Active Learning for Gaussian Process Considering Uncertainties With Application to Shape Control of Composite Fuselage

Xiaowei Yue, *Member, IEEE*, Yuchen Wen, Jeffrey H. Hunt, and Jianjun Shi

Abstract—In the machine learning domain, active learning is an iterative data selection algorithm for maximizing information acquisition and improving model performance with limited training samples. It is very useful, especially for industrial applications where training samples are expensive, time-consuming, or difficult to obtain. Existing methods mainly focus on active learning for classification, and a few methods are designed for regression, such as linear regression or Gaussian process (GP). Uncertainties from measurement errors and intrinsic input noise inevitably exist in the experimental data, which further affects the modeling performance. The existing active learning methods do not incorporate these uncertainties for GP. In this article, we propose two new active learning algorithms for the GP with uncertainties, which are variance-based weighted active learning algorithm and D-optimal weighted active learning algorithm. Through numerical study, we show that the proposed approach can incorporate the impact of uncertainties and realize better prediction performance. This approach has been applied to improving the predictive modeling for automatic shape control of composite fuselage.

Note to Practitioners—This article was motivated by automatic shape control of composite fuselage. The main objective is to realize active learning for predictive analytics, which means maximizing information acquisition with limited experimental samples. This kind of need for active learning is very common in the industrial systems where it is expensive, time-consuming, or difficult to obtain experimental data. Existing approaches, from either a machine learning perspective or a statistics perspective, mainly focus on active learning for classification or regression models without incorporating impacts from intrinsic input uncertainties. However, intrinsic uncertainties widely exist in industrial systems. This article develops two active learning algorithms for the Gaussian process with uncertainties. The algorithms take variance-based information measure and Fisher information measure into consideration. The proposed algorithms can also be applied in other active learning scenarios, specifically for predictive models with multiple uncertainties.

Index Terms—Active learning, advanced manufacturing, composite fuselage, Gaussian process (GP), machine learning, uncertainty.

I. INTRODUCTION

ACTIVE learning is a type of iterative supervised learning, which focuses on maximizing information acquisition with limited samples. In statistics literature, this process is also called optimal experimental design or sequential design. The main idea of active learning is to iteratively pose "query" or "design" to explore the most informative new experimental samples according to the information obtained from the current samples.

In many machine learning applications, especially in some industrial systems, the explanatory data X are rich and easy to get, but the response data Y are very expensive, time-consuming, or difficult to obtain. For example, when training autonomous driving algorithms, a lot of media (e.g., images and videos) require that oracle users mark them with particular labels, such as "vehicle," "street sign," or "road lines." It can be tedious, redundant, and time-consuming to annotate a lot of these instances. In composite fuselage shape control problem, it is very expensive and time-consuming to collect the dimensional shape under specific actuators' forces. Other examples, such as speech recognition, factual information extraction, and computational biology, can be found [1]. In these scenarios, active learning is well-motivated because it can reduce the samples needed as well as obtain sufficient information for parameter inference of predictive modeling.

Different sampling strategies have been proposed to realize active learning. In machine learning literature, these strategies are classified into query synthesis, stream-based selective sampling, and pool-based sampling [1]. In statistics literature, space-filling designs, criterion-based designs, and expected improvement (EI) algorithm have been proposed for sequential design. For the comparison between active learning and sequential design, they have several common characteristics. For example, they both iteratively pose "query" or "design" to explore the most informative new samples according to the information obtained from the current samples and some criteria are similar for active learning and sequential design. The uncertainty sampling in active learning and sequential designs

is based on certain optimality criteria (e.g., mean-squared prediction error (MSPE) and maximum entropy) that are mathematically equivalent [1], [2].

However, there exist some differences between active learning and sequential design. First, sequential design literature is mainly for regression and prediction problems. While active learning literature mainly focuses on classification problems in machine learning, and a few articles focus on regression with extrinsic uncertainties but not incorporating impacts from intrinsic input uncertainties. Second, in sequential design, usually, the experiment will be conducted at the selected points from an input space, whereas in active learning, the experiment sampling may be from a large pool of existent unlabeled data, and the learner can conduct the experiment at a selected point or discard it. In Section II, we will review the literature from two perspectives: machine learning (computational learning) and statistics (sequential design) in detail. Other than Section II, we use active learning as a consistent terminology.

In this article, we develop two active learning algorithms for Gaussian process (GP) model with uncertainties. Uncertainties inevitably exist in the input and output data of any system. From the predictive modeling perspective, several approaches have been developed to incorporate input and output uncertainties. Ankenman *et al.* [3] separated the uncertainties into intrinsic uncertainty inherent to stochastic simulation and extrinsic uncertainty about unknown response surfaces. They proposed the stochastic Kriging model for stochastic computer experiments. Cervone and Pillai [4] investigated GP regression with input uncertainty from measurement errors and showed that approximate methods for incorporating location measurement error are essential to valid parameter inference. Wang *et al.* [5] compared the best linear unbiased predictor (BLUP) and stochastic Kriging predictor and proved their asymptotic properties. For the sampling perspective, the existing active learning methods cannot be straightforwardly extended to GP with uncertainties. In this article, our main contribution is to propose active learning algorithms for GP regression considering both intrinsic and extrinsic uncertainties. Specifically, the proposed active learning algorithms could maximize the acquisition of various information (for instance, actuators' uncertainty, composite parts variability, modeling nonlinearity, and measurement errors).

The proposed active learning approach is applied to improving the predictive modeling for the automatic shape control system of composite fuselage. The automatic shape control system is essential to reducing the dimensional deviations of composite fuselage as well as the flow time of the aircraft assembly process. Ten actuators are installed cross the edge of the composite fuselage. These actuators can provide push and pull forces to change the in-plane shape of the composite fuselage. The automatic shape control system effectively and efficiently adjusts composite fuselages to the optimal configuration [6]. Precise prediction of the dimensional shape of composite fuselage is essential for the shape control, while it is challenging to develop predictive modeling because of multiple uncertainties. A surrogate model considering part uncertainty, actuator uncertainty, and model uncertainty was developed to link the relationship between actuators' forces

and dimensional deviations [7]. This approach assumes that the training samples are sufficiently large, and it does not consider efficient experimental sampling and effective information acquisition. However, obtaining the experimental samples in the composite fuselage assembly process is time-consuming and expensive. To find an approach to minimize the sample size and sequentially improve predictive modeling would be beneficial. In this article, we propose active learning algorithms for maximizing information extractions for the GP model considering uncertainties. We first derive the log likelihood for the general GP model considering uncertainties, including stochastic model and surrogate model considering uncertainties. Then, we take two kinds of information measure into consideration: variance-based information measure and Fisher information measure. Based on the information measure, we proposed the variance-based weighted active learning (VWAL) algorithm and D-optimal weighted active learning (DOWAL) algorithm. The initial design and stopping criteria have also been explored. The proposed active learning strategies have been applied to improving the predictive modeling of composite fuselage shape control.

The remainder of this article is organized as follows. Section II illustrates the literature review about active learning and sequential design from the machine learning perspective and statistics perspective. Section III describes the predictive modeling based on the GP and then proposes the active learning approach for the GP model considering uncertainties. Section IV presents the case study, including an introduction for automatic shape control of composite fuselage, validation procedure, evaluation criteria, and comparison between the performance of our proposed method and benchmark methods. Finally, a brief summary is provided in Section V.

II. LITERATURE REVIEW

In this section, we review the existing active learning and sequential design methodologies from machine learning (computational learning) and statistics (sequential design) perspectives.

A. Literature From Machine Learning Domain

In the computational learning theory literature, active learning is sometimes called "query learning." There are three kinds of scenarios: query synthesis, stream-based selective sampling, and pool-based sampling [1]. In query synthesis, the learner may request any unlabeled data instance in the input space, while in stream-based selective sampling, each unlabeled instance is typically drawn one at a time from the input space, and then, the learner decides whether to query or discard it. For pool-based sampling, there are a set of labeled data and a large pool of unlabeled data available, and queries are selected from the pool. The detailed descriptions of these three scenarios can be found in [1].

Settles [1] and Olsson [8] provided a very detailed literature review on active learning for classification problems, while in this article, we mainly focus on the literature review on active learning for regression problems.

