

# A survey of adaptive sampling for global metamodeling in support of simulation-based complex engineering design

Haitao Liu<sup>1</sup> · Yew-Soon Ong<sup>2</sup> · Jianfei Cai<sup>3</sup>

Received: 2 January 2017 / Revised: 13 April 2017 / Accepted: 7 June 2017 / Published online: 27 June 2017  
© Springer-Verlag GmbH Germany 2017

**Abstract** Metamodeling is becoming a rather popular means to approximate the expensive simulations in today's complex engineering design problems since accurate metamodels can bring in a lot of benefits. The metamodel accuracy, however, heavily depends on the locations of the observed points. Adaptive sampling, as its name suggests, places more points in regions of interest by learning the information from previous data and metamodels. Consequently, compared to traditional space-filling sampling approaches, adaptive sampling has great potential to build more accurate metamodels with fewer points (simulations), thereby gaining increasing attention and interest by both practitioners and academicians in various fields. Noticing that there is a lack of reviews on adaptive sampling for global metamodeling in the literature, which is needed, this article categorizes, reviews, and analyzes the state-of-the-art single-/multi-response adaptive sampling approaches for global metamodeling in support of simulation-based engineering design. In addition, we also review and discuss some important issues that affect the success of an adaptive sampling approach as well as providing brief remarks on adaptive sampling for other purposes. Last, challenges and future research directions are provided and discussed.

**Keywords** Adaptive sampling · Global metamodeling · Simulation-based engineering design

✉ Haitao Liu  
htliu@ntu.edu.sg

<sup>1</sup> Rolls-Royce@NTU Corporate Lab, Singapore, Singapore

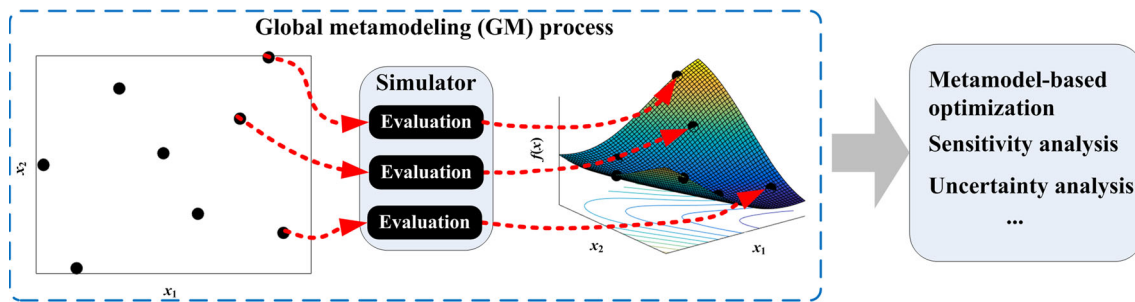
<sup>2</sup> Data Science and Artificial Intelligence Research Center, Nanyang Technological University, Singapore, Singapore

<sup>3</sup> School of Computer Science and Engineering, Nanyang Technological University, Singapore, Singapore

## 1 Introduction

Nowadays, computer simulation models, e.g., computational fluid dynamics (CFD) and finite element analysis (FEA), are gaining widespread use in various types of engineering design problems, e.g., global optimization, domain exploration, and sensitivity/uncertainty analysis. The simulation models can approximate detailed information of real-world physical problems, but require huge computational resources. To mitigate the computing cost, metamodels, also known as surrogates, response surfaces, or model-of-the-model, have gained intensive attention. Metamodels can mimic the behavior of simulation models with analytical expressions. Various types of metamodeling techniques, e.g., Kriging (also known as Gaussian process, GP) (Cressie 1988; Rasmussen 2006), radial basis functions (RBF) (Dyn et al. 1986), polynomial response (PR) and support vector regression (SVR) (Clarke et al. 2005), have been continuously developed. For more information, please refer to Refs. (Wang and Shan 2007; Kleijnen 2009; Razavi et al. 2012b; Viana et al. 2014).

The scope of this survey is confined to *global metamodeling* (GM), whose goal is to build a global metamodel as accurate as possible for deterministic simulation-based problems with reasonable computational budget. Figure 1 illustrates the global metamodeling process for approximating a target function  $f$ , which consists of two main steps: (1) sampling (also known as Design of Experiments, DoE), wherein a set of points is generated over the domain; and (2) metamodeling, wherein a metamodel  $\hat{f}$  is fitted to the observed points. The high prediction quality of a metamodel is beneficial for further applications, e.g., the convergence speedup in metamodel-based optimization, and the accurate identification of important inputs in sensitivity analysis. It is found that the choice of observed points by DoE is crucial to the prediction quality of a metamodel. Thus,



**Fig. 1** Global metamodeling process and its applications

sampling approaches that deal with how to gather informative experiments for better understanding a given phenomenon have been intensively studied. For global metamodeling, it is to build an accurate global metamodel with as few points (simulations) as possible.

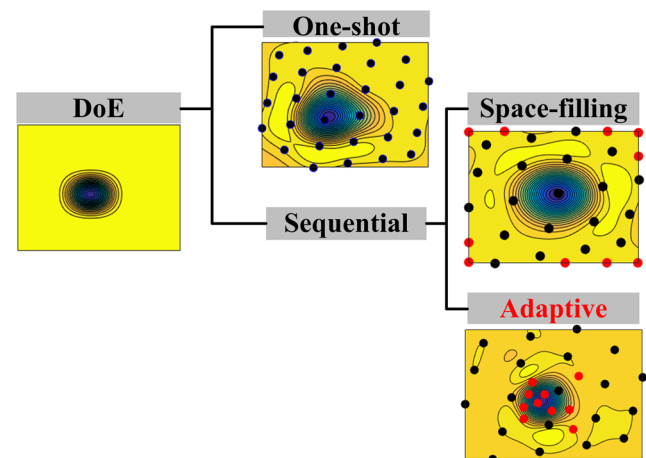
In general, as shown in Fig. 2, the conventional sampling approaches can be classified into two categories: *one-shot sampling* and *sequential sampling*. One-shot sampling approaches determine the sample size and points in a single stage. However, for a target function with no prior knowledge, it is hard to predetermine an optimal or appropriate sample size. Hence, flexible sequential sampling approaches have been introduced, which sequentially determine the points using the information from previous iterations. The sequential sampling approaches can be classified into two categories: *space-filling sequential sampling* and *adaptive sequential sampling*. “space-filling” means that the generated points spread over the entire domain evenly. The space-filling sequential sampling approaches are usually developed from some well-known one-shot sampling criteria by simulating the points in a sequential manner.

Compared to the space-filling sequential sampling, the adaptive sequential sampling, also known as active learning (Settles 2010), allows to choose informative points via the metamodel or data that it learns, and consequently, performs better with fewer points. For a complex function, rather than sampling the points evenly, it has been suggested that more points should be placed in regions where the model has large prediction errors. That is, the sampling process should adapt to the properties of the target function. These interesting regions are denoted as “current region of interest” (Barton 1997), “continuous and multimodal region” (Li et al. 2010), “region of interest” (Lin et al. 2004), or “irregular region” (Farhang-Mehr and Azarm 2005). The adaptive sampling strategy is particularly interesting for simulation-based problems, since it is potential to build accurate global metamodels with fewer points than the space-filling sampling, thereby saving the cost of expensive simulations. The adaptive sampling strategy has been demonstrated in many successful real-world engineering applications, including design of a heat exchanger (Aute et al. 2013), aerodynamic data modeling (Mackman and Allen 2010b; Rosenbaum and Schulz 2012), parametric

macromodeling of a microwave antenna (Deschrijver et al. 2011), design of an industrial naphtha cracking furnace (Jin et al. 2016) and others.

In the past two decades, many adaptive sampling approaches, together with the metamodeling techniques, have been continuously developed for various purposes. To the best of our knowledge, however, there has been a lack of reviews on the adaptive sampling for global metamodeling in the literature. Recently related reviews, on the other hand, have mainly focused on traditional space-filling sampling approaches (Kleijnen 2008; Pronzato and Müller 2012; Kleijnen 2015; Damblin et al. 2013; Sanchez and Wan 2015). Pickett and Turner (2011) reviewed various adaptive sampling criteria for the development of NURBs-based metamodels. In the field of machine learning, Settles (2010) and Fu et al. (2013) reviewed the active learning approaches mainly for classification problems.

Therefore, the objective of this article is to categorize, review, and analyze the literatures regarding adaptive sampling for deterministic global metamodeling, and examine the considerations and characteristics of adaptive sampling in support of complex engineering design. The main contributions of this review come in three-folds. First, this article categorizes the adaptive sampling approaches according to the sampling criterion for identifying interesting regions, and comprehensively reviews the diverse single-/multi-response adaptive



**Fig. 2** Conventional sampling categories

sampling approaches for global metamodeling occurred in the literature. Second, this article reviews and analyzes the design considerations for the completeness and success of an adaptive sampling approach, and provides brief remarks on adaptive sampling for other purposes. Last, based on the reviews and analyses, some attractive research directions are highlighted for addressing the rising complexity of modern engineering design problems.

The remaining of this article is organized as follows. Section 2 first offers a non-exhaustive review regarding space-filling sampling for global metamodeling. With the formulation of adaptive sampling for global metamodeling in Section 3, we extensively review the state-of-the-art single-response adaptive sampling approaches and their design considerations in Section 4 and Section 5. Furthermore, Section 6 reviews the multi-response adaptive sampling approaches for global metamodeling in three scenarios. Some brief remarks on adaptive sampling for other purposes are also provided in Section 7. Finally, Section 8 and Section 9 discuss further research directions and concluding remarks.

## 2 Conventional space-filling sampling for global metamodeling

Classical sampling approaches that are mainly developed for physical experiments have a long history in the DoE field. The widely used classical sampling approaches include Full Factorial design, Fractional Factorial design, Central Composite design, Box-Behnken design, and Plackett-Burman design (Wang and Shan 2007). For classical sampling approaches, (1) they determine the sample locations in order to reduce the impact of random errors on physical experiments; (2) they usually sample more points around the boundary regions rather than the interior regions; and (3) they are *model-dependent*, i.e., the sampling process is based on a pre-determined regression model. But for deterministic computer codes, a good sampling approach prefers filling the entire domain (i.e., space-filling) rather than concentrating on boundary (Sacks et al. 1989; Jin et al. 2001).

The space-filling sampling tends to spread the points over the entire domain evenly in order to gather most information about the underlying function. Table 1 lists the basic space-filling approaches and the related references. Among them, the Latin Hypercube Design (LHD) has gained much popularity in various fields. According to the survey (Viana 2013) from Google Scholar, there is strong growth in the number of papers using LHD when compared to other space-filling sampling approaches. One main reason for the popularity of LHD is that (1) it has the *space-filling* property to make points fill up the entire domain; meanwhile, (2) it has the *non-collapsing* property; that is, if a few inputs are useless, the resulting points still follow a LHD type.

To further improve the space-filling property of LHD, the so-called optimal LHD has been developed by combining with some space-filling strategies, e.g., the maximin criterion (van Dam et al. 2007, 2009) and the orthogonal arrays criterion (Joseph and Hung 2008; Lin et al. 2009; Loepky et al. 2012). However, as a sophisticated optimization problem, obtaining an optimal LHD is a non-trivial task. For example, the maximin-LHD has been shown to be a discrete optimization problem believed to be NP-hard (Rimmel and Teytaud 2014). Hence, different time-consuming optimization strategies, e.g., columnwise-pairwise (Stocki 2005), genetic algorithm (Liefvendahl and Stocki 2006), simulated annealing (Pholdee and Bureerat 2015), enhanced stochastic evolutionary algorithm (Jin et al. 2005), and iterated local search heuristics (Grosso et al. 2009), have been presented to account for the non-collapsing property under different optimal criteria, e.g., maximin distance criterion,  $\phi_p$  criterion, entropy criterion, and centered  $L_2$ -discrepancy criterion (Jin et al. 2005). Recent comparison studies of these search algorithms for optimal LHD can be found in (Damblin et al. 2013; Rimmel and Teytaud 2014). To further alleviate the optimization budget, a near-optimal LHD is achieved by the translational propagation idea (Viana et al. 2010b; Pan et al. 2014a). It first creates small blocks containing a good seed design with a few points, and then translates them over the entire hypercube. This approach is much more efficient than the formal optimization based approaches, while still yielding comparable LHD points.

The above space-filling approaches, however, usually run in a one-shot fashion; that is, all the points are generated in a single stage. Sequential space-filling sampling further improves the one-shot space-filling by sequentially adding new points for saving computational budget. For example, to obtain  $m$  points with  $n$  dimensions, the one-shot sampling criterion needs to solve an  $m \times n$  dimensional optimization problem, which becomes harder with the increase of dimensionality and sample size. While the sequential sampling decomposes this large problem into a sequence of  $n$  dimensional sub-problems.

Many of current space-filling approaches can be easily modified to run sequentially. For sequential LHD, it needs specific strategies for ensuring the non-collapsing property. Wang (2003) partitioned the domain along each dimension according to the required sample size, and then sampled new points in the intervals with no existing ones. But due to the distribution of the observed points, the proper domain partition cannot always be successfully achieved. A more general idea (van Dam et al. 2007) is to treat the sequential LHD process as a set of optimization problems, with the objective being some space-filling criteria, and the constraints being a set of one-dimensional distance thresholds. This quasi-optimal sequential LHD strategy thereafter has been widely studied and extended (Xiong et al. 2009; Crombecq et al. 2011b; Liu et al. 2015a).

Space-filling sampling has also been studied and applied in some interesting scenarios. For example, the *nested space-*

**Table 1** Space-filling sampling approaches

Approach	Ref.	Approach	Ref.
Uniform design	Fang et al. (2000)	Sobol's sequence	Sobol' (1979)
Maximin/Minimax design	Johnson et al. (1990)	Halton sequence	Halton (1960)
Orthogonal array	Owen (1992)	Latin Hypercube Design	McKay et al. (1979)
Audze-Eglais	Audze and Eglais (1977)	Lattice sampling	Patterson (1954)
Hammersley sequences	Kalagnanam and Diwekar (1997)	Centroidal Voronoi tessellation	Saka et al. (2007)

*filling design* (Husslage et al. 2005; van Dam et al. 2010; Rennen et al. 2010; Qian et al. 2009; Qian and Ai 2010; Haaland and Qian 2010) were proposed for multiple responses with linking parameters, sequential evaluation and multistage model fitting. “nested” means the design consists of two sample sets, with one being a subset of the other. This type of sampling approach has been widely used for multi-fidelity modeling, see Section 6.3. Besides, if the domain is no longer a regular hypercube, some space-filling sampling approaches for irregular regions have been presented by using, e.g., the maximin criterion (Stinstra et al. 2003; Auffray et al. 2012), the non-collapsing sampling criterion (Draguljić et al. 2012), and the central composite discrepancy criterion (Chuang and Hung 2010; Chen et al. 2014). Such constrained space-filling sampling approaches are especially useful for problems with irregular input regions.

