

ADAPTIVE DESIGN FOR GAUSSIAN PROCESS REGRESSION UNDER CENSORING

BY JIALEI CHEN^{1,a}, SIMON MAK^{2,d}, V. ROSHAN JOSEPH^{1,b} AND CHUCK ZHANG^{1,c}

¹H. Milton Stewart School of Industrial & Systems Engineering Georgia Institute of Technology, ^ajialei.chen@gatech.edu,
^broshan@gatech.edu, ^cchuck.zhang@gatech.edu

²Department of Statistical Science, Duke University, ^dsm769@duke.edu

A key objective in engineering problems is to predict an unknown experimental surface over an input domain. In complex physical experiments this may be hampered by response censoring which results in a significant loss of information. For such problems, experimental design is paramount for maximizing predictive power using a small number of expensive experimental runs. To tackle this, we propose a novel adaptive design method, called the integrated *censored* mean-squared error (ICMSE) method. The ICMSE method first estimates the posterior probability of a new observation being censored, then adaptively chooses design points that minimize predictive uncertainty under censoring. Adopting a Gaussian process regression model with product correlation function, the proposed ICMSE criterion is easy to evaluate which allows for efficient design optimization. We demonstrate the effectiveness of the ICMSE design in two real-world applications on surgical planning and wafer manufacturing.

1. Introduction. In many engineering problems a key objective is to predict an unknown experimental surface over an input domain. However, for complex physical experiments one can encounter *censoring*, that is, the experimental response is missing or partially measured. Censoring arises from a variety of practical experimental constraints, including limits in measurement devices, safety considerations of experimenters, and a fixed experimental time budget. Figure 1 provides an illustration: in such cases the experimental response of interest is latent, and the observed measurement is subject to censoring. Here, censoring can result in significant loss of information which leads to poor predictive performance (Brooks (1982)). For example, suppose an engineer wishes to explore how pressure in a nuclear reactor changes under different control settings. Due to safety concerns, experiments are forced to stop if the pressure hits a certain upper limit, leading to censored responses. To further complicate matters, the input region, which results in censoring, is typically *unknown* prior to experiments and needs to be estimated from data.

Given the presence of censoring in physical experiments, it is, therefore, of interest to carefully design experimental runs and to best model and predict the experimental response surface. We present a new integrated *censored* mean-squared error (ICMSE) method which

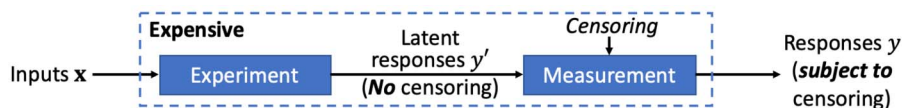


FIG. 1. An illustration of response censoring in the measurement process. The goal is to predict the response surface for the experiment prior to censoring.

Received October 2019; revised June 2021.

Key words and phrases. Adaptive sampling, censored experiments, experimental design, kriging, multifidelity modeling.

sequentially selects physical experimental runs to minimize predictive uncertainty under *censoring*. ICMSE leverages a Gaussian process model (GP; Sacks et al. (1989))—a flexible Bayesian nonparametric model—for the response surface to obtain an easy-to-evaluate design criterion that maximizes GP’s predictive power under censoring. We consider two flavors of ICMSE. The first is a “single-fidelity” ICMSE method for sequentially designing (potentially) censored physical experiments. The second is a “bifidelity” ICMSE method for sequentially designing (potentially) censored physical experiments, given auxiliary computer simulation data. The two settings are motivated from the following two applications.

1.1. 3D-printed aortic valves for surgical planning. The first motivating problem concerns the design of 3D-printed tissue-mimicking aortic valves for heart surgeries. With advances in additive manufacturing (Gibson, Rosen and Stucker (2014)), 3D-printed medical prototypes (Rengier et al. (2010)) play an increasingly important role in presurgical studies (Qian et al. (2017)). They are particularly helpful in complicated heart diseases, for example, aortic stenosis, where 3D-printed aortic valves can be used to select the best surgical option with minimal postsurgical complication (Chen et al. (2018b)). The printed aortic valve (see Figure 2(a)) contains a biomimetic substructure: an enhancement polymer (white) is embedded in a substrate polymer (clear); this is known as *metamaterial* (Wang et al. (2016)) in the materials engineering literature. The goal is to build a model to understand how the *stiffness* of the metamaterials is affected by the *geometry* of the enhancement polymer (see Figure 2(b)). This model can then be used by doctors to select a polymer geometry that mimics the target stiffness of the specific patient—a procedure known as “tissue-mimicking” (Chen et al. (2021)). An accurate tissue-mimicking is paramount for surgery success, since inaccurate stiffness may lead to severe postsurgery complications and death.

Using earlier terminology, this is a bifidelity modeling problem involving two types of experiments: a preconducted database of computer simulations and patient-specific physical experiments. The physical experiments here are very *costly*: we need to 3D print each metamaterial sample, then physically test its stiffness using a load cell. Furthermore, the measurement from physical experiments may be *censored*, due to an inherent upper limit of the testing machine. This is shown in Figure 2(c): if the metamaterial sample is stiffer than the load cell (i.e., a spring), the experiment is forced to stop to prevent breakage of the load cell. One workaround is to use a stiffer load cell; however, it is oftentimes *not* a preferable option: a stiffer load cell with a broader measurement range can be very expensive, costing over a hundred times more than the standard integrated load cells. Here, the proposed ICMSE method can adaptively design experimental runs to maximize the predictive power of a GP model under censoring. We show later in Section 4.2, ICMSE can lead to better predictions for younger patients (with stiff tissues which can be *censored* in experiments) and older patients (with soft tissues which are *uncensored* in experiments) (Sicard et al. (2018)). This

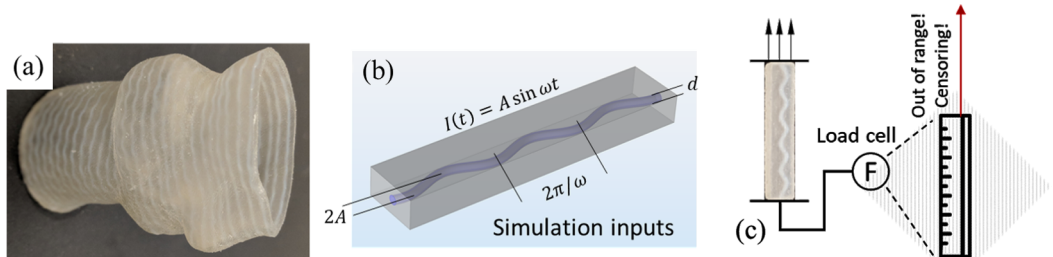


FIG. 2. Illustrating the surgical planning application: (a) a 3D-printed aortic valve with enhanced metamaterial, (b) simulation inputs in the computer experiment, (c) visualizing the physical experiment and the measurement censoring of the load cell (labeled “F”).

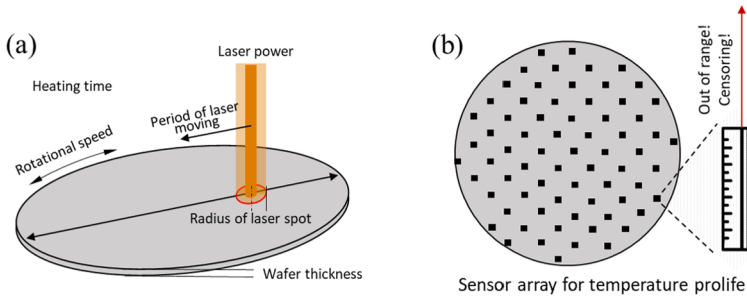


FIG. 3. Illustrating the wafer manufacturing application: (a) visualizing the thermal processing procedure with the six input parameters, (b) visualizing the measurement censoring of the temperature sensor array.

then leads to greatly improved tissue-mimicking performance for personalized printed valves (Chen et al. (2021)) which is crucial for improving heart surgery success rate (Qian et al. (2017)). Our method is particularly valuable in urgent heart surgeries, where one can perform only a small number of runs prior to the actual surgery.

1.2. Thermal processing in wafer manufacturing. The second problem considers the design of the semiconductor wafer manufacturing process (Jin, Chang and Shi (2012), Quirk and Serda (2001)). Wafer manufacturing involves processing silicon wafers in a series of refinement stages to be used as circuit chips. Among these stages, thermal processing is one of the most important stages (Singh, Fakhruddin and Poole (2000)), since it facilitates the necessary chemical reactions and allows for surface oxidation. Figure 3(a) illustrates the typical thermal processing procedure: a laser beam (in orange) is moved back and forth over a rotating wafer. The output of interest here is the *minimal* temperature over the whole wafer after heating; a higher minimal temperature facilitates better completeness of chemical reactions which leads to better quality of the final wafer product (Goodson et al. (1993), Van Gorp and Palmén (1998)). However, higher temperatures may result in higher energy costs for heating. With the fitted predictive model on minimal wafer temperature, industrial engineers can then use this to optimize a heating process which not only is economical (i.e., conserves heating power) but also meets target quality requirements.

However, laser heating experiments are quite costly, involving high material and operation costs. In industrial settings the minimal wafer temperature is often subject to *censoring*, due to the nature of measurement procedures. This is shown in Figure 3(b): the wafer temperature is typically measured by either an array of temperature sensors or a thermal camera, both of which have upper measurement limits (Feteira (2009)). The minimal temperature is censored when the whole sensor array reaches the measurement limits. While more sophisticated sensors exist, they are much more expensive and may lead to tedious do-overs of experiments. The proposed single-fidelity ICMSE method can be used to adaptively design experimental runs that maximize the predictive power of a GP model under censoring. We show later in Section 4.1 that the resulting model using ICMSE enjoys improved predictive performance for high wafer temperatures (that are potentially censored) and low temperatures (that are not censored) to ensure flexibility for different quality requirements. The fitted model can then be used to find an optimal thermal processing setting which minimizes operation costs subject to target quality requirements.

1.3. Literature. GP regression (or *kriging*; see Mathéron (1963)) is widely used as a predictive model for expensive experiments (Sacks et al. (1989)) and has been applied in cosmology (Kaufman et al. (2011)), aerospace engineering (Mak et al. (2018)), healthcare (Chen

et al. (2021)), and other applications. The key appeals of GPs are the flexible nonparametric model structure and closed-form expressions for prediction and uncertainty quantification (Santner, Williams and Notz (2018)). In the engineering literature, GPs have been used for modeling expensive physical experiments (Ankenman, Nelson and Staum (2010)), integrating computer and physical experiments (Kennedy and O'Hagan (2001)), and incorporating various constraints (Da Veiga and Marrel (2012), Ding, Mak and Wu (2019), Groot et al. (2012), Henkenjohann et al. (2005), López-Lopera et al. (2018)). We will adapt in this work a recent censored GP model (Cao et al. (2018)) which integrates censored physical experimental data.

There have been several works in the literature on experimental design under response censoring; see, for example, Borth (1996), Monroe and Pan (2008). These methods, however, presume a parametric form for the response surface, which may be a dangerous assumption for black-box experiments, hence the recent shift for more nonparametric models, such as GPs. Existing design methods for GPs can be divided into two categories—space-filling and model-based designs. Space-filling designs aim to fill empty gaps in the input space; this includes minimax designs (Johnson, Moore and Ylvisaker (1990)), maximin designs (Morris and Mitchell (1995)), and maximum projection designs (Joseph, Gul and Ba (2015)). Model-based designs instead maximize an optimality criterion based on an *assumed* GP model; this includes integrated mean-squared error designs (Sacks et al. (1989)) and maximum entropy designs (Shewry and Wynn (1987)). Such designs can also be implemented sequentially in an adaptive manner; see Bect, Bachoc and Ginsbourger (2019), Chen, Wang and Wu (2017), Lam (2008), Xiong, Qian and Wu (2013). Recently, Binois et al. (2019) proposed a design method for a heteroscedastic GP model (i.e., under input-dependent noise); this provides a flexible framework that allows for different correlation functions, closed-form gradients for optimization, and batch sequential implementation.