Lewis and Catlett [9] proposed one active learning strategy called uncertainty sampling. In this strategy, the learner regards the unlabeled samples with the largest uncertainty as the most informative samples. Therefore, the learner queries these unlabeled samples in order to generalize information from the current samples. In regression, the learner can query the unlabeled samples for which the model has the highest prediction variance. Under the Gaussian assumption, this prediction variance-based uncertainty sampling approach is equivalent to the entropy-based uncertainty sampling because the entropy is a monotonic function of its variance [1]. Other than uncertainty-based criteria, some algorithms introduced diversity constraints to avoid selecting similar samples for model training. For example, Wang *et al.* [10], [11] proposed an active learning algorithm to integrate uncertainty and diversity via sparse modeling in the sample selection. Yang *et al.* [12] also proposed an uncertainty sampling with diversity maximization to improve multiclass active learning. Cohn *et al.* [13] proposed an active learning strategy that selects the design points by minimizing the integrated average variance of the learner. Sugiyama [14] proposed a new active learning method called Active Learning using the Importance-weighted least-squares learning based on Conditional Expectation of the generalization error (ALICE). This method predicts conditional expectation of the generalization error given training input points, whereas most of the existing methods predict the full expectation of the generalization error. Burbidge *et al.* [15] investigated an active learning strategy for regression based on query by committee, which considers choosing sequential points according to the average expected variance over the reference points. Sugiyama and Rubens [16] developed a new ensemble active learning approach for solving active learning and model selection in linear regression simultaneously. However, this approach is mainly designed for linear regression and not easy to extend to other models, such as the GP. Pasolli and Melgani [17] proposed two active learning strategies for GP regression. One is based on adding samples that have large kernel distance from the current training samples, which considers space-filling properties. The other exploits an intrinsic GP regression outcome to pick up the samples with the largest variance. Cai *et al.* [18] proposed a new active learning framework for regression called expected model change maximization (EMCM), which aims to choose the examples that lead to the largest change to the current model. However, the EMCM is sensitive to outliers, which may result in nonstationary parameter estimations. Schreiter *et al.* [19] proposed a safe exploration of active learning with GPs. A differential entropy criterion was used to explore the relevant data regions. These existing active learning methods do not straightforwardly extend to incorporate uncertainties in the automatic shape control system of composite fuselages.

B. Literature From Statistics Domain

In the statistics domain, the sequential design is to propose experiment designs at a limited number of times, and inputs/responses from the previous design may impact the

following design. The basic idea behind the sequential design is to select input points that will allow us to model and minimize the discrepancy between the output from the computer model and predictions from the surrogate model.

For experimental designs relevant to computer experiments, there are two categories: space-filling designs and criterion-based designs. Space-filling designs (e.g., Latin hypercube designs, maximin designs, and Sobol's sequence [20]) assume that samples provide information equally across the entire input space, which encourages the exploration among the whole input space. However, these designs are not adaptive to the information from the response surface. Designs based on certain optimality criteria, such as MSPE [21] and entropy [22], make full use of information from both inputs and outputs. However, they do not consider the shape of the response surface.

The EI algorithm is a global optimization algorithm proposed by Mockus *et al.* [23] and then brought to the field of computer experiments [24]. The main idea of the EI algorithm is to identify the nature of input–output relationships and then subsequently choose the design points one at a time, or in groups, to maximize the EI on the objective. Williams *et al.* [25] modified the EI algorithm by considering both control and environmental variables, which computed the posterior EI over the current optimum for each untested point and then selected the next point to maximize modified EI. They further extended the EI algorithm to a bivariate modified EI algorithm, which realizes sequential design for computer experiments where there is a bivariate response [26]. Vazquez and Bect [27] investigated the convergence properties of the EI algorithm. Provided that the objective function belongs to the reproducing kernel Hilbert space, the EI algorithm produces a dense sequence of evaluation points in the search domain.

In addition, Lam [2] proposed a modified integrated mean-squared prediction error (IMSPE) criterion by imposing a penalty to prevent the additional design points from clustering together. This method can realize the tradeoff between exploration and exploitation. Deng *et al.* [28] pointed out that there were two kinds of approaches to generate sequential designs: stochastic approximation and optimal design. The optimal design approach has a better performance when the assumed model is the true model, but it is not robust to model assumptions. They used a combination of stochastic approximation and D-optimal designs to judiciously select the design points [28]. By maximizing the estimated variance of the hyperplane, they placed the next point at the location of the greatest uncertainty. The proposed method improved the process of money laundering detection. Crombecq *et al.* [29] proposed a hybrid sequential design strategy that used a Monte Carlo-based approximation of a Voronoi tessellation for exploration and local linear approximations of the simulator for exploitation. The advantage of this method is that it is independent of the model type and can be used in heterogeneous modeling environments.

From an engineering perspective, there are several literature works presenting the sequential strategies for other objectives, such as measurement and detection. Jin *et al.* [30] proposed a sequential measurement strategy for efficient wafer geometric

profile estimation. This strategy reduced the number of samples measured in wafers as well as provided an adequate accuracy for quality feature estimation. The sequential samples are chosen based on the gradient and error of profile prediction. Yan [31] proposed a sequential sampling strategy called adaptive kernelized maximum–minimum distance (AKM2D) to speed up inspection and anomaly detection process. The proposed method realized the tradeoff between space-filling sampling (exploration) and focused sampling near the anomalous region (exploitation). Xiang *et al.* [32] developed a novel distribution-free control chart via cucconi statistic for joint monitoring of location and scale parameters with extremely small samples. However, these methods are mainly focused on efficient detection, not for predictive modeling.

III. ACTIVE LEARNING FOR GP CONSIDERING UNCERTAINTIES

GP models have been widely used as surrogate models of expensive deterministic computer simulations. It has many advantages, such as good prediction performance, good flexibility from multiple correlation choices, complete mathematical properties in both Bayesian and frequentist statistical framework, and capability of uncertainty quantification. In Section II, several active learning strategies relevant to the GP models have been reviewed, including uncertainty sampling [9], [17], entropy-based active learning [19], and EI algorithms [24]–[26]. These strategies work well for the data acquisition of the deterministic computer simulations, while in some other cases, there exist multiple input or output uncertainties in the data sets. To the best of our knowledge, there lacks a tailored active learning strategy for GP considering uncertainties. In this article, we first review two GP models considering uncertainties for automatic shape control of composite fuselage. One is the stochastic Kriging, and the other is the surrogate model considering uncertainties. These models can provide a very accurate prediction for composite fuselage shape control. Then, we proposed new active learning strategies for GP considering uncertainties.

A. GP With Nugget Effects: Stochastic Kriging

Cressie [33] represented the measurement error or uncertainties as a nugget effect for GP. Ankenman *et al.* [3] integrated the intrinsic uncertainty and extrinsic uncertainty with their stochastic Kriging model. Cervone and Pillai [4] investigated GP regression with input uncertainty from measurement errors and showed that approximate methods for incorporating location measurement error are essential to valid parameter inference. In this section, we summarize the GP model with nugget effects for composite fuselage shape control.

Consider a set of design setting $\mathcal{X}$, which includes pairs $\{F_t, n_t\}$, $t = 1, 2, \ldots, k$, where F_t is the input vector (actuators' forces) with a dimension of $1 \times q$ and n_t is the number of replications for the design point F_t. Consider that the p-dimensional output (dimensional deviations) for the design point F_t is $Y(F_t)$, where each element $Y_{ij}(F_t)$ denotes the jth variable of the output vector under the ith replication at the input F_t. We use the set $\mathcal{D}$ to represent the input–output pairs