Finally, it is worth noting that most of the space-filling sampling approaches reviewed here are *model-independent*, i.e., the determination of sample points is independent of the model or output characteristics. As a result, the sampling process cannot gain benefits from any new findings from the subsequent modeling process. In contrast, the adaptive sampling to be reviewed next utilizes the characteristics of both inputs and outputs for choosing informative points. Besides, for some variance based sampling approaches in (Jin et al. 2002), we prefer reviewing them in the context of adaptive sampling since they slightly adapt to the output responses by utilizing the variance information.

This section offers a non-exhaustive review of conventional space-filling sampling approaches for global metamodeling. For more information about space-filling sampling, one can refer to the recent books, reviews and comparison studies (Kleijnen 2008; Pronzato and Müller 2012; Kleijnen 2015; Damblin et al. 2013; Sanchez and Wan 2015), and the interesting website <https://spacefillingdesigns.nl/>.

### 3 Problem formulation: Adaptive sampling for global metamodeling

Adaptive sampling is designed for scenarios where the output responses of sample points are very expensive to obtain. In the

field of machine learning, adaptive sampling, known as active learning, is described as “... *the key hypothesis is that if the learning algorithm is allowed to choose the data from which it learns—to be “curious”, if you will—it will perform better with less training.*” (Settles 2010). For global metamodeling, adaptive sampling seeks the sequential collection of informative points for building accurate global metamodels with less computing cost.

Figure 3 depicts a general adaptive sampling process for global metamodeling. For the target function  $f$  defined in  $D \in \mathbb{R}^n$ , the sampling process begins with a set of initial points  $\mathbf{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_m\}^T$ , and uses them to fit a metamodel  $\hat{f}$ . Then based on an adaptive sampling criterion, it solves an auxiliary optimization problem to iteratively add new points  $\mathbf{X}_{\text{new}}$  and update the metamodel until a stopping criterion is met. Suppose that an adaptive sampling approach runs in a one-by-one fashion, generally the new point is selected by maximizing a score function as:

$$\mathbf{x}_{\text{new}} = \underset{\mathbf{x} \in D}{\operatorname{argmax}} \operatorname{Score}(\operatorname{local}(\mathbf{x}), \operatorname{global}(\mathbf{x})) \quad (1)$$

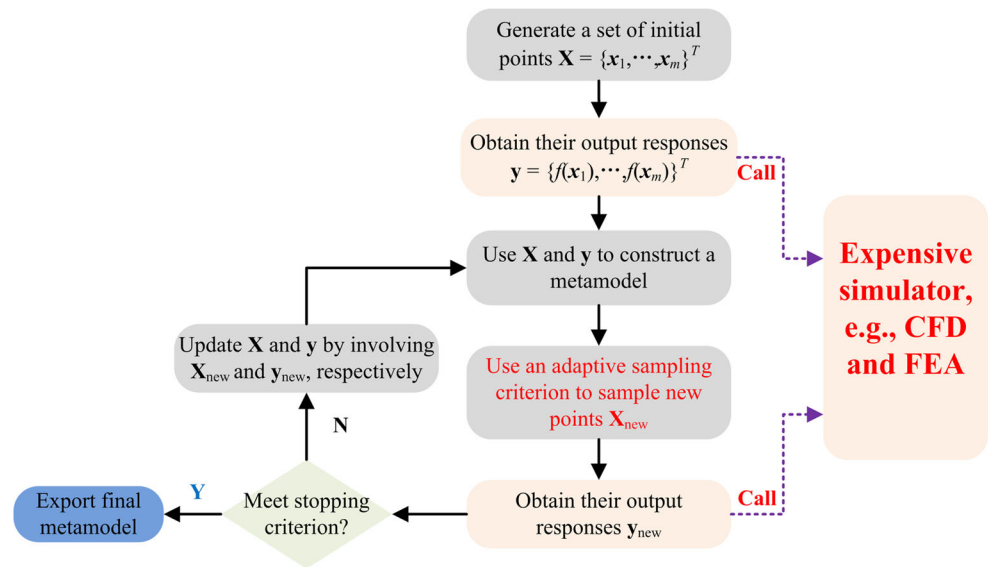
where  $\operatorname{local}(\mathbf{x})$  and  $\operatorname{global}(\mathbf{x})$  respectively represent the local exploitation term and global exploration term, which will be explained below. It is notable that we need to execute the expensive simulator to obtain the true response  $f(\mathbf{x}_i)$  at point  $\mathbf{x}_i$ . The adaptive sampling process here is illustrated for single-response problems that contain only a single target function. It however can be modified to handle multi-response problems with multiple target functions, which will be discussed and reviewed in Section 6.

For efficiently improving the overall model accuracy, an adaptive sampling approach should contain two conflicting parts (Deschrijver et al. 2011; Liu et al. 2016a):

- *Local exploitation*: This part plays a key role in adaptive sampling, since it helps find the interesting regions with large prediction errors to sample more points. Note that the actual prediction error in practice is unknown a priori. Alternatively, we can use the prediction variance, cross-validation error, local nonlinearity, or local gradients to represent it. Thus, the most important task in adaptive



**Fig. 3** General flowchart of an adaptive sampling approach for global metamodeling



sampling is to design an appropriate local exploitation criterion for identifying interesting regions.

- *Global exploration:* This part, which helps discover the interesting regions that have not yet been detected before, makes the adaptive sampling process complete for global metamodeling. A pure exploitation based adaptive sampling approach is often biased due to having an incomplete view of the entire domain. To address this issue, the global exploration term based on, e.g., some distance criteria, is required.

Finally, recall Fig. 2 that we use different kinds of sampling approaches to build a Kriging model for the function with a center unimodal region. We use the translational propagation LHD (TPLHD) sampling approach (Viana et al. 2010b) to generate thirty one-shot points; while we use the sequential maximin (MM) sampling approach (Johnson et al. 1990) to generate ten space-filling sequential points (the red circles), and use the CV-Voronoi sampling approach (Xu et al. 2014) to generate ten adaptive sequential points. By using the root mean square error (RMSE) to assess the model accuracy, it is found that the adaptive sampling approach yields more accurate predictions ( $RMSE_{CVV} = 0.026$ ,  $RMSE_{MM} = 0.095$ ,  $RMSE_{TPLHD} = 0.072$ ) using the same number of points.

#### 4 Single-response adaptive sampling for global metamodeling

We begin with the single-response adaptive sampling because most of current adaptive sampling approaches are designed for this scenario. As stated earlier, for discovering interesting regions, the key to an adaptive sampling approach is to design an effective local exploitation criterion that well estimates the actual prediction errors  $e = |f - \hat{f}|$  over the domain. Based on the way of representing the actual prediction errors, we classify current adaptive sampling approaches into four categories, as shown in Table 2:

- The variance based adaptive sampling asserts that regions with large prediction variances (i.e., large uncertainty) estimated by the statistical model is most likely to contain large prediction errors;
- The query-by-committee (QBC) based adaptive sampling is a semi-variance based strategy that uses the response variances estimated simultaneously by several competing metamodels to represent the prediction errors;
- The cross-validation (CV) based adaptive sampling directly estimates the prediction errors using the cheap-to-run cross-validation process; and

**Table 2** Classification of current single-response adaptive sampling approaches

Type	Local exploitation	Global exploration	Model-dependent
Variance based	Variance, adjusted variance using some error metrics, or some modified expected improvement criteria	Distance based variance	Y
Query-by-committee based	Response variance	Distance criteria	N
Cross-validation based	Cross-validation errors and some variants	Distance criteria or some implicit error-pursuing mechanism	N
Gradient based	Gradient information or geometric information	Distance criteria	Y/N

- Finally, the gradient based adaptive sampling employs the local gradient information to implicitly represent the prediction errors.

Note that as explained in Section 2, “model-dependent” in Table 2 characterizes whether an adaptive sampling approach utilizes some inner properties of a specific metamodel type or is independent of the choice of model type (i.e., model-independent). In what follows, the four categories of adaptive sampling in current literature will be reviewed and analyzed in detail.

#### 4.1 Variance based adaptive sampling

The variance based sampling approaches are deeply combined with the Kriging or GP model which considers the output response as the realization of a Gaussian process. Through using the prior information from observed points, this type of model constructed through the Bayesian rule offers a posterior Gaussian distribution  $y(\mathbf{x}) \sim \text{GP}(\hat{f}(\mathbf{x}), \hat{\sigma}^2(\mathbf{x}))$  where  $\hat{f}(\mathbf{x})$  is the prediction response and  $\hat{\sigma}^2(\mathbf{x})$  is the prediction variance (also known as mean square error (MSE)). The prediction variance, regarded as an estimation of the actual prediction error, is widely employed to assist the sampling process for reducing model uncertainty, thus improving the prediction quality.

In the statistics community, many variance based sampling approaches have been proposed. The intuitive idea is to sample a new point by maximizing the mean square error (MMSE) (Jin et al. 2002) as

$$\mathbf{x}_{\text{new}} = \underset{\mathbf{x} \in D}{\operatorname{argmax}} \hat{\sigma}(\mathbf{x}) \quad (2)$$

Moreover, Sacks et al. (1989) considered the integrated mean square error (IMSE) criterion by averaging the MSE value over the entire domain as  $\int_{\mathbf{x} \in D} \hat{\sigma}(\mathbf{x}) d\mathbf{x}$ . The IMSE criterion selects a new point that maximizes the average reduction in prediction variance over the entire domain after adding it in the sample set. Besides, Shewry and Wynn (1987) proposed the maximum entropy (ME) criterion that selects new points via maximizing the determinant of correlation matrix in the Bayesian framework. Morris et al. (1993) further combined the ME criterion with the known first derivatives to improve the sampling performance. Interestingly, Jin et al. (2002) pointed that the MMSE criterion is equivalent to the one-by-one ME criterion. These traditional variance based sampling approaches can be integrated into the Kriging metamodeling to form a procedure called *Design and Analysis of Computer Experiments* (DACE) (Sacks et al. 1989; Viana et al. 2014). Recently, an efficient sequential sampling approach was

introduced, which selects new points by maximizing the mutual information over the design space (Beck and Guillas 2016).

It is notable that the above variance based sampling approaches utilize the property of the Kriging model that follows a stationary assumption, in which the covariance structure is identical over the entire domain. Consequently, the points generated by them are primarily space-filling with a slight adaption to the relative variability along each coordinate direction. From this view, the prediction variance of the Kriging model can be further improved to better represent the actual prediction error.

To this end, one way is to introduce some non-stationary models, e.g., the Bayesian treed GP model (Gramacy and Lee 2009) and the non-stationary covariance-based Kriging model (Xiong et al. 2007), and then apply the variance based sampling criteria to them (Gramacy and Lee 2006, 2009). The other way, which is the focus of this survey article, is to develop some adaptive sampling criteria by adjusting the prediction variance. We here employ the well-known *bias-variance decomposition* (Geman et al. 1992)

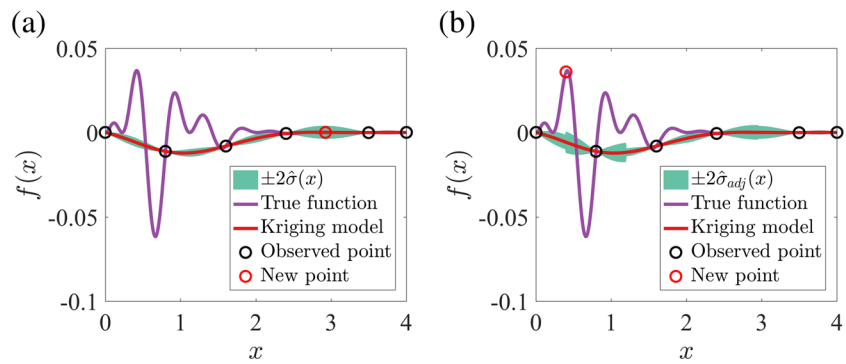
$$\begin{aligned} \mathbb{E}\{[y(\mathbf{x}) - f(\mathbf{x})]^2\} &= [\mathbb{E}(y(\mathbf{x})) - f(\mathbf{x})]^2 \\ &+ \mathbb{E}\{[y(\mathbf{x}) - \mathbb{E}(y(\mathbf{x}))]^2\} \end{aligned} \quad (3)$$

in the field of machine learning to explain the potential of adjusting prediction variance. The first term in the right-hand side of (3), denoted as *bias*, means the difference between the prediction response and the actual response; while the second expectation term, denoted as *variance*, is actually the prediction variance  $\hat{\sigma}^2(\mathbf{x})$  of the metamodel. For quickly reducing the generalization error of the metamodel,  $\int_{\mathbf{x}} \mathbb{E}\{[y(\mathbf{x}) - f(\mathbf{x})]^2\} d\mathbf{x}$ , the selection of new points should reduce not only the prediction variance but also the bias. In practice because of the unknown actual response  $f(\mathbf{x})$ , we need to estimate the bias information for adjusting the prediction variance.

Figure 4 depicts a 1D example to select a new point by maximizing the prediction variance and the adjusted prediction variance of Kriging, respectively. The considered function is  $f(x) = (\sin(7x) + \cos(14x))x^2e^{-4x}$ ,  $x \in [0, 4]$ , with the multimodal behavior in the left-hand side. The Kriging model is built by five points, and the shadow region represents 95% confidence interval. The prediction variance in Fig. 4(b) is adjusted by the approach in (Le Gratiet and Cannamela 2015). It is found that maximizing  $\hat{\sigma}(\mathbf{x})$  makes the new point fall into the largest interval, while maximizing  $\hat{\sigma}_{\text{adj}}(\mathbf{x})$  makes the new point fall into the region with large prediction errors.

Generally, there are two ways to adjust the prediction variance. The first way, denoted as *internal* variance based adaptive sampling, is to incorporate the bias information into the

**Fig. 4** A 1D example to select a new point by maximizing (a) the prediction variance and (b) the adjusted prediction variance of Kriging, respectively



prediction variance. Lin et al. (2004) pointed that the correlation function  $R(\cdot)$  between two arbitrary points in the Kriging model depends solely on the relative distance. For adaptive sampling, they adjusted the prediction variance via some error information as

$$\begin{aligned}\hat{\sigma}_{\text{adj}}^2(\mathbf{x}_i, \mathbf{x}_j) &= \sigma^2 R_{\text{adj}}(\mathbf{x}_i, \mathbf{x}_j) \\ &= \sigma^2 \prod_{k=1}^n \exp\left(-\eta_i \eta_j \theta_k |d_k|^2\right)\end{aligned}\quad (4)$$

where the adjusted correlation function  $R_{\text{adj}}(\cdot)$  uses the Gaussian formula as an example to describe the relationship between  $\mathbf{x}_i$  and  $\mathbf{x}_j$ ;  $\theta_k$  and  $d_k$  describe the correlation and the Euclidian distance between two points along the  $k$ th direction, respectively. Equation (4) achieves the adaption by introducing two adjust factors  $\eta_i$  and  $\eta_j$  that consider the error information at points  $\mathbf{x}_i$  and  $\mathbf{x}_j$ . As a result, the prediction variance now depends on not only the relative distance between two points but also their error information. In their approach the factors  $\eta_i$  and  $\eta_j$  are estimated using the prediction errors at both the observed points and some additional validation points. The validation points, however, are usually unavailable in practice. Hence, Liu et al. (2016a) suggested using the cross-validation error to formulate the adjust factors  $\eta_i$  and  $\eta_j$ . Similarly, Le Gratiet and Cannamela (2015) used both the cross-validation error and the cross-validation prediction variance to adjust the prediction variance. Besides, Farhang-Mehr and Azarm (2005) adjusted the prediction variance by the locations of local optima on the Kriging model in order to identify irregular regions. This strategy, however, heavily depends on the quality of Kriging model. A poor model may lead to an erroneous identification of interesting regions (Liu et al. 2016a). Different from the above approaches, the variance in the hierarchical variance sampling (HVS) approach (de Oliveira et al. 2012) is no longer estimated by Kriging. The HVS first adopts the classification and regression trees (CART) partition algorithm (Steinberg and Colla 2009) and the analysis of variance (ANOVA) splitting algorithm (Atkinson and Therneau 2000) to divide the entire domain. Then it calculates a statistical correction which depends on the sample size to derive an upper bound of the true variance

with a  $1-\alpha$  confidence. Finally, new points are selected by considering both the estimated variance upper bound and the region size.