The above GP design methods, however, do not consider potential response *censoring*. The key challenge in incorporating censoring information is that an experimenter does not know which inputs may lead to censoring prior to experimentation, since the response surface is black-box. The proposed ICMSE method addresses this by leveraging a GP model on the unknown response surface: it first estimates the posterior probability of a potential observation being censored and then finds design points that minimize predictive uncertainty under censoring. Under product correlation functions our method admits an easy-to-evaluate design criterion which allows for efficient sequential sampling. We show that ICMSE can yield considerably improved predictive performance over existing design methods (which do not consider censoring) in both motivating applications.

1.4. Structure. Section 2 presents the ICMSE design method for the single-fidelity setting with only physical experiment data. Section 3 extends the ICMSE method for the bifidelity setting, where auxiliary computer simulation data are available. Section 4 demonstrates the effectiveness of ICMSE in the two motivating applications. Section 5 concludes the work.

2. ICMSE design. We now present the ICMSE design method for the single-fidelity setting; a more elaborate bifidelity setting is discussed later in Section 3. We first review the GP model for censored data, and derive the proposed ICMSE design criterion. We then visualize this via a one-dimensional (1D) example and provide some insights.

2.1. Modeling framework. We adopt the following model for physical experiments. Let $\mathbf{x}_i \in [0, 1]^p$ be a vector of p input variables (each normalized to $[0, 1]$), and let y'_i be its latent response from the physical experiment *prior to* potential censoring (see Figure 1). We assume

$$(2.1) \quad y'_i = \xi(\mathbf{x}_i) + \epsilon_i, \quad i = 1, 2, \dots, n,$$

where $\xi(\mathbf{x}_i)$ is the mean of the latent response y'_i at input $\mathbf{x}_i$ and ϵ_i is the corresponding measurement error. Since $\xi(\cdot)$ is unknown, we further assign to it a GP prior with mean μ_ξ , variance σ_ξ^2 , and correlation function $R_{\theta_\xi}(\cdot, \cdot)$ with parameters θ_ξ . This is denoted as

$$(2.2) \quad \xi(\cdot) \sim \text{GP}(\mu_\xi, \sigma_\xi^2 R_{\theta_\xi}(\cdot, \cdot)).$$

The experimental noise $\epsilon_i \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_\epsilon^2)$ is assumed to be i.i.d. normally distributed and independent of $\xi(\cdot)$.

For simplicity, we consider only the case of right-censoring below, that is, censoring of the response only when it exceeds some *known* upper limit (this is the setting for both motivating applications). All equations and insights derived in the paper hold analogously for the general case of interval censoring, albeit with more cumbersome notation. Suppose, from n experiments, n_o responses are observed without censoring, and n_c responses are right-censored at limit c , where $n_o + n_c = n$. The training set experimental data can then be written as the set $\mathcal{Y}_n = \{\mathbf{y}_o, \mathbf{y}'_c \geq \mathbf{c}\}$, where $\mathbf{y}_o$ is a vector of *observed* responses at inputs $\mathbf{x}_o = \mathbf{x}_{1:n_o} = \{\mathbf{x}_1, \dots, \mathbf{x}_{n_o}\}$, $\mathbf{y}'_c$ is the latent response vector for inputs in censored regions $\mathbf{x}_c = \mathbf{x}_{(n_o+1):n}$ prior to censoring, and $\mathbf{c} = [c, \dots, c]^T$ is the vector of the right-censoring limit. Assuming known model parameters, a straightforward adaptation of the equations (11) and (12) in [Cao et al. \(2018\)](#) gives the following expressions for the conditional mean and variance of $\xi(\mathbf{x}_{\text{new}})$ at new input $\mathbf{x}_{\text{new}}$:

$$(2.3) \quad \hat{\xi}(\mathbf{x}_{\text{new}}) = \mathbb{E}[\xi(\mathbf{x}_{\text{new}})|\mathcal{Y}_n] = \mu_\xi + \boldsymbol{\gamma}_{n,\text{new}}^T \boldsymbol{\Gamma}_n^{-1}([\mathbf{y}_o, \hat{\mathbf{y}}_c]^T - \mu_\xi \cdot \mathbf{1}_n),$$

$$(2.4) \quad s^2(\mathbf{x}_{\text{new}}) = \text{Var}[\xi(\mathbf{x}_{\text{new}})|\mathcal{Y}_n] = \sigma_\xi^2 - \boldsymbol{\gamma}_{n,\text{new}}^T (\boldsymbol{\Gamma}_n^{-1} - \boldsymbol{\Gamma}_n^{-1} \boldsymbol{\Sigma} \boldsymbol{\Gamma}_n^{-1}) \boldsymbol{\gamma}_{n,\text{new}}.$$

Here, $\boldsymbol{\Gamma}_n = \sigma_\xi^2 [R_{\theta_\xi}(\mathbf{x}_i, \mathbf{x}_j)]_{i=1}^n [j=1]^n + \sigma_\epsilon^2 \mathbf{I}_n$, $\boldsymbol{\gamma}_{n,\text{new}} = \sigma_\xi^2 [R_{\theta_\xi}(\mathbf{x}_1, \mathbf{x}_{\text{new}}), \dots, R_{\theta_\xi}(\mathbf{x}_n, \mathbf{x}_{\text{new}})]^T$, $\mathbf{1}_n$ is a one-vector of length n , and $\mathbf{I}_n$ is an $n \times n$ identity matrix. Furthermore, $\hat{\mathbf{y}}_c = \mathbb{E}[\mathbf{y}'_c|\mathcal{Y}_n]$ is the expected response for the latent vector $\mathbf{y}'_c$, given the dataset $\mathcal{Y}_n$, $\boldsymbol{\Sigma}_c = \text{Var}[\mathbf{y}'_c|\mathcal{Y}_n]$ is its conditional variance, and $\boldsymbol{\Sigma} = \text{diag}(\mathbf{0}_{n_o}, \boldsymbol{\Sigma}_c)$. The computation of these quantities will be discussed later in Section 3.3. The conditional mean (2.3) is used to predict the mean experimental response at an untested input $\mathbf{x}_{\text{new}}$, and the conditional variance (2.4) is used to quantify predictive uncertainty.

In the case of no censoring (i.e., $\mathcal{Y}_n = \{\mathbf{y}_o\}$), equations (2.3) and (2.4) reduce to

$$(2.5) \quad \hat{\xi}(\mathbf{x}_{\text{new}}) = \mathbb{E}[\xi(\mathbf{x}_{\text{new}})|\mathcal{Y}_n] = \mu_\xi + \boldsymbol{\gamma}_{n,\text{new}}^T \boldsymbol{\Gamma}_n^{-1}(\mathbf{y}_o - \mu_\xi \cdot \mathbf{1}_n), \quad \text{and}$$

$$(2.6) \quad s^2(\mathbf{x}_{\text{new}}) = \text{Var}[\xi(\mathbf{x}_{\text{new}})|\mathcal{Y}_n] = \sigma_\xi^2 - \boldsymbol{\gamma}_{n,\text{new}}^T \boldsymbol{\Gamma}_n^{-1} \boldsymbol{\gamma}_{n,\text{new}}.$$

These are precisely the conditional mean and variance expressions for the standard GP regression model ([Santner, Williams and Notz \(2018\)](#)) which is as expected.

2.2. Design criterion. Now, given data $\mathcal{Y}_n$ from n experiments (n_o of which are observed exactly; n_c of which are censored), we propose a new design method that accounts for the posterior probability of a potential observation being censored. Let $\mathbf{x}_{n+1}$ be a potential next input for experimentation, Y'_{n+1} be its latent response *prior to* censoring, and $Y_{n+1} = Y'_{n+1}(1 - \mathbb{1}_{\{Y'_{n+1} \geq c\}}) + c\mathbb{1}_{\{Y'_{n+1} \geq c\}}$ be its corresponding observation *after* censoring, with $\mathbb{1}_{\{\cdot\}}$ denoting the indicator function. The proposed method chooses the next input $\mathbf{x}_{n+1}^*$ as

$$(2.7) \quad \begin{aligned} \mathbf{x}_{n+1}^* &= \underset{\mathbf{x}_{n+1}}{\text{argmin}} \text{ICMSE}(\mathbf{x}_{n+1}) \\ &:= \underset{\mathbf{x}_{n+1}}{\text{argmin}} \int_{[0,1]^p} \mathbb{E}_{Y_{n+1}|\mathcal{Y}_n} [\text{Var}(\xi(\mathbf{x}_{\text{new}})|\mathcal{Y}_n, Y_{n+1})] d\mathbf{x}_{\text{new}}. \end{aligned}$$

The design criterion $\text{ICMSE}(\mathbf{x}_{n+1})$ can be understood in two parts. First, the term $\text{Var}(\xi(\mathbf{x}_{\text{new}})|\mathcal{Y}_n, Y_{n+1})$ quantifies the predictive variance (i.e., mean-squared error, MSE) of the mean response at an untested input $\mathbf{x}_{\text{new}}$, given both the training data $\mathcal{Y}_n$ and the potential observation Y_{n+1} . This is a reasonable quantity to minimize for design, since we wish to find which new input $\mathbf{x}_{n+1}$ can minimize predictive uncertainty. Second, note that this MSE term cannot be used directly as a criterion, since it depends on the potential observation Y_{n+1} , which is yet to be observed. One way around this is to take the conditional expectation $\mathbb{E}_{Y_{n+1}|\mathcal{Y}_n}[\cdot]$ (more on this below). Finally, the integral over $[0, 1]^p$ yields the average predictive uncertainty over the entire design space.

The proposed criterion in (2.7) can be viewed as an extension of the sequential integrated mean-squared error (IMSE) design (Lam (2008), Santner, Williams and Notz (2018)) for the censored response setting. Assuming no censoring (i.e., $\mathcal{Y}_n = \{\mathbf{y}_o\}$), the sequential IMSE design chooses the next input $\mathbf{x}_{n+1}^*$ by minimizing

$$(2.8) \quad \min_{\mathbf{x}_{n+1}} \text{IMSE}(\mathbf{x}_{n+1}) := \min_{\mathbf{x}_{n+1}} \int_{[0,1]^p} \text{Var}(\xi(\mathbf{x}_{\text{new}})|\mathcal{Y}_n, Y'_{n+1}) d\mathbf{x}_{\text{new}}.$$

Note that, in the *uncensored* setting, the MSE term $\text{Var}(\xi(\mathbf{x}_{\text{new}})|\mathcal{Y}_n, Y'_{n+1})$ in (2.8) does *not* depend on the potential observation Y'_{n+1} which allows the criterion to be easily computed in practice. However, in the *censored* setting at hand, not only does this MSE term *depend* on Y'_{n+1} but such an observation may not be directly observed due to censoring. The conditional expectation $\mathbb{E}_{Y_{n+1}|\mathcal{Y}_n}[\cdot]$ in (2.7) addresses this by accounting for the posterior probability of censoring in Y'_{n+1} .

