performed) out of a total of I distinct design points; $r_j(\mathbf{w}_i)$ denotes the observed response from the j th replication at $\mathbf{w}_i$, and n is the number of replications at each design point. The sample average at $\mathbf{w}_i$ can then be calculated as

$$\bar{r}(\mathbf{w}_i) = \frac{1}{n} \sum_{j=1}^n r_j(\mathbf{w}_i), \quad (8)$$

and

$$\bar{\mathbf{r}} = \left(\bar{r}(\mathbf{w}_1), \bar{r}(\mathbf{w}_2), \dots, \bar{r}(\mathbf{w}_I) \right)^T \quad (9)$$

is the $I \times 1$ vector of sample averages at the I distinct design points.

3.2. Model fitting

Given the experimental data (7), in which the sensor responses are subject to random errors, the method developed by Ankeman et al. [31] is employed to fit the GP calibration model (3). The fitting steps [31] are adapted and given in Fig. 1 for readers' convenience. For the GP model (3), the unknown parameters to be fitted from the sample data include: σ^2 , the variance of the random error ϵ ; μ , the mean of the GP $M(\cdot)$; τ^2 , the variance of $M(\cdot)$; and $\theta = (\theta_1, \theta_2, \dots, \theta_d)$, the parameters in the correlation function (4). To solve the optimization problem in Step (iii) of Fig. 1, the *fmincon* function in Matlab was adopted for the empirical case in Section 5.

With the fitted parameters $(\hat{\sigma}^2, \hat{\mu}, \hat{\tau}^2, \hat{\theta})$ for the GP calibration model, the expected sensor response at an arbitrary exposure condition $\mathbf{w}_0$ can be estimated as [31]:

$$\hat{r}(\mathbf{w}_0) = \hat{F}(\mathbf{w}_0) = \hat{\mu} + \hat{\tau}^2 \mathbf{v}(\mathbf{w}_0; \hat{\theta}) \left[\hat{\tau}^2 \mathbf{R}(\hat{\theta}) + \frac{\hat{\sigma}^2}{n} \mathcal{I}_I \right]^{-1} (\bar{\mathbf{r}} - \mu \mathbf{1}_I), \quad (12)$$

with $\mathcal{I}_I$ being the identity matrix of size I , and $\bar{\mathbf{r}}$ the sampling averages (9).

As explained in Section 2.1, the calibration model $\hat{F}$ can be inverted to estimate analyte concentrations for observed sensor responses and environmental factors in operational use: $\hat{c}_0 = \hat{F}^{-1}(r_0, \mathbf{x}_0)$. In this work, this GP-based inverse computation is performed numerically through the cubic spline interpolation [33].

3.3. Statistical inference issues

As noted above, the fitted GP calibration model $\hat{F}$ allows for the point estimation in both forward and inverse directions: (i) for any $\mathbf{w}_0 = (c_0, \mathbf{x}_0^\top)^\top$, the expected sensor response can be estimated as $\hat{r}(\mathbf{w}_0)$ in (12); (ii) for observed r_0 and $\mathbf{x}_0$, the estimated analyte concentration can be obtained as $\hat{c}_0$ from the GP-based inverse computation. In this subsection, we discuss the uncertainty of these point estimates.

What directly concerns end users is $Var[\hat{c}_0]$, which measures the uncertainty of $\hat{c}_0$, the point estimate of the analyte concentration in an environment of interest. Since $\hat{c}_0 = \hat{F}^{-1}(r_0, \mathbf{x}_0)$, $Var[\hat{c}_0]$ stems from two sources. The first source lies in the random error to which the sensor response r_0 is subject. Since $\hat{c}_0$ is a function of r_0 , $Var[\hat{c}_0]$ is affected by $Var[r_0]$. The variability $Var[r_0]$ is inherent of the sensor device, and the control of it is outside of the scope of this modeling work; The second source comes from the uncertainty of the model $\hat{F}$, which is fitted from the calibration data randomly sampled at selected design points (exposure conditions). The uncertainty of $\hat{F}$ is determined by the design of calibration experiments, which specifies the location of design points and the number of replications assigned to each point. It is worthy of noting that the observed $\mathbf{x}_0$ is considered error-free, since the environmental factors such as temperature can be accurately measured.

Hence, in this work we focus on the uncertainty of the model $\hat{F}$ and $Var[\hat{c}_0]$, which are discussed in details below.

3.3.1. Quantifying the model uncertainty—As recognized in the literature [31,34], the uncertainty of a fitted GP model $\hat{F}$ is well embodied by $\{Var[\hat{r}(\mathbf{w}_0)]; \mathbf{w}_0 \in \mathcal{W}\}$, the variability of the response estimates (12) over the feasible region $\mathcal{W}$.

Given the sample data (7), it has been derived that [31]

$$\widehat{Var}[\hat{r}(\mathbf{w}_0)] = \hat{\tau}^2 - \hat{\tau}^4 \mathbf{v}(\mathbf{w}_0, \hat{\theta})^\top \left[\hat{\tau}^2 \mathbf{R}(\hat{\theta}) + \frac{\hat{\sigma}^2}{n} \mathcal{J}_I \right]^{-1} \mathbf{v}(\mathbf{w}_0, \hat{\theta}) + \eta^\top \eta \left(\mathbf{1}_I^\top \left[\hat{\tau}^2 \mathbf{R}(\hat{\theta}) + \frac{\hat{\sigma}^2}{n} \mathcal{J}_I \right]^{-1} \mathbf{1}_I \right)^{-1} \quad (13)$$

$$\text{where } \eta = \mathbf{1} - \mathbf{1}_I^\top \left[\hat{\tau}^2 \mathbf{R}(\hat{\theta}) + \frac{\hat{\sigma}^2}{n} \mathcal{J}_I \right]^{-1} \mathbf{v}(\mathbf{w}_0, \hat{\theta}) \hat{\tau}^2.$$

The model uncertainty, which is measured by $\{Var[\hat{r}(\mathbf{w}_0)]; \mathbf{w}_0 \in \mathcal{W}\}$, is used to drive the DOE for sensor calibration (as will be seen in Section 4.1.2), since it is determined by the design of calibration data.

3.3.2. Quantifying the uncertainty of concentration estimates—Through the inverse computation based on the fitted GP, the concentration estimate $\hat{c}_0 = \hat{F}^{-1}(r_0, \mathbf{x}_0)$ can be obtained, and its uncertainty has a clear interpretation to end users. Analytical expressions, such as the one in (13), are not available for the uncertainty of $\hat{c}_0$ [30,35]. Hence, we adapted a bootstrapping method [36–40] to evaluate $Var[\hat{c}_0]$ through resampling.