$\{F_t, Y(F_t)\}$. Considering both a noise term ε_i and a stochastic process term $z_j(F)$ that are regarded as intrinsic uncertainty (measurement error) and extrinsic uncertainty, respectively, the GP with nuggets effects (or called the stochastic Kriging) model can be developed as

$$Y_{ij}(F) = FS_j + z_j(F) + \varepsilon_{ij} \tag{1}$$

where S_j represents the sensitivity matrix corresponding to the jth response, $j = 1, 2, \ldots, p$. $z_j(F)$ is a stochastic process that represents the extrinsic uncertainty relevant to functional mapping. Specifically, the stochastic process $z_j(F)$ is assumed to be a GP $GP(0, \Sigma_{zj})$ with covariance matrix Σ_{zj} with dimension $k \times k$. For any two vectors of the actuators' forces, the covariance is $\Sigma_{zj}(F_m, F_n) = \text{cov}[z_j(F_m), z_j(F_n)]$. Let $\Sigma_{zj}(F_0, \cdot) = (\text{cov}[z_j(F_0), z_j(F_1)], \ldots, \text{cov}[z_j(F_0), z_j(F_k)])$ as the covariance between z_j values at design points and new actuators' forces F_0. The intrinsic noise ε_{ij} is assumed to be independent and identically distributed with a normal distribution $\varepsilon_{ij} \sim \mathbb{N}(0, \sigma_{\varepsilon j})$. The noise covariance matrix is $\Sigma_{\varepsilon j}$, and a $k \times k$ covariance matrix with (a, b) element is $\text{Cov}[\sum_{i=1}^{n_a} \varepsilon_{ij}(F_a)/n_a, \sum_{i=1}^{n_b} \varepsilon_{ij}(F_b)/n_b]$. The sample mean at F_t as $\bar{Y}_j(F_t) = \sum_{i=1}^{n_t} Y_{ij}(F_t)/n_t$. Let $\bar{Y}_j = (\bar{Y}_j(F_1), \ldots, \bar{Y}_j(F_k))^T$ and $F_{DOE} = (F_1; \ldots; F_k)$ denote all the design points of the actuators' forces. Let $R_j(F_0, \cdot) = (\text{Cov}[Y_j(F_0), \bar{Y}_j(F_1)]; \ldots; \text{Cov}[Y_j(F_0), \bar{Y}_j(F_k)])$. The covariance between vector a and vector b satisfies $\Sigma_{zj}(\tau_j^2, \theta, a, b) = \tau_j^2 R_{zj}(\theta, a - b)$, where $R_{zj}(\theta, a - b)$ is one of the correlation functions. With spatial correlation, $z_j(F_m)$ and $z_j(F_n)$ tend to be similar (e.g., $R_{zj,mn} = R_{zj}(\theta, F_m, F_n)$ tends to be large) if F_m and F_n are close.

According to the literature [3], [4], [7], we know the best MSPE linear unbiased predictor as

$$\hat{Y}_j(F_0|\mathcal{D}) = F_0 S_j + R_j(F_0, \cdot)^T [\Sigma_{zj} + \Sigma_{\varepsilon j}]^{-1} (\bar{Y}_j - F_{DOE} S_j). \tag{2}$$

B. Surrogate Model Considering Uncertainties

The stochastic Kriging model uses a nugget effect to approximate both input uncertainties and output uncertainties. Yue *et al.* [7] proposed a surrogate model considering part uncertainty, actuator uncertainty, and model uncertainty, which is a GP model with consideration of uncertainties in detail. A surrogate model considering uncertainties was proposed as

$$Y_{ij}(F_t) = F_t S_j + F_t \tilde{S}_j + \tilde{F}_t S_j + z_j(F_t) + \varepsilon_{ij}(F_t) \tag{3}$$

where $i = 1, 2, \ldots, n_t$; $j = 1, 2, \ldots, p$, with p the number of output responses (key dimensional features). F_t is the target actuators' forces vector, and $\tilde{F}_t$ is a vector of an additional random deviation of actuators' forces that result from the actuators' uncertainty with distribution $\mathbb{N}(0, \Sigma_F)$. It can be obtained from the tolerance of actuators instruction; $F_t + \tilde{F}_t$ represents the true actuators' force vector. S_j is an ideal sensitivity vector (column vector) and $\tilde{S}_j$ represents the random sensitivity vector variability from the part uncertainty, which is assumed to follow $\mathbb{N}(0, \Sigma_S)$. Both S_j and Σ_S are unknown. $z_j(F_t)$ is assumed to be a stationary GP $z_j \sim GP(0, \Sigma_{zj})$. $\varepsilon_{ij}(F_t)$ is assumed to follow an independent

normal distribution $\varepsilon_{ij}(\boldsymbol{F}_t) \sim \mathbb{N}(0, \sigma_{\varepsilon j}^2(\boldsymbol{F}_t))$, which represents the inherent simulation variability in a stochastic simulation, or measurement errors in a physical experiment. ε_{ij}, $\tilde{\boldsymbol{F}}_t$, $\tilde{\boldsymbol{S}}_j$, and z_j are assumed to be mutually independent, and their higher order interaction term $\tilde{\boldsymbol{F}}_t \cdot \tilde{\boldsymbol{S}}_j$ is assumed to be zero. The model can be interpreted as a decomposition of the response $Y_{ij}(\boldsymbol{F}_t)$ into three parts: a regression term $\boldsymbol{F}_t \boldsymbol{S}_j + \boldsymbol{F}_t \tilde{\boldsymbol{S}}_j + \tilde{\boldsymbol{F}}_t \boldsymbol{S}_j$, a GP term $z_j(\boldsymbol{F}_t)$, and a noise term $\varepsilon_{ij}(\boldsymbol{F}_t)$.

Assume that $\boldsymbol{S}_j$, $\boldsymbol{\Sigma}_{zj}$, $\boldsymbol{\Sigma}_F$, $\boldsymbol{\Sigma}_\varepsilon$, and $\boldsymbol{\Sigma}_S$ are known, and the best MSPE linear unbiased predictor can be derived as

$$\hat{\boldsymbol{Y}}_j(\boldsymbol{F}_0|\mathcal{D}) = \boldsymbol{F}_0 \boldsymbol{S}_j + \boldsymbol{R}_j^T(\boldsymbol{F}_0, \cdot) \boldsymbol{R}_j^{-1} (\bar{\boldsymbol{Y}}_j - \boldsymbol{F}_{DOE} \boldsymbol{S}_j) \quad (4)$$

where $\boldsymbol{R}_j = \boldsymbol{F}_{DOE} \boldsymbol{\Sigma}_S \boldsymbol{F}_{DOE}^T + \boldsymbol{\Sigma}_{zj} + \boldsymbol{S}_j^T \boldsymbol{\Sigma}_F \boldsymbol{S}_j \cdot \boldsymbol{I} + \boldsymbol{\Sigma}_{\varepsilon j}$, and $\boldsymbol{R}_j(\boldsymbol{F}_0, \cdot) = (\boldsymbol{F}_0 \boldsymbol{\Sigma}_S \boldsymbol{F}_1^T + \tau_j^2 R_{zj}(\boldsymbol{\theta}, \boldsymbol{F}_0 - \boldsymbol{F}_1); \ldots; \boldsymbol{F}_0 \boldsymbol{\Sigma}_S \boldsymbol{F}_k^T + \tau_j^2 R_{zj}(\boldsymbol{\theta}, \boldsymbol{F}_0 - \boldsymbol{F}_k))$. The best MSPE linear unbiased predictor is also simply called a BLUP. More detail related to the algorithm for the surrogate model considering uncertainties can be found in [7].

The surrogate model considering uncertainties in (4) analyzes the different sources of uncertainties in the composite fuselage shape control system, whereas the stochastic Kriging predictor in (2) approximates all the uncertainties by introducing a nugget effect. Wang *et al.* [5] proved that the stochastic Kriging and GP with input location errors asymptotically converge to the same limit. In our active learning strategy design, we will analyze both cases.

C. Information Measure

Active learning is an iterative data selection algorithm for maximizing information acquisition and improving model performance with limited training samples. First, we need to propose the information measure for GP considering uncertainties.

Suppose that $\mathfrak{R}_j(\boldsymbol{F}_0, \cdot)$ and $\mathfrak{R}_j$ describe the covariance between $\boldsymbol{F}_0$ and historical samples and the covariance among historical samples for general GP models considering uncertainties. When the model is the surrogate model considering uncertainties, $\mathfrak{R}_j(\boldsymbol{F}_0, \cdot) = (\boldsymbol{F}_0 \boldsymbol{\Sigma}_S \boldsymbol{F}_1^T + \tau_j^2 R_{zj}(\boldsymbol{\theta}, \boldsymbol{F}_0 - \boldsymbol{F}_1); \ldots; \boldsymbol{F}_0 \boldsymbol{\Sigma}_S \boldsymbol{F}_k^T + \tau_j^2 R_{zj}(\boldsymbol{\theta}, \boldsymbol{F}_0 - \boldsymbol{F}_k))$, $\mathfrak{R}_j = \boldsymbol{F}_{DOE} \boldsymbol{\Sigma}_S \boldsymbol{F}_{DOE}^T + \boldsymbol{\Sigma}_{zj} + \boldsymbol{S}_j^T \boldsymbol{\Sigma}_F \boldsymbol{S}_j \cdot \boldsymbol{I} + \boldsymbol{\Sigma}_{\varepsilon j}$; when the model is the stochastic Kriging model, $\mathfrak{R}_j(\boldsymbol{F}_0, \cdot) = (\tau_j^2 R_{zj}(\boldsymbol{\theta}, \boldsymbol{F}_0 - \boldsymbol{F}_1); \ldots; \tau_j^2 R_{zj}(\boldsymbol{\theta}, \boldsymbol{F}_0 - \boldsymbol{F}_k))$, $\mathfrak{R}_j = \boldsymbol{\Sigma}_{zj} + \boldsymbol{\Sigma}_{\varepsilon j}$. Suppose that $\boldsymbol{\Sigma}_{\varepsilon j} = \sigma_j^2 \boldsymbol{I}$ and $\boldsymbol{\Sigma}_S = \varphi_j^2 \boldsymbol{I}$. Let Θ represent the key parameter set, for example, in the stochastic Kriging model, $\Theta = \{\tau_j^2, \boldsymbol{\theta}_j, \sigma_j^2\}$, while in the surrogate model considering uncertainties, $\Theta = \{\tau_j^2, \boldsymbol{\theta}_j, \sigma_j^2, \varphi_j^2\}$.