One attractive feature of the ICMSE criterion (2.7) is that it will be *adaptive* to the experimental responses from data. The criterion (2.7) inherently hinges on whether the potential observation Y_{n+1} is censored (i.e., $Y'_{n+1} \geq c$) or not (i.e., $Y'_{n+1} < c$), but this censoring behavior needs to be estimated from experimental data. Viewed this way, the ICMSE criterion can be broken down into two steps; it: (i) estimates the posterior probability of a new observation being censored from data and then (ii) samples the next point that minimizes the *average* predictive uncertainty under censoring. We will show how our method adaptively incorporates the posterior probability of censoring Y_{n+1} for sequential design, in contrast to the existing IMSE method (2.8).

2.2.1. No censoring in training data. To provide some intuition, consider a simplified scenario with no censoring in the *training* set, that is, $\mathcal{Y}_n = \{\mathbf{y}_o\}$ (censoring may still occur for the new Y_{n+1}). In this case the following proposition gives an explicit expression for the ICMSE criterion.

PROPOSITION 1. *Suppose there is no censoring in training data, that is, $\mathcal{Y}_n = \{\mathbf{y}_o\}$. Then, the ICMSE criterion (2.7) has the explicit expression*

$$(2.9) \quad \text{ICMSE}(\mathbf{x}_{n+1}) = \int_{[0,1]^p} \sigma_{\text{new}}^2 - h_c(\mathbf{x}_{n+1}) \rho_{\text{new}}^2(\mathbf{x}_{n+1}) \sigma_{\text{new}}^2 d\mathbf{x}_{\text{new}}, \quad \text{where}$$

$$h_c(\mathbf{x}_{n+1}) = h(z_c) = \Phi(z_c) - z_c \phi(z_c) + \frac{\phi^2(z_c)}{1 - \Phi(z_c)}, \quad z_c = \frac{c - \mu_{n+1}}{\sigma_{n+1}}.$$

Here, $\sigma_{\text{new}}^2 = \text{Var}[\xi(\mathbf{x}_{\text{new}})|\mathcal{Y}_n]$, $\rho_{\text{new}}(\mathbf{x}_{n+1}) = \text{Corr}[\xi(\mathbf{x}_{n+1}), \xi(\mathbf{x}_{\text{new}})|\mathcal{Y}_n]$, $\mu_{n+1} = \mathbb{E}[\xi(\mathbf{x}_{n+1})|\mathcal{Y}_n]$, and $\sigma_{n+1}^2 = \text{Var}[\xi(\mathbf{x}_{n+1})|\mathcal{Y}_n]$ follow from (2.5) and (2.6). $\phi(\cdot)$, and $\Phi(\cdot)$ are the probability density and cumulative distribution functions for the standard normal distribution.

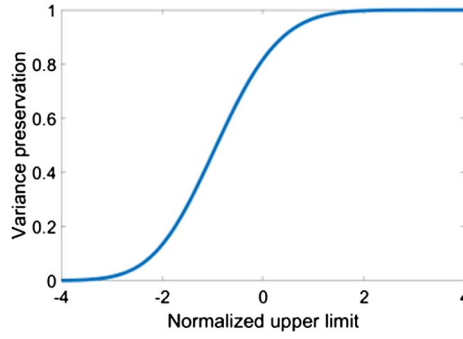


FIG. 4. Visualizing the censoring adjustment function $h(z_c)$, where z_c is the normalized right-censoring limit.

In words, μ_{n+1} is the predictive mean at $\mathbf{x}_{n+1}$, given data $\mathcal{Y}_n$, σ_{n+1}^2 and σ_{new}^2 are the predictive variances at $\mathbf{x}_{n+1}$ and $\mathbf{x}_{\text{new}}$, respectively, and $\rho_{\text{new}}(\mathbf{x}_{n+1})$ is the posterior correlation between $\xi(\mathbf{x}_{n+1})$ and $\xi(\mathbf{x}_{\text{new}})$. Note that the p -dimensional integral in (2.9) can also be efficiently computed in practice; we provide more discussion later in Corollary 1. The proof of this proposition can be found in Appendix A.2 of the Supplementary Material.

To glean intuition from the criterion (2.9), we compare it with the existing sequential IMSE criterion (2.8). Under no censoring in training data (i.e., $\mathcal{Y}_n = \{\mathbf{y}_o\}$), (2.8) can be rewritten as

$$(2.10) \quad \text{IMSE}(\mathbf{x}_{n+1}) = \int_{[0,1]^p} \sigma_{\text{new}}^2 - \rho_{\text{new}}^2(\mathbf{x}_{n+1}) \sigma_{\text{new}}^2 d\mathbf{x}_{\text{new}}.$$

Comparing (2.10) with (2.9), we note a key distinction in the ICMSE criterion: the presence of $h_c(\mathbf{x}_{n+1}) = h(z_c)$, where z_c is the normalized right-censoring limit under the posterior distribution at $\mathbf{x}_{n+1}$. We call $h(\cdot)$ the *censoring adjustment* function. Figure 4 visualizes $h(z_c)$ for different choices of z_c . Consider first the case of z_c large. From the figure we see that $h(z_c) \rightarrow 1$ as $z_c \rightarrow \infty$ in which case the proposed ICMSE criterion (2.9) reduces to the standard IMSE criterion (2.10). This makes sense intuitively: a large value of z_c (i.e., a high right-censoring limit) means that a new observation at $\mathbf{x}_{n+1}$ has little posterior probability of being censored at c . In this case the ICMSE criterion (which minimizes predictive variance *under* censoring) should then reduce to the IMSE criterion (which minimizes predictive variance *ignoring* censoring). Consider next the case of z_c small. From the figure, we see that $h(z_c) \rightarrow 0$ as $z_c \rightarrow -\infty$, and the proposed criterion (2.9) reduces to the integral of σ_{new}^2 . Again, this makes intuitive sense: a small value of z_c (i.e., a low right-censoring limit) means a new observation at $\mathbf{x}_{n+1}$ has a high posterior probability of being censored. In this case the ICMSE criterion reduces to the predictive variance of the testing point $\mathbf{x}_{\text{new}}$, given only the first n training data points, meaning a new design point at $\mathbf{x}_{n+1}$ offers little reduction in predictive variance. Viewed this way, the proposed ICMSE criterion modifies the standard IMSE criterion by accounting for the posterior probability of censoring via the censoring adjustment function $h(z_c)$.

Equation (2.9) also reveals an important trade-off for the proposed design under censoring. Consider first the standard IMSE criterion (2.10) which minimizes predictive uncertainty under no censoring. Since the first term σ_{new}^2 does not depend on the new design point $\mathbf{x}_{n+1}$, this uncertainty minimization is achieved by maximizing the second term $\rho_{\text{new}}^2(\mathbf{x}_{n+1}) \sigma_{\text{new}}^2$. This can be interpreted as the variance reduction from observing Y'_{n+1} (Gramacy and Apple (2015)). Consider next the proposed ICMSE criterion (2.9) which maximizes the term $h(z_c) \rho_{\text{new}}^2(\mathbf{x}_{n+1}) \sigma_{\text{new}}^2$. This can further be broken down into: (i) the maximization of variance reduction term $\rho_{\text{new}}^2(\mathbf{x}_{n+1}) \sigma_{\text{new}}^2$ and (ii) the maximization of the censoring adjustment function $h(z_c)$. Objective (i) is the same as for the standard IMSE criterion; it minimizes predictive uncertainty assuming no response censoring. Objective (ii), by maximizing the

censoring adjustment function $h(z_c)$, aims to minimize the posterior probability of the new design point being censored. Putting both parts together, the ICMSE criterion (2.9) features an important trade-off: it aims to find a new design point that jointly minimizes predictive uncertainty (in the absence of censoring) and the posterior probability of being censored.

2.2.2. Censoring in training data. We now consider the general case of censored training data $\mathcal{Y}_n = \{\mathbf{y}_o, \mathbf{y}'_c \geq \mathbf{c}\}$. The following proposition gives an explicit expression for the ICMSE criterion.

PROPOSITION 2. *Given the censored data $\mathcal{Y}_n = \{\mathbf{y}_o, \mathbf{y}'_c \geq \mathbf{c}\}$, we have*

$$(2.11) \quad \text{ICMSE}(\mathbf{x}_{n+1}) = \int_{[0,1]^p} \sigma_{\text{new}}^2 - \boldsymbol{\gamma}_{n+1,\text{new}}^T \boldsymbol{\Gamma}_{n+1}^{-1} \mathbf{H}_c(\mathbf{x}_{n+1}) \boldsymbol{\Gamma}_{n+1}^{-1} \boldsymbol{\gamma}_{n+1,\text{new}} d\mathbf{x}_{\text{new}},$$

where $\sigma_{\text{new}}^2 = \text{Var}[\xi(\mathbf{x}_{\text{new}})|\mathcal{Y}_n]$ and $\boldsymbol{\gamma}_{n+1,\text{new}}$ and $\boldsymbol{\Gamma}_{n+1}$ follow from (2.3) and (2.4). The matrix $\mathbf{H}_c(\mathbf{x}_{n+1})$ has an easy-to-evaluate expression given in Appendix A.3 of the Supplementary Material.

Here, σ_{new}^2 is the predictive variance at point $\mathbf{x}_{\text{new}}$, conditional on the data $\mathcal{Y}_n$. The full expression for $(n+1) \times (n+1)$ matrix $\mathbf{H}_c(\mathbf{x}_{n+1})$, while easy-to-evaluate, is quite long and cumbersome; this expression is provided in Appendix A.3 of the Supplementary Material. The key computation in calculating $\mathbf{H}_c(\mathbf{x}_{n+1})$ is evaluating several orthant probabilities from a multivariate normal distribution. The proof for this proposition can be found in Appendix A.3. Section 3.3 and Appendix C provide further details on computation.

While this general ICMSE criterion (2.11) is more complex, its interpretation is quite similar to the earlier criterion; its integrand contains a posterior variance term conditional on data $\mathcal{Y}_n$ and a variance reduction term from the potential observation Y_{n+1} . The matrix $\mathbf{H}_c(\mathbf{x}_{n+1})$ on the variance reduction term serves a similar purpose to the censoring adjustment function. A large value of $\mathbf{H}_c(\mathbf{x}_{n+1})$ (in a matrix sense) suggests a low posterior probability of censoring for a new point $\mathbf{x}_{n+1}$, whereas a small value suggests a high posterior probability of censoring. This again results in the important trade-off for sequential design under censoring: the proposed ICMSE criterion aims to find the next design point which not only: (i) minimizes predictive uncertainty of the fitted model in the absence of censoring but also (ii) minimizes the posterior probability that the resulting observation is censored. The posterior probability is adaptively learned from the training data and is not considered by the standard IMSE criterion.

2.3. An illustrative example. We illustrate the ICMSE criterion using a 1D example. Suppose the mean response of the physical experiment is

$$(2.12) \quad \xi(x) = 0.5 \sin(10(x - 1.02)^2) - 1.25(x - 0.75)(2x - 0.25) + 0.2,$$

with measurement noise variance $\sigma_\epsilon^2 = 0.1^2$. Further, suppose censoring occurs above an upper limit of $c = 0.55$. The initial design consists of six equally-spaced runs which result in five observed runs and one censored run. The Gaussian correlation function is used for R_{θ_ξ} , with model parameters estimated via maximum likelihood.