Given the real calibration data collected $\{(\mathbf{w}_i, r_j(\mathbf{w}_i)) ; i = 1, 2, \dots, I, j = 1, 2, \dots, n\}$, we denote e_{ij} as the j th residual at the design point $\mathbf{w}_i$, which is calculated as

$$e_{ij} = r_j(\mathbf{w}_i) - \hat{r}(\mathbf{w}_i). \quad (14)$$

Under the assumption that the error terms in the calibration model (1) are i.i.d., the pool of residuals $\{e_{ij} ; i = 1, 2, \dots, I, j = 1, 2, \dots, n\}$ are employed in the bootstrap algorithm (Fig. 2). Bootstrap resampling is performed in both Step 1 and 3 of this algorithm, following the forward direction as the real sampling does. The fitted GP calibration model $\hat{f}$ and the resampled residuals are used to simulate sensor responses mimicking real responses. The resampling in Step 1 accounts for the uncertainty of the fitted GP model, while the resampling in Step 3 simulates the randomness of the observed sensor response. Following the recommendations in the bootstrapping literature [41, Chapter 17, p. 572] [36, Chapter 6, p. 50], we set the resampling size B_1 as 50 in Step 1, and B_2 as 25 in Step 3. Both sources of uncertainty are reflected in the bootstrapping outputs

$\{\hat{c}_{0,b,k}^* ; b=1, 2, \dots, B_1, k=1, 2, \dots, B_2\}$, which can be used to estimate $Var[\hat{c}_0]$ as shown in (15), Fig. 2.

The statistical inference issues discussed in this section provide the necessary basis for the DOE procedure, as will be seen in Section 4.

4. The batch sequential procedure

There are two questions that a general DOE typically addresses: (i) At what design points should samples be allocated? (ii) At each design point, how many replications should be assigned? For the physical and/or chemical experiments involved in sensor calibration, it is a common practice to have a predetermined and fixed number of replications assigned to a design point, each time that exposure condition is selected; this is to ensure the reliability of the response measurements which are subject to random errors. Hence, Question (ii) can be answered based on the random variability of the sensor device, as in the experimental practice for calibrating a sensor. Usually, the number of replications n assigned to each design point ranges from 3 to 8, depending on the variability of sensor responses.

Hence, in our DOE for sensor calibration, the focus is on addressing Question (i). The answer to this question aims at using the least experimental effort to achieve the calibration model of the highest quality. In the existing literature [42,43], classic DOE methods such as factorial designs [12, Chapter 5, p. 170] were adopted to determine the locations of the design points in the high-dimensional space spanned by exposure-condition factors (i.e., the analyte concentration and environmental factors). These classic designs require the selection of several (typically, 3–5 evenly-spaced) levels for each factor, assign all the samples once and for all, typically lead to a large number of design points in the high-dimensional space, and are usually not optimal for nonlinear calibration relationships. For nonlinear calibration, the optimum DOE depends on the underlying relationship, which is unknown when designing experiments. To circumvent this dilemma, a sequential framework is adopted in the general DOE literature [34,44,45][25, Chap. 17, p. 257]; such a framework allows for a

learning process, and typically leads to savings in sample size, time and cost. To the best of our knowledge, the recent work by Rodríguez-Luján et al. [45] represents the existing effort of a batch sequential DOE procedure for calibrating nonlinear sensor arrays. Their DOE is performed with the support vector machine being the calibration model for analyte classification. In this work, our batch sequential DOE procedure is based on using Gaussian process as the calibration model for analyte quantification, and can be considered as a parallel to that in Rodríguez-Luján et al. [45].

The inputs for the batch sequential procedure are listed in Fig. 3. The pre-specified upper limit $\mathcal{P}$ is the target precision level for the estimated analyte concentrations. More specifically, the sequential procedure is expected to lead to a fitted calibration model $\hat{F}(c, \mathbf{x})$ that is able to achieve the following goal when coupled with the sensor device for monitoring. For an unknown environment with observed measurements $(r_0, \mathbf{x}_0)$, the calibration model $\hat{F}(c, \mathbf{x})$ can provide two estimates to the sensor user:

- the point estimate of the analyte concentration $\hat{c}_0 = \hat{F}^{-1}(r_0, \mathbf{x}_0)$;
- the interval estimate for the underlying analyte concentration

$$\left[\hat{c}_0 - z_{1-\alpha/2} \cdot \delta, \hat{c}_0 + z_{1-\alpha/2} \cdot \delta \right], \quad (16)$$

where $1 - \alpha$ is the probability level, and $z_{1-\alpha/2}$ is the $100(1 - \alpha/2)$ th percentile of the standard normal distribution. For $\alpha = 0.05$, $z_{1-\alpha/2} = 1.96$. The validity of the interval (16) relies on approximately normal assumption, which appears to hold based on our empirical experience. The interval (16) can be interpreted as follows: The true analyte concentration c_0 of the environment falls within this interval with a chance of $100(1 - \alpha) \%$.

The procedure is outlined in Fig. 4, which is initiated by a batch of samples allocated to D_0 design points. Based on the data collected so far, the GP calibration model is fitted and evaluated. If the stopping criteria are not met, then a new stage/batch of experiments are called for; the design of the new batch takes advantage of the information derived from all the experimental data already obtained, and seeks to minimize the uncertainty of the resulting GP calibration model (Section 3.3.1). To the newly expanded data, the GP model is fitted and evaluated again, and such a process is repeated until experimental budget N_{max} has been exhausted or the desired precision level $\mathcal{P}$ has been achieved for the estimated analyte concentrations. (Recall that the uncertainty of concentration estimates is discussed in Section 3.3.2.)

4.1. Design of experiments

As explained earlier, the task of DOE is to find the locations of the design points at which calibration experiments are to be performed. All the design points are required to fall within the feasible region $\mathcal{W}$, and $\mathcal{W}$ is typically a hypercube specified by the feasible ranges of the exposure-condition factors $\mathbf{w} = (c, \mathbf{x}^\top)^\top$.

For evaluation purpose (as will be seen in Sections 4.1.2 and 4.2), we further define a set of grid points $\mathcal{W}_{Grid}$ by dividing $\mathcal{W}$ into a fine grid of equally-sized hyper-rectangles. The set of grids is denoted as

$$\mathcal{W}_{Grid} = \left\{ \mathbf{w}_g^{(Grid)}; g=1, 2, \dots, G \right\}, \quad (17)$$

with G being the total number of grid points.