Under the multivariate normal distribution, the log-likelihood function of $(\boldsymbol{S}_j, \tau_j^2, \boldsymbol{\theta}, \boldsymbol{\Sigma}_S)$ is

$$\mathcal{L}(\boldsymbol{S}_j, \boldsymbol{\Theta}|\mathcal{D})$$
$$= -\frac{1}{2} \ln[(2\pi)^k]$$
$$\quad - \frac{1}{2} \ln[\det(\mathfrak{R}_j)] - \frac{1}{2} (\bar{\boldsymbol{Y}}_j - \boldsymbol{F}_{DOE} \boldsymbol{S}_j)^T \mathfrak{R}_j^{-1} (\bar{\boldsymbol{Y}}_j - \boldsymbol{F}_{DOE} \boldsymbol{S}_j).$$
$$(5)$$

We can derive the estimated parameter $\hat{\boldsymbol{S}}_j$ by making the first derivative of $\mathcal{L}(\boldsymbol{S}_j, \tau_j^2, \boldsymbol{\theta}_j, \boldsymbol{\Sigma}_S)$ to $\boldsymbol{S}_j$ to be equal to zero.

Then, we can get the generalized least-square estimation of the parameter $\hat{\boldsymbol{S}}_j$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{S}_j} = (\bar{\boldsymbol{Y}}_j - \boldsymbol{F}_{DOE} \boldsymbol{S}_j)^T \mathfrak{R}_j^{-1} \boldsymbol{F}_{DOE} = \boldsymbol{0}$$

$$\hat{\boldsymbol{S}}_j(\boldsymbol{\Theta}) = \left(\boldsymbol{F}_{DOE}^T \mathfrak{R}_j^{-1} \boldsymbol{F}_{DOE}\right)^{-1} \boldsymbol{F}_{DOE}^T \mathfrak{R}_j^{-1} \boldsymbol{F}_{DOE} \bar{\boldsymbol{Y}}_j. \quad (6)$$

One straightforward and widely used measure is variance. The variance of the predictor $\hat{\boldsymbol{Y}}_j(\boldsymbol{F}_0|\mathcal{D})$ is $\mathrm{Var}_j(\boldsymbol{F}_0|\mathcal{D}) = \tau_j^2 - \mathfrak{R}_j^T(\boldsymbol{F}_0, \cdot) \mathfrak{R}_j^{-1} \mathfrak{R}_j(\boldsymbol{F}_0, \cdot)$.

The other information measure is the Fisher information matrix $I(\Theta) \in \mathbb{R}^{(m+2) \times (m+2)}$, which is calculated by

$$I_{ab}(\boldsymbol{\Theta}|\mathcal{D}) = -\mathbb{E}\left[\frac{\partial^2}{\partial \Theta_a \partial \Theta_b} \mathcal{L}(\boldsymbol{\Theta})|\boldsymbol{\Theta}\right] \quad (7)$$

where $a, b = 1, \ldots, m + 2$.

Based on (6) and (7), the Fisher information matrix of $\hat{\boldsymbol{\Theta}}$ is

$$I_{ab}(\hat{\boldsymbol{\Theta}}|\mathcal{D}) = \left[\frac{\partial \boldsymbol{S}_j}{\partial \Theta_a}\right]^T \boldsymbol{F}_{DOE}^T \mathfrak{R}_j^{-1} \boldsymbol{F}_{DOE} \left[\frac{\partial \boldsymbol{S}_j}{\partial \Theta_b}\right]$$
$$\quad + \frac{1}{2} Tr\left[\mathfrak{R}_j^{-1} \frac{\partial \mathfrak{R}_j}{\partial \Theta_a} \mathfrak{R}_j^{-1} \frac{\partial \mathfrak{R}_j}{\partial \Theta_b}\right] \quad (8)$$

where $\mathfrak{R}_j$ is the covariance matrix of the GP, and the first-order derivative of the coefficients $\boldsymbol{S}_j$ with respect to each entry in the parameter vector can be represented as

$$\frac{\partial \boldsymbol{S}_j}{\partial \Theta_a}(\hat{\boldsymbol{\Theta}}) = -\left(\boldsymbol{F}_{DOE}^T \mathfrak{R}_j^{-1} \boldsymbol{F}_{DOE}\right)^{-1}$$
$$\quad \times \left[\boldsymbol{F}_{DOE}^T \mathfrak{R}_j^{-1} \frac{\partial \mathfrak{R}_j}{\partial \Theta_a} \mathfrak{R}_j^{-1} (\bar{\boldsymbol{Y}}_j - \boldsymbol{F}_{DOE} \boldsymbol{S}_j)\right]. \quad (9)$$

According to the Cramér–Rao inequality explanation for the Fisher information matrix [34], the inverse of Fisher information sets a lower bound on the variance of the model's parameter estimates. Maximizing the Fisher information is equivalent to minimizing the lower bound on the variance of parameter estimations. When the Fisher information is a matrix, we minimize the determinant of the inverse information matrix. This is called D-optimality in the optimal experimental design.

D. Active Learning

In this section, we propose algorithms to select the next sample sequentially, which is the main implementation of active learning. Suppose the next samples can be selected from a pool of candidates. We denote these candidate samples by $\mathcal{F} = \{\tilde{\boldsymbol{F}}_1, \tilde{\boldsymbol{F}}_2, \ldots, \tilde{\boldsymbol{F}}_N\}$. N denotes the size of the pool $\mathcal{F}$. It is worth mentioning that the choice of candidate pool $\mathcal{F}$ is very important. The best active learning strategy should perform a tradeoff between exploitation and exploration, where the exploitation suggests selecting samples in regions that were previously identified to be interesting. On the other hand, the exploration involves selecting samples in unrepresented regions of the design space. In our active learning strategy, the exploitation is relevant to algorithms of selecting the next samples from the candidate pool, while the candidate pool determines the exploration. In our algorithms, the maximin

Latin hypercube design [35], which demonstrates good space-filling properties and first-dimension projection properties, is implemented to obtain the samples in the candidate pool.

There are two information measures according to Section III-C, the output variance of its prediction and the Fisher information matrix. First, we develop a VWAL algorithm for GP considering uncertainties, meaning that the next sample is selected based on

$$
\begin{aligned}
F_{\text{new}} &= \operatorname*{argmax}_{F \in \{\tilde{F}_1, \tilde{F}_2, \ldots, \tilde{F}_N\}} \sum_{j=1}^{p} W_j \cdot \operatorname{Var}_j(F | \mathcal{D}) \\
&= \operatorname*{argmax}_{F \in \{\tilde{F}_1, \tilde{F}_2, \ldots, \tilde{F}_N\}} \sum_{j=1}^{p} \tau_j^2 - \mathfrak{R}_j^T(F_0, \cdot) \mathfrak{R}_j^{-1} \mathfrak{R}_j(F_0, \cdot)
\end{aligned}
\tag{10}
$$

where F_{new} is the next sample to be queried, $\operatorname{Var}_j(F | \mathcal{D})$ is the variance of the predictor $\hat{Y}_j(F_0 | \mathcal{D})$ at the jth critical dimension of the composite fuselage, and W_j is weight coefficient for the jth critical dimension and we suppose $\sum_j W_j = 1$. $W_j (j = 1, 2, \ldots, p)$ is determined by the engineering-domain knowledge. Here, W_j describes the relative degree of importance of critical dimensional feature j. For instance, the lower part of the fuselage may have higher weights under some scenarios. Under Gaussian assumption, the entropy of a random variable is a monotonic function of its variance.