We compare the ICMSE method (2.11) with IMSE methods. Note that, from Section 2.2, the standard IMSE criterion (2.8) cannot be directly applied here, since it depends on the potential observation Y'_{n+1} which is unobserved. We adopt the following two variants of IMSE for the censored setting. The first method, “IMSE-Impute,” is a simple baseline which *imputes* the censored responses in the training data with the known measurement limit c . The new design point is then optimized, assuming its corresponding response Y'_{n+1} is *not* subject

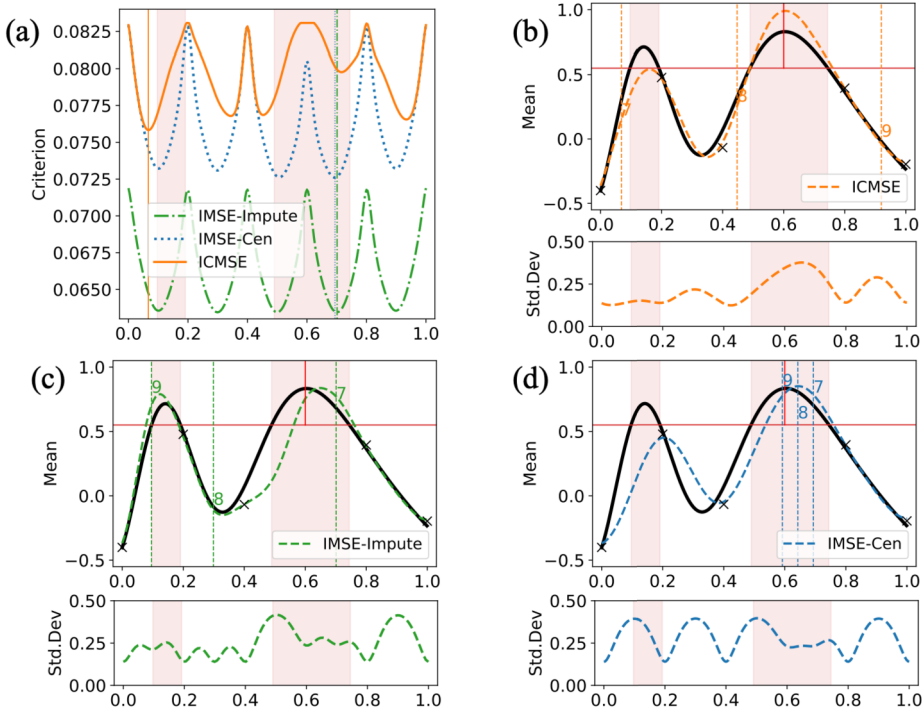


FIG. 5. The 1D illustrative example in (2.12): (a) shows the design criteria of the next run x_7 for the considered methods. (b), (c), and (d) show the three sequential runs (x_7^* , x_8^* , x_9^*) using ICMSE, IMSE-Impute, and IMSE-Cen, respectively, with the censored regions shaded in red. Top plots show the true function $\xi(\cdot)$ (black line) and the predictor $\hat{\xi}(\cdot)$ (dashed line) with initial observed runs (black crosses, with censored runs indicated by red vertical lines) and sequential points (numbered). Bottom plots show the corresponding predictive standard deviation.

to censoring. The second method, “IMSE-Cen,” integrates the censored runs (in training data) using the *censored* GP model (2.3)–(2.4). The new design point is again optimized, assuming its response Y'_{n+1} is *not* subject to censoring. In contrast, the proposed ICMSE method (2.11) considers *both* censored training data and the possibility of censoring in the new observation Y'_{n+1} within the design criterion. For a fair comparison we use the censored GP model (2.3)–(2.4) for evaluating predictive performance for all design methods.

Figure 5(a) shows the proposed criterion for the ICMSE method (in orange). It selects the next design point at $x_7^* = 0.068$ which balances the two desired properties from the ICMSE criterion. First, it avoids regions with high posterior probabilities of response censoring, due to the presence of $\mathbf{H}_c(\cdot)$ in (2.11). The next point x_7^* , which minimizes (2.11), subsequently *avoids* the censored regions (shaded red), as desired. In contrast, Figure 5(a) also shows the design criteria for IMSE-Impute (green) and IMSE-Cen (blue). We see that both IMSE methods choose the next point *within* the censored regions, as the IMSE design criterion does not consider the probability of a new observation being censored. Second, the next point x_7^* chosen by ICMSE minimizes the overall predictive uncertainty for the mean function $\xi(\cdot)$, since the ICMSE criterion is small in regions *away* from existing design points. This can be seen within the region $[0.2, 0.5]$, where local minima of the ICMSE criterion are found between training points.

The top plots in Figure 5(b)–(d) show the next three design points (x_7^* , x_8^* , x_9^*) from the three considered design methods as well as the final predictor $\hat{\xi}(\cdot)$, using the censored GP model (2.3)–(2.4) with all nine points. The bottom plots in Figure 5(b)–(d) show the corresponding predictive standard deviation. We see that ICMSE yields noticeably better predictive performance, compared to the two IMSE methods. One reason is that the proposed

TABLE 1

Predictive performance (in RMSE) for three sequential runs in the 1D example (2.12), using ICMSE and the two IMSE baselines (IMSE-Impute and IMSE-Cen)

	IMSE-Impute	IMSE-Cen	ICMSE
6 runs	0.260	0.260	0.260
7 runs	0.214	0.214	0.119
8 runs	0.172	0.236	0.102
9 runs	0.153	0.203	0.096

criterion makes use of the censored GP model for both modeling and design, whereas the two baselines do not. Table 1 shows the root mean-squared error (RMSE) after the three sequential runs over a test set of 1000 equally-spaced points. The proposed ICMSE method achieves much smaller errors compared to the two IMSE baselines. We will provide a more comprehensive comparison of predictive performance in Section 3.4.

3. ICMSE design for bifidelity modeling. Next, we extend the ICMSE design to the bifidelity setting, where auxiliary computer experiment data are available. We first present the GP framework for bifidelity modeling and extend the earlier ICMSE criterion. We then present an algorithmic framework for efficient implementation and investigate its performance on two illustrative examples.

3.1. Modeling framework. Let $f(\mathbf{x})$ denote the *computer* experiment output at input $\mathbf{x}$. We model $f(\cdot)$ as the GP model

$$(3.1) \quad f(\cdot) \sim \text{GP}\{\mu_f, \sigma_f^2 R_{\theta_f}(\cdot, \cdot)\}.$$

Following Section 2.1, let $\xi(\mathbf{x})$ denote the latent mean response for *physical* experiments at input $\mathbf{x}$. We assume that $\xi(\cdot)$ takes the form

$$(3.2) \quad \xi(\mathbf{x}) = f(\mathbf{x}) + \delta(\mathbf{x}),$$

where $\delta(\mathbf{x})$ is the so-called *discrepancy* function, quantifying the difference between computer and physical experiments at input $\mathbf{x}$. Following Kennedy and O’Hagan (2001), we model this discrepancy using a zero-mean GP model,

$$(3.3) \quad \delta(\cdot) \sim \text{GP}\{0, \sigma_\delta^2 R_{\theta_\delta}(\cdot, \cdot)\},$$

where the prior on $\delta(\cdot)$ is independent of $f(\cdot)$. Here, physical experiments are observed with experimental noise, as in Section 2.1, whereas computer experiments are observed without noise.

Suppose $(n - m)$ computer experiments and m physical experiments (n experiments in total) are conducted at inputs $\mathbf{x}_{1:n} = \{\mathbf{x}_{1:(n-m)}^f, \mathbf{x}_{1:m}^\xi\}$, yielding data $\mathbf{f} = [f_1, \dots, f_{n-m}]$ and $\mathcal{Y}_m = \{\mathbf{y}_o, \mathbf{y}_c' \geq \mathbf{c}\}$. Note that censoring occurs only in physical experiments, since computer experiments are conducted via numerical simulations. Assuming all model parameters are known (parameter estimation is discussed later in Section 3.3), the mean response $\xi(\mathbf{x}_{\text{new}})$ at a new input $\mathbf{x}_{\text{new}}$ has the following conditional mean and variance:

$$(3.4) \quad \hat{\xi}(\mathbf{x}_{\text{new}}) = \mathbb{E}[\xi(\mathbf{x}_{\text{new}})|\mathbf{f}, \mathcal{Y}_m] = \mu_f + \boldsymbol{\gamma}_{n,\text{new}}^T \boldsymbol{\Gamma}_n^{-1} ([\mathbf{f}, \mathbf{y}_o, \hat{\mathbf{y}}_c]^T - \mu_f \mathbf{1}_n),$$

$$(3.5) \quad s^2(\mathbf{x}_{\text{new}}) = \text{Var}[\xi(\mathbf{x}_{\text{new}})|\mathbf{f}, \mathcal{Y}_m] = \sigma_f^2 + \sigma_\delta^2 - \boldsymbol{\gamma}_{n,\text{new}}^T (\boldsymbol{\Gamma}_n^{-1} - \boldsymbol{\Gamma}_n^{-1} \boldsymbol{\Sigma} \boldsymbol{\Gamma}_n^{-1}) \boldsymbol{\gamma}_{n,\text{new}},$$

where $\boldsymbol{\gamma}_{n,\text{new}} = \sigma_f^2 [R_{\theta_f}(\mathbf{x}_i, \mathbf{x}_{\text{new}})]_{i=1}^n + \sigma_\delta^2 [\mathbf{0}_{n-m}, R_{\theta_\delta}(\mathbf{x}_i, \mathbf{x}_{\text{new}})]_{i=1}^m$ is the covariance vector and $\boldsymbol{\Gamma}_n = \sigma_f^2 [R_{\theta_f}(\mathbf{x}_i, \mathbf{x}_j)]_{i=1}^n [R_{\theta_f}(\mathbf{x}_i, \mathbf{x}_j)]_{j=1}^n + \text{diag}(\mathbf{0}_{n-m}, \sigma_\epsilon^2 \mathbf{I}_m + \sigma_\delta^2 \times [R_{\theta_\delta}(\mathbf{x}_i, \mathbf{x}_j)]_{i=1}^m [R_{\theta_\delta}(\mathbf{x}_i, \mathbf{x}_j)]_{j=1}^m)$ is the

covariance matrix. Here, $\hat{\mathbf{y}}_c = \mathbb{E}[\mathbf{y}'_c | \mathbf{f}, \mathbf{y}_o, \mathbf{y}'_c \geq \mathbf{c}]$ is the expected response for latent vector $\mathbf{y}'_c$, given data $\{\mathbf{f}, \mathcal{Y}_m\}$, and $\Sigma_c = \text{Var}[\mathbf{y}'_c | \mathbf{f}, \mathbf{y}_o, \mathbf{y}'_c \geq \mathbf{c}]$ is its conditional variance, with $\Sigma = \text{diag}(\mathbf{0}_{n-n_c}, \Sigma_c)$. While such equations appear quite involved, they are simply the bifidelity extensions of the earlier GP modeling equations (2.3) and (2.4). For simplicity, we have overloaded some notations from (2.3) and (2.4) here; the difference should be clear from the context.

3.2. Bifidelity design criterion. Now, we extend the ICMSE design to the bifidelity setting. The goal is to design *physical* experiment runs (which may be censored), given auxiliary computer experiment data (which are not censored).

Under the above bifidelity GP model the following proposition gives an explicit expression for the ICMSE design criterion.