4.1.1. Initial design of calibration experiments—At Stage $s = 0$, the design for the initial batch of experiments includes D_0 distinct design points. In sensor calibration, it is a common practice to assign $n = 3$ replications to each design point. The purpose of the initial design is to provide a decent coverage of the region $\mathcal{W}$, so that the initial sample data allows for a sketching of the target response surface over $\mathcal{W}$; such information will be used to guide the DOE in the next stage. In light of these, the design points in Stage 0 are selected out of the following considerations: First, to allow for an adequate coverage of $\mathcal{W}$, the number of design points D_0 is chosen to be greater than five times the dimension of $\mathcal{W}$. This many design points also provide the necessary data base for the GP modeling to sketch out the response surface [46,47]. Second, all the vertex points of the hypercube $\mathcal{W}$ are included to avoid extrapolation, which may cause severe estimation biases [26,27]. Third, the non-vertex design points are determined by employing space filling designs such as Latin hypercube sampling (LHS) [48], which seeks to evenly distribute the design points in $\mathcal{W}$. Lastly, with the experimental budget N_{max} , we set $n \times D_0$ as $1/4 - 1/3$ of N_{max} , following the general guidelines in Santner et al. [24].

4.1.2. Design augmentation in subsequent stages—The DOE for a subsequent stage s ($s \geq 1$) is performed based on all the calibration data collected in the previous stages. The data set already obtained is denoted as $\{(\mathbf{w}_i, r_j(\mathbf{w}_i)); i = 1, 2, \dots, D, j = 1, 2, \dots, n\}$, with a total of $D = D_0 + D_1 + \dots + D_{s-1}$ design points. Given the information derived via the GP modeling of the existing data, the design augmentation at Stage s aims at finding the best locations of the D_s additional design points so that the newly expanded data lead to GP calibration model of the highest quality. Hence, such a follow-up DOE is an optimization problem: Optimizing the quality of the resulting GP model w.r.t. the design points to be added.

In the literature, some research efforts have been devoted to batch sequential designs for GP modeling [49,32,50–53]. For the optimization objective in those works, a range of design criteria were explored, which include DISTANCE [50], ENTROPY [49], IMSE, MAX MSE [32], etc. As articulated in Section 3.3, the DOE determines the uncertainty of the fitted GP model, through which affects the uncertainty of the estimated analyte concentration (Section 3.3.2) that primarily concerns sensor users. Thus, for our sensor calibration, the uncertainty of the fitted GP is adopted, and used to derive the MAX MSE criterion. The details follow.

The design optimization at Stage s is formulated as

$$\begin{aligned}
 & \min_{\{\mathbf{w}_{D+1}, \mathbf{w}_{D+2}, \dots, \mathbf{w}_{D+D_s}\}} \max_{\mathbf{w} \in \mathcal{W}_{Grid}} \widehat{Var}[\hat{r}(\mathbf{w})] \\
 & \text{Subject to } D = D_0 + D_1 + \dots + D_{s-1} \text{ design points have already} \\
 & \text{been sampled as } \{(\mathbf{w}_i, r_j(\mathbf{w}_i)); i=1, 2, \dots, D, j=1, 2, \dots, n\}.
 \end{aligned} \tag{18}$$

The objective function $\max_{\mathbf{w} \in \mathcal{W}_{Grid}} \widehat{Var}[\hat{r}(\mathbf{w})]$ represents the maximum model (prediction) uncertainty among the exposure conditions included in the set of dense grids $\mathcal{W}_{Grid}$. Recall Eq. (13) for the quantification of $\widehat{Var}[\hat{r}(\mathbf{w})]$ at any values of $\mathbf{w}$. The prediction uncertainty $\widehat{Var}[\hat{r}(\mathbf{w})]$ is also referred to as MSE (mean square error), and the objective is to minimize the MAX MSE w.r.t. the Stage- s design.

To solve the design problem (18), an optimization algorithm such as Tabu search (TS) or genetic algorithm (GA) is needed to generate design candidates in search of the best/good solutions. In this work, the *ga* function in Matlab, which implements a genetic algorithm, is adopted for the optimization search. Readers can refer to Cormen et al. [54] for guides regarding the selection of suitable algorithms. As the search moves along, each candidate design is evaluated and compared with other candidates in terms of its objective value. The evaluation of a candidate design, denoted as $\{\mathbf{w}_{D+1}, \mathbf{w}_{D+2}, \dots, \mathbf{w}_{D+D_s}\}$, is performed as in Fig. 5.

4.2. Model evaluation and stopping criteria

At the end of Stage s ($s = 0, 1, 2, \dots$), calibration data have been sampled at a total of $D = D_0 + D_1 + \dots + D_s$ design points, and the data is denoted as $\{(\mathbf{w}_i, r_j(\mathbf{w}_i)); i = 1, 2, \dots, D, j = 1, 2, \dots, n\}$. From the collected data, the GP model is fitted as $\hat{F}(\mathbf{w})$ and the residuals calculated by using (14).

Is the current model $\hat{F}(\mathbf{w})$ able to provide concentration estimates of desired precision? The procedure given in Fig. 6 determines whether or not to proceed with the next-batch experimentation by evaluating the estimation uncertainty of analyte concentrations. As shown in Fig. 6, the uncertainty is evaluated for the estimated concentrations over the wide range of exposure conditions represented by the grid points in $\mathcal{W}_{Grid}$. The experimentation is terminated only when the desired precision has been achieved for all the grid points, or the experimental budget has been used up.

5. Empirical results

The GP-based procedure was applied to calibrate a chemiresistor sensor, whose response can be substantially drifted by the fluctuations in humidity and temperature. The effectiveness of the calibration procedure is illustrated, and its efficiency over the traditional DOE method is demonstrated.

5.1. The simulated chemiresistor sensor

The evaluation of a statistical method such as our calibration procedure requires an extremely large amount of validation data (as will be exemplified in Sections 5.3 and 5.4), and is usually performed based on simulation, as opposed to real experiments, which could be prohibitively expensive. Built from real data, simulation models are able to capture any important features of real data [55], and at the same time can be used to generate data via computer experiments. Hence, in this work, a simulated chemiresistor was developed to evaluate the proposed procedure.

The simulation model is derived from the experimental data of a real chemiresistor given in Hossein-Babaei and Ghafarinia [6]. The chemiresistor sensor's expected response, which is denoted as $E[r(\mathbf{w})] = E[r(c, x_1, x_2)]$, depends on the analyte (methanol) concentration c as well as the environmental factors $\mathbf{x} = (x_1, x_2)^\top$: the ambient humidity x_1 and temperature x_2 . The functional expression of $E[r(c, x_1, x_2)]$ is given in Eqs. (A.1), A; it takes the form of a single-layer neural network, and is estimated from the real calibration data provided by Hossein-Babaei and Ghafarinia [6]. Fig. 7 gives a graphic illustration of the expected response curves generated by the neural network model; each graph in the figure plots the expected sensor response against the analyte concentration at an example setting of the environmental factors $\mathbf{x} = (x_1, x_2)^\top$.