The second way, denoted as *external* variance based adaptive sampling, is to add an external bias term for adjusting the prediction variance. For instance, Lam (2008) modified the expected improvement (EI) criterion (Jones et al. 1998) for global fit. This approach considers the nearest response variation in order to sample informative regions. The expected improvement over the nearest observed point is

$$EI_{\text{GF}}(\mathbf{x}) = \left(\hat{f}(\mathbf{x}) - f(\mathbf{x}^*)\right)^2 + \hat{\sigma}^2(\mathbf{x}) \quad (5)$$

where the first term in the right-hand side, playing as local exploitation, tends to be large at a point where its response is much different from that of the nearest observed point  $\mathbf{x}^*$ ; the second term, playing as global exploration, is large at a point where the metamodel has a large amount of uncertainty. Zhao et al. (2009) presented an insertion criterion (IC) as  $2Z_{(1+\alpha)/2} / 2\hat{\sigma}(\mathbf{x}) / |\hat{f}(\mathbf{x})|$  where  $Z_{(1+\alpha)/2}$  is the  $\alpha$ -level quantile of standard normal distribution. By maximizing the IC criterion, it can find the “weakest point” as the new point at which we have the least confidence on the prediction. This criterion, however, cannot handle the case where  $\hat{f}(\mathbf{x}) = 0$ . Different from the continuous formulation in the above approaches, Busby et al. (2007) used the adaptive gridding algorithm to decompose the domain into disjoint cells, the edges of which are of the order of the correlation lengths of different variables. The approach then applies the cross-validation criterion and maximum entropy criterion (Morris et al. 1993) to decide “bad cells” with large prediction errors or of empty for further sampling. Moreover, Busby (2009) modified this approach by replacing the maximum entropy criterion with the maximin distance criterion (Johnson et al. 1990).

It is worth noting that the above reviewed variance based adaptive sampling criteria have a complete form of local exploitation and global exploration conceptually described as

$$\mathbf{x}_{\text{new}} = \operatorname{argmax}_{\mathbf{x} \in D} \operatorname{Score}\left(\hat{\sigma}_{\text{adj}}(\mathbf{x}), \hat{\sigma}(\mathbf{x})\right). \quad (6)$$

The local exploitation is achieved by the variance term  $\hat{\sigma}_{\text{adj}}$  adjusted by, e.g., the prediction errors from validation points, the cross-validation errors, the locations of local optima, and the nearest response variation; the global exploration is usually achieved using the original prediction variance  $\hat{\sigma}$  provided by Kriging. Last, it is observed that the model-dependent variance based adaptive sampling approaches heavily rely on the inner properties of the statistical metamodel. Thus, they are hard to extend to other metamodel types, e.g., RBF.

#### 4.2 Query-by-committee based adaptive sampling

The query-by-committee (QBC) strategy was first introduced in Refs. (Seung et al. 1992; Freund et al. 1993) for classification problems in the field of machine learning. For global metamodeling, this strategy first maintains a committee  $\mathcal{C} = \{\hat{f}_i\}_{1 \leq i \leq t}$  of different metamodels representing competing hypotheses. Then each metamodel is required to predict the response at a candidate point  $\mathbf{x}$ . The point at which the committee members have maximal disagreement is selected as the new point. The level of disagreement at point  $\mathbf{x}$  is usually defined as the variance of the predictions of  $t$  committee members (Krogh and Vedelsby 1995). The simplest form is as

$$\hat{\sigma}_{\text{QBC}}^2(\mathbf{x}) = \frac{1}{t} \sum_{i=1}^t \left( \hat{f}_i(\mathbf{x}) - \bar{\hat{f}}(\mathbf{x}) \right)^2 \quad (7)$$

where  $\bar{\hat{f}}(\mathbf{x}) = \frac{1}{t} \sum_{i=1}^t \hat{f}_i(\mathbf{x})$  is the average value of  $t$  predictions at  $\mathbf{x}$ . Note that the QBC variance in (7) is estimated using the predictions from different metamodels. Thus, QBC has the capacity to identify the regions with large prediction errors. For instance, we build four Kriging models with the Gaussian basis function, exponential basis function, spline basis function and cubic spline basis function, respectively, for a 1D example in Fig. 5. The shadow region represents the prediction response plus/minus twice the QBC variance. It is found that maximizing the QBC variance makes the new point fall into the region with large prediction errors. Besides, compared to the variance based sampling approaches reviewed before, the QBC adaptive sampling is more generic since it is model-independent, i.e., it does not impose on the type of metamodel that can be used. Last, it is worth noting that QBC maximizes (7) to collect new points for reducing the variance over not only the output predictions but also the model parameters (Settles 2010).

Here, we use the *ambiguity decomposition* in the field of ensemble learning (Mendes-Moreira et al. 2012) to explain why QBC is able to quickly reduce the generalization error of the committee metamodels. We have

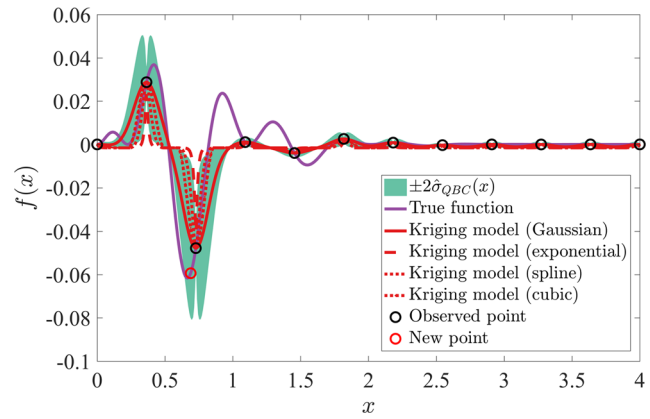


Fig. 5 A 1D example to show the QBC variance

$$\begin{aligned} \left( \bar{\hat{f}}(\mathbf{x}) - f(\mathbf{x}) \right)^2 &= \frac{1}{t} \sum_{i=1}^t \left( \hat{f}_i(\mathbf{x}) - f(\mathbf{x}) \right)^2 - \frac{1}{t} \\ &\quad \times \sum_{i=1}^t \left( \hat{f}_i(\mathbf{x}) - \bar{\hat{f}}(\mathbf{x}) \right)^2 \end{aligned} \quad (8)$$

where the first term in the right-hand side is the difference between the single metamodel prediction and the actual response, which reveals the prediction quality of the committee metamodels; while the second term is the QBC variance  $\hat{\sigma}_{\text{QBC}}^2(\mathbf{x})$  in (7). Once we choose the model types for the committee members, i.e., the first term of (8) is fixed, for reducing the generalization error the new points should be selected to maximize the QBC variance, which is what the QBC adaptive sampling attempts to do. It is found that the key in the QBC strategy is that there must be some disagreement among the committee models, called *QBC diversity*. In the context of QBC diversity, we classify current approaches into *homogeneous* QBC based adaptive sampling and *heterogeneous* QBC based adaptive sampling.

“homogeneous” means that the committee models are built using the same model type. In this context, to achieve the QBC diversity, we can generate multiple different sample sets  $\{\mathbf{X}_i\}_{1 \leq i \leq t}$  from the original sample set  $\mathbf{X}$ . To this end, one way is to utilize the resampling methods, e.g., the bootstrap (Efron and Tibshirani 1994) that generates  $t$  sample sets by selecting randomly with replacement from the set  $\mathbf{X}$ . For instance, Raychaudhuri and Hamey (1995) measured the disagreement among ten neural networks, each of which is built on random selection of half the sample set. The other way is to use the cross-validation method to partition the sample set into  $t$  groups, with each of the committee models built on  $t-1$  of the  $t$  groups. For example, Burbidge et al. (2007) used the cross-validation strategy to investigate the performance of QBC for global metamodeling, and pointed that this strategy works when the committee member’s bias is small. Besides, Kleijnen and Van Beers (2004) proposed a jackknifing criterion based on the leave-one-out cross-validation. In the jackknifing criterion, the committee models are trained using different subsamples. This kind of



approach can also be found in Refs. (Eason and Cremaschi 2014; Golzari et al. 2015). Gazut et al. (2008) applied both the bootstrap and the cross-validation to a QBC type adaptive sampling approach. Finally, similar to the homogeneous QBC adaptive sampling but without the perturbation of sample set  $\mathbf{X}$ , Ajdari and Mahlooji (2014) presented a quasi-QBC adaptive sampling approach. It first partitions the domain into a mesh of triangles, and then samples a new point in the triangle whose three vertices yield the function responses with maximal variation.

“heterogeneous” means that the committee models are built using different model types. In this context, the QBC diversity is achieved by the different model types built on the same sample set. For example, we can build the committee models by Kriging, RBF, SVR and so on; or simply, we build the committee models by Kriging but with different basis functions, see Fig. 5. Douak et al. (2012) used three different regression models to form the QBC committee, and found that this strategy outperforms the space-filling maximin criterion (Johnson et al. 1990) and the random (passive) sampling criterion. De Geest et al. (1999) suggested selecting a new point by maximizing a reflective function measured by the response difference between the best and second best metamodels. Furthermore, the sampling approach presented by Hendrickx and Dhaene (2005) constructs a committee containing three best metamodels, and simultaneously selects the new points at which two of the committee members have the maximal disagreement. Li et al. (2009) trained the multiple additive regression trees (MART) model on the sample set 20 times with different random seeds, and suggested using the coefficient of variance  $\hat{\sigma}/\mu$  (where  $\mu$  represents the mean) instead of the QBC variance to better rank the candidate points.

It is found that the QBC strategy plays the role of greedy local exploitation. With no constraints, the new points tend to cluster in regions with large prediction difference (Li et al. 2009). This can be confirmed in Fig. 5 that the new point is close to the observed point. For the improvement of global accuracy, a distance term  $d$ , which plays a role of global exploration, is often required to discard some points that are close to each other (Hendrickx and Dhaene 2005; Li et al. 2009; Eason and Cremaschi 2014). To summarize, the QBC based adaptive sampling process iteratively selects the new point as

$$\mathbf{x}_{\text{new}} = \underset{\mathbf{x} \in D}{\operatorname{argmax}} \operatorname{Score} \left( \hat{\sigma}_{\text{QBC}}(\mathbf{x}), d(\mathbf{x}) \right) \quad (9)$$

where the local exploitation is achieved by the QBC prediction variance, while the global exploration is achieved via the distance term.

### 4.3 Cross-validation based adaptive sampling

It is known that the Kriging model offers the prediction variance and thus there emerges a series of variance based

adaptive sampling approaches. But such model-dependent adaptive sampling approaches use specific model properties that exclude them from being used with other model types. This is an undesirable property and calls for more generic (model-independent) adaptive sampling approaches. Hence, in order to estimate the actual prediction error, the well-known cross-validation approach has attracted intensive attention in adaptive sampling, because it requires only the information at observed points.

For the commonly used leave-one-out cross-validation approach, it selects  $\mathbf{x}_i$  as a validation point and uses the remaining  $m-1$  points to fit a metamodel  $\hat{f}^{-i}$  for obtaining the cross-validation (CV) error as

$$\hat{e}(\mathbf{x}_i) = \left| f(\mathbf{x}_i) - \hat{f}^{-i}(\mathbf{x}_i) \right|, \quad 1 \leq i \leq m \quad (10)$$

This procedure is repeated at  $m$  observed points to obtain the generalized mean square cross-validation error (GMSE), where  $GMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m \hat{e}(\mathbf{x}_i)^2}$ . An accurate metamodel prefers a small GMSE value. The GMSE usually overestimates the actual metamodel accuracy, and is able to successfully estimate the model accuracy given sufficient observed data (Viana et al. 2009; Liu et al. 2016b). Besides, the cross-validation approach is capable of estimating the optimal model parameters (Rippa 1999; Sundararajan and Keerthi 2001), selecting the model type (Meckesheimer et al. 2002; Acar and Rais-Rohani 2009), simultaneously estimating the model parameters and selecting the model type (Acar 2014; Liu et al. 2016b), and constructing conservative metamodels (Viana et al. 2010a).

It is found that the CV errors are able to estimate the local prediction errors to some extent. A small  $\hat{e}(\mathbf{x}_i)$  implies that the model accuracy is insensitive to the loss of  $\mathbf{x}_i$ , that is, the metamodel has been well fitted around  $\mathbf{x}_i$ ; while on the contrary, a large  $\hat{e}(\mathbf{x}_i)$  indicates that the region around  $\mathbf{x}_i$  does not contain enough points such that the model accuracy is significantly affected by the loss of  $\mathbf{x}_i$ . Note that (10) cannot be used to estimate the errors at unobserved points. To address this issue for better conducting local exploitation, Li et al. (2010) and Aute et al. (2013) fitted an error model  $\hat{f}_e(\mathbf{x})$  to the CV errors at observed points. Besides, Jin et al. (2002) presented a more general CV error formula as  $\hat{e}(\mathbf{x}) = \sqrt{\frac{1}{m} \sum_{i=1}^m \left( f(\mathbf{x}) - \hat{f}^{-i}(\mathbf{x}) \right)^2}$ , which was also used by Kim et al. (2009). Moreover, Jiang et al. (2015) extended this formula to a weighted-sum cross-validation error where the weights reflect the influence of the observed points. According to the estimation of CV errors, we classify this type of sampling strategy into *continuous* CV based adaptive sampling and *discrete* CV based adaptive sampling.

In the context of continuous CV based adaptive sampling, the estimated CV errors help identify interesting regions for guiding local exploitation in adaptive sampling process (Li

et al. 2010; Aute et al. 2013; Jiang et al. 2015). It has been widely pointed that the obtained new points, however, would cluster around some observed points if one directly maximizes the CV based criteria, i.e., the focus is on pure local exploitation (Jin et al. 2002; Li et al. 2010; Aute et al. 2013). Hence, a distance based space-filling term  $d(\mathbf{x})$  in terms of global exploration is recommended to avoid the clustering phenomenon. A common way is to define a distance constraint to prevent the points from being too close to each other. The auxiliary optimization problem can be generally formulated as

$$\begin{aligned} & \text{Find } \mathbf{x}_{\text{new}} \\ & \max \hat{f}_e(\mathbf{x}) \leftarrow \mathbf{X}; \hat{e}(\mathbf{x}_1), \dots, \hat{e}(\mathbf{x}_m) \\ & \text{s.t. } \|\mathbf{x} - \mathbf{x}_i\| \geq d, \mathbf{x} \in D \end{aligned} \quad (11)$$

In practice, however, an appropriate threshold  $d$  value is hard to determine (Xu et al. 2014; Jiang et al. 2015). Too small  $d$  value cannot help avoid the clustering phenomenon, while too large  $d$  value forces the new points to spread over the entire domain evenly, which consequently makes the sampling approach incapable of exploiting the interesting regions. Li et al. (2010) suggested using the average of the minimal distance of each observed point as the threshold  $d$  value. While Aute et al. (2013) suggested the  $d$  value as the maximum of the minimal distance of each observed point, which leads to more emphasis on global exploration.