For the implementation of the active learning algorithm, first, we estimate model parameters for the GP model considering uncertainties by maximizing the log-likelihood function (5). Then, we calculate the variance of predictors for each sample in the candidate pool. Next, the new sample point F_{new} can be selected based on solving (10). The experiment will be conducted to collect the oracle response $Y(F_{\text{new}})$. The existing sample set $\mathcal{D}$ will be updated by adding the new sample point and the corresponding response. The parameters will be iteratively updated and the new samples will be selected actively until the iteration number reaches the maximum iteration N_{iter} or the error of the model E_{GP} is smaller than a specific threshold. The pseudocode of this active learning algorithm is summarized in Algorithm 1.

Based on another information measure, the Fisher information matrix, we also develop a DOWAL algorithm for GP considering uncertainties, meaning that the next sample is selected based on

$$
F_{\text{new}} = \operatorname*{argmax}_{F \in \{\tilde{F}_1, \tilde{F}_2, \ldots, \tilde{F}_N\}} \sum_{j=1}^{p} W_j \cdot \det\left(\left[I_{ab}(\hat{\Theta} | \mathcal{D}, F) \right]^{-1} \right). \tag{11}
$$

The inverse of Fisher information sets a lower bound on the variance of the model's parameter estimates, which is known as the Cramér–Rao inequality. By minimizing the determinant of the inverse Fisher information matrix (D-optimality), we can maximize the information acquisition in each step of the active learning process. Furthermore, the D-optimality is relevant to minimizing the differential posterior entropy of the parameter estimation [36].

Similarly, we can summarize the pseudocode of this DOWAL algorithm for the GP model considering uncertainties in Algorithm 2.

Algorithm 1 VWAL for GP Considering Uncertainties

Require: $\mathcal{D}, W_j$
1: $i = 1$
2: Estimate parameters by maximizing log-likelihood function (5) on $\mathcal{D}$
3: Calculate the variance of the predictor $\hat{Y}_j(F_0 | \mathcal{D})$, $(j = 1, 2, \ldots, p)$ for each sample
4: **while** $(i \leq N_{iter})$ && $(E_{GP} > \text{threshold})$ **do**
5:     get F_{new} from solving Equation (10)
6:     Implement experiment for F_{new} and obtain $Y(F_{new})$
7:     $\mathcal{D} \leftarrow \mathcal{D} \bigcup (F_{new}, Y(F_{new}))$
8:     Estimate parameters by maximizing log-likelihood function (5) on $\mathcal{D}$
9:     Calculate the variance of the predictor $\hat{Y}_j$ $(F_0 | \mathcal{D}, (j = 1, 2, \ldots, p)$ for each sample
10:     $i = i + 1$
11: **end while**

Algorithm 2 DOWAL for GP Considering Uncertainties

Require: $\mathcal{D}, W_j$
1: $i = 1$
2: Estimate parameters by maximizing log-likelihood function (5) on $\mathcal{D}$
3: Calculate the Fisher information matrix by solving Equations (8) and (9)
4: **while** $(i \leq N_{iter})$ && $(E_{GP} > \text{threshold})$ **do**
5:     get F_{new} from solving Equation (11)
6:     Implement experiment for F_{new} and obtain $Y(F_{new})$
7:     $\mathcal{D} \leftarrow \{\mathcal{D} \bigcup (F_{new}, Y(F_{new}))$
8:     Estimate parameters by maximizing log-likelihood function (5) on $\mathcal{D}$
9:     Calculate the Fisher information matrix by solving Equations (8) and (9)
10:     $i = i + 1$
11: **end while**

In this section, we proposed two active learning algorithms for the GP model considering uncertainties. In the next sections, we will explore the initial design, evaluation criteria, and stopping criteria for these active learning algorithms.

E. Initial Design

The initial design defines the preliminary parameter estimation and has a significant effect on the efficiency and accuracy of the model. It is critical to determine initial samples with a suitable size. The initial sample points should explore the entire space with the very good space-filling property. Maximin distance criterion can be used to choose a good design with the space-filling property. In addition, the initial samples need to provide as much information as possible. Therefore, the Latin hypercube design with good projection property is helpful. We choose the maximin Latin hypercube design to ensure the good exploration and exploitation performance for the initial samples. In addition, the size of the initial samples is very important. Underselection of initial

samples may result in insufficient parameter estimation for the proposed model, whereas overselection of initial samples may reduce the efficiency and significance of the active learning algorithms. We choose the initial design by the rule of thumb of predictive modeling in composite fuselage shape control. When we run the predictive modeling in [7], we found that at least ten training samples are needed to get a reasonable prediction performance. Therefore, we selected a maximin Latin hypercube design with 11 samples as an initial design.

F. Stopping Criterion

A potentially important element of interactive learning applications, in general, is to determine when to stop learning [1]. We take modeling for composite fuselage shape control as an example. We hope to use a minimum number of training samples to get a model with satisfied prediction accuracy. It is quite complex and time-consuming to collect experimental samples in the assembly process. Therefore, it is important to know that how to recognize when the accuracy of the model can satisfy the prediction requirement, and acquiring more samples is likely increasing the flow time without much improvement on the model. Active learning provides us with the capability to accurately balance model accuracy and reducing the cost of obtaining samples.

There are several stopping criteria for active learning, such as stopping based on an intrinsic measure of stability or self-confidence within the learner [37]. In our algorithms, we check the model prediction error is consistently smaller than the engineering specifications for several continuous steps. In addition, we check that the total iteration number is smaller than a specific threshold to prevent an infinite iteration loop due to algorithm errors.

IV. Case Study

We conducted a case study to demonstrate the effectiveness of the proposed active learning algorithms. First, we introduced the automatic shape control of composite fuselage. Next, we discussed the parameter estimation algorithms for predictive modeling. Three evaluation criteria have been used to evaluate the performance of active learning methods. By comparing our proposed methods with benchmark methods, we conclude that the proposed active learning methods can obtain better performance for predictive modeling. In addition, the proposed methods provide us with a decision point for stopping the collection of experimental samples.

A. Automatic Shape Control of Composite Fuselage

In current practice, experienced engineers adjust the shape of composite fuselage multiple times by trial-and-error until the deviations between the real shape and the target shape are smaller than a specific engineering specification. This approach brings large uncertainties for the fuselage assembly process, and it may only reach an acceptable shape but not the optimal one. For the automatic shape control, the system can measure the real dimensional shape of the fuselage by a laser metrology system, then compute the optimal actuators'

forces to minimize the dimensional deviations of the current composite fuselage to the target one, and finally implement the shape adjustment. As shown in Fig. 1(a) [6], ten actuators are installed at the edge of the composite fuselage. These actuators can push or pull the fuselage to adjust its shape to the target shape. Du *et al.* [38] investigated the optimal placement of these actuators via sparse learning. The automatic shape control system can be further used for improving the dimensional quality [39] and reducing the residual stress [40] in the multistage composite structures assembly processes (e.g., automotive and aerospace). One of the most challenging tasks for the automatic shape control system is to develop a predictive model with limited experimental samples. Thus, we proposed active learning strategies to maximize information acquisition for predictive modeling and provide a stopping criterion for experiments.

In order to validate our proposed methodology economically, we developed a finite-element model of the composite fuselage, with software ANSYS Composite PrepPost [6]. The finite-element model, shown in Fig. 1(b), exactly mimics the fabrication process of composite fuselages, including material (carbon fiber and epoxy resin) introduction, stackup/sublaminates design, material orientation, geometrical setting, and fixture setup. Physical experiments with a fuselage from the sponsor company were conducted to calibrate and validate the accuracy of the developed computer simulation model. We applied an effective model calibration approach via sensible variable identification and adjustment [41]. After calibration, the finite-element analysis (FEA) simulation results and the physical experimental outputs are quite consistent [6], [41]. One simulation result of the total deformation is shown in Fig. 1(c). More details about the simulation and experimental setup can be found in [6] and [41].

After obtaining an accurate finite-element model, we generate the training and testing data sets by computer experiments. Then, GP models considering uncertainties are trained to predict models considering uncertainties can dimensional deviations under specific actuators' forces. More details about the performance of the GP be found in [7] .