At any $\mathbf{w}$, the sensor response $r(\mathbf{w})$ is simulated as follows to mimic real sensor signals subject to random errors:

$$r(\mathbf{w}) = E[r(\mathbf{w})] + \epsilon \quad (19)$$

where ϵ is the random error that follows normal distribution with mean zero and standard deviation 0.1. The simulation model (19) was used to generate data as needed for the empirical evaluations in Sections 5.2–5.4.

The simulation model serves two purposes in the remainder of this section: (i) to generate calibration data, and (ii) to provide the true benchmark for evaluating the calibration models fitted from the data.

5.2. Calibrating the simulated sensor

To apply the batch sequential procedure (Fig. 4) for the calibration of the chemiresistor, the inputs for the procedure (Fig. 3) are specified as follows.

- $\mathcal{W}$ is defined by the following three ranges: the methanol concentrations $c \in (100, 600)$ ppm, the temperature $x_1 \in (25, 40)$ °C, and the absolute humidity $x_2 \in (0.5\%, 5.0\%)$.
- The desired precision $\delta^2 = 0.05^2$.
- $D_s = 15$ ($s = 0, 1, 2, \dots$), that is, the number of distinct design points in each batch of data is set as 15.
- The number of replications $n = 5$ for a selected design point.

- The experimental budget $N_{max} = \infty$, which means that the procedure will terminate only if the desired precision is achieved.

To facilitate the GP modeling [27], the three factors (c, x_1, x_2) are all re-scaled to be within $[0, 1]$. To obtain the grid points, the scaled range of each factor is discretized with a step size of 0.1, leading to $G = 11^3 = 1331$ grid points $\{\mathbf{w}_g^{(Grid)}; g=1, 2, \dots, G\}$.

Following the initial design strategy (Section 4.1.1), the $D_0 = 15$ point design includes the eight vertices of the hypercube $\mathcal{W}$, and seven inner points provided by the LHS design. The *lhsdesign* function in Matlab is used to perform the LHS design. At each design point, $n = 5$ replications of simulation experiments were carried out based on Model (19). The Stage-0 data set is denoted as $\{(\mathbf{w}_i, r_j(\mathbf{w}_i)); i = 1, 2, \dots, 15, j = 1, 2, \dots, 5\}$. Based on the initial data, the GP fitting (Fig. 1) was performed. The quality of the fitted model was evaluated following the procedure in Fig. 5, and a further stage of experiments is needed.

At Stage 1, given the Stage-0 data and the fitted GP model, the selection of the additional $D_1 = 15$ design points was performed following the design augmentation method (Section 4.1.2). After sampling at these newly-added design points, the GP fitting was carried out again based on all the samples obtained, and the evaluation results of the updated GP model suggest that the desired precision $\hat{\sigma}^2 = 0.05^2$ has not been satisfied yet. Each follow-up stage was performed in the same manner as Stage 1, and the procedure terminated at the end of Stage 2. The final fitted calibration model $\hat{F}(\mathbf{w})$ is expected to be able to estimate the analyte concentrations with the desired precision, and the fitted GP parameters are obtained as:

$$\begin{aligned} \hat{\sigma}^2 &= 0.010; & \hat{\mu} &= 5.292; \\ \hat{\tau}^2 &= 27.316; & (\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3) &= (0.995, 0.101, 0.032). \end{aligned} \quad (20)$$

The three batches of calibration data are given in Table Table B.1, Appendix B. The fitted calibration model specified by the parameters (20) can be used to predict the sensor response at an arbitrary setting of the analyte concentration and environmental variables. The quality of this forward prediction provided by the fitted model is evaluated in Appendix C.

5.3. Effectiveness evaluation of the calibration procedure

The calibration procedure aims at delivering a calibration model of the desired quality, which is specified by the precision level $\hat{\sigma}^2$ of the analyte concentration estimates provided by the calibration model. To evaluate if the procedure is able to achieve this goal, the following steps are taken to perform a total of $M = 100$ macro-replications, which involves 100 independent applications of the procedure.

- *Step 0.* Set the total number of macro-replications $M = 100$; initialize the index $m = 1$.
- *Step 1.* Apply the calibration procedure as in Section 4, and obtain the fitted GP calibration model $\hat{F}(c, \mathbf{x}_0)$.

- *Step 2.* Generate validation data sets (VDS) by running simulation experiments, which are independent of and aside from those for the calibration data in Step 1. Specifically, at each grid point $\mathbf{w}_g^{(Grid)} \in \mathcal{W}_{Grid}$, $J=5$ replications are carried out leading to the VDS represented as:

$$\{(c_{mgj}, \mathbf{x}_{mgj}, r_{mgj}); g=1, 2, \dots, G, j=1, 2, \dots, J\}. \quad (21)$$

Suppose that a pair of observations $(r_{mgj}, \mathbf{x}_{mgj})$ has been obtained for an environment of interest. The analyte concentration is estimated as

$\hat{c}_{mgj} = \hat{F}^{-1}(r_{mgj}, \mathbf{x}_{mgj})$, and the interval estimate is obtained as $\hat{c}_{mgj} \pm 2\delta$ with a confidence level of 95%.

- *Step 3.* Evaluate the point estimate $\hat{c}_{mgj}$ by calculating the deviations of the estimates from their true values:

$$\{\hat{c}_{mgj} - c_{mgj}; g=1, 2, \dots, G, j=1, 2, \dots, J\}. \quad (22)$$

Evaluate the interval estimates (16) by calculating

$$\Lambda_{mg} = \sum_{j=1,2,\dots,J} \text{IF}(c_{mgj} \in [\hat{c}_{mgj} \pm 2\delta]). \quad (23)$$

In (23), the IF function returns a value of 1 if the condition within the parentheses holds, and it returns 0 otherwise. Hence, (23) counts the number of intervals that include the true concentration values at the g th grid from the m th macro-replications.

- *Step 4.* If $m = M$, then STOP; otherwise set $m = m + 1$, repeat Steps 1–3.

The deviations of the concentration estimates from their true values (22) are collected for all the macro-replications, and plotted in Fig. 8. 98.95% of the deviations are within 2δ , which suggests that the point estimates of the analyte concentrations are within the desired distance to their true values.