In the context of discrete CV based adaptive sampling, a common way is to partition the design space in order to promise the cooperation with the discrete CV errors. An early work (Devabhaktuni and Zhang 2000) used a set of test points to discover the region with the largest prediction error and split it into  $2^n$  regions. Although the approach is simple and efficient, it can lead to over-sampling for high-dimensional problems; meanwhile, in practice the test points are usually unavailable. Similarly, Braconnier et al. (2011) partitioned the domain into a set of hypercubes and selected the one with the largest CV error, and subsequently used the quad-tree algorithm (Finkel and Bentley 1974) to divide the hypercube into  $2^n$  equal hypercubes. This approach still has the over-sampling problem in high dimensions. It is notable that the approach presented by Xu et al. (2014) is very simple and effective. This approach first uses the Voronoi diagram algorithm (Aurenhammer 1991) to partition the domain into Voronoi cells based on the observed points. Then it calculates the CV errors of the cells and selects the one with the largest error as the sensitive cell. Finally, it searches a space-filling new point in the sensitive cell to do local exploitation and meanwhile avoid the local clustering phenomenon. An adaptive sampling approach similar to (Xu et al. 2014) has also been presented for non-intrusive POD-based metamodels (Vasile et al. 2013). The difference is that this approach determines the new point by considering both the POD basis and the design space. Finally, it is worth noting that compared to the continuous CV based

adaptive sampling approaches, these discrete CV based adaptive sampling approaches (1) discretize the auxiliary optimization problem via space partition such that the computational cost is relieved; and (2) keep an implicit global exploration via dynamic shift of the regions with the largest CV error (Xu et al. 2014).

To summarize, the general form of a CV based adaptive sampling process can be expressed as

$$\mathbf{x}_{\text{new}} = \operatorname{argmax}_{\mathbf{x} \in D} \operatorname{Score} \left( \hat{e}(\mathbf{x}), d(\mathbf{x}) \right) \quad (12)$$

where the local exploitation is achieved by the CV errors, while the global exploration is achieved via the distance constraints, the areas of partitioned polygons, or the implicit error-pursuing mechanism;  $D$  is a continuous or discrete design space.

#### 4.4 Gradient based adaptive sampling

Geometrically, the function responses in regions with large gradients have dramatic changes. As a result, the metamodel is hard to fit well and is easy to produce high prediction errors in these regions. Thus, the gradient information can help us discover the interesting regions for adaptive sampling. For conventional metamodels, e.g., the RBF model with smooth kernel functions, the gradient information at any point can be cheaply derived.

Rumpfkeil et al. (2011) assumed that the gradient information at observed points are available, and presented an adaptive sampling approach for refining the gradient-enhanced Kriging model. At each candidate point, this approach builds a local metamodel via Dutch Intrappolation and measures the discrepancy between the global Kriging model and local model. The new point is the one with the largest discrepancy. This approach, however, has a substantial restriction because of the requirement of available derivatives.

A practical way is to use the gradient information provided by the metamodel itself. Yao et al. (2009) used the first-order gradient information  $\partial \hat{f} / \partial \mathbf{x}$  of the RBFNN model to exploit the interesting regions. Meanwhile, to maintain the global model, they employed the optimal LHD approach (Kenny et al. 2000; Xiong et al. 2009) to add space-filling points if no accuracy improvement can be achieved using local exploitation. Moreover, since the curvature is a measure of how “curved” a curve is (Wei et al. 2012), and usually provides more information than the first-order gradients, some studies (Mackman and Allen 2010a, b; Wei et al. 2012; Mackman et al. 2013) used the curvature of the RBF model, i.e., the Hessian matrix  $H(\mathbf{x}) = \partial^2 \hat{f} / \partial \mathbf{x}^2$ , to guide the local exploitation. In these sampling approaches, global exploration is often achieved by a distance term. The gradient based adaptive sampling has also been applied to other model types. For adaptively refining NURBs-based metamodels (Turner et al.

2007; Pickett and Turner 2011), locally, a slope criterion is presented to choose a new point at which the model yields the most dramatic changes; globally, a proximity criterion is adopted to compute a parabolic span between adjacent control points. Then, the next point is selected as the one with the maximum parabola depth. Shahsavani and Grimvall (2009) proposed a simple adaptive sampling approach. It begins with a cuboid-shaped domain, and then identifies the cuboid containing the maximal roughness (nonlinearity), which is measured by the coefficients of a local second-order polynomial model, and finally splits the cuboid into two halves along the direction with the maximal curvature.

Yao et al. (2009) thought that though a metamodel built with a limited sample set is just a rough approximation of the target function, it describes the trend of  $f$  to some extent, and thus can guide the sampling procedure effectively. But Deschrijver et al. (2011) argued that the above reviewed gradient based adaptive sampling approaches highly rely on the quality of the metamodel. The model type and model parameters may affect the estimation of the gradient information, which influences the selection of new points. Therefore, more generic (model-independent) gradient based adaptive sampling approaches are needed.

To this end, a model-independent gradient based adaptive sampling approach, called LOLA-Voronoi (Crombecq et al. 2011a), and a variant (Deschrijver et al. 2011) and a faster version based on fuzzy logic (van der Hertten et al. 2015) have been continuously proposed, with the new point selected as

$$\mathbf{x}_{\text{new}} = \operatorname{argmax}_{\mathbf{x} \in D} G(\mathbf{x}) + V(\mathbf{x}) \quad (13)$$

In (13), the gradient term  $G(\mathbf{x})$ , measured by a local linear approximation built using a set of nearby points, is used to do local exploitation; while the sampling density  $V(\mathbf{x})$ , measured by the areas of the partitioned Voronoi cells, is used to do global exploration. This approach has been involved in a MATLAB toolbox *SURrogate MOdeling* (SUMO) (Gorissen et al. 2010) for adaptively refining metamodels. More simply, Lovison and Rigoni (2011) used the adjacent points to estimate the local Lipschitz constant at  $\mathbf{x}$  to represent the local nonlinearity of the target function.

Except the gradient information, other geometric features have also been considered in adaptive sampling. For instance, Pan et al. (2014b) used the locations of the local optima on the metamodel to guide local exploitation. The RBF model with multi-quadric basis function is selected in this sampling approach since it is twice differentiable and can easily obtain the first- and second-order gradients, which is beneficial for the search of local optima. The approach also uses a density function (Kitayama et al. 2011) constructed by the RBF network for global exploration. Dos Santos and dos Santos (2008) employed the delaunay triangulation algorithm to mesh the domain according to the observed points, and selected the

new point as the middle of the longest side of the largest triangle.

To summarize, the general form of the gradient based adaptive sampling can be described as

$$\mathbf{x}_{\text{new}} = \operatorname{argmax}_{\mathbf{x} \in D} \operatorname{Score} \left( \hat{g}(\mathbf{x}), d(\mathbf{x}) \right) \quad (14)$$

where the local exploitation term  $\hat{g}$  represents the estimated first-/second-order gradients based on some specific models, or it represents some other geometric features; while the global exploration is achieved by some distance based criteria.

## 5 Design considerations of single-response adaptive sampling

This section provides a review on the design considerations in single-response adaptive sampling for global metamodeling. In particular, the design of initial points and stopping criteria, the trade-off between local exploitation and global exploration, and the selection of new points, which should be considered for completing the sampling process in Fig. 3 and more importantly, for improving the sampling performance, are analyzed and discussed in what follows.

### 5.1 Initial data and stopping criteria

An adaptive sampling approach generally begins with the generation of some initial points that are recommended to fill the entire domain evenly. A space-filling initial set ensures that everywhere in the domain has equal chance to be detected. Consequently, the metamodel can better represent the underlying function and guide the subsequent adaptive sampling process to quickly find the interesting regions. Many of the one-shot sampling approaches and space-filling sequential sampling approaches reviewed in Section 2 can be used here.

The above discussions, however, have an implicit assumption that we can confidently determine how to generate initial points. Practical problems, however, often have scenarios where there already exists a set of initial points that may not be space-filling. An extreme case is that all the initial points gather together in a local region. To handle such scenarios, the adaptive sampling is required to possess an effective trade-off strategy between local exploitation and global exploration, which will be discussed in details later. For instance, in the extreme case, an effective trade-off strategy is able to identify the bad representation of current metamodel, and then the sampling approach turns to focus on global exploration.

Another important factor affecting the performance of an adaptive sampling approach is the number of initial points,  $N$ . It has been pointed out that if the sampling approach starts with an initial size that is too small, poor metamodel



predictions can lead the sampling process to focus at inappropriate locations (Kim et al. 2009; Ghoreyshi et al. 2009). On the contrary, if the sampling approach starts with an initial size that is too large, most of the computational budget can be wasted on space-filling points, which could have been spent more effectively on other adaptive sample points (Crombecq et al. 2011a). The proper initial sample size, however, depends on the dimensionality and complexity of the target function, the computational budget, and even the characteristics of the metamodeling techniques. Some metamodeling literatures suggested several empirical formulas to determine the proper initial size through extensive numerical experiments, as summarized in Table 3, where  $n$  represents the number of inputs and  $N_{\max}$  denotes the maximal number of function evaluations. It is worth noting that when using a GP model for approximation, Loepky et al. (2009) showed that the appropriate initial sample size should be ten times the dimensionality of the target function.

In simulation-based engineering design problems, the limited computational budget, i.e., the maximal number of function evaluations, is commonly used as a stopping criterion for adaptive sampling. Another common stopping criterion is the successive relative improvement (SRI) measured by the difference of some error criteria, e.g., the cross-validation error, the jackknifing variance (Kleijnen and Van Beers 2004), or the relative absolute error (Kim et al. 2009), in the last several iterations. The adaptive sampling process terminates as soon as there is no positive gain in the SRI.

## 5.2 Trade-off between local exploitation and global exploration

Recall that an adaptive sampling approach usually consists of two conflicting parts: local exploitation in interesting regions and global exploration over the entire domain. For improving the sampling performance, these two parts should be organized in a balanced manner, which is conceptually expressed as:

$$\text{Score} = w_{\text{local}} \times \text{local} + w_{\text{global}} \times \text{global} \quad (15)$$

where  $w_{\text{local}}$  and  $w_{\text{global}}$ , which are assumed to satisfy  $w_{\text{local}} + w_{\text{global}} = 1$ , are the weights for local exploitation and global exploration, respectively.

Liu et al. (2016a) showed that a valid trade-off not only conducts adequate local exploitation in identified interesting regions, but also duly offers global exploration to avoid missing some undetected interesting regions. Many of current adaptive sampling approaches, however, use a fixed rule to balance local exploitation and global exploration. For instance, the SEED approach (Lin et al. 2004) adopts a balance factor  $\lambda$  but fixes it at 2; the LOLA-Voronoi approach (Crombecq et al. 2011a) assigns the same weight for local exploitation and global exploration.

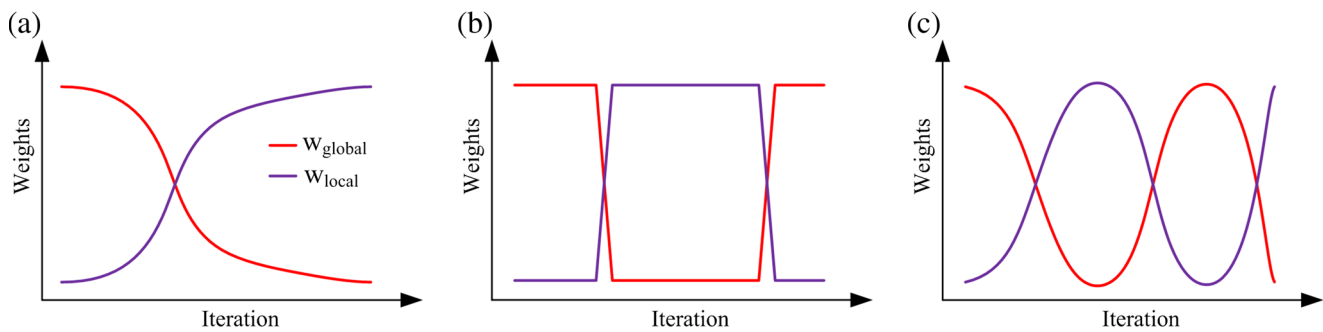
**Table 3** Empirical formulas to determine the initial sample size

Initial sample size	Property
$N = 10n$ (Jones et al. 1998; Loepky et al. 2009)	Depend on dimensionality
$N = 5n$ (Xu et al. 2014; Liu et al. 2016a)	
$N = 2^n$ (Gutmann 2001)	
$N = 2(n + 1)$ (Regis and Shoemaker 2007)	Depend on metamodeling property
$N = \max[2(n + 1), 0.1N_{\max}]$ (Razavi et al. 2012a)	
$N = (n + 2)(n + 1)/2 + 10$ (Busby 2009)	Depend on computing budget
$N = 0.35N_{\max}$ (Sóbester et al. 2005)	
$N = 0.5N_{\max}$ (Aute et al. 2013)	

Some researchers attempt to balance the two conflicting parts in a flexible way. In our view, the balance strategies between local exploitation and global exploration can be classified into three categories, as shown in Fig. 6. The first decreasing strategy (Kim et al. 2009; Singh et al. 2013; Turner et al. 2007) in Fig. 6(a) starts with a  $w_{\text{global}}$  value close to one to fully explore the entire domain in order to discover the interesting regions. Then with the sampling process progresses,  $w_{\text{global}}$  decreases iteratively, and the process converges with a  $w_{\text{local}}$  value close to one, which implies a complete local exploitation in identified interesting regions. For instance, by comparing the relative absolute errors of the metamodels in successive iterations, Kim et al. (2009) presented a decreasing law of the weights according to several error thresholds. It is noteworthy that instead of using a single global-local criterion, Turner et al. (2007) proposed using two pure global criteria and two pure local criteria to formulate a cooling schedule. This schedule uses Bernstein basis functions to decide the dominant criterion in different sampling stages. However, a main problem of the decreasing strategy is that in the final sampling stage, it almost ignores the global exploration. But since in practice many target functions are black-box, it is hard to ensure that the main function characteristics have been fully explored so that we can absolutely drop off the global exploration.