B. Validation Procedure of Active Learning Algorithms

To validate the performance of active learning algorithms, we use a GP model trained by the finite-element simulation data sets as an oracle model. The oracle model is to imitate the real engineering system. We generate the initial design of input by maximin Latin hypercube design. Outputs of experimental samples are generated by the oracle model with input uncertainties. For each step of active learning algorithms, we conduct parameter estimation for GP considering uncertainties by maximizing the log-likelihood function in (5). The detailed procedures of active learning algorithms are shown in Algorithms 1 and 2.

In the active learning algorithms, the candidate pool is generated by maximin Latin hypercube design with input bounds [−450 lbf, 450 lbf]. The size of the candidate pool is 200.
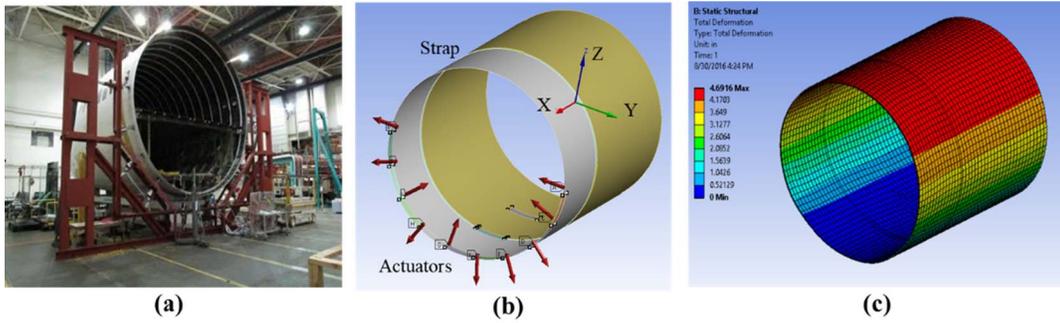
Fig. 1. Shape control of composite fuselage. (a) Fuselage shape control [6]. (b) Finite-element model of shape control. (c) One simulation result.

C. Evaluation Criteria

To evaluate the performance of active learning algorithms, we introduce three evaluation scores: mean of mean absolute deviations (mean MAD), maximum of mean absolute deviations (max MAD), and mean square error (MSE) of cross validation (cross-validation MSE). All these three evaluation scores are calculated based on the evaluation pool, which includes 200 samples explored in the whole input space. The outputs are generated by the oracle model.

The mean MAD for each iteration can be calculated by the following equation:

$$\text{MAD}_{\text{mean}} = \underset{j}{\text{mean}} \left\{ \frac{1}{N_{\text{eva}}} \sum_{i=1}^{N_{\text{eva}}} \left| \hat{Y}_j(F_i) - Y_j^*(F_i) \right| \right\} \quad (12)$$

where N_{eva} is the size of evaluation samples, $\hat{Y}_j(F_i)$ is the predictive response of the jth critical dimension at F_i for each iteration. It is worth mentioning that the predictive model is developed based on all samples at each iteration. $Y_j^*(F_i)$ is the oracle output of the jth critical dimension at F_i.

Similarly, the max MAD can be calculated by the following equation:

$$\text{MAD}_{\text{max}} = \underset{j}{\text{max}} \left\{ \frac{1}{N_{\text{eva}}} \sum_{i=1}^{N_{\text{eva}}} \left| \hat{Y}_j(F_i) - Y_j^*(F_i) \right| \right\}. \quad (13)$$

The two evaluation scores mentioned earlier are based on the mean absolute deviations (MADs). MAD is an important index to check the model performance in composite fuselage shape control. We also introduce the cross-validation MSE. We use leave-one-out cross validation in which the accuracy measures are obtained by using the following equation:

$$\text{MSE}_{cv} = \frac{1}{N_{\text{eva}}} \sum_{i=1}^{N_{\text{eva}}} \underset{j}{\text{mean}} \left\{ \left\| \hat{Y}_j^{[i]}(F_i) - Y_j^*(F_i) \right\|^2 \right\} \quad (14)$$

where $\hat{Y}_j^{[i]}(F_i)$ denotes the predicted response of the jth critical dimension at F_i, whose predictive model is trained by all the residual samples except the $\underline{i}$th one.

Cross validation is primarily a way of measuring the predictive performance of a statistical model. The procedure of calculating cross-validation MSE at each iteration includes the following.

1) Letting sample i forms the testing data, and $Y_j^*(F_i)$ is the oracle output of this testing sample.
2) Train the predictive model by all the residual samples except the ith one at this iteration, then predict the response at F_i, and next calculate the MSE.
3) Repeat steps 1) and 2) for all samples at this iteration and get the mean of obtained MSEs from step 2). This mean is cross-validation MSE.

D. Comparison With Benchmark Methods

In this section, we compare the performance of our proposed active learning algorithms with several benchmark methods. The first benchmark method is to obtain design samples by running the design of the experiment for each sample size. In this method, we do not use active learning strategy. The second benchmark method is a random selection from the candidate pool, which is the most basic pool-based sampling strategy. The third benchmark method is to select the next sample that has the largest maximin distance from the current samples. This method makes full use of the space-filling information in the current input variables but not utilize the information in the response. The fourth method is the EI approach. The EI algorithm identifies the nature of input–output relationships and then subsequently choose design point one at a time to maximize the EI on the objective. It is widely used in a sequential design, especially for computer experiments.

The evaluation criteria for these active learning methods are introduced in Section IV-C. Three evaluation scores, mean MAD, max MAD, and cross-validation MSE, are calculated as the increase of sample size. These active learning curves are shown in Figs. 2–4. In the figures, the four benchmark methods are represented by a black dashed line (no active learning), a blue dashed–dotted line (random selection), a green solid line (selection based on maximin distance), and a magenta dotted line (EI algorithm). The proposed active learning algorithms are represented by a red solid line with asterisk marker (the proposed VWAL algorithm) and a black solid line with a plus-sign marker (the proposed DOWAL algorithm).

In Figs. 2–4, we can find that as the number of samples increases, the mean MAD becomes small for most of these methods. It makes sense because more samples tend to provide more information for training of predictive models. Without
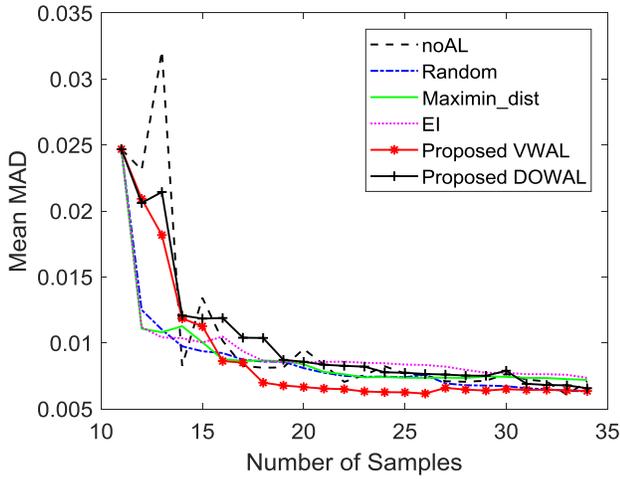
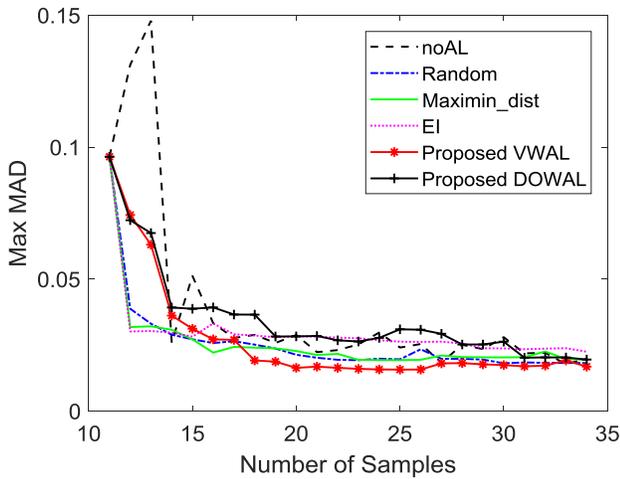Fig. 2. Active learning curves for the mean MAD of different methods.