Let $\Lambda_g = \sum_{m=1}^M \Lambda_{mg}$, then each element in

$$\{\Lambda_g / (MJ); g=1, 2, \dots, G\} \quad (24)$$

represents the estimated coverage probability (i.e., the percentage of interval estimates that cover the true concentrations) at the g th grid point. Among the G coverage probabilities (24), the minimum, which is denoted as $\min \{\Lambda_g / (MJ); g = 1, 2, \dots, G\}$, is 0.95; and at 75% of the grids, the corresponding coverage probability falls between 0.98 and 1. These statistics suggest that the interval estimates of analyte concentrations provided by our procedure are able to achieve the target nominal level 0.95 across all the grids (the range of exposure conditions considered).

5.4. Efficiency evaluation of the procedure

In this part, we compare the efficiency of our batch sequential procedure (SP) with that of a full factorial design (FD) [12], which determines the experimental conditions all at once.

The same experimental effort used in SP is assigned to the FD method. Specifically, $45 = 5 \times 3 \times 3$ distinct design points is used in FD, with 5 evenly-spaced levels for the analyte concentration and 3 for each of the two environmental factors; 5 replications are allocated to each design point. For FD, $M = 100$ macro-replications can also be performed following the steps described in the previous subsection, except that Step 1 needs to be replaced by applying FD to obtain the fitted calibration model.

For each method, the macro-replications allow for the computation of the following two performance measures: the root mean square error (RMSE)

$$RMSE_g = \sqrt{\frac{1}{MJ} \sum_{m=1}^M \sum_{j=1}^J (\hat{c}_{mgj} - c_{mgj})^2} \quad g=1, 2, \dots, G, \quad (25)$$

and the absolute value of the maximum prediction error (AME),

$$AME_g = \max_{\substack{m=1, 2, \dots, M \\ j=1, 2, \dots, J}} |\hat{c}_{mgj}^{(Grid)} - c_{mgj}^{(Grid)}| \quad g=1, 2, \dots, G. \quad (26)$$

SP and FD are compared in terms of RMSE and AME in Fig. 9(a) and (b), respectively. Each box in Fig. 9 corresponds to a method, and is generated based on one of the two performance metrics associated with that method. As can be seen from the box plots, the boxes for SP are narrower and closer to zero compared to those of FD.

6. Summary

To efficiently calibrate sensors in drifting environments, a Gaussian process (GP)-based batch sequential procedure is developed. A GP form is employed for the calibration model, and used to capture the possibly comprehensive and nonlinear relationship between the sensor's response and its exposure condition including the target analyte concentrations as well as other environmental factors. Based on GP modeling, a bootstrap resampling method is developed to quantify the uncertainty of the calibration model fitted from experimental data. Such uncertainty quantification serves as the necessary basis of optimizing the design of experiments (DOE) for sensor calibration. The GP-based modeling, uncertainty quantification and design optimization are embedded in each stage of the sequential procedure to fit the calibration model from the collected data, and to guide the DOE of the new batch of experiments. The procedure is terminated when the estimation uncertainty is within the pre-specified level or the total experimental budget has been spent. The effectiveness of the procedure and its efficiency over classic DOE have been demonstrated through simulation studies.

Although primarily motivated by sensor calibration with environmental/external drifts, the GP-based methods can also be extended to calibrate sensors with both external and internal drifts, which is part of our on-going work.

Acknowledgments

Research reported in this publication was partially supported by the National Science Foundation under Award Number CMMI-1068131 and by the National Institute Of Neurological Disorders And Stroke of the National Institutes of Health (NIH) under Award Number R15NS087515. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF or NIH. We thank Drs. Hossein-Babaei and Ghafarinia for providing the original experimental data on the chemiresistor sensor.

Appendix A: Simulation model for the chemiresistor

The chemiresistor sensor's expected response $E[r(\mathbf{w})]$ is simulated via a single-layer neural network, which take inputs as the methanol concentrations c , the temperature x_1 and the absolute humidity x_2 (all factors rescaled to $[0, 1]$). The neural network model is estimated from the real calibration data provided by Hossein-Babaei and Ghafarinia [6], and given as follows.

$$\begin{aligned}
 E[r(\mathbf{w})] &= E[r(c, x_1, x_2)] \\
 &= 47.6453 + \frac{0.0009}{1 + \exp(71.01 - 32.31c - 62.50x_1 - 9.61x_2)} \\
 &\quad - \frac{22.9161}{1 + \exp(-2.58 - 10.54c - 1.32x_1 - 0.13x_2)} \\
 &\quad + \frac{5.9423}{1 + \exp(1.54 + 0.60c + 0.83x_1 + 0.36x_2)} \\
 &\quad - \frac{24.4458}{1 + \exp(-2.35 - 14.71c - 15.67x_1 - 1.59x_2)}
 \end{aligned} \tag{A.1}$$

Appendix B: Calibration data for the case study

The sample size for the final calibration data set is 225, allocated to 45 distinct design points as shown in Table Table B.1. The column $\bar{r}$ represents the sample average of 5 replications at each design point. A graphic illustration of the three-stage design points is also given in Fig. B.1: The Stage-0 points are depicted as “•”, Stage-1 as “.”, and Stage-2 as “▼”. The initial (Stage-0) design points seek to provide a decent coverage in the design space (Section 4.1.1). The subsequential (Stage-1 and Stage-2) designs are driven by the prediction uncertainty of sensor responses, and depend on the existing samples as well as the calibration model estimated from the existing samples (Section 4.1.2). In this case, the uncertainty in sensor responses is mainly attributed by analyte concentrations; hence, as can be seen from Fig. B.1, the distribution of the design points is not even across the three dimensions: loose on temperature and humidity, but intense on concentration.

Table B.1

Sampling data for calibrating the chemiresistor.