PROPOSITION 3. *With experimental data $\{\mathbf{f}, \mathcal{Y}_m\}$, the proposed ICMSE criterion has the following explicit expression:*

$$\begin{aligned} \text{ICMSE}(\mathbf{x}_{n+1}) &= \int_{[0,1]^p} \mathbb{E}_{Y_{n+1} | \mathbf{f}, \mathcal{Y}_m} [\text{Var}(\xi(\mathbf{x}_{\text{new}}) | \mathbf{f}, \mathcal{Y}_m, Y_{n+1})] d\mathbf{x}_{\text{new}} \\ (3.6) \quad &= \int_{[0,1]^p} \sigma_{\text{new}}^2 - \boldsymbol{\gamma}_{n+1, \text{new}}^T \boldsymbol{\Gamma}_{n+1}^{-1} \mathbf{H}_c(\mathbf{x}_{n+1}) \boldsymbol{\Gamma}_{n+1}^{-1} \boldsymbol{\gamma}_{n+1, \text{new}} d\mathbf{x}_{\text{new}}, \end{aligned}$$

where $\sigma_{\text{new}}^2 = \text{Var}[\xi(\mathbf{x}_{\text{new}}) | \mathbf{f}, \mathcal{Y}_m]$ and $\boldsymbol{\gamma}_{n+1, \text{new}}$ and $\boldsymbol{\Gamma}_{n+1}$ follow from (3.4) and (3.5). The matrix $\mathbf{H}_c(\mathbf{x}_{n+1})$ has an easy-to-evaluate expression given in Appendix B.1 of the Supplementary Material (Chen et al. (2022)).

The proof can be found in Appendix B.1. The following corollary gives a simplification of (3.6) under a product correlation structure.

COROLLARY 1. *Suppose $R_{\theta_f}(\cdot, \cdot)$ and $R_{\theta_\delta}(\cdot, \cdot)$ are product correlation functions,*

$$(3.7) \quad R_{\theta_f}(\mathbf{x}, \mathbf{x}') = \prod_{l=1}^p R_{\theta_f}^{(l)}(x_l, x'_l), \quad R_{\theta_\delta}(\mathbf{x}, \mathbf{x}') = \prod_{l=1}^p R_{\theta_\delta}^{(l)}(x_l, x'_l),$$

with $\mathbf{x} = [x_1, \dots, x_p]^T$. Then, the ICMSE criterion (3.6) can be further simplified as

$$(3.8) \quad \text{ICMSE}(\mathbf{x}_{n+1}) = \bar{\sigma}^2 - \text{tr}(\boldsymbol{\Gamma}_{n+1}^{-1} \mathbf{H}_c(\mathbf{x}_{n+1}) \boldsymbol{\Gamma}_{n+1}^{-1} \boldsymbol{\Lambda}),$$

where $\bar{\sigma}^2 = \int \sigma_{\text{new}}^2 d\mathbf{x}_{\text{new}}$ and $\boldsymbol{\Lambda}$ is an $(n+1) \times (n+1)$ matrix with (i, j) th entry:

$$(3.9) \quad \Lambda_{ij} = \prod_{l=1}^p \left[\int_0^1 \zeta^{(l)}(x_{i,l}, x) \zeta^{(l)}(x_{j,l}, x) dx \right], \quad \text{and}$$

$$\zeta^{(l)}(z, x) = R_{\theta_f}^{(l)}(z, x) + \mathbb{1}_{\{i > (n-m)\}} R_{\theta_\delta}^{(l)}(z, x).$$

The key simplification from Corollary 1 is that it reduces the p -dimensional integral in the ICMSE criterion (3.6) to a product of 1D integrals which are more easily computed. Furthermore, if Gaussian correlation functions are used, these integrals can be reduced to error functions which yield an easy-to-evaluate design criterion for ICMSE (see Appendix B.2 for details). Given the computational complexities of censored data, this simplification allows for efficient design optimization. Corollary 1 is motivated by the simplification of the IMSE criterion in Sacks, Schiller and Welch (1989). The proof can be found in Appendix B.2 of the Supplementary Material (Chen et al. (2022)).

The interpretation of the bifidelity ICMSE criterion (3.6) is analogous to that of the single-fidelity ICMSE criterion (2.11). Similar to the censoring adjustment function, the matrix $\mathbf{H}_c(\cdot)$ factors in the posterior probability of censoring over the input space and is used to adjust the variance reduction term in the criterion. Viewed this way, the ICMSE criterion (3.6) provides the same design trade-off as before: the next design point should jointly: (i) avoid censored regions by adaptively identifying such regions from data at hand and (ii) minimize predictive uncertainty from the GP model.

3.3. An adaptive algorithm for sequential design. We present next an adaptive algorithm ICMSE for implementing the proposed ICMSE design. Algorithm 1 applies for both the single-fidelity setting (with flag $I_{\text{BF}} = 0$) in Section 2 and the bifidelity setting (with flag $I_{\text{BF}} = 1$) in Section 3. First, an initial n_{ini} -point design is set up for initial experimentation: physical experiments for the single-fidelity setting and computer experiments for the bifidelity setting. In our implementation we used the maximum projection (MaxPro) design, proposed by Joseph, Gul and Ba (2015), which provides good projection properties and thereby good GP predictive performance. Next, the following two steps are performed iteratively: (i) using observed data $\{\mathbf{f}, \mathcal{Y}_m\}$, the GP model parameters are estimated using maximum likelihood; (ii) the next design point $\mathbf{x}_{n+1}^*$ is then obtained by minimizing the ICMSE criterion (equation (2.11) for the single-fidelity setting, equation (3.6) for the bifidelity setting), along with its corresponding response Y_{n+1} . This is then repeated until a desired number of samples is obtained.

To optimize the ICMSE criterion, we use standard numerical optimization methods in the R package `nloptr` (Ypma, Borchers and Eddelbuettel (2014)), in particular, the Nelder–Mead method (Nelder and Mead (1965)). The main computational bottleneck in optimization is evaluating moments of the truncated multivariate normal distribution for $\mathbf{H}_c(\cdot)$ (see equations (A.10) and (B.3) in the Supplementary Material, Chen et al. (2022)). In our implementation these moments are efficiently computed using the R package `tmvtnorm` (Wilhelm and Manjunath (2010)). Appendix C details further computational steps for speeding-up design optimization, involving an approximation of the expected variance term via a plug-in estimator. Similar to the standard IMSE criterion, the ICMSE criterion can be quite multimodal. We, therefore, suggest performing multiple random restarts of the optimization and taking the solution with the best objective value as the new design point.

Algorithm 1 ICMSE($n_{\text{ini}}, n_{\text{seq}}, c, I_{\text{BF}}$): Adaptive design under censoring

- 1: **if** $I_{\text{BF}} = 0$, **then** ▷ Single-fidelity
 - 2: Generate an n_{ini} -run initial MaxPro design $\mathbf{x}_{1:n_{\text{ini}}}$
 - 3: Collect initial data $\mathcal{Y}_{n_{\text{ini}}}$ at inputs $\mathbf{x}_{1:n_{\text{ini}}}$ from physical experiments
 - 4: Estimate model parameters $\{\mu_\xi, \sigma_\xi^2, \boldsymbol{\theta}_\xi\}$ using MLE from initial data $\mathcal{Y}_{n_{\text{ini}}}$
 - 5: **else** ▷ Bifidelity
 - 6: Generate an n_{ini} -run initial MaxPro design $\mathbf{x}_{1:n_{\text{ini}}}$
 - 7: Collect initial data $\mathbf{f}$ at inputs $\mathbf{x}_{1:n_{\text{ini}}}$ from computer experiments
 - 8: Estimate model parameters $\{\mu_f, \sigma_f^2, \boldsymbol{\theta}_f\}$ using MLE from $\mathcal{Y}_{n_{\text{ini}}}$, and let $\sigma_\delta^2 = 0$
 - 9: **for** $k = n_{\text{ini}} + 1, \dots, n_{\text{ini}} + n_{\text{seq}}$ **do** ▷ n_{seq} sequential runs
 - 10: **if** $I_{\text{BF}} = 0$, **then**
 - 11: Obtain new design point $\mathbf{x}_k^*$ by minimizing ICMSE criterion (2.11)
 - 12: **else**
 - 13: Obtain new design point $\mathbf{x}_k^*$ by minimizing ICMSE criterion (3.6)
 - 14: Perform experiment at $\mathbf{x}_k^*$, and collect response Y_k (which may be censored)
 - 15: Update model parameter estimates using new data
-

3.4. *Illustrative examples with adaptive design algorithm.* We first illustrate the proposed algorithm ICMSE on a 1D bifidelity example. Suppose the computer simulation is given by

(3.10)
$$f(x) = 0.5 \sin(10(x - 1.02)^2) + 0.1$$

with the same physical experiment settings as in Section 2.3. We begin with an $n_{\text{ini}} = 6$ -run equally-spaced points $x_{1:6}^f = \{(i - 1)/5\}_{i=1}^6$ for computer experiments. We then perform a sequential $n_{\text{seq}} = 20$ -run design for physical experiments using the algorithm ICMSE. The Gaussian correlation function is used for both GPs. In addition to the two IMSE methods in Section 2.3, we consider an additional “sequential MaxPro” method (Joseph (2016)) which implements a sequential space-filling design. Again, for fair comparison we use the censored GP model (2.3)–(2.4) for evaluating predictions for all design methods. The simulation is replicated 20 times.

We consider two evaluation metrics for predictive performance: RMSE and the interval score proposed in Gneiting and Raftery (2007). The first assesses predictive accuracy, and the second assesses uncertainty quantification. The $(1 - \alpha)\%$ interval score is defined as

(3.11)
$$\text{IS}(\xi_l, \xi_u; \xi) = (\xi_u - \xi_l) + \frac{2}{\alpha}(\xi_l - \xi)_+ + \frac{2}{\alpha}(\xi - \xi_u)_+,$$

where $(a)_+ = \max(a, 0)$, ξ is the ground truth and $[\xi_l, \xi_u]$ is an $(1 - \alpha)\%$ predictive interval. Here, we set $1 - \alpha = 68\%$, with predictive interval $[\hat{\xi} - \sqrt{s^2}, \hat{\xi} + \sqrt{s^2}]$, where $\hat{\xi}$ and s^2 are obtained from (3.4) and (3.5). The mean interval score (MIS) is then computed over the entire test set. We also compared computation time on a 1.4 GHz Quad-Core Intel Core i5 laptop.

Figure 6 shows the log-RMSE (a), log-MIS (b), and log-computation time (c) for the four considered methods. The ICMSE method yields noticeable improvements over the IMSE and sequential MaxPro methods with smaller RMSE and MIS values for most sequential run sizes. One reason for this is that the proposed method integrates the possibility of a new observation being censored directly within the design criterion which allows it to minimize predictive uncertainty by avoiding censored regions. While ICMSE requires more computation time, compared to the two baseline methods, the computation complexities appear to be comparable. Here, the IMSE-Cen method is terminated early after 12 sequential runs, due to numerical instabilities (and, thereby, expensive computation) in evaluating the predictive equations. This is because, by ignoring censoring, IMSE-Cen overestimates the potential variance reduction in censored regions, leading to many sequential points very close together in such regions.

Figure 7 shows the sequential design points and the predicted mean responses $\hat{\xi}(\cdot)$ for a single replication. Compared to existing methods, ICMSE yields visually improved prediction in both the censored (shaded) and uncensored (clear) regions. One reason for this is that the ICMSE criterion chooses points which jointly: (i) avoid censored regions and (ii)

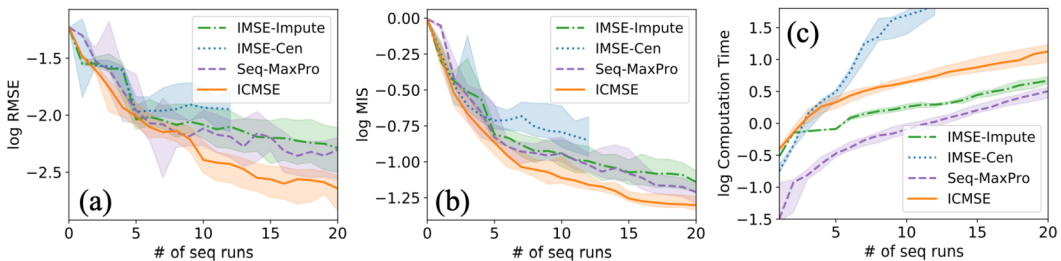


FIG. 6. The 1D bifidelity example in (3.10): (a), (b), and (c) show the log-RMSE, log-MIS, and log-computation time (in seconds) over the number of sequential runs for each method. Solid lines mark the median over the 20 replications, and shaded regions mark the 25%–75% quantiles.