To address this issue, the second switch strategy (Sasena et al. 2002; Sasena 2002) or the greedy strategy (Singh et al. 2013) is proposed, as shown in Fig. 6(b). Singh et al. (2013) suggested a simple greedy strategy by using a threshold value  $\varepsilon$ . If a random value  $r < \varepsilon$ , the sampling process switches to pure global exploration; otherwise, it switches to pure local exploitation. The greedy strategy uses a fixed threshold value to choose global exploration and local exploitation, which sometimes is unreasonable. Hence, Sasena et al. (2002) and Sasena (2002) proposed a switch strategy that is similar to a cooling schedule used in the stochastic simulated annealing algorithm (Černý 1985). Initially, this switch strategy drives the global exploration to reduce the uncertainty of the entire domain. Then it switches to





**Fig. 6** Conceptual illustration of three balance strategies between local exploitation and global exploration

pure local exploitation in interesting regions for improving the model accuracy efficiently. If the predictions in current interesting regions are relatively accurate, it switches to pure global exploration to identify undetected interesting regions. Iteratively, the strategy switches back and forth between global exploration and local exploitation in order to improve the metamodel with as few points as possible.

The third adaptive strategy (Singh et al. 2013), as shown in Fig. 6(c), is a generalized version of the switch strategy. In successive sampling iterations, the local and global weights are adaptively changed in a dynamic way through comparing the information among successive iterations, e.g., the deviation between the model errors. Preliminary numerical experiments by Singh et al. (2013) revealed that the adaptive strategy outperforms the greedy and decreasing strategies. In a spirit similar to the adaptive strategy, Liu et al. (2016a) employed a novel balance strategy to perform adaptive sampling by circularly looping through a search pattern that contains several weights from global to local.

Except using an explicit way like (15) to balance local exploitation and global exploration, there exist some adaptive sampling approaches (Kleijnen and Van Beers 2004; Xu et al. 2014; Lin et al. 2004; Lam 2008) in which the trade-off is buried deep. Similar to the third strategy, the trade-off here usually changes in a dynamic way when the adaptive sampling progresses continuously.

It is worth noting that the trade-off between local exploitation and global exploration should gain sufficient attention, since it greatly influences the performance of adaptive sampling. In particular, the promising adaptive balance strategy needs more research efforts.

### 5.3 Single selection or batch selection

During the adaptive sampling process, if only a new point is selected in each iteration, we denote it as *single selection*; on the other hand, it is denoted as *batch selection* if more than one new point is added in each iteration. Most of current adaptive sampling approaches adopt single selection, because they usually use an adaptive sampling criterion to select the new point

that is the global optimum of an auxiliary optimization problem.

It is found that the batch sampling process can be easily deployed in a parallel fashion to improve computational efficiency, which is meaningful for expensive simulation-based problems. To achieve batch selection, a naive strategy is to use a sampling criterion to rank the candidate points, and select the  $q$  best points. This simple “ $q$ -best” strategy, however, is usually inappropriate because it does not consider the information overlapping of the  $q$  new points, and may lead to the clumping of batch points. The  $q$  new points in batch selection should be both *informative* and *diverse* to make the best use of computational resources. “informative” means that the new points should be sampled in the interesting regions; while “diverse” on the contrary requires that the  $q$  new points should be far away from themselves and the observed points in order to gather more information about the target function.

For space-filling batch sampling, the informative requirement and the diverse requirement can be merged, since this type of sampling advocates that the informative points should be space-filling. Recently, Loeppky et al. (2010) investigated different space-filling sampling criteria with single selection and batch selection, respectively. Thereafter, several works (Williams et al. 2011; Atamturktur et al. 2013) also paid attention on batch selection with different space-filling sampling criteria. It is notable that in batch sampling, because of selecting  $q$  points jointly, the dimensionality of the auxiliary optimization problem has greatly increased. To improve optimization efficiency, we often employ the modified exchange algorithm (Fedorov 1972) that replaces a point by the one that improves the sampling criterion at each iteration.

For adaptive batch sampling, the informative requirement and the diverse requirement should be considered jointly. For some partition based adaptive sampling approaches (Devabhaktuni and Zhang 2000; Shahsavani and Grimvall 2009; Braconnier et al. 2011), they usually run with batch selection: it partitions the domain into hypercubes and samples a few new points along different directions in the identified hypercube. But this type of batch selection considers only the informative requirement, since it is conducted in a single hypercube, leading to the clumping of

new points. Instead, Gramacy and Lee (2009) used a treed maximum entropy sampling approach that spaces the  $q$  new points apart from the observed points and themselves. Huang et al. (2015) modified the well-known DIRECT optimization criterion (Jones et al. 1993) for building an accurate SVM based high dimensional model representation (HDMR) model. The employed DIRECT idea can select a batch of new points from a set of potential hypercubes via the identified potential Lipschitz constants that consider both local exploitation and global exploration. Note that for the purpose of global metamodeling, now the objective for DIRECT is not the function values but the prediction errors at observed points. The DIRECT-type sampling is inherently parallelizable, but the number of new points in each iteration may be different and thus uncontrollable. Quan (2014) explored several approaches combining the EI idea (Jones et al. 1998) and some space-filling criteria for adaptive batch sampling. Numerical results revealed that there is no evidence that the usage of batch selection provides potential performance improvements, but it is meaningful for improving computational efficiency by taking advantage of parallelization.

In short, for practical implementation, the single selection is usually more informative than the batch selection, since it will update the information after involving a new point. However, if the simulation process is significantly time-consuming, batch selection is a meaningful way to take advantage of parallelization to save computational cost. Besides, it is worth noting that mostly the choice of batch size  $q$  is determined by the available computer resource (e.g., the number of processors) rather than the sampling criterion.

## 6 Multi-response adaptive sampling for global metamodeling

Most of current adaptive sampling approaches are designed for single-response problems. But in practice, a multi-response system  $\mathcal{F} = \{f_1, \dots, f_k\}$  containing  $k$  uncorrelated/correlated output responses is frequently encountered. Some traditional space-filling sampling approaches, e.g., LHD and maximin, can be naturally extended to handle the multi-response system, since they determine the sample distributions without considering the characteristics of the output responses. For adaptive sampling

approaches, however, due to the adaption to the characteristics of a specific output response, they cannot directly handle multi-response systems. Figure 7 illustrates two two-response cases where the two functions are uncorrelated and have multimodal behaviors in different regions in Fig. 7(a), while the two functions are highly correlated with the main characteristics located in the same region in Fig. 7(b). In the following, we review the existing multi-response adaptive sampling approaches in three scenarios.

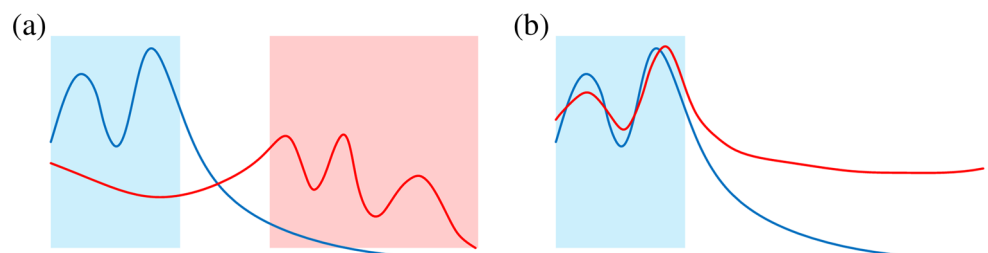
### 6.1 General multi-response adaptive sampling

For the general multi-response adaptive sampling, it usually does not assume that the multiple responses are correlated. That is, it considers all the  $k$  output responses jointly, but ignores their correlations. In this case, we usually model the responses individually.

If we already know the degree of importance of the  $k$  responses, it is then nature to assign different weights to them, which turns the multi-response adaptive sampling into single-response sampling (Lin 2004). However, the importance information in practice is usually unavailable. Hence, we need more general multi-response sampling framework. To this end, Reichart et al. (2008) investigated two sampling strategies for two-response problems. The first strategy, called *alternating selection*, is an intuitive idea that handles the two responses successively: in the  $i$ th iteration, it samples a new point for the first response via a sampling criterion; then in the  $(i + 1)$ th iteration, it samples a new point for the second response. The second strategy, called *rank combination*, evaluates the scores of a candidate point for all the responses, and then sums the scores as the final score of this candidate point. Finally, the candidate point with the highest score is selected as the new point.

The rank combination strategy is commonly used in the general multi-response sampling framework, and can be classified into three types. Type 1 sums or multiplies the scores of all the  $k$  responses at a candidate point (Rosenbaum 2013). In contrast, type 2 takes the maximum/minimum of the scores for the  $k$  responses at a candidate point (Crombecq 2011; Lovison and Rigoni 2011). This type is found to perform well for cases with many responses. For example, suppose that among the many responses, there is a response that has nonlinear behavior everywhere while the remaining responses have linear behavior. When taking the sum/multiple type 1, every region will almost have the

**Fig. 7** **a** Uncorrelated and **b** correlated two-response cases



same score, while type 2 can identify the most nonlinear response. Finally, the recently developed type 3 attempts to combine type 1 and type 2 with the fusion by a weighted-sum term and a maximum term (Liu et al. 2016c). It has been shown that this hybrid type makes the general multi-response adaptive sampling perform well in different scenarios.

Besides, Aute (2009) turned the multi-response sampling process into a multi-objective optimization problem, and proposed several ways to choose new points from the Pareto set. For example, it can choose all the Pareto points as new points; or it can choose the Pareto points that have the highest scores for each of the responses.

It is found that the general multi-response adaptive sampling criteria are usually not the best choice for each response alone but provide a compromise for all the responses. The advantage of the general multi-response adaptive sampling is that they have no assumption about the multi-response problem and thus gain wide applicability. The disadvantages, however, are apparent: (1) the metamodels are built individually for the responses. Hence, the recently developed multi-response modeling approaches cannot be integrated to utilize the response correlations; and (2) furthermore, because of the ignorance of response correlations, the obtained new points need to be simulated for all the responses, which might waste computing resources.

## 6.2 Symmetric multi-response adaptive sampling

Different from the general multi-response adaptive sampling, symmetric multi-response adaptive sampling considers the response correlations, and treats the  $k$  responses equally. This kind of correlated multi-response case is frequently encountered in practice, e.g., the drag response and the lift response simulated simultaneously from an airfoil design problem (Liu et al. 2014). In what follows, we first review current symmetric multi-response modeling approaches, and then discuss the adaptive sampling approaches under such modeling framework.

By utilizing the response correlations, we can model the  $k$  responses jointly, with the aim of improving over modeling them individually. Some commonly used symmetric multi-response modeling approaches include, for example, the artificial neural networks (ANNs) that are extended to support multi-response by defining how inputs are shared across responses (Caruana 1995; Collobert and Weston 2008), the multi-response Gaussian process framework that captures the response correlations through cross-covariance functions built using linear model of coregionalization or convolved process (Bonilla et al. 2007; Alvarez and Lawrence 2009; Alvarez et al. 2012), and the general stacked single-target framework and ensemble of regressor chains (Spyromitros-Xioufis et al. 2016). Among them, the multi-response Gaussian process, also called multi-task Gaussian process (MTGP), has been widely used because it provides not only prediction mean but also prediction variance under the Bayesian framework.

Contrary to the continuous emergence of symmetric multi-response modeling approaches, the development of the associated sampling approaches, especially the adaptive sampling approaches, is lagging behind. To the best of our knowledge, only a few works have focused on the topic of symmetric multi-response adaptive sampling, most of which are studied in the MTGP framework. For a valid symmetric multi-response adaptive sampling approach, it should address two issues: (1) *how* to query the most informative point under the symmetric multi-response modeling framework; and (2) *which* responses should be used to evaluate the obtained new point.

To address the first issue, a common way is to extend current sampling approaches, especially the variance based sampling approaches, under the symmetric multi-response modeling framework. For example, Li et al. (2006) extended the maximum entropy criterion by using the stacking strategy, i.e., treating the output responses as extra inputs. Romero et al. (2006) and Romero et al. (2012) systematically investigated the extensions of current single-response sampling approaches, including maximin-LHD, maximum entropy sampling and maximum cross-validation variance, for multi-response cases. It is found that, however, the above sampling approaches simulate the obtained new points for all the responses, which is not an efficient manner.

To address the second issue, the simplest way is to simulate the new point for the response that assigns the highest score. Along this line, Osborne et al. (2012) extended the maximum mean square error sampling approach for multi-response cases. It is found that compared to the single-response sampling, the multi-response sampling requires less points to achieve the same model accuracy by using the correlations captured through the MTGP framework. Taking the two functions in Fig. 7(b) for example, suppose that we sample a new point in the shadow region and only simulate it by the blue function. After involving the new point, the symmetric multi-response modeling approach, e.g., MTGP, can jointly improve the models for the two functions because they are highly correlated. As a result, the computational budget (number of simulations) has been reduced. Recently, Zhang et al. (2016) presented an entropy based adaptive sampling approach under the sparse convolved MTGP framework (Alvarez and Lawrence 2009), which is developed for handling large datasets. Given the sparse model structure, the proposed adaptive sampling criterion introduces an exploitation-exploration trade-off expression to query the most informative point, and selects the proper response to evaluate it. Numerical experiments revealed that this approach performs well in different scenarios.

## 6.3 Asymmetric multi-response adaptive sampling

“Asymmetric” here means among the  $k$  responses there is a *primary response*. The goal of asymmetric multi-response adaptive sampling is to enhance the modeling performance for the

primary response given the other *secondary responses*. The asymmetric multi-response problem is usually referred as transfer learning (Pan and Yang 2010) in the machine learning community, while it is restricted to multi-fidelity modeling (MFM), variable-fidelity modeling (VFM), or data fusion in the engineering community (Fernández-Godino et al. 2016; Peherstorfer et al. 2016). Here we focus on multi-fidelity cases where the basic assumption is that for a target function there exist correlated simulation codes with different levels of fidelity. For the widely studied two-level fidelity cases, the high-fidelity simulation provides accurate predictions but requires huge computational budget, while the low-fidelity simulation is cheap to run but provides coarse predictions. In this context, the goal of MFM is to have computational gains by transferring knowledge from the low-fidelity simulation to enhance the modeling results of high-fidelity simulation. The MFM frameworks, e.g., scaling function based modeling (Han et al. 2013; Zhou et al. 2017) and Bayesian multi-fidelity modeling (also called Co-Kriging) (Kennedy and O'Hagan 2000, 2001; Forrester et al. 2007; Qian and Wu 2008), have gained popularity in multidisciplinary design, optimization and uncertainty quantification. Some reviews and comparison studies regarding MFM can be found in (Toal 2015; Fernández-Godino et al. 2016; Peherstorfer et al. 2016; Park et al. 2017).

Similar to the symmetric multi-response case, the MFM approaches have been extensively studied while the associated sampling approaches, especially the adaptive sampling, have rarely been studied. For constructing accurate global multi-fidelity models, a sampling approach should also answer the “how” and “which” questions. Here, the “which” question is significantly important because the high-fidelity simulation is more time-consuming than the low-fidelity simulation.

Most of current MFM processes employ the space-filling sampling approaches, e.g., optimal LHD (Jin et al. 2005), to generate high/low-fidelity points. Particularly, in the Co-Kriging modeling process, it requires that the high-fidelity points should be a subset of the low-fidelity points. This requirement allows to estimate the high/low-fidelity model hyperparameters individually (Kennedy and O'Hagan 2000). Hence, the *nested sampling* (Qian 2009; Haaland and Qian 2010; Rennen et al. 2010) reviewed in Section 2 and the *nearest neighbor sampling* (Le Gratiet and Garnier 2014) have been widely used. But these space-filling sampling approaches usually run in a one-shot fashion. Besides, it is worth noting that if we estimate the hyperparameters jointly in Co-Kriging like (Han et al. 2010b), the nesting requirement can be relaxed, thus allowing easier and more widespread implementation of current sampling approaches.

Furthermore, several studies investigated the adaptive sampling under the MFM framework. Qian et al. (2014) proposed an asymmetric nested lattice sampling approach. This approach attempts to refine the samples along some inputs that have more significant effects on the responses. But it does not

consider the information of low/high-fidelity responses. Han et al. (2010a) presented an adaptive sampling strategy based on the assumption that with the adding of new high-fidelity points, the multi-fidelity model and the high-fidelity model will converge to the same results. Hence, the point at which the two models yield the maximum discrepancy is selected as the new high-fidelity point. This pure error based sampling strategy, however, may lead to excessive local exploitation. Instead, except minimizing the GMSE value of the multi-fidelity model, Zhou et al. (2016) employed an extra distance penalty term to drive the new high-fidelity point far away from the existing points. The adaptive sampling approach developed in (Benamara et al. 2016) also relies on an error measurement derived under the Gappy-POD based MFM. Note that this approach evaluates the new point through both the low- and high-fidelity simulations.

It is found that most of current asymmetric multi-response adaptive sampling approaches obtain the new point as high-fidelity point, because they follow an underlying assumption that the low fidelity model has been well fitted. In practice, however, (1) the assumption may not be satisfied, especially in high dimensions; and (2) in terms of saving computing cost, we are interested in the case that adding a cheap low-fidelity point can also improve the multi-fidelity model. Along this line, Le Gratiet and Cannamela (2015) proposed an adaptive sampling approach for Co-Kriging. To address the “how” issue, this new approach utilizes the leave-one-out cross-validation error to adjust the original prediction variance so that it can sample more points in interesting regions. To address the “which” issue, the Co-Kriging process is organized in a recursive formulation (Le Gratiet and Garnier 2014) which offers the contribution of each level of simulation to the final model. As a result, the adaptive sampling criterion can take into account the computing ratios between different levels of simulation.

So far, we have provided a review of the state-of-the-art multi-response adaptive sampling approaches in three scenarios. In practical implementation, for uncorrelated multiple responses like Fig. 7(a), the general multi-response adaptive sampling is recommended; for correlated responses, or furthermore, for responses with obvious hierarchy of dependence, the symmetric and asymmetric multi-response adaptive sampling approaches become a good choice. Defining adaptive sampling strategies in multi-response modeling framework is of interest and is still an open problem that needs more research efforts.

## 7 Brief remarks on adaptive sampling for other purposes

As has been depicted in Fig. 1, adaptive sampling can be further applied to other applications, e.g., metamodel-based



global optimization and sensitivity analysis. To showcase the potential of adaptive sampling, this section gives some remarks on adaptive sampling for metamodel-based global optimization and sensitivity analysis.

Metamodel-based global optimization (MBGO) uses metamodels to facilitate the optimization, especially for simulation-based problems. Figure 8 illustrates two commonly used MBGO strategies from (Wang and Shan 2007): sequential strategy and adaptive strategy. For the sequential strategy, in order to sequentially refine the metamodels for optimization, the space-filling and the adaptive sampling approaches reviewed before can be used to gather informative points in the design space. Then the new points together with the optimal points are used to improve the metamodels. This strategy is simple but not efficient since the new points are generated for improving global metamodeling instead of optimization.

The adaptive strategy now is widely used for MBGO. In this strategy, the DoE techniques are used to build initial metamodels. The metamodels then are used to guide the optimization sampling process towards global optimum. In some literatures, such metamodel-based optimization sampling process is also denoted as adaptive sampling. This type of adaptive sampling for global optimization also contains local exploitation and global exploration. But different from the adaptive sampling for global metamodeling, here the local exploitation attempts to sample more points in regions that may contain global optimum, while the global exploration attempts to find undetected interesting regions.

Table 4 provides a partial list of the basic adaptive sampling ideas for MBGO, and classifies them into five categories and describes their properties. All these criteria consist of local exploitation, global exploration and trade-off between them. Besides, it is found that the variance based criteria are model-dependent since they utilize the inner properties of a statistical model; the local model based criteria employ local metamodels to facilitate the local exploitation.

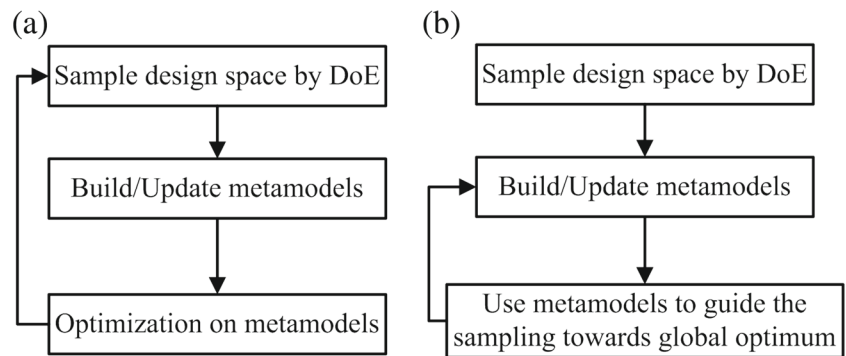
These basic adaptive sampling strategies in Table 4 can be further extended to handle various types of optimization problems, e.g., constrained optimization, parallel optimization, multi-objective optimization, multi-disciplinary optimization, and robust optimization. An exhaustive review of metamodel-based optimization is beyond the scope of this article. For more information one can refer to recent review studies (Queipo et al. 2005; Forrester and Keane 2009; Shan and Wang 2010b; Parr et al. 2010; Jin 2011; Martins and Lambe 2013; Haftka et al. 2016) and the references therein.

Sensitivity analysis (SA) measures the contribution of an individual input to the total uncertainty of the output response. The SA is a very broad topic and has been widely reviewed in different fields, see (Kleijnen 2005; Saltelli et al. 2008; Tian 2013; Janouchová and Kučerová 2013; Iooss and Lemaître 2015; Borgonovo and Plischke 2016). Considering the studied

expensive simulation-based scenarios where the sample size in reasonable time is rather limited, we focus on sampling-based global sensitivity analysis (GSA) (Janouchová and Kučerová 2013) and metamodel-based global sensitivity analysis (Tian 2013). For these two types of GSA, the role of sampling remains the same as that for global metamodeling, i.e., aiming to gather informative points over the domain in order to build accurate global metamodels (if required) for the subsequent GSA process.

The sampling-based GSA first samples a set of points in the domain and then calculates the sensitivity indicators based on the sample set. One of the key issues in sampling-based GSA is to gather informative points in the design space, which is the topic of this article, for the subsequent calculation of sensitivity indicators, e.g., Spearman's rank correlation coefficient (SRCC) and Sobol's indices (Sobol 1993). Again note that here we focus on the sampling approaches rather than the sensitivity indicators for the sampling-based and metamodel-based GSA. Since the sampling-based GSA does not utilize the information from metamodels, the conventional space-filling sampling approaches are suitable here. The LHD gains popularity for GSA because it provides a possibility to represent prescribed probabilistic distribution of particular variables (Helton et al. 2006). Helton et al. (2005) compared the random sampling and LHD sampling for the SA on results from a model of two-phase fluid flow, and showed a preference for LHD. Furthermore, Janouchová and Kučerová (2013) applied optimal LHDs under eight criteria to GSA, and pointed that the modified  $L_2$ -discrepancy criterion provides the best results in sensitivity predictions.

Compared to the sampling-based GSA, the recently developed metamodel-based GSA has an intermediate step wherein a metamodel is built using the sample set. Then the subsequent calculation of sensitivity indicators can be massively conducted on the metamodel for saving computational cost. Hence, both the conventional space-filling and the adaptive sampling approaches can be employed here. Currently, the LHD approach (Kleijnen 2005) and the MMSE sampling approach (if the metamodel is selected as a statistical model, e.g., GP) (Le Gratiet et al. 2017) are widely used. The only work using adaptive sampling for metamodel-based GSA is presented in (Guenther et al. 2015). This work extended the MMSE sampling approach under the treed GP modeling framework (Gramacy and Lee 2008) for adaptively selecting a batch of  $q$  points. For a set of candidate points, it first selects the one with the largest MSE, then it selects a subset of candidate points by a modified distance criterion that considers the total sensitivity information. Finally, the remaining  $q-1$  points are randomly selected from the subset via their MSE values. The adaptive sampling approaches reviewed in this article, however, have not yet been applied to metamodel-based GSA. It is meaningful to investigate the performance of adaptive sampling for metamodel-based GSA.

**Fig. 8** a Sequential and b adaptive MBGO strategies

## 8 Challenges and future researches

Though many efficient adaptive sampling approaches and successful real-world applications are constantly emerging to date, from the present review, several notable challenges still remain.

### 8.1 High-dimensional adaptive sampling

It is found that for global metamodeling, the design space, the computational demand, and the required number of points grow exponentially with dimensions, known as the “curse of

dimensionality”. In practice, the high-dimensional problems are frequently encountered, and thus there have some high-dimensional modeling works (Shan and Wang 2010a, b, 2011; Li et al. 2012; Cai et al. 2016; Liu et al. 2017a). However, there is little work in literature regarding the topic of high-dimensional adaptive sampling. Kupresanin and Johannesson (2011) studied and compared several sequential sampling approaches for the GP model up to thirty-two dimensions. The results showed that there is no clear winner among those sampling approaches for global metamodeling. For the data from a stationary Gaussian process, the MMSE approach (Jin et al. 2002) seems to yield more accurate

**Table 4** A partial list of the basic adaptive sampling ideas for MBGO

Type	Criterion	Property
Variance based	Expected improvement (Jones et al. 1998) Upper confidence bound (Srinivas et al. 2010) Entropy search (Villemonteix et al. 2009) Predictive entropy search (Hernández-Lobato et al. 2014)	Model-dependent; global model; minimize prediction response to do local exploitation while maximize prediction variance to do global exploration; trade-off between global & local is automatically determined
Lipschitz based	Extended DIRECT (Liu et al. 2015b; Liu et al. 2017b)	Model-independent; global model; identify a set of potential Lipschitz constants with small values corresponding to local exploitation while large values corresponding to global exploration; trade-off between global & local is automatically determined
Distance based	Constrained optimization using response surfaces (Regis and Shoemaker 2005) RBF for global optimization (Gutmann 2001) Stochastic response surface (Regis and Shoemaker 2007)	Model-independent; global model; minimize prediction response to do local exploitation while maximize a distance criterion to do global exploration; trade-off between global & local is achieved through a balance pattern
Probability based	Mode-pursuing sampling (Wang et al. 2004)	Model-independent; global model; use a probability function to take into account local exploitation and global exploration; trade-off between global & local is randomly determined
Local model based	Trust-region (Alexandrov et al. 1998) Space exploration and unimodal region elimination (Younis and Dong 2010)	Model-independent; local model; visit local region to do local exploration while move or expand local region to do global exploration

predictions, while for the data from a non-stationary Gaussian process, the EI based adaptive sampling approach (Lam 2008) stands out.

Conducting adaptive sampling in high dimensions is actually an intractable task:

- Most of current adaptive sampling approaches use an adaptive sampling criterion to obtain the new points by solving an auxiliary optimization problem with a sophisticated surface. Due to the “curse of dimensionality”, solving such a high-dimensional optimization problem becomes very difficult and computationally expensive; and
- Also due to the “curse of dimensionality”, the number of points required for constructing a reliable metamodel becomes massive. The sample size, however, is usually limited for practical simulation-based engineering problems. Consequently, poor metamodels built with limited sample data in high dimensions may erroneously guide the adaptive sampling for global metamodeling.

To mitigate the effects of “curse of dimensionality”, one may use some efficient high-dimensional optimization algorithms (Gao and Wang 2007; Yang et al. 2007; Wang et al. 2011) to accelerate the adaptive sampling process. Besides, some effective strategies tackling high-dimensionality, for example, decomposing the design problems into sub-problems, identifying important variables, mapping, and reducing design space that have been extensively reviewed in (Shan and Wang 2010b), can be adopted to ease the challenges of high-dimensional adaptive sampling.

## 8.2 Incorporating more information

The reviewed adaptive sampling approaches typically identify new points and obtain their responses via a mathematical model, e.g., FEA and CFD. That is, they consider only a single data source from computer simulations. Intuitively, more information brings about a more informed selection of the new points, and thereafter more accurate metamodels. The multi-response adaptive sampling reviewed in Section 6 can be regarded as a case of incorporating more information from correlated responses. More broadly, in practice for a specific problem, there are multiple data sources with varying fidelity and property, e.g., the physical experiment results, low-/high-fidelity simulation results, and even user’s a priori knowledge. Completely and effectively utilizing these different kinds of data is expected to improve the sampling performance.

The first attempt to integrate different information together into sequential sampling is made by the Q2S2 sampling approach (Rai and Campbell 2008). This approach divides the data into quantitative (QT) information like the physical results and simulation results, and qualitative (QL) information like user guesses and sensitivity information. Then a confidence function

is used to merge the QT and QL information together into the sequential sampling process. Three examples have showed that Q2S2 performs well and may have distinct advantages in high dimensions. Recently, Huang et al. (2016) proposed an optimal clustered-sliced LHD sampling approach for problems with both QT and QL variables. This approach is designed under a GP modeling framework modified for both QT and QL variables (Qian et al. 2008).

To the best of our knowledge, we have not found any work incorporating QT and QL information in the context of adaptive sampling. In order to enlarge the application area, more research efforts are needed for adaptive sampling that considers multiple information sources.

## 9 Conclusions

To mitigate the issue of high computing requirements, there have been increasing research interests and studies on adaptive sampling approaches for building accurate yet efficient global metamodels, which is particularly prominent in the field of simulation-based complex engineering design. To the best of our knowledge, however, there has been a lack of reviews on adaptive sampling for global metamodeling to date. This article has thus attempted to categorize, review and analyze the current research and development in adaptive sampling so as to offer a comprehensive review and also a reference for practitioners working in the field.

The reviews and discussions in this article do not attempt to give definite answers to what the best adaptive sampling strategy is for global metamodeling. This is because all the reviewed adaptive sampling strategies represent the possible ways to select points that provide the information about the target function as much as possible, and no strategy always outperforms the others. But from our numerical experience, some recommendations can be given: if a specific metamodel type is used, e.g. the GP model, the variance based adaptive sampling will be a good choice; otherwise, the model-independent adaptive sampling is preferred.

By conducting a comprehensive review of the diverse single-/multi-response adaptive sampling approaches, we hope that this review has highlighted the need for research efforts on adaptive sampling, and identified new important research directions such as how to address the increasing complexity of simulation-based engineering design which involves growing ‘Big’ dimensionality and information.

**Acknowledgements** This work was conducted within the Rolls-Royce@NTU Corporate Lab with support from the National Research Foundation (NRF) Singapore under the Corp Lab@University Scheme. It is also partially supported by the Data Science and Artificial Intelligence Research Center (DSAIR) and the School of Computer Science and Engineering at Nanyang Technological University.

## References

- Acar E (2014) Simultaneous optimization of shape parameters and weight factors in ensemble of radial basis functions. *Struct Multidiscip Optim* 49(6):969–978
- Acar E, Rais-Rohani M (2009) Ensemble of metamodels with optimized weight factors. *Struct Multidiscip Optim* 37(3):279–294
- Ajdari A, Mahlooji H (2014) An adaptive exploration-exploitation algorithm for constructing metamodels in random simulation using a novel sequential experimental design. *Commun Stat Simul Comput* 43(5):947–968
- Alexandrov NM, Dennis JE Jr, Lewis RM, Torczon V (1998) A trust-region framework for managing the use of approximation models in optimization. *Struct Optimization* 15(1):16–23
- Alvarez M, Lawrence ND (2009) Sparse convolved Gaussian processes for multi-output regression. In: *Advances in Neural Information Processing Systems*, pp 57–64
- Alvarez MA, Rosasco L, Lawrence ND (2012) Kernels for vector-valued functions: a review. *Found Trends Mach Learn* 4(3):195–266
- Atamturktur S, Williams B, Egeberg M, Unal C (2013) Batch sequential design of optimal experiments for improved predictive maturity in physics-based modeling. *Struct Multidiscip Optim* 48(3):549–569
- Atkinson EJ, Therneau TM (2000) An introduction to recursive partitioning using the RPART routines. Mayo Foundation, Rochester
- Audze P, Eglais V (1977) New approach for planning out of experiments. *Problems of Dynamics and Strengths* 35:104–107
- Auffray Y, Barbillon P, Marin J-M (2012) Maximin design on non hypercube domains and kernel interpolation. *Stat Comput* 22(3):703–712
- Aurenhammer F (1991) Voronoi diagrams—a survey of a fundamental geometric data structure. *ACM Comput Surv* 23(3):345–405
- Aute VC (2009) Single and multiresponse adaptive design of experiments with application to design optimization of novel heat exchangers. University of Maryland, College Park, City of College Park
- Aute V, Saleh K, Abdelaziz O, Azarm S, Rademacher R (2013) Cross-validation based single response adaptive design of experiments for kriging metamodeling of deterministic computer simulations. *Struct Multidiscip Optim* 48(3):581–605
- Barton RR (1997) Design of experiments for fitting subsystem metamodels. In: *Proceedings of the 29th conference on Winter simulation*, Atlanta, Georgia, USA. IEEE, pp 303–310
- Beck J, Guillas S (2016) Sequential design with mutual information for computer experiments (MICE): emulation of a tsunami model. *SIAM/ASA J Uncertain Quantif* 4(1):739–766
- Benamara T, Breikopf P, Lepot I, Sainvitu C (2016) Adaptive infill sampling criterion for multi-fidelity optimization based on Gappy-POD. *Struct Multidiscip Optim* 54(4):843–855
- Bonilla EV, Chai KMA, Williams CK (2007) Multi-task Gaussian process prediction. In: *NIPS*, pp 153–160
- Borgonovo E, Plischke E (2016) Sensitivity analysis: a review of recent advances. *Eur J Oper Res* 248(3):869–887
- Braconnier T, Ferrier M, Jouhaud J-C, Montagnac M, Sagaut P (2011) Towards an adaptive POD/SVD surrogate model for aeronautic design. *Comput Fluids* 40(1):195–209
- Burbidge R, Rowland JJ, King RD (2007) Active learning for regression based on query by committee. In: *International Conference on Intelligent Data Engineering and Automated Learning*. Springer, pp 209–218
- Busby D (2009) Hierarchical adaptive experimental design for Gaussian process emulators. *Reliab Eng Syst Saf* 94(7):1183–1193
- Busby D, Farmer CL, Iske A (2007) Hierarchical nonlinear approximation for experimental design and statistical data fitting. *SIAM J Sci Comput* 29(1):49–69
- Cai X, Qiu H, Gao L, Yang P, Shao X (2016) An enhanced RBF-HDMR integrated with an adaptive sampling method for approximating high dimensional problems in engineering design. *Struct Multidiscip Optim* 53(6):1209–1229
- Caruana R (1995) Learning many related tasks at the same time with backpropagation. In: *Advances in Neural Information Processing Systems*. pp 657–664
- Černý V (1985) Thermodynamical approach to the traveling salesman problem: an efficient simulation algorithm. *J Optim Theory Appl* 45(1):41–51
- Chen R-B, Hsu Y-W, Hung Y, Wang W (2014) Discrete particle swarm optimization for constructing uniform design on irregular regions. *Comput Stat Data Anal* 72:282–297
- Chuang S, Hung Y (2010) Uniform design over general input domains with applications to target region estimation in computer experiments. *Comput Stat Data Anal* 54(1):219–232
- Clarke SM, Griebisch JH, Simpson TW (2005) Analysis of support vector regression for approximation of complex engineering analyses. *J Mech Des* 127(6):1077–1087
- Collobert R, Weston J (2008) A unified architecture for natural language processing: deep neural networks with multitask learning. In: *Proceedings of the 25th International conference on machine learning*. ACM, p 160–167
- Cressie N (1988) Spatial prediction and ordinary kriging. *Math Geol* 20(4):405–421
- Crombecq K (2011) Surrogate modeling of computer experiments with sequential experimental design. Ghent University, Antwerpen
- Crombecq K, Gorissen D, Deschrijver D, Dhaene T (2011a) A novel hybrid sequential design strategy for global surrogate modeling of computer experiments. *SIAM J Sci Comput* 33(4):1948–1974
- Crombecq K, Laermans E, Dhaene T (2011b) Efficient space-filling and non-collapsing sequential design strategies for simulation-based modeling. *Eur J Oper Res* 214(3):683–696
- Damblin G, Couplet M, Iooss B (2013) Numerical studies of space-filling designs: optimization of Latin hypercube samples and subprojection properties. *J Simul* 7(4):276–289
- De Geest J, Dhaene T, Faché N, De Zutter D (1999) Adaptive CAD-model building algorithm for general planar microwave structures. *IEEE Trans Microwave Theory Tech* 47(9):1801–1809
- de Oliveira Castro P, Petit E, Beyler JC, Jalby W (2012) ASK: adaptive sampling kit for performance characterization. In: *European conference on parallel processing*, Rhodes Island, Greece. Springer, pp 89–101
- Deschrijver D, Crombecq K, Nguyen HM, Dhaene T (2011) Adaptive sampling algorithm for macromodeling of parameterized-parameter responses. *IEEE Trans Microwave Theory Tech* 59(1):39–45
- Devabhaktuni VK, Zhang Q-J (2000) Neural network training-driven adaptive sampling algorithm for microwave modeling. In: *2000 30th European microwave conference*, Paris, France. IEEE, pp 1–4
- dos Santos MIR, dos Santos PMR (2008) Sequential experimental designs for nonlinear regression metamodels in simulation. *Simul Model Pract Theory* 16(9):1365–1378
- Douak F, Melgani F, Alajlan N, Pasolli E, Bazi Y, Benoudjit N (2012) Active learning for spectroscopic data regression. *J Chemom* 26(7): 374–383
- Draguljić D, Santner TJ, Dean AM (2012) Noncollapsing space-filling designs for bounded nonrectangular regions. *Technometrics* 54(2): 169–178
- Dyn N, Levin D, Rippa S (1986) Numerical procedures for surface fitting of scattered data by radial functions. *SIAM J Sci Stat Comput* 7(2): 639–659
- Eason J, Cremaschi S (2014) Adaptive sequential sampling for surrogate model generation with artificial neural networks. *Comput Chem Eng* 68:220–232
- Efron B, Tibshirani RJ (1994) An introduction to the bootstrap. CRC press, Florida



- Fang K-T, Lin DK, Winker P, Zhang Y (2000) Uniform design: theory and application. *Technometrics* 42(3):237–248
- Farhang-Mehr A, Azarm S (2005) Bayesian meta-modelling of engineering design simulations: a sequential approach with adaptation to irregularities in the response behaviour. *Int J Numer Methods Eng* 62(15):2104–2126
- Fedorov VV (1972) *Theory of optimal experiments*. Academic Press, New York
- Fernández-Godino MG, Park C, Kim N-H, Haftka RT (2016) Review of multi-fidelity models. *arXiv preprint arXiv:160907196*
- Finkel RA, Bentley JL (1974) Quad trees a data structure for retrieval on composite keys. *Acta Informatica* 4(1):1–9
- Forrester AI, Keane AJ (2009) Recent advances in surrogate-based optimization. *Prog Aerosp Sci* 45(1):50–79
- Forrester AI, Sobester A, Keane AJ (2007) Multi-fidelity optimization via surrogate modelling. *Proc R Soc Lond A Math Phys Eng Sci* 463(2088):3251–3269
- Freund Y, Seung HS, Shamir E, Tishby N (1993) Information, prediction, and query by committee. In: *Advances in Neural Information Processing Systems*, pp 483–483
- Fu Y, Zhu X, Li B (2013) A survey on instance selection for active learning. *Knowl Inf Syst* 35(2):249–283
- Gao Y, Wang Y-J (2007) A memetic differential evolutionary algorithm for high dimensional functions' optimization. In: *Third International Conference on Natural Computation*. IEEE, pp 188–192
- Gazut S, Martinez J-M, Dreyfus G, Oussar Y (2008) Towards the optimal design of numerical experiments. *IEEE Trans Neural Netw* 19(5):874–882
- Geman S, Bienenstock E, Doursat R (1992) Neural networks and the bias/variance dilemma. *Neural Comput* 4(1):1–58
- Ghoreishi M, Badcock K, Woodgate M (2009) Accelerating the numerical generation of aerodynamic models for flight simulation. *J Aircr* 46(3):972–980
- Golzari A, Sefat MH, Jamshidi S (2015) Development of an adaptive surrogate model for production optimization. *J Pet Sci Eng* 133:677–688
- Gorissen D, Couckuyt I, Demeester P, Dhaene T, Crombecq K (2010) A surrogate modeling and adaptive sampling toolbox for computer based design. *J Mach Learn Res* 11:2051–2055
- Gramacy RB, Lee HK (2006) Adaptive design of supercomputer experiments. The Statistical Laboratory, University of Cambridge, UK
- Gramacy RB, Lee HKH (2008) Bayesian treed Gaussian process models with an application to computer modeling. *J Am Stat Assoc* 103(483):1119–1130
- Gramacy RB, Lee HK (2009) Adaptive design and analysis of supercomputer experiments. *Technometrics* 51(2):130–145
- Grosso A, Jamali A, Locatelli M (2009) Finding maximin latin hypercube designs by iterated local search heuristics. *Eur J Oper Res* 197(2):541–547
- Guenther J, Lee HK, Gray GA (2015) Sequential design for achieving estimated accuracy of global sensitivities. *Appl Stoch Model Bus Ind* 31(6):782–800
- Gutmann H-M (2001) A radial basis function method for global optimization. *J Glob Optim* 19(3):201–227
- Haaland B, Qian PZ (2010) An approach to constructing nested space-filling designs for multi-fidelity computer experiments. *Stat Sin* 20(3):1063
- Haftka RT, Villanueva D, Chaudhuri A (2016) Parallel surrogate-assisted global optimization with expensive functions—a survey. *Struct Multidiscip Optim* 54(1):3–13
- Halton JH (1960) On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numer Math* 2(1):84–90
- Han Z-H, Görtz S, Hain R (2010a) A variable-fidelity modeling method for aero-loads prediction. In: *New results in numerical and experimental fluid mechanics VII*. Springer, pp 17–25
- Han Z-H, Zimmermann R, Goretz S (2010b) A new cokriging method for variable-fidelity surrogate modeling of aerodynamic data. In: *48th AIAA Aerospace sciences meeting including the new horizons forum and Aerospace exposition*, p 1225
- Han Z-H, Görtz S, Zimmermann R (2013) Improving variable-fidelity surrogate modeling via gradient-enhanced kriging and a generalized hybrid bridge function. *Aerosp Sci Technol* 25(1):177–189
- Helton JC, Davis F, Johnson JD (2005) A comparison of uncertainty and sensitivity analysis results obtained with random and Latin hypercube sampling. *Reliab Eng Syst Saf* 89(3):305–330
- Helton JC, Johnson JD, Oberkampf W, Sallaberry CJ (2006) Sensitivity analysis in conjunction with evidence theory representations of epistemic uncertainty. *Reliab Eng Syst Saf* 91(10):1414–1434
- Hendrickx W, Dhaene T (2005) Sequential design and rational metamodeling. In: *Proceedings of the 37th conference on Winter simulation*, Orlando, Florida. ACM, pp 290–298
- Hernández-Lobato JM, Hoffman MW, Ghahramani Z (2014) Predictive entropy search for efficient global optimization of black-box functions. In: *Advances in Neural Information Processing Systems*. pp 918–926
- Huang Z, Qiu H, Zhao M, Cai X, Gao L (2015) An adaptive SVR-HDMR model for approximating high dimensional problems. *Eng Comput* 32(3):643–667
- Huang H, Lin DK, Liu M-Q, Yang J-F (2016) Computer experiments with both qualitative and quantitative variables. *Technometrics* 58(4):495–507
- Husslage B, Van Dam E, Den Hertog D (2005) Nested maximin Latin hypercube designs in two dimensions. *CentER discussion paper no.* 2005-79
- Iooss B, Lemaître P (2015) A review on global sensitivity analysis methods. In: *Uncertainty management in simulation-optimization of complex systems*. Springer, pp 101–122
- Janouchová E, Kučerová A (2013) Competitive comparison of optimal designs of experiments for sampling-based sensitivity analysis. *Comput Struct* 124:47–60
- Jiang P, Shu L, Zhou Q, Zhou H, Shao X, Xu J (2015) A novel sequential exploration-exploitation sampling strategy for global metamodeling. *IFAC-PapersOnLine* 48(28):532–537
- Jin Y (2011) Surrogate-assisted evolutionary computation: recent advances and future challenges. *Swarm Evol Comput* 1(2):61–70
- Jin R, Chen W, Simpson TW (2001) Comparative studies of metamodeling techniques under multiple modelling criteria. *Struct Multidiscip Optim* 23(1):1–13
- Jin R, Chen W, Sudjianto A (2002) On sequential sampling for global metamodeling in engineering design. In: *ASME 2002 International design engineering technical Conferences and Computers and information in engineering conference*, Montreal, Canada. ASME, pp 539–548
- Jin R, Chen W, Sudjianto A (2005) An efficient algorithm for constructing optimal design of computer experiments. *J Stat Plan Inference* 134(1):268–287
- Jin Y, Li J, Du W, Qian F (2016) Adaptive sampling for surrogate modeling with artificial neural network and its application in an industrial cracking furnace. *Can J Chem Eng* 94(2):262–272
- Johnson ME, Moore LM, Ylvisaker D (1990) Minimax and maximin distance designs. *J Stat Plan Inference* 26(2):131–148
- Jones DR, Perttunen CD, Stuckman BE (1993) Lipschitzian optimization without the Lipschitz constant. *J Optim Theory Appl* 79(1):157–181
- Jones DR, Schonlau M, Welch WJ (1998) Efficient global optimization of expensive black-box functions. *J Glob Optim* 13(4):455–492
- Joseph VR, Hung Y (2008) Orthogonal-maximin Latin hypercube designs. *Stat Sin* 18(1):171–186
- Kalagnanam JR, Diwekar UM (1997) An efficient sampling technique for off-line quality control. *Technometrics* 39(3):308–319

- Kennedy MC, O'Hagan A (2000) Predicting the output from a complex computer code when fast approximations are available. *Biometrika* 87(1):1–13
- Kennedy MC, O'Hagan A (2001) Bayesian calibration of computer models. *J R Stat Soc Ser B Stat Methodol* 63(3):425–464
- Kenny QY, Li W, Sudjianto A (2000) Algorithmic construction of optimal symmetric Latin hypercube designs. *J Stat Plan Inference* 90(1):145–159
- Kim B, Lee Y, Choi D-H (2009) Construction of the radial basis function based on a sequential sampling approach using cross-validation. *J Mech Sci Technol* 23(12):3357–3365
- Kitayama S, Arakawa M, Yamazaki K (2011) Sequential approximate optimization using radial basis function network for engineering optimization. *Optim Eng* 12(4):535–557
- Kleijnen JP (2005) An overview of the design and analysis of simulation experiments for sensitivity analysis. *Eur J Oper Res* 164(2):287–300
- Kleijnen JP (2008) Design and analysis of simulation experiments. Springer, New York
- Kleijnen JP (2009) Kriging metamodeling in simulation: a review. *Eur J Oper Res* 192(3):707–716
- Kleijnen JP (2015) Kriging metamodels and their designs. In: Design and analysis of simulation experiments. Springer, pp 179–239
- Kleijnen JP, Van Beers WC (2004) Application-driven sequential designs for simulation experiments: kriging metamodeling. *J Oper Res Soc* 55(8):876–883
- Krogh A, Vedelsby J (1995) Neural network ensembles, cross validation, and active learning. In: Advances in Neural Information Processing Systems, pp 231–238
- Kupresanin A, Johannesson G (2011) Comparison of sequential designs of computer experiments in high dimensions. Technical report LLNL-TR-491692, Lawrence Livermore National Laboratory (LLNL), Livermore
- Lam CQ (2008) Sequential adaptive designs in computer experiments for response surface model fit. Ph.D. thesis. The Ohio State University, Columbus
- Le Gratiet L, Cannamela C (2015) Cokriging-based sequential design strategies using fast cross-validation techniques for multi-fidelity computer codes. *Technometrics* 57(3):418–427
- Le Gratiet L, Garnier J (2014) Recursive co-kriging model for design of computer experiments with multiple levels of fidelity. *Int J Uncertain Quantif* 4(5):365–386
- Le Gratiet L, Marelli S, Sudret B (2017) Metamodel-based sensitivity analysis: polynomial chaos expansions and Gaussian processes. In: Ghanem R, Higdon D, Owahdi H (eds) Handbook of uncertainty quantification. Springer International Publishing, Cham, pp 1289–1325. doi:10.1007/978-3-319-12385-1\_38
- Li G, Azarm S, Farhang-Mehr A, Diaz A (2006) Approximation of multiresponse deterministic engineering simulations: a dependent metamodeling approach. *Struct Multidiscip Optim* 31(4):260–269
- Li B, Peng L, Ramadass B (2009) Accurate and efficient processor performance prediction via regression tree based modeling. *J Syst Archit* 55(10):457–467
- Li G, Aute V, Azarm S (2010) An accumulative error based adaptive design of experiments for offline metamodeling. *Struct Multidiscip Optim* 40(1):137–155
- Li E, Wang H, Li G (2012) High dimensional model representation (HDMR) coupled intelligent sampling strategy for nonlinear problems. *Comput Phys Commun* 183(9):1947–1955
- Liefvendahl M, Stocki R (2006) A study on algorithms for optimization of Latin hypercubes. *J Stat Plan Inference* 136(9):3231–3247
- Lin Y (2004) An efficient robust concept exploration method and sequential exploratory experimental design. Georgia Institute of Technology, Atlanta
- Lin Y, Mistree F, Allen JK, Tsui K-L, Chen VC (2004) A sequential exploratory experimental design method: development of appropriate empirical models in design. In: ASME 2004 International design engineering technical Conferences and Computers and information in engineering conference, Salt Lake City, Utah, USA. ASME, pp 1021–1035
- Lin CD, Mukerjee R, Tang B (2009) Construction of orthogonal and nearly orthogonal Latin hypercubes. *Biometrika* 96(1):243–247
- Liu X, Zhu Q, Lu H (2014) Modeling multiresponse surfaces for airfoil design with multiple-output-Gaussian-process regression. *J Aircr* 51(3):740–747
- Liu H, Xu S, Wang X (2015a) Sequential sampling designs based on space reduction. *Eng Optim* 47(7):867–884
- Liu H, Xu S, Wang X, Wu J, Song Y (2015b) A global optimization algorithm for simulation-based problems via the extended DIRECT scheme. *Eng Optim* 47(11):1441–1458
- Liu H, Xu S, Ma Y, Chen X, Wang X (2016a) An adaptive Bayesian sequential sampling approach for global metamodeling. *J Mech Des* 138(1):011404
- Liu H, Xu S, Wang X, Meng J, Yang S (2016b) Optimal weighted pointwise ensemble of radial basis functions with different basis functions. *AIAA J* 3117–3133
- Liu H, Xu S, Wang X, Yang S, Meng J (2016c) A multi-response adaptive sampling approach for global metamodeling. *Proc Inst Mech Eng Pt C J Mechan Eng Sci*. doi:10.1177/0954406216672250
- Liu H, Wang X, Xu S (2017a) Generalized radial basis function-based high-dimensional model representation handling existing random data. *J Mech Des* 139(1):011404
- Liu H, Xu S, Chen X, Wang X, Ma Q (2017b) Constrained optimization via a DIRECT-type constraint-handling technique and an adaptive metamodeling strategy. *Struct Multidiscip Optim* 55(1):155–177
- Loeppky JL, Sacks J, Welch WJ (2009) Choosing the sample size of a computer experiment: a practical guide. *Technometrics* 51(4):366–376
- Loeppky JL, Moore LM, Williams BJ (2010) Batch sequential designs for computer experiments. *J Stat Plan Inference* 140(6):1452–1464
- Loeppky JL, Moore LM, Williams BJ (2012) Projection array based designs for computer experiments. *J Stat Plan Inference* 142(6):1493–1505
- Lovison A, Rigoni E (2011) Adaptive sampling with a Lipschitz criterion for accurate metamodeling. *Commun Appl Ind Math* 1(2):110–126
- Mackman T, Allen C (2010a) Aerodynamic data modelling using multi-criteria adaptive sampling. In: 13th AIAA/ISSMO multidisciplinary analysis optimization conference, Ft. worth. AIAA, pp AIAA 2010-9194
- Mackman T, Allen C (2010b) Investigation of an adaptive sampling method for data interpolation using radial basis functions. *Int J Numer Methods Eng* 83(7):915–938
- Mackman T, Allen C, Ghoreyshi M, Badcock K (2013) Comparison of adaptive sampling methods for generation of surrogate aerodynamic models. *AIAA J* 51(4):797–808
- Martins JR, Lambe AB (2013) Multidisciplinary design optimization: a survey of architectures. *AIAA J* 51(9):2049–2075
- McKay MD, Beckman RJ, Conover WJ (1979) Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* 21(2):239–245
- Meckesheimer M, Booker AJ, Barton RR, Simpson TW (2002) Computationally inexpensive metamodel assessment strategies. *AIAA J* 40(10):2053–2060
- Mendes-Moreira J, Soares C, Jorge AM, Sousa JFD (2012) Ensemble approaches for regression: a survey. *ACM Comput Surv* 45(1):10
- Morris MD, Mitchell TJ, Ylvisaker D (1993) Bayesian design and analysis of computer experiments: use of derivatives in surface prediction. *Technometrics* 35(3):243–255
- Osborne MA, Roberts SJ, Rogers A, Jennings NR (2012) Real-time information processing of environmental sensor network data using bayesian gaussian processes. *ACM Trans Sens Netw* 9(1):1
- Owen AB (1992) Orthogonal arrays for computer experiments, integration and visualization. *Stat Sin* 2(2):439–452

- Pan SJ, Yang Q (2010) A survey on transfer learning. *IEEE Trans Knowl Data Eng* 22(10):1345–1359
- Pan G, Ye P, Wang P (2014a) A novel Latin hypercube algorithm via translational propagation. *Sci World J* 2014:163949
- Pan G, Ye P, Wang P, Yang Z (2014b) A sequential optimization sampling method for metamodels with radial basis functions. *Sci World J* 2014:192862
- Park C, Haftka RT, Kim NH (2017) Remarks on multi-fidelity surrogates. *Struct Multidiscip Optim* 55(3):1029–1050
- Parr J, Holden CM, Forrester AI, Keane AJ (2010) Review of efficient surrogate infill sampling criteria with constraint handling. In: 2nd International conference on engineering optimization, pp 1–10
- Patterson H (1954) The errors of lattice sampling. *J R Stat Soc Ser B Methodol* 16(1):140–149
- Peherstorfer B, Willcox K, Gunzburger M (2016) Survey of multifidelity methods in uncertainty propagation, inference, and optimization. Department of Aeronautics & Astronautics, MIT, Cambridge
- Pholdee N, Bureerat S (2015) An efficient optimum Latin hypercube sampling technique based on sequencing optimisation using simulated annealing. *Int J Syst Sci* 46(10):1780–1789
- Pickett B, Turner CJ (2011) A review and evaluation of existing adaptive sampling criteria and methods for the creation of nurbs-based metamodels. In: ASME 2011 International design engineering technical Conferences and Computers and information in engineering conference, Washington, DC, USA. ASME, p 609–618
- Pronzato L, Müller WG (2012) Design of computer experiments: space filling and beyond. *Stat Comput* 22(3):681–701
- Qian PZG (2009) Nested Latin hypercube designs. *Biometrika* 96(4):957–970
- Qian PZ, Ai M (2010) Nested lattice sampling: a new sampling scheme derived by randomizing nested orthogonal arrays. *J Am Stat Assoc* 105(491):1147–1155
- Qian PZ, Wu CJ (2008) Bayesian hierarchical modeling for integrating low-accuracy and high-accuracy experiments. *Technometrics* 50(2):192–204
- Qian PZG, Wu H, Wu CJ (2008) Gaussian process models for computer experiments with qualitative and quantitative factors. *Technometrics* 50(3):383–396
- Qian PZ, Tang B, Wu CJ (2009) Nested space-filling designs for computer experiments with two levels of accuracy. *Stat Sin* 9(1):287–300
- Qian PZ, Ai M, Hwang Y, Su H (2014) Asymmetric nested lattice samples. *Technometrics* 56(1):46–54
- Quan A (2014) Batch sequencing methods for computer experiments. The Ohio State University, Columbus
- Queipo NV, Haftka RT, Shyy W, Goel T, Vaidyanathan R, Tucker PK (2005) Surrogate-based analysis and optimization. *Prog Aerosp Sci* 41(1):1–28
- Rai R, Campbell M (2008) Q2S2: a new methodology for merging quantitative and qualitative information in experimental design. *J Mech Des* 130(3):031103
- Rasmussen CE (2006) Gaussian processes for machine learning. The MIT press, London
- Raychaudhuri T, Hamey LG (1995) Minimisation of data collection by active learning. In: IEEE International Conference on Neural Networks. IEEE, p 1338–1341
- Razavi S, Tolson BA, Burn DH (2012a) Numerical assessment of metamodeling strategies in computationally intensive optimization. *Environ Model Softw* 34:67–86
- Razavi S, Tolson BA, Burn DH (2012b) Review of surrogate modeling in water resources. *Water Resour Res* 48(7):1–32
- Regis RG, Shoemaker CA (2005) Constrained global optimization of expensive black box functions using radial basis functions. *J Glob Optim* 31(1):153–171
- Regis RG, Shoemaker CA (2007) A stochastic radial basis function method for the global optimization of expensive functions. *INFORMS J Comput* 19(4):497–509
- Reichart R, Tomanek K, Hahn U (2008) Rappoport a multi-task active learning for linguistic annotations. In: ACL, p 861–869
- Rennen G, Husslage B, Van Dam ER, Den Hertog D (2010) Nested maximin Latin hypercube designs. *Struct Multidiscip Optim* 41(3):371–395
- Rimmel A, Teytaud F (2014) A survey of meta-heuristics used for computing maximin latin hypercube. In: European Conference on Evolutionary Computation in Combinatorial Optimization. Springer, p 25–36
- Rippa S (1999) An algorithm for selecting a good value for the parameter  $c$  in radial basis function interpolation. *Adv Comput Math* 11(2):193–210
- Romero DA, Amon CH, Finger S (2006) On adaptive sampling for single and multi-response bayesian surrogate models. In: ASME 2006 International design engineering technical Conferences and Computers and information in engineering conference, Philadelphia. ASME, p 393–404
- Romero DA, Amon CH, Finger S (2012) Multiresponse metamodeling in simulation-based design applications. *J Mech Des* 134(9):091001
- Rosenbaum B (2013) Efficient global surrogate models for responses of expensive simulations. Universität Trier, Trier
- Rosenbaum B, Schulz V (2012) Comparing sampling strategies for aerodynamic kriging surrogate models. *J Appl Math Mech* 92(11–12):852–868
- Rumpfkeil M, Yamazaki W, Dimitri M (2011) A dynamic sampling method for kriging and cokriging surrogate models. In: 49th AIAA Aerospace sciences meeting including the new horizons forum and Aerospace exposition, p 883
- Sacks J, Welch WJ, Mitchell TJ, Wynn HP (1989) Design and analysis of computer experiments. *Stat Sci* 4(4):409–423
- Saka Y, Gunzburger M, Burkardt J (2007) Latinized, improved LHS, and CVT point sets in hypercubes. *Int J Numer Anal Model* 4(3–4):729–743
- Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, Saisana M, Tarantola S (2008) Global sensitivity analysis: the primer. Wiley, UK
- Sanchez SM, Wan H (2015) Work smarter, not harder: a tutorial on designing and conducting simulation experiments. In: Proceedings of the 2015 Winter Simulation conference. IEEE Press, p 1795–1809
- Sasena MJ (2002) Flexibility and efficiency enhancements for constrained global design optimization with kriging approximations. University of Michigan, Ann Arbor
- Sasena M, Parkinson M, Goovaerts P, Papalambros P, Reed M (2002) Adaptive experimental design applied to ergonomics testing procedure. In: ASME 2002 International design engineering technical Conferences and Computers and information in engineering conference, Montreal. ASME, p 529–537
- Settles B (2010) Active learning literature survey. Computer science technical report 1648, University of Wisconsin, Madison
- Seung HS, Oppor M, Sompolinsky H (1992) Query by committee. In: Proceedings of the fifth annual workshop on computational learning theory, Pittsburgh. ACM, pp 287–294
- Shahsavani D, Grimvall A (2009) An adaptive design and interpolation technique for extracting highly nonlinear