Stage 0		Stage 1		Stage 2	
(c, x_1, x_2)	$\bar{r}$	(c, x_1, x_2)	$\bar{r}$	(c, x_1, x_2)	$\bar{r}$
(0.4, 0.3, 0.3)	1.101	(0.7, 0.0, 0.0)	0.658	(0.0, 0.5, 1.0)	2.346
(0.5, 0.2, 0.6)	0.811	(0.4, 1.0, 1.0)	0.620	(1.0, 0.5, 1.0)	0.172
(0.6, 0.6, 0.4)	0.509	(0.5, 1.0, 0.0)	0.588	(0.0, 0.5, 0.0)	3.451
(0.3, 0.4, 0.9)	1.104	(0.7, 0.0, 1.0)	0.427	(0.8, 0.4, 0.0)	0.423
(0.9, 0.5, 0.8)	0.203	(0.3, 0.0, 0.0)	2.076	(0.6, 0.5, 1.0)	0.422
(0.2, 0.7, 0.2)	1.648	(0.7, 1.0, 1.0)	0.285	(0.1, 0.0, 0.4)	3.106
(0.7, 0.8, 0.5)	0.384	(0.2, 1.0, 1.0)	1.082	(0.2, 1.0, 0.0)	1.455
(0.0, 0.0, 0.0)	4.982	(0.2, 0.0, 1.0)	1.739	(0.5, 0.0, 0.0)	1.126
(0.0, 0.0, 1.0)	3.278	(0.8, 1.0, 0.0)	0.423	(0.0, 1.0, 0.5)	2.199
(0.0, 1.0, 0.0)	2.610	(0.0, 0.5, 0.5)	2.848	(0.8, 1.0, 1.0)	0.165
(0.0, 1.0, 1.0)	1.874	(0.9, 0.0, 0.5)	0.353	(1.0, 0.0, 0.1)	0.347
(1.0, 0.0, 0.0)	0.293	(0.2, 1.0, 0.4)	1.261	(0.2, 0.4, 0.0)	2.103
(1.0, 0.0, 1.0)	0.127	(1.0, 0.5, 0.0)	0.376	(0.8, 0.0, 1.0)	0.358
(1.0, 1.0, 0.0)	0.161	(0.5, 0.0, 1.0)	0.751	(0.0, 0.0, 0.9)	3.290
(1.0, 1.0, 1.0)	0.114	(1.0, 1.0, 0.5)	0.178	(0.5, 0.6, 0.0)	0.828

Fig. B.1.

Three-stage design points: “•” for Stage 0, “■” Stage 1, “▼” Stage 2.

Appendix C: Evaluation of the fitted calibration model

Fig. C.1 illustrates the prediction quality of the fitted calibration model obtained in Section 5.2: The predicted sensor responses are depicted as “○”, along with the “true” expected concentration-response curves as given in Fig. 7. It is clear from Fig. C.1 that predicted responses are very close to their “true” values (the curves).

Fig. C.1.
The predicted sensor responses and their “true” values.

Biographies

Zongyu Geng is a Ph.D candidate in the Industrial and Management Systems Engineering Department at West Virginia University. His research work has been focused on statistics and chemometrics.

Feng Yang is currently an associate professor in the Industrial and Management Systems Engineering Department at West Virginia University. She received her Ph.D. degree in Industrial Engineering and Management Sciences from Northwestern University in 2006. Her research interests include stochastic simulation and metamodeling, design of experiments, and applied statistics.

Xi Chen is an assistant professor in the Department of Industrial and Systems Engineering at Virginia Tech. Her research interests include stochastic modeling and simulation, applied probability and statistics, computer experiment design and analysis, and simulation optimization.

Nianqiang (Nick) Wu received a Ph.D degree in materials science and engineering in 1997. He worked at Keck Interdisciplinary Surface Science Center in Northwestern University from 2001 to 2005. Currently he is a professor of materials science in Mechanical Engineering and Aerospace Engineering, West Virginia University. His research interest lies in low-dimensional nanomaterials, chemical sensors and biosensors, photocatalysts, photoelectrochemical cells and solar cells.

References

- [1]. Fonollosa J, Rodriguez-Lujan I, Shevade AV, Homer ML, Ryan MA, Huerta R. Human activity monitoring using gas sensor arrays. *Sens. Actuators B: Chem.* 2014; 199:398–402.
- [2]. Geng Z, Yang F, Wu N. Optimum design of sensor arrays via simulation-based multivariate calibration. *Sens. Actuators B: Chem.* 2011; 156:854–862.

- [3]. Docquier N, Candel S. Combustion control and sensors: a review. *Prog. Energy Combust. Sci.* 2002; 28:107–150.
- [4]. Martinelli G, Carotta MC. Thick-film gas sensors. *Sens. Actuators B.* 1995; 23:157–161.
- [5]. Meixner H, Lampe U. Metal oxide sensors. *Sens. Actuators B.* 1996; 33:198–202.
- [6]. Hossein-Babaei F, Ghafarinia V. Compensation for the drift-like terms caused by environmental fluctuations in the responses of chemoresistive gas sensors. *Sens. Actuators B.* 2010; 143:641–648.
- [7]. Vergara A, Benkstein KD, Montgomery CB, Semancik S. Demonstration of fast and accurate discrimination and quantification of chemically similar species utilizing a single cross-selective chemiresistor. *Anal. Chem.* 2014; 86(14):6753–6757. [PubMed: 24931319]
- [8]. Yamazoe N, Fuchigami J, Kishikawa M, Seiyama T. Interactions of tin oxide with O₂, H₂O and H₂. *Surf. Sci.* 1979; 86:335–344.
- [9]. Mielle P. Managing dynamic thermal exchanges in commercial semiconducting gas sensors. *Sens. Actuators B.* 1996; 34:533–538.
- [10]. Boyle JF, Jones KA. The effects of CO water vapor and surface temperature on the conductivity of a SnO₂ gas sensor. *J. Electron. Mater.* 1997; 6:717–733.
- [11]. Barsan N, Weimar U. Understanding the fundamental principles of metal oxide based gas sensors; the example of CO sensing with SnO₂ sensors in the presence of humidity. *J. Phys.: Condens. Matter.* 2003; 15:R813–R839.
- [12]. Montgomery, DC. *Design and Analysis of Experiments*. 5th Ed. Wiley; New York, NY: 2001.
- [13]. Holmberg, M.; Artursson, T. *Handbook of Machine Olfaction: Electronic Nose Technology*. Wiley-VCH; 2003. Drift compensation, standards, and calibration methods; p. 325–346.
- [14]. Marco S, Gutiérrez-Gálvez A. Signal and data processing for machine olfaction and chemical sensing: a review. *IEEE Sens. J.* 2012; 12:3189–3213.
- [15]. Haugen JE, Tomic O, Kvaal K. A calibration method for handling the temporal drift of solid state gas-sensors. *Anal. Chim. Acta.* 2000; 407:23–39.
- [16]. Romain AC, Nicolas J. Long term stability of metal oxide-based gas sensors for e-nose environmental applications: an overview. *Sens. Actuators B.* 2010; 246:502–506.
- [17]. Padilla M, Perera A, Montoliu I, Chaudry A, Persaud K, Marco S. Drift compensation of gas sensor array data by orthogonal signal correction. *Chemomet. Intell. Lab. Syst.* 2010; 100(1):28–35.
- [18]. Ziyatdinov A, Marco S, Chaudry A, Persaud K, Caminal P, Perera A. Drift compensation of gas sensor array data by common principal component analysis. *Sens. Actuators B: Chem.* 2010; 146(2):460–465.
- [19]. Ghasemi-Varnamkhasti M, Mohtasebi SS, Lozano MSJ, Ahmadi H, Razavi SH, Dicko A. Aging fingerprint characterization of beer using electronic nose. *Sens. Actuators B: Chem.* 2011; 159(1):51–59.
- [20]. Frank ML, Fulkerson MD, Patton BR, Dutta PK. TiO₂-based sensor arrays modeled with nonlinear regression analysis for simultaneously determining CO and O₂ concentrations at high temperatures. *Sens. Actuators B.* 2002; 87:471–479.
- [21]. Vergara A, Vembu S, Ayhan T, Ryan MA, Homer ML, Huerta R. Chemical gas sensor drift compensation using classifier ensembles. *Sens. Actuators B: Chem.* 2012; 166:320–329.
- [22]. Curry B, Morgan PH. Model selection in neural networks: some difficulties. *Eur. J. Oper. Res.* 2006; 170(2):567–577.
- [23]. Zhang P, Lee C, Verweij H, Akbar S, Hunter G, Dutta P. High temperature sensor array for simultaneous determination of O₂, CO, and CO₂ with kernel ridge regression data analysis. *Sens. Actuators B.* 2007; 123:950–963.
- [24]. Santner, TJ.; Williams, BJ.; Notz, WI. *The Design and Analysis of Computer Experiments*. Springer; New York: 2003.
- [25]. Atkinson, A.; Donev, A. *Optimum Experimental Designs*. Oxford University Press; 2007.
- [26]. Myers, RH.; Montgomery, DC. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*. 3rd Ed. Wiley-Interscience; 2009.