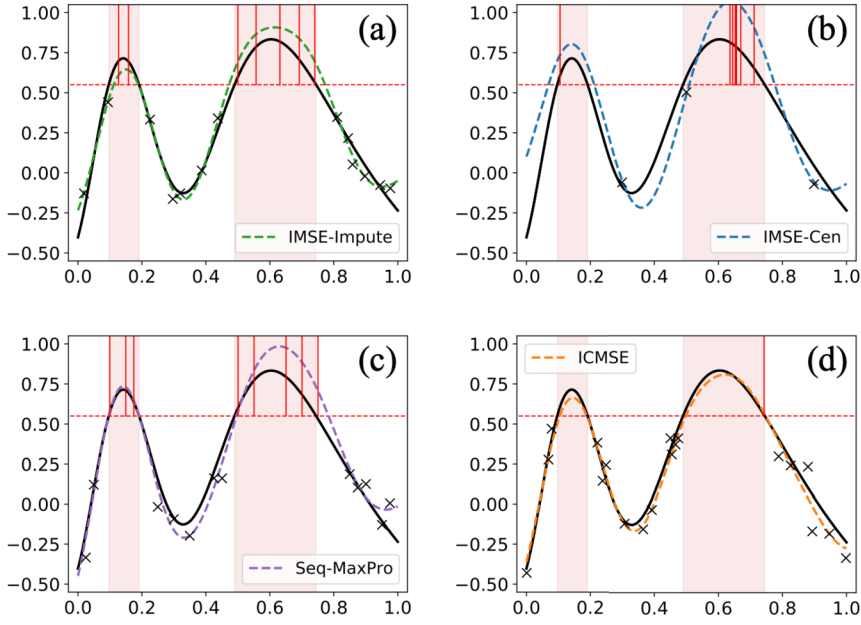


FIG. 7. One replication in the 1D bifidelity example (3.10): (a)-(d) show the observed runs (black crosses, with censored runs indicated by red vertical lines) and the predictor $\hat{\xi}(\cdot)$ (dotted lines), using the 4 considered methods. Here, black lines mark the mean physical experiment $\xi(\cdot)$, and shaded regions mark the censored regions.

minimize predictive uncertainty. For (i), note that only $1/20 = 5\%$ of sequential runs are censored for ICMSE, whereas $7/20 = 35\%$, $8/20 = 40\%$, and $9/12 = 75\%$ of sequential runs are censored for IMSE-Impute, MaxPro, and IMSE-Cen, respectively. This shows that ICMSE effectively estimates the posterior probability of censoring, and avoids regions with high probabilities for sampling. For (ii), Figure 7(d) shows that the sequential runs from ICMSE are far away from existing points and also concentrated near the boundary of the censored region. Intuitively, this minimizes predictive uncertainty by ensuring design points well explore the input space while avoiding losing information due to censoring.

Next, we conduct a 2D simulation. The computer simulation and mean physical experiment functions are taken from [Xiong, Qian and Wu \(2013\)](#),

$$(3.12) \quad f(\mathbf{x}) = \frac{1}{4}\xi\left(x_1 + \frac{1}{20}, x_2 + \frac{1}{20}\right) + \frac{1}{4}\xi\left(x_1 + \frac{1}{20}, \left(x_2 - \frac{1}{20}\right)_+\right) \\ + \frac{1}{4}\xi\left(x_1 - \frac{1}{20}, x_2 + \frac{1}{20}\right) + \frac{1}{4}\xi\left(x_1 - \frac{1}{20}, \left(x_2 - \frac{1}{20}\right)_+\right),$$

$$(3.13) \quad \xi(\mathbf{x}) = \left[1 - \exp\left(-\frac{1}{2x_2}\right)\right] \frac{2300x_1^3 + 1900x_1^2 + 2092x_1 + 6}{100x_1^3 + 500x_1^2 + 4x_1 + 20},$$

with measurement variance $\sigma_\epsilon^2 = 1$ and a right censoring limit of $c = 10$. We begin with an initial $n_{\text{ini}} = 12$ -run MaxPro design for the computer experiment, then add $n_{\text{seq}} = 40$ sequential runs for physical experiments using ICMSE. This is then replicated 20 times.

Table 2 summarizes the median RMSE, MIS, and computation time after five, 15, and 40 sequential runs. We see that ICMSE yields noticeably lower RMSE and MIS, suggesting the proposed design method gives a better predictive performance. While computationally more expensive than IMSE-Impute and MaxPro, the proposed ICMSE method appears more effective at integrating censoring information for sequential design which leads to improved predictive performance.

TABLE 2

The median RMSE, MIS, and computation time, under different sequential run sizes for the three considered design methods in a 2D bifidelity example (3.12)

Sequential runs	RMSE			MIS			Computation Time (in s)		
	5	15	40	5	15	40	5	15	40
IMSE-Impute	1.62	1.38	1.14	4.91	4.03	3.29	4.14	12.18	64.34
IMSE-Cen	1.61	1.46	–	4.57	4.27	–	25.13	121.24	–
Seq-MaxPro	1.74	1.36	1.12	5.58	4.01	3.22	3.03	10.18	57.24
ICMSE	1.40	1.21	0.97	4.58	3.80	3.01	9.77	25.01	95.67

4. Case studies. We now return to the two motivating applications. For the wafer manufacturing problem (which only has physical experiments), we use the single-fidelity ICMSE method in Section 2. For the surgical planning application (which has both computer and physical experiments), we use the bifidelity ICMSE method in Section 3.

4.1. Thermal processing in wafer manufacturing. Consider first the wafer manufacturing application in Section 1.2, where an engineer is interested in how a wafer chip’s heating performance is affected by six process input variables that control wafer thickness, rotation speed, heating laser (i.e., its moving speed, radius, and power), and heating time. The response of interest $\xi(\mathbf{x})$ is the minimum temperature over the wafer which provides an indication of the wafer’s quality after thermal processing. Standard industrial temperature sensors have a measurement limit of $c = 350^\circ\text{C}$ (Thermo Electric Company (2010)), and temperatures greater than this limit are censored in the experiment.

As mentioned earlier, certain physical experiments are not only costly (e.g., wafers and laser operation can be expensive), but also time consuming to perform (e.g., each experiment requires a recalibration of thermal sensors as well as a warmup and cooldown of the laser beam). To compare the sequential performance of these methods over a large number of runs, we mimic the costly physical experiments¹ with COMSOL Multiphysics simulations (Figure 8(a)) which provides a realistic representation of heat diffusion physics (Dickinson, Ekström and Fontes (2014)). Measurement noise is then added, following an i.i.d. zero-mean normal distribution with standard deviation $\sigma_\epsilon = 1.0^\circ\text{C}$.

The set-up is as follows. We start with an $n_{\text{ini}} = 30$ -run initial experiment, then perform $n_{\text{seq}} = 45$ sequential runs. Note that the total number of $n_{\text{ini}} + n_{\text{seq}} = 75$ runs is slightly

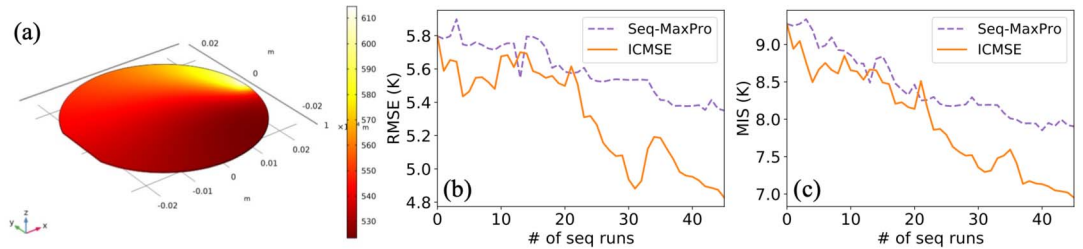


FIG. 8. (a) The temperature contour over the wafer chip, simulated using COMSOL Multiphysics. (b) and (c) show the RMSE and MIS of the fitted GP models over the sequential design size, respectively, for the two design methods.

¹The surgical planning application in Section 4.2 performs actual physical experiments but provides fewer sequential runs, due to the expensive nature of such experiments.

more than the rule-of-thumb sample size of $10p$ recommended by Loeppky, Sacks and Welch (2009); this is to ensure good predictive accuracy under censoring. Due to the limited budget, the proposed ICMSE method is compared with only the sequential MaxPro method. This is because, from simulations in Section 3.4, it provides the best predictive performance and is the fastest among the three baseline methods.

The fitted GP models are then tested on temperature data generated (without noise) on a 200-run Sobol' sequence (Sobol' (1967)). Of these 200 test samples, 25 samples have minimum temperatures, which exceed the censoring limit of $c = 350^\circ\text{C}$, suggesting that, roughly, 12.5% of the design space leads to censoring. It is important to note that predictive accuracy is desired for *both* censored and uncensored test runs, since the engineering objective is to predict the experimental response surface prior to censoring. This allows industrial engineers to explore a wide range of quality requirements in manufacturing wafers with low temperatures (which are uncensored in experimentation) and high temperatures (which may potentially be censored).

4.1.1. Predictive performance. Figure 8 compares the RMSE and MIS after $n_{\text{seq}} = 45$ sequential runs. While both sequential methods provide relatively steady improvements in RMSE and MIS, the proposed ICMSE method gives a greater predictive improvement over MaxPro. In particular, with 45 sequential runs ICMSE achieves an RMSE reduction of $(5.8 - 4.8)/5.8 = 17.2\%$ over the initial 30 runs which is greater than the RMSE reduction of $(5.8 - 5.35)/5.8 = 7.8\%$ for MaxPro. Similarly, for MIS, ICMSE achieves a reduction of $(9.28 - 6.95)/9.28 = 25.1\%$, compared to $(9.28 - 7.9)/9.28 = 14.8\%$ for MaxPro. This can again be explained by the fact that ICMSE jointly avoids censoring and minimizes predictive uncertainty. Here, ICMSE yields no censored measurements, whereas MaxPro yields five censored measurements (a censoring rate of $5/45 = 11.1\%$). Moreover, ICMSE *adaptively* chooses points that minimize predictive uncertainty of the GP model under censoring. This can be seen from Figure 8(b) and (c): the ICMSE yields progressively lower RMSE and MIS values as sample size increases.

While the ICMSE provides noticeable improvements over the sequential MaxPro, the reductions in RMSE for both methods are only moderate in magnitude. One reason may be that the underlying response surface for minimum temperature is quite nonsmooth over the parameter space, which makes it difficult to learn with a limited number of experimental runs, particularly in censored regions. It is also worth noting, however, that even moderate improvements in predictive accuracy can lead to significant improvements in wafer manufacturing. As mentioned in Section 1.2, the fitted GP model is used to find process settings that jointly minimize operational costs while meeting target quality requirements. The improved predictive model using ICMSE can cut down waste in heating power and reduce the number of wafers to be remanufactured which results in significant cost reductions in the wafer manufacturing process.