Fig. 3. Active learning curves for the max MAD of different methods.
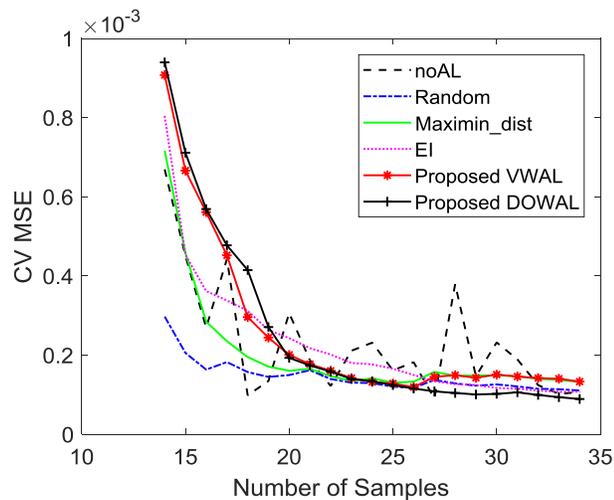

Fig. 4. Active learning curves for the cross-validation MSE of different methods.

active learning, the learning curve has pretty large fluctuations. We can also find that as the number of samples increases, the MAD or MSE may become higher, e.g., the number of

samples from 26 to 28. One potential reason may be that sample 27 is out of the typical response surface. It means that simply increasing sample size sometimes may not be a good idea. The quality of the sample needs to be ensured. Further investigations to introduce an evaluation for sample quality will be conducted for future work. In Figs. 2 and 3, among all these active learning methods, the proposed VWAL algorithm realizes the best performance and it has the smallest mean MAD and max MAD when the number of samples becomes larger than 17. However, the proposed DOWAL algorithm does not have superior performance. It shows the VWAL algorithm is the best choice if the main objective is to realize the smallest mean or max MAD.

If we take cross-validation MSE as the evaluation criterion, the proposed DOWAL algorithm realizes the best performance and has the smallest cross-validation MSE, as shown in Fig. 4. The proposed VWAL is very sensitive to sample 27, which makes prediction worse. It means that simply increasing sample size sometimes may not be a good idea, especially when there is large nonlinearity in the response surface and the quality of samples cannot be fully ensured. From Figs. 2–4, we can also find the benchmark methods (e.g., random selection, selection based on maximin distance, and EI algorithm) can realize good performance. The main reason is that the candidate pool is well chosen according to the maximin Latin hypercube design, and the GP model considering uncertainties can capture the main information structure and response surface within the data sets. Besides, cross-validation MSE is a widely used evaluation metric in the active learning domain, whereas MAD is a general metric for advanced aircraft assembly applications. From the case study, we found that the performance of active learning algorithms may be closely related to the evaluation criterion we selected.

As shown in Section III-F, another important element of active learning applications knows when to stop collecting new experimental samples. For a specific threshold from the engineering domain knowledge, e.g., mean MAD is smaller than 0.007 in, we know if we collect 20 experimental samples, and we can train a predictive model that has enough accuracy. This shows that active learning algorithms can not only improve the accuracy of predictive modeling but also reduce the flow time for experimental sampling.

V. Summary

To realize automatic shape control of composite fuselage, it is critical to develop an accurate predictive model. However, getting an accurate model requires many experiments by obtaining dimensional shape under different actuators' forces which are expensive and time-consuming. Thus, there is an urgent need for active learning in this application to maximize information extraction with limited experimental samples. In practice, an industrial system inevitably has numerous uncertainties, such as input uncertainties, measurement errors, modeling uncertainties, and uncertainties from system parameters. Current active learning literature mainly focuses on the classification problems in machine learning, and a few articles focus on regression with extrinsic uncertainties but not incorporating impacts from intrinsic input uncertainties. Therefore,

there is a gap between the existing active learning methods and the need to incorporate various intrinsic and extrinsic uncertainties in the advanced aircraft assembly process.

The main contribution of this article is to propose two active learning algorithms for the GP model considering uncertainties. The proposed algorithms investigate two predictive models with uncertainties: stochastic Kriging model and surrogate model considering uncertainty. We take two kinds of information measure, variance-based information and Fisher information, into consideration. Two active learning algorithms are proposed to obtain the most informative samples for GP modeling considering uncertainties. We also explored the initial design, stopping criteria for active learning algorithms. To validate the performance of the proposed algorithms, we introduced three evaluation criteria, including mean MAD, max MAD, and cross-validation MSE. The proposed approaches are compared with four benchmark methods in the case study. It shows that the proposed VWAL algorithm can realize the best MAD performance, and the proposed DOWAL algorithm can realize the best cross-validation MSE performance. The proposed active learning algorithms can also be used to trigger the stopping point of experimental sampling. These strategies can be extended to other regression models with uncertainties.

References

[1] B. Settles, "Active learning," *Synth. Lectures Artif. Intell. Mach. Learn.*, vol. 6, no. 1, pp. 1–114, 2012.

[2] C. Q. Lam, "Sequential adaptive designs in computer experiments for response surface model fit," Ph.D. dissertation, Ohio State Univ., Columbus, OH, USA, 2008.

[3] B. Ankenman, B. L. Nelson, and J. Staum, "Stochastic Kriging for simulation metamodeling," *Oper. Res.*, vol. 58, no. 2, pp. 371–382, Apr. 2010.

[4] D. Cervone and N. S. Pillai, "Gaussian process regression with location errors," 2015, *arXiv:1506.08256*. [Online]. Available: http://arxiv.org/abs/1506.08256

[5] W. Wang, X. Yue, B. Haaland, and C. F. Jeff Wu, "Gaussian processes with input location error and applications to the composite parts assembly process," 2020, *arXiv:2002.01526*. [Online]. Available: http://arxiv.org/abs/2002.01526

[6] Y. Wen, X. Yue, J. H. Hunt, and J. Shi, "Feasibility analysis of composite fuselage shape control via finite element analysis," *J. Manuf. Syst.*, vol. 46, pp. 272–281, Jan. 2018.

[7] X. Yue, Y. Wen, J. H. Hunt, and J. Shi, "Surrogate model-based control considering uncertainties for composite fuselage assembly," *J. Manuf. Sci. Eng.*, vol. 140, no. 4, Apr. 2018, Art. no. 041017.

[8] F. Olsson, "A literature survey of active machine learning in the context of natural language processing," Swedish Inst. Comput. Sci., Kista, Sweden, SICS Tech. Rep. T2009:06, 2009. [Online]. Available: http://eprints.sics.se/3600/

[9] D. D. Lewis and J. Catlett, "Heterogeneous uncertainty sampling for supervised learning," in *Machine Learning Proceedings*. Amsterdam, The Netherlands: Elsevier, 1994, pp. 148–156.

[10] G. Wang, J.-N. Hwang, C. Rose, and F. Wallace, "Uncertainty sampling based active learning with diversity constraint by sparse selection," in *Proc. IEEE 19th Int. Workshop Multimedia Signal Process. (MMSP)*, Oct. 2017, pp. 1–6.

[11] G. Wang, J.-N. Hwang, C. Rose, and F. Wallace, "Uncertainty-based active learning via sparse modeling for image classification," *IEEE Trans. Image Process.*, vol. 28, no. 1, pp. 316–329, Jan. 2019.

[12] Y. Yang, Z. Ma, F. Nie, X. Chang, and A. G. Hauptmann, "Multi-class active learning by uncertainty sampling with diversity maximization," *Int. J. Comput. Vis.*, vol. 113, no. 2, pp. 113–127, Jun. 2015.

[13] D. A. Cohn, Z. Ghahramani, and M. I. Jordan, "Active learning with statistical models," *J. Artif. Intell. Res.*, vol. 4, pp. 129–145, Mar. 1996.

[14] M. Sugiyama, "Active learning in approximately linear regression based on conditional expectation of generalization error," *J. Mach. Learn. Res.*, vol. 7, pp. 141–166, Jan. 2006.

[15] R. Burbidge, J. J. Rowland, and R. D. King, "Active learning for regression based on query by committee," in *Proc. Int. Conf. Intell. Data Eng. Automated Learn.* Berlin, Germany: Springer, 2007, pp. 209–218.

[16] M. Sugiyama and N. Rubens, "Active learning with model selection in linear regression," in *Proc. SIAM Int. Conf. Data Mining*, Apr. 2008, pp. 518–529.

[17] E. Pasolli and F. Melgani, "Gaussian process regression within an active learning scheme," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2011, pp. 3574–3577.