- [27]. Rasmussen, CE.; Williams, CKI. Gaussian Processes for Machine Learning. The MIT Press; Boston: 2006.
- [28]. Bates, DM.; Watts, DG. Nonlinear Regression Analysis and Its Application. Wiley; New York: 1998.
- [29]. Lei H, Pitt WG, McGrath LK, Hob CK. Modeling carbon black/polymer composite sensors. *Sensors and Actuators B*. 2007; 125:396–407.
- [30]. Geng Z, Yang F, Li M. A bootstrapping-based statistical procedure for multivariate calibration of sensor arrays. *Sens. Actuators B*. 2013; 188:440–453.
- [31]. Ankenman B, Nelson BL, Staum J. Stochastic kriging for simulation metamodeling. *Oper. Res*. 2010; 58(2):371–382.
- [32]. Sacks J, Welch WJ, Mitchell TJ, Wynn HP. Design and analysis of computer experiments. *Stat. Sci.* 1989; 4(4):409–423.
- [33]. Hastie, T.; Tibshirani, R.; Friedman, JH. The Elements of Statistical Learning. Vol. 1. Springer; New York: 2001.
- [34]. Loepky J, Moore L, Williams B. Batch sequential designs for computer experiments. *J. Stat. Plan. Infer.* 2010; 140:1452–1464.
- [35]. Jones G. Bootstrapping in controlled calibration experiments. *Technometrics*. 1999; 41:224–233.
- [36]. Efron, B.; Tibshirani, RJ. An Introduction to the Bootstrap. Chapman and Hall; New York: 1993.
- [37]. Malzahn D, Oppel M. Learning curves and bootstrap estimates for inference with gaussian processes: A statistical mechanics study. *Complexity*. 2003; 8:57–63.
- [38]. den Hertog D, Kocijan J. Customized sequential designs for random simulation experiments: Kriging metamodeling and bootstrapping. *Eur. J. Oper. Res.* 2008; 186(3):1099–1113.
- [39]. van Beers WC, Kleijnen JP. Customized sequential designs for random simulation experiments: Kriging metamodeling and bootstrapping. *Eur. J. Oper. Res.* 2008; 186(3):1099–1113.
- [40]. Kirk PDW, Stumpf MPH. Gaussian process regression bootstrapping: exploring the effects of uncertainty in time course data. *Bioinformatics*. 2009; 25(10):1300–1306. [PubMed: 19289448]
- [41]. Srivastava, M. Methods of Multivariate Statistics. Wiley-Interscience; 2002.
- [42]. Crary, SB. Optimal design of experiments for sensor calibration. Proceeding of 1991 International Conference on Solid-State Sensors and Actuators; San Francisco. 1991; p. 404–407.
- [43]. D'Apuzzo M, D'Arco M, Pasquino N. Design of experiments and data-fitting techniques applied to calibration of high-frequency electromagnetic field probes. *Measurement*. 2011; 44:1153–1165.
- [44]. Jennison, C.; Turnbull, BW. Group Sequential Methods with Applications to Clinical Trials. 1st Ed. Chapman and Hall/CRC; 1999.
- [45]. Rodríguez-Luján I, Fonollosa J, Vergara A, Homer M, Huerta R. On the calibration of sensor arrays for pattern recognition using the minimal number of experiments. *Chemomet. Intell. Lab. Syst.* 2014; 130:123–134.
- [46]. Bernardo MC, Buck R, Liu L, Nazaret WA, Sacks J, Welch WJ. Integrated circuit design optimization using a sequential strategy. *IEEE Trans. Comp.-aided Des.* 1992; 11(3):361–372.
- [47]. Jones DR, Schonlau M, Welch WJ. Efficient global optimization of expensive black-box functions. *J. Global Optim.* 1998; 13:455–492.
- [48]. McKay MD, Beckman RJ, Conover WJ. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*. 1979; 21:239–245.
- [49]. Shewry MC, Wynn HP. Maximum entropy sampling. *J. Appl. Stat.* 1987; 14:165–170.
- [50]. Johnson ME, Moore LM, Ylvisaker D. Minimax and maximin distance designs. *J. Stat. Plan. Infer.* 1990; 26:131–148.
- [51]. Kleijnen J, van Beers W. Application-driven sequential designs for simulation experiments: Kriging metamodeling. *J. Oper. Res. Soc.* 2004; 55:876–883.
- [52]. Picheny V, Ginsbourger G, Roustant O, Haftka RT, Kim N. Adaptive designs of experiments for accurate approximation of a target region. *J. Mech. Design*. 2010; 132(7):071008-1–071008-9.
- [53]. Bect J, Ginsbourger D, Li L, Picheny V, Vazquez E. Sequential design of computer experiments for the estimation of a probability of failure. *Stat. Comput.* 2012; 22(3):773–793.