4.2. 3D-printed aortic valves for surgical planning. Consider next the surgical planning application in Section 1.1 which uses state-of-the-art 3D printing technology to mimic biological tissues. Here, doctors are interested in predicting the stiffness of the printed organs with different metamaterial geometries. We will consider three design inputs $\mathbf{x} = (A, \omega, d)$, which parametrize a standard sinusoidal form of the substructure curve $I(t) = A \sin(\omega t)$, with diameter d (see Figure 2(b) for a visualization). This parametric form has been shown to provide effective tissue-mimicking performance in prior studies (Chen et al. (2018a), Wang et al. (2016)). The response of interest $\xi(\mathbf{x})$ is the elastic modulus at a strain level of 8% which quantifies the stiffness at a similar load situation inside the human body (Wang et al. (2016)).

We use the bifidelity ICMSE design framework in Section 3, since a preconduted database of computer simulations is available, and we are interested in the sequential design of physical experiments. Computer simulations were performed with finite element analysis (Zienkiewicz et al. (1977)) using COMSOL Multiphysics. Physical experiments were performed in two steps: the aortic valves were first 3D-printed by the Connex 350 machine (Stratasys Ltd.), and then its stiffness was measured by a load cell using uniaxial tensile tests (see Figure 2(c); Wang et al. (2016)). Here, physical experiments are very costly, requiring expensive material and printing costs as well as several hours of an experimenter’s time per sample. Censoring is also present in physical experiments; this happens when the force measurement of the load cell exceeds the standard limit of $15N$, corresponding to a modulus upper limit of $c = 0.23MPa = 15N(\text{force})/8mm^2(\text{area})/8\%(\text{deformation})$.

The following design set-up is used. We start with an $n_{\text{ini}} = 25$ -run initial computer experiment design and then perform $n_{\text{seq}} = 8$ sequential runs using physical experiments. The limited number of sequential runs is due to the urgent demand of the patients; in such cases, only one to two days of surgical planning can be afforded (Chen et al. (2018a)). Since physical experiments require tedious 3D printing and a tensile test (around 1.5 hours per run), this means only a handful of runs can be performed in urgent cases. As before, we compare the proposed ICMSE method with the MaxPro method. The fitted GP models from both methods are tested on the physical experiment data from a 20-run Sobol’ sequence. Among these 20 runs, five of them are censored due to the load cell limit; in such cases we reperform the experiment using a different testing machine with a wider measurement range. The re-experimentation is typically *not* feasible in urgent surgical scenarios, since it requires even more time-consuming tests and higher material costs.

4.2.1. Predictive performance. Figure 9 compares the predictive performance of the two design methods over $n_{\text{seq}} = 8$ sequential runs. While MaxPro shows some stagnation in RMSE and MIS improvement, ICMSE yields more noticeable improvements, as sample size increases. More specifically, ICMSE achieves an RMSE reduction of roughly $(0.0315 - 0.0235)/0.0315 = 25.4\%$ over the initial GP model (fitted using 25 computer experiment runs) which is much greater than the RMSE reduction of $(0.0315 - 0.0288)/0.0315 = 8.57\%$ for MaxPro. Similar improvements can be seen by inspecting MIS. This can again be attributed to the key design trade-off. ICMSE adaptively identifies and avoids censored regions on the design space using the fitted bifidelity model (3.4). Here, the proposed method yields no censored measurements, whereas MaxPro yields three censored measurements (a censoring rate of $3/8 = 37.5\%$). Furthermore, in contrast to MaxPro, which encourages physical runs to be “space-filling” to the initial computer experiment runs, ICMSE instead incorporates censoring information within an adaptive design scheme which allows for improved predictive performance.

We investigate next the predictive performance of both designs within the *censored* region. This region (corresponding to stiff valves) is important for prediction, since such valves can

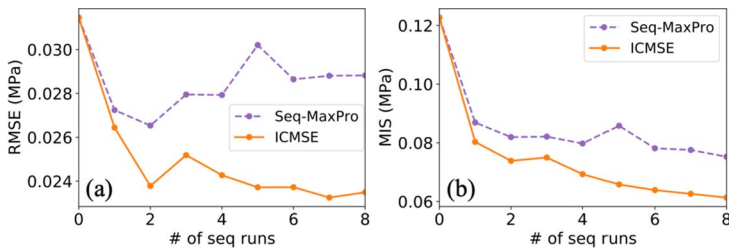


FIG. 9. RMSE (a) and MIS (b) for the two sequential design methods, over the number of sequential runs.

TABLE 3

RMSE on the full test set, the five censored runs and the 15 observed runs for the two sequential design methods

	MaxPro	ICMSE
Full	0.0288	0.0235
Censored	0.0462	0.0416
Observed	0.0199	0.0126

be used to mimic older patients (Sicard et al. (2018)). We divide the test set (20 runs in total) into two categories: observed runs (15 in total) and censored runs (five in total). The responses for the latter are obtained via new experiments on a stiffer load cell (which, as mentioned in Section 1.1, is typically not feasible in practice). Table 3 compares the RMSE of the two methods for the censored and uncensored test runs. For both methods the RMSE for observed test runs is much smaller than that for censored test runs, which is as expected. For censored test runs, ICMSE also performs slightly better than MaxPro, with $(0.0462 - 0.0416)/0.0462 = 9.9\%$ lower RMSE. One reason for this is that ICMSE encourages new runs near (but not within) censored regions (see Figure 6), to maximize information under censoring. Because of this adaptivity, ICMSE achieves better predictive performance within the censored region without putting any sequential runs in this region.

The improved performance for ICMSE can greatly improve the surgery planning procedure. As discussed in Section 1.1, the fitted GP model is used to optimize a polymer geometry which mimics a patient's tissue stiffness. An improved predictive model leads to better tissue-mimicking performance of the printed valves which then translates to improved surgery success rates. Indeed, in a recent study Chen et al. (2021) it was shown that a predictive model with 42% improvement in predictive performance leads to six-fold error reduction for tissue-mimicking. We would expect a similar improvement in tissue-mimicking performance here, when comparing ICMSE with the baseline design methods. The resulting improved artificial valves from ICMSE then leads to improved success rates for heart surgeries.

4.2.2. Discrepancy modeling. The ICMSE method can also yield valuable insights on the discrepancy between computer simulation and reality. The learning of this discrepancy from data is important for several reasons: it allows doctors to: (i) pinpoint where simulations may be unreliable, (ii) identify potential root causes for this discrepancy, and (iii) improve the simulation model to better mimic reality. In our modeling framework this discrepancy can be estimated as

$$(4.1) \quad \hat{\delta}(x) = \hat{\xi}(x) - \hat{f}(x),$$

where $\hat{\xi}(x)$ is the predictor for the physical experiment mean, fitted using 25 initial computer experiment runs and eight physical experiment runs, and $\hat{f}(x)$ is the computer experiment model fitted using only the 25 initial runs.

Figure 10 shows the fitted discrepancy $\hat{\delta}(x)$ as a function of each pair of design inputs, with the third input fixed. These plots reveal several interesting insights. First, when the diameter d is moderate (i.e., $d \in [0.2, 0.7]$), Figure 10(a) and (b) show that the discrepancy is quite small; however, when d is small (i.e., $[0, 0.2]$) or large (i.e., $[0.7, 1]$), the discrepancy can be quite large. This is related to the limitations of finite element modeling. When diameter d is small, the simulations can be inaccurate, since the mesh size would be relatively large compared to d . When diameter d is large, simulations can again be inaccurate, due to the violation of the perfect interface assumption between the two printed polymers. Second, from Figure 10, model discrepancy also appears to be largest when all design inputs are large (i.e., close to

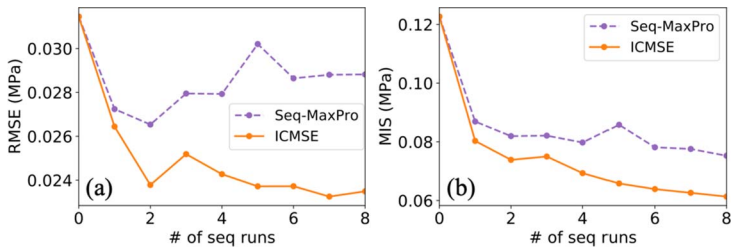


FIG. 10. Visualization of the estimated discrepancy $\hat{\delta}(\cdot)$ (a) over d and A with fixed $\omega = 1$, (b) over d and ω with fixed $A = 1$, and (c) over A and ω with fixed $d = 1$.

1). This suggests that simulations can be unreliable, when the stiff material is both thick ($d \approx 1$) and fluctuating ($\omega \approx 1$, $A \approx 1$). Finally, the model discrepancy is mostly positive over the design domain, revealing smaller stiffness evaluation via simulation, compared to physical evaluation. This may be caused by the hardening of 3D-printed samples, due to exposure to natural light, as an aging property for the polymer family (e.g., see Liao et al. (1998)). Therefore, the printed aortic valves should be stored in dark storage cells for surgical planning to minimize exposure to light.

5. Conclusion. In this paper we proposed a novel integrated censored mean-squared error (ICMSE) method for adaptively designing physical experiments under response censoring. The ICMSE method iteratively performs two steps: it first estimates the posterior probability of a new observation being censored and then selects the next design point which yields the greatest reduction in predictive uncertainty under censoring. We derived easy-to-evaluate expressions for the ICMSE design criterion, in both the single-fidelity and bifidelity settings, and presented an adaptive design for efficient implementation. We then demonstrated the effectiveness of the proposed ICMSE method over existing methods in real-world applications on 3D-printed aortic valves for surgical planning and thermal processing in wafer manufacturing. An R package is currently in development and will be released soon.

Looking ahead, there are several interesting directions to be explored. In this work the censoring limit c is assumed to be known. While this is true for the two motivating applications, there are other problems where c is unknown and needs to be learned from data; it would be useful to extend ICMSE for such problems. Another direction is to explore the connection between the ICMSE method and the multipoints expected improvement method (Ginsbourger, Le Riche and Carraro (2010)) which may speed up design optimization via rejection sampling. Finally, for the bifidelity ICMSE it would be interesting to explore more elaborate design schemes that allow for additional computer experiments to be added sequentially.

Funding. This work is supported by NSF CSSI Frameworks 2004571, NSF CMMI grant 1921646, and Piedmont Heart Institute.

SUPPLEMENTARY MATERIAL

Additional proofs and implementation details (DOI: 10.1214/21-AOAS1512SUPP;.pdf). We provide in this supplement further details on technical results.