[18] W. Cai, Y. Zhang, and J. Zhou, "Maximizing expected model change for active learning in regression," in *Proc. IEEE 13th Int. Conf. Data Mining*, Dec. 2013, pp. 51–60.

[19] J. Schreiter, D. Nguyen-Tuong, M. Eberts, B. Bischoff, H. Markert, and M. Toussaint, "Safe exploration for active learning with Gaussian processes," in *Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discovery Databases*. Cham, Switzerland: Springer, 2015, pp. 133–149.

[20] T. J. Santner, B. J. Williams, and W. I. Notz, *The Design and Analysis of Computer Experiments*. New York, NY, USA: Springer, 2013.

[21] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn, "Design and analysis of computer experiments," *Stat. Sci.*, vol. 4, no. 4, pp. 409–423, 1989.

[22] C. Currin, T. Mitchell, M. Morris, and D. Ylvisaker, "Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments," *J. Amer. Stat. Assoc.*, vol. 86, no. 416, pp. 953–963, Dec. 1991.

[23] J. Mockus, V. Tiesis, and A. Zilinskas, *Toward Global Optimization, Chapter Bayesian Methods for Seeking the Extremum*, vol. 2. Amsterdam, The Netherlands: North Holland, 1978.

[24] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black-box functions," *J. Global Optim.*, vol. 13, no. 4, pp. 455–492, 1998.

[25] B. J. Williams, T. J. Santner, and W. I. Notz, "Sequential design of computer experiments to minimize integrated response functions," *Statistica Sinica*, vol. 4, no. 4, pp. 1133–1152, 2000.

[26] B. J. Williams, T. J. Santner, W. I. Notz, and J. S. Lehman, "Sequential design of computer experiments for constrained optimization," in *Statistical Modelling and Regression Structures*. Berlin, Germany: Springer, 2010, pp. 449–472.

[27] E. Vazquez and J. Bect, "Convergence properties of the expected improvement algorithm with fixed mean and covariance functions," *J. Stat. Planning Inference*, vol. 140, no. 11, pp. 3088–3095, Nov. 2010.

[28] X. Deng, V. R. Joseph, A. Sudjianto, and C. F. J. Wu, "Active learning through sequential design, with applications to detection of money laundering," *J. Amer. Stat. Assoc.*, vol. 104, no. 487, pp. 969–981, Sep. 2009.

[29] K. Crombecq, D. Gorissen, D. Deschrijver, and T. Dhaene, "A novel hybrid sequential design strategy for global surrogate modeling of computer experiments," *SIAM J. Sci. Comput.*, vol. 33, no. 4, pp. 1948–1974, Jan. 2011.

[30] R. Jin, C.-J. Chang, and J. Shi, "Sequential measurement strategy for wafer geometric profile estimation," *IIE Trans.*, vol. 44, no. 1, pp. 1–12, Jan. 2012.

[31] H. Yan, "High dimensional data analysis for anomaly detection and quality improvement," Ph.D. dissertation, Georgia Inst. Technol., Atlanta, GA, USA, 2017.

[32] D. Xiang, S. Gao, W. Li, X. Pu, and W. Dou, "A new nonparametric monitoring of data streams for changes in location and scale via cucconi statistic," *J. Nonparametric Statist.*, vol. 31, no. 3, pp. 743–760, Jul. 2019.

[33] N. Cressie, "Statistics for spatial data," *Terra Nova*, vol. 4, no. 5, pp. 613–617, 1992.

[34] M. L. Stein, *Interpolation of Spatial Data: Some Theory for Kriging*. New York, NY, USA: Springer, 1999.

[35] V. R. Joseph and Y. Hung, "Orthogonal-maximin Latin hypercube designs," *Statistica Sinica*, vol. 18, no. 1, pp. 171–186, 2008.

[36] K. Chaloner and I. Verdinelli, "Bayesian experimental design: A review," *Stat. Sci.*, vol. 10, no. 3, pp. 273–304, Aug. 1995.

[37] F. Olsson and K. Tomanek, "An intrinsic stopping criterion for committee-based active learning," in *Proc. 13th Conf. Comput. Natural Lang. Learn. (CoNLL)*, 2009, pp. 138–146.

[38] J. Du, X. Yue, J. H. Hunt, and J. Shi, "Optimal placement of actuators via sparse learning for composite fuselage shape control," *J. Manuf. Sci. Eng.*, vol. 141, no. 10, Oct. 2019, Art. no. 101004.

[39] X. Yue and J. Shi, "Surrogate model–based optimal feed-forward control for dimensional-variation reduction in composite parts' assembly processes," *J. Qual. Technol.*, vol. 50, no. 3, pp. 279–289, Jul. 2018.

[40] Y. Wen, X. Yue, J. H. Hunt, and J. Shi, "Virtual assembly and residual stress analysis for the composite fuselage assembly process," *J. Manuf. Syst.*, vol. 52, pp. 55–62, Jul. 2019.

[41] Y. Wang, X. Yue, R. Tuo, J. H. Hunt, and J. Shi, "Effective model calibration via sensible variable identification and adjustment, with application to composite fuselage simulation," 2019, *arXiv:1912.12569*. [Online]. Available: http://arxiv.org/abs/1912.12569

Xiaowei Yue (Member, IEEE) received the B.S. degree in mechanical engineering from the Beijing Institute of Technology, Beijing, China, in 2011, the M.S. degree in power engineering and engineering thermophysics from Tsinghua University, Beijing, in 2013, and the M.S. degree in statistics and the Ph.D. degree in industrial engineering with minor in machine learning from the Georgia Institute of Technology, Atlanta, GA, USA, in 2016 and 2018, respectively.

He is currently an Assistant Professor with the Grado Department of Industrial and Systems Engineering, Virginia Tech, Blacksburg, VA, USA. His research interest is focused on engineering-driven data analytics for advanced manufacturing.

Dr. Yue is a member of the American Society of Mechanical Engineers (ASME), the Institute for Operations Research and the Management Sciences (INFORMS), and the Institute of Industrial and Systems Engineers (IISE). He was a recipient of the IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING Best Paper Award and several other best paper awards.

Yuchen Wen received the B.S. degree in computer science from Shanghai Jiao Tong University, Beijing, China, and the University of Michigan, Ann Arbor, MI, USA, in 2014, and the M.S. degree in statistics from the Georgia Institute of Technology, Atlanta, GA, USA, in 2018, where he is currently pursuing the Ph.D. degree in industrial and systems engineering.

His research interest is focused on developing machine learning algorithms for advanced manufacturing systems, especially composite fuselage assembly process.

Mr. Wen is a member of the Institute for Operations Research and the Management Sciences (INFORMS) and the Institute of Industrial and Systems Engineers (IISE).

Jeffrey H. Hunt received the B.S. degree in physics from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 1979, and the M.A. degree in physics and the Ph.D. degree in physics from the University of California at Berkeley, Berkeley, CA, USA, in 1981 and 1988, respectively.

He is currently a Principal Scientist and a Senior Technical Fellow with Boeing Company, El Segundo, CA, USA. His career has included physics-based projects in condensed matter physics, quantum information sciences, surface science, and nonlinear optics and work on diverse applications, including both in defense sciences and commercial air and space technologies. He has published over 30 articles, three books, and two encyclopedia articles on condensed matter sciences. He holds 61 U.S. patents. His main research areas are wide scientific and technical challenges in aviation and aerospace industry, with particular applications in condensed matter sciences and nonlinear optics, composite aircraft assembly, and so on.

Dr. Hunt is a fellow of the American Physical Society and the Optical Society of America.

Jianjun Shi received the B.S. and M.S. degrees in electrical engineering from the Beijing Institute of Technology, Beijing, China, in 1984 and 1987, respectively, and the Ph.D. degree in mechanical engineering from the University of Michigan, Ann Arbor, MI, USA, in 1992.

He is currently a Carolyn J. Stewart Chair Professor with the H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, USA. His research interests include the fusion of advanced statistical and domain knowledge to develop methodologies for modeling, monitoring, diagnosis, and control for manufacturing systems.

Dr. Shi is a fellow of the Institute of Industrial and Systems Engineers (IISE), the American Society of Mechanical Engineers (ASME), and the Institute for Operations Research and the Management Sciences (INFORMS), an Academician of the International Academy for Quality, an Elected Member of the International Statistical Institute (ISI), a Life Member of the American Statistical Association (ASA), and a member of the National Academy of Engineering (NAE).