- [54]. Cormen, TH.; Leiserson, CE.; Rivest, RL.; Stein, C. Introduction to Algorithms. MIT press; 1990.
- [55]. Law, A. McGraw-Hill Series in Industrial Engineering and Management Science. 4th Ed. McGraw-Hill Education; 2006. Simulation Modeling and Analysis with Expertfit Software.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Algorithm GP-Fitting

Input: The sample data $\{(\mathbf{w}_i, r_j(\mathbf{w}_i)); i = 1, 2, \dots, I, j = 1, 2, \dots, n\}$.

Process:

(i) Based on the sample data, calculate the sample averages $\bar{\mathbf{r}}$, as given in (9).

(ii) Estimate the variance of the random error as

$$\hat{\sigma}^2 = \frac{1}{I(n-1)} \sum_{i=1}^I \sum_{j=1}^n (r_j(\mathbf{w}_i) - \bar{r}(\mathbf{w}_i))^2. \quad (10)$$

(iii) Obtain $(\hat{\mu}, \hat{\tau}^2, \hat{\boldsymbol{\theta}})$ by maximizing the log-likelihood function:

$$\begin{aligned} \mathcal{L}(\mu, \tau^2, \boldsymbol{\theta}) = & -\ln[(2\pi)^{I/2}] - \frac{1}{2} \ln |\tau^2 \mathbf{R}(\boldsymbol{\theta}) + \frac{\hat{\sigma}^2}{n} \mathcal{I}_I| \\ & - \frac{1}{2} (\bar{\mathbf{r}} - \mu \mathbf{1}_I)^\top [\tau^2 \mathbf{R}(\boldsymbol{\theta}) + \frac{\hat{\sigma}^2}{n} \mathcal{I}_I]^{-1} (\bar{\mathbf{r}} - \mu \mathbf{1}_I), \end{aligned} \quad (11)$$

where $\mathcal{I}_I$ is the identity matrix of size I , and $\mathbf{1}_I$ is an $(I \times 1)$ vector of ones.

Output: Estimated model parameters $(\hat{\sigma}^2, \hat{\mu}, \hat{\tau}^2, \hat{\boldsymbol{\theta}})$.

Fig. 1.
The fitting algorithm of the Gaussian process (GP) model.

Fig. 2.
The Gaussian process-based bootstrapping algorithm.

The inputs for the procedure:

$\mathcal{W}$, the design space (or feasible region) of exposure condition $\mathbf{w}$.

δ^2 , the desired upper limit of variance for the estimated analyte concentrations.

$\{D_0, D_1, D_2, \dots, D_S\}$, with D_s representing the number of distinct design points to be newly included in Batch s ($s = 0, 1, \dots, S$).

n , the number of replications assigned to a design point, each time experiments are to be performed at that point.

N_{max} , the maximum number of samples allowed, which can be estimated based on the monetary and time budget; and $N_{max} \geq n \times \sum_{s=0}^S D_s$.

Fig. 3.
The list of inputs for the batch sequential procedure.

Fig. 4.
Flowchart of the batch sequential procedure.

Algorithm Design-Evaluation

Inputs: (a) the existing design points $\{\mathbf{w}_i; i = 1, 2, \dots, D\}$, at which real data have been collected from previous stages, with $D = D_0 + D_1 + \dots, D_{s-1}$; (b) the GP model fitted from the existing data and specified by the estimated model parameters $(\hat{\sigma}^2, \hat{\mu}, \hat{\tau}^2, \hat{\theta})$; (c) the set of grid points $\mathcal{W}_{Grid} = \{\mathbf{w}_g^{(Grid)}; g = 1, 2, \dots, G\}$; (d) the candidate design $\{\mathbf{w}_{D+1}, \mathbf{w}_{D+2}, \dots, \mathbf{w}_{D+D_s}\}$

Do for $g = 1, 2, \dots, G$:

- (i) Based on the newly expanded design $\{\mathbf{w}_i; i = 1, 2, \dots, D + D_s\}$, construct $\mathbf{R}(\hat{\theta})$ and $\mathbf{v}(\mathbf{w}_g^{(Grid)}, \hat{\theta})$.
- (ii) Compute $\widehat{\text{Var}}[\hat{\mathbf{w}}_g^{(Grid)}]$ using Formula (13).

Output: the value of $\max_{\mathbf{w} \in \mathcal{W}_{Grid}} \widehat{\text{Var}}[\hat{r}(\mathbf{w})]$ associated with the newly expanded design $\{\mathbf{w}_i; i = 1, 2, \dots, D + D_s\}$.

Fig. 5.
The evaluation of a candidate design.

Algorithm Model-Evaluation

Inputs: (a) the design points $\{\mathbf{w}_i; i = 1, 2, \dots, D\}$, at which real data have been collected with $D = D_0 + D_1 + \dots + D_s$; (b) the GP calibration model $\hat{F}(\mathbf{w})$, fitted from the real data; (c) the residual pool $\{e_{ij}; i = 1, 2, \dots, D, j = 1, 2, \dots, n\}$; (d) the re-sampling size B_1 and B_2 ; (e) the set of grid points $\mathcal{W}_{Grid} = \{\mathbf{w}_g^{(Grid)}; g = 1, 2, \dots, G\}$; (f) the desired precision δ^2 ; (g) N_{max} , the maximum number of samples allowed.

Process: Initialize $TERMINATE = False$

IF $n \times D \geq N_{max}$, set $TERMINATE = True$ and exit the algorithm.

IF $n \times D < N_{max}$, then for each grid point $\mathbf{w}_g^{(Grid)}$, $g = 1, 2, \dots, G$,

- Set $\mathbf{w}_0 = \mathbf{w}_g^{(Grid)}$.
- Run Algorithm GP-Bootstrapping (Figure 2) by providing the inputs: $\mathbf{w}_0 = (c_0, \mathbf{x}_0^\top)^\top$, and (a)–(d) as given above; and obtain the returned value of $\widehat{\text{Var}}[\hat{c}_0]$.
- If $\widehat{\text{Var}}[\hat{c}_0] > \delta^2$, set $TERMINATE = True$ and exit the algorithm.

Output: The value of the boolean variable $TERMINATE$. If $TERMINATE$ is equal to *False*, then an additional batch of calibration experiments will be performed; otherwise, the sequential procedure (Figure 4) will be terminated.

Fig. 6.
The evaluation of a fitted GP calibration model.

Fig. 7.
“True” expected concentration–response curves under different environmental conditions.

Fig. 8.
Histogram of the deviations with $\delta = 0.05$ and $2\delta = 0.1$.

Fig. 9.
Box-plots of RMSE and AME.