REFERENCES

ANKENMAN, B., NELSON, B. L. and STAUM, J. (2010). Stochastic kriging for simulation metamodeling. *Oper. Res.* **58** 371–382. MR2674803 <https://doi.org/10.1287/opre.1090.0754>

BECT, J., BACHOC, F. and GINSBOURGER, D. (2019). A supermartingale approach to Gaussian process based sequential design of experiments. *Bernoulli* **25** 2883–2919. MR4003568 <https://doi.org/10.3150/18-BEJ1074>

- BINOIS, M., HUANG, J., GRAMACY, R. B. and LUDKOVSKI, M. (2019). Replication or exploration? Sequential design for stochastic simulation experiments. *Technometrics* **61** 7–23. MR3933655 <https://doi.org/10.1080/00401706.2018.1469433>
- BORTH, D. M. (1996). Optimal experimental designs for (possibly) censored data. *Chemom. Intell. Lab. Syst.* **32** 25–35.
- BROOKS, R. J. (1982). On the loss of information through censoring. *Biometrika* **69** 137–144. MR0655678 <https://doi.org/10.1093/biomet/69.1.137>
- CAO, F., BA, S., BRENNEMAN, W. A. and JOSEPH, V. R. (2018). Model calibration with censored data. *Technometrics* **60** 255–262. MR3804253 <https://doi.org/10.1080/00401706.2017.1345704>
- CHEN, R.-B., WANG, W. and WU, C. F. J. (2017). Sequential designs based on Bayesian uncertainty quantification in sparse representation surrogate modeling. *Technometrics* **59** 139–152. MR3635038 <https://doi.org/10.1080/00401706.2016.1172027>
- CHEN, J., WANG, K., ZHANG, C. and WANG, B. (2018a). An efficient statistical approach to design 3D-printed metamaterials for mimicking mechanical properties of soft biological tissues. *Addit. Manuf.* **24** 341–352.
- CHEN, J., XIE, Y., WANG, K., WANG, Z. H., LAHOTI, G., ZHANG, C., VANNAN, M. A., WANG, B. and QIAN, Z. (2018b). Generative invertible networks (GIN): Pathophysiology-interpretable feature mapping and virtual patient generation. arXiv preprint. Available at [arXiv:1808.04495](https://arxiv.org/abs/1808.04495).
- CHEN, J., MAK, S., JOSEPH, V. R. and ZHANG, C. (2021). Function-on-function kriging, with applications to 3D printing of aortic tissues. *Technometrics* **63** 384–395.
- CHEN, J., MAK, S., JOSEPH, V. R. and ZHANG, C. (2022). Supplement to “Adaptive design for Gaussian process regression under censoring.” <https://doi.org/10.1214/21-AOAS1512SUPP>
- THERMO ELECTRIC COMPANY (2010). Wafer sensors. Available at <http://www.te-direct.com/products/silicon-wafers/>.
- DA VEIGA, S. and MARREL, A. (2012). Gaussian process modeling with inequality constraints. *Ann. Fac. Sci. Toulouse Math.* (6) **21** 529–555. MR3076411 <https://doi.org/10.5802/afst.1344>
- DICKINSON, E. J., EKSTRÖM, H. and FONTES, E. (2014). COMSOL multiphysics®: Finite element software for electrochemical analysis. A mini-review. *Electrochem. Commun.* **40** 71–74.
- DING, L., MAK, S. and WU, C. F. J. (2019). Bdrygp: A new Gaussian process model for incorporating boundary information. arXiv preprint. Available at [arXiv:1908.08868](https://arxiv.org/abs/1908.08868).
- FETEIRA, A. (2009). Negative temperature coefficient resistance (NTCR) ceramic thermistors: An industrial perspective. *J. Amer. Ceram. Soc.* **92** 967–983.
- GIBSON, I., ROSEN, D. W. and STUCKER, B. (2014). *Additive Manufacturing Technologies* **17**. Springer, Berlin.
- GINSBOURGER, D., LE RICHE, R. and CARRARO, L. (2010). Kriging is well-suited to parallelize optimization. In *Computational Intelligence in Expensive Optimization Problems* 131–162. Springer, Berlin.
- GNEITING, T. and RAFTERY, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *J. Amer. Statist. Assoc.* **102** 359–378. MR2345548 <https://doi.org/10.1198/016214506000001437>
- GOODSON, K., FLIK, M., SU, L. and ANTONIADIS, D. A. (1993). Annealing-temperature dependence of the thermal conductivity of lpcvd silicon-dioxide layers. *IEEE Electron Device Lett.* **14** 490–492.
- GRAMACY, R. B. and APLEY, D. W. (2015). Local Gaussian process approximation for large computer experiments. *J. Comput. Graph. Statist.* **24** 561–578. MR3357395 <https://doi.org/10.1080/10618600.2014.914442>
- GROOT, P., LUCAS, P., CANO, A., GÓMEZ-OLMEDO, M. and NIELSEN, T. (2012). Gaussian process regression with censored data using expectation propagation. In *PGM 2012: Proceedings of the Sixth European Workshop on Probabilistic Graphical Models, PGM’12* (A. Cano, M. Gómez-Olmedo and T. D. Nielsen, eds.) 115–122. DECSAI, Granada.
- HENKENJOHANN, N., GÖBEL, R., KLEINER, M. and KUNERT, J. (2005). An adaptive sequential procedure for efficient optimization of the sheet metal spinning process. *Qual. Reliab. Eng. Int.* **21** 439–455.
- JIN, R., CHANG, C.-J. and SHI, J. (2012). Sequential measurement strategy for wafer geometric profile estimation. *IIE Trans.* **44** 1–12.
- JOHNSON, M. E., MOORE, L. M. and YLVIKAKER, D. (1990). Minimax and maximin distance designs. *J. Statist. Plann. Inference* **26** 131–148. MR1079258 [https://doi.org/10.1016/0378-3758\(90\)90122-B](https://doi.org/10.1016/0378-3758(90)90122-B)
- JOSEPH, V. R. (2016). Rejoinder. *Qual. Eng.* **28** 42–44.
- JOSEPH, V. R., GUL, E. and BA, S. (2015). Maximum projection designs for computer experiments. *Biometrika* **102** 371–380. MR3371010 <https://doi.org/10.1093/biomet/asv002>
- KAUFMAN, C. G., BINGHAM, D., HABIB, S., HEITMANN, K. and FRIEMAN, J. A. (2011). Efficient emulators of computer experiments using compactly supported correlation functions, with an application to cosmology. *Ann. Appl. Stat.* **5** 2470–2492. MR2907123 <https://doi.org/10.1214/11-AOAS489>
- KENNEDY, M. C. and O’HAGAN, A. (2001). Bayesian calibration of computer models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **63** 425–464. MR1858398 <https://doi.org/10.1111/1467-9868.00294>
- LAM, C. Q. (2008). *Sequential Adaptive Designs in Computer Experiments for Response Surface Model Fit*. ProQuest LLC, Ann Arbor, MI. Thesis (Ph.D.)—The Ohio State University. MR2711949

- LIAO, K., SCHULTESIZ, C. R., HUNSTON, D. L. and BRINSON, L. C. (1998). Long-term durability of fiber-reinforced polymer-matrix composite materials for infrastructure applications: A review. *J. Adv. Mater.* **30** 3–40.
- LOEPPKY, J. L., SACKS, J. and WELCH, W. J. (2009). Choosing the sample size of a computer experiment: A practical guide. *Technometrics* **51** 366–376. MR2756473 <https://doi.org/10.1198/TECH.2009.08040>
- LÓPEZ-LOPERA, A. F., BACHOC, F., DURRANDE, N. and ROUSTANT, O. (2018). Finite-dimensional Gaussian approximation with linear inequality constraints. *SIAM/ASA J. Uncertain. Quantificat.* **6** 1224–1255. MR3857898 <https://doi.org/10.1137/17M1153157>
- MAK, S., SUNG, C.-L., WANG, X., YEH, S.-T., CHANG, Y.-H., JOSEPH, V. R., YANG, V. and WU, C. F. J. (2018). An efficient surrogate model for emulation and physics extraction of large eddy simulations. *J. Amer. Statist. Assoc.* **113** 1443–1456. MR3902221 <https://doi.org/10.1080/01621459.2017.1409123>
- MATHÉRON, G. (1963). Principles of geostatistics. *Econ. Geol.* **58** 1246–1266.
- MONROE, E. M. and PAN, R. (2008). Experimental design considerations for accelerated life tests with nonlinear constraints and censoring. *J. Qual. Technol.* **40** 355–367.
- MORRIS, M. D. and MITCHELL, T. J. (1995). Exploratory designs for computational experiments. *J. Statist. Plann. Inference* **43** 381–402.
- NELDER, J. A. and MEAD, R. (1965). A simplex method for function minimization. *Comput. J.* **7** 308–313. MR3363409 <https://doi.org/10.1093/comjnl/7.4.308>
- QIAN, Z., WANG, K., LIU, S., ZHOU, X., RAJAGOPAL, V., MEDURI, C., KAUTEN, J. R., CHANG, Y.-H., WU, C. et al. (2017). Quantitative prediction of paravalvular leak in transcatheter aortic valve replacement based on tissue-mimicking 3D printing. *JACC Cardiovasc. Imaging* **10** 719–731.
- QUIRK, M. and SERDA, J. (2001). *Semiconductor Manufacturing Technology* **1**. Prentice Hall, Upper Saddle River, NJ.
- RENGIER, F., MEHNDIRATTA, A., VON TENGG-KOBLIGK, H., ZECHMANN, C. M., UNTERHINNINGHOFEN, R., KAUCZOR, H.-U. and GIESEL, F. L. (2010). 3D printing based on imaging data: Review of medical applications. *Internat. J. Comput. Assisted Radiol. Surg.* **5** 335–341.
- SACKS, J., SCHILLER, S. B. and WELCH, W. J. (1989). Designs for computer experiments. *Technometrics* **31** 41–47. MR0997669 <https://doi.org/10.2307/1270363>
- SACKS, J., WELCH, W. J., MITCHELL, T. J. and WYNN, H. P. (1989). Design and analysis of computer experiments. *Statist. Sci.* **4** 409–435. MR1041765
- SANTNER, T. J., WILLIAMS, B. J. and NOTZ, W. I. (2018). *The Design and Analysis of Computer Experiments. Springer Series in Statistics*. Springer, New York. MR3887662
- SHEWRY, M. C. and WYNN, H. P. (1987). Maximum entropy sampling. *J. Appl. Stat.* **14** 165–170.
- SICARD, D., HAAK, A. J., CHOI, K. M., CRAIG, A. R., FREDENBURGH, L. E. and TSCHUMPERLIN, D. J. (2018). Aging and anatomical variations in lung tissue stiffness. *Am. J. Physiol., Lung Cell. Mol. Physiol.* **314** L946–L955. <https://doi.org/10.1152/ajplung.00415.2017>
- SINGH, R., FAKHRUDDIN, M. and POOLE, K. (2000). Rapid photothermal processing as a semiconductor manufacturing technology for the 21st century. *Appl. Surf. Sci.* **168** 198–203.
- SOBOL', I. M. (1967). On the distribution of points in a cube and the approximate evaluation of integrals. *Ž. Vyčisl. Mat. Mat. Fiz.* **7** 784–802.
- VAN GURP, M. and PALMEN, J. (1998). Time-temperature superposition for polymeric blends. *Rheol. Bull.* **67** 5–8.
- WANG, K., WU, C., QIAN, Z., ZHANG, C., WANG, B. and VANNAN, M. A. (2016). Dual-material 3D printed metamaterials with tunable mechanical properties for patient-specific tissue-mimicking phantoms. *Addit. Manuf.* **12** 31–37.
- WILHELM, S. and MANJUNATH, B. G. (2010). tmvtnorm: A package for the truncated multivariate normal distribution. *R J.* **2** 25–29.
- XIONG, S., QIAN, P. Z. G. and WU, C. F. J. (2013). Sequential design and analysis of high-accuracy and low-accuracy computer codes. *Technometrics* **55** 37–46. MR3038483 <https://doi.org/10.1080/00401706.2012.723572>
- YPMA, J., BORCHERS, H. and EDELBUETTEL, D. (2014). nloptr: R interface to nlopt. *R J.* **1**.
- ZIENKIEWICZ, O. C., TAYLOR, R. L., ZIENKIEWICZ, O. C. and TAYLOR, R. L. (1977). *The Finite Element Method* **36**. McGraw-Hill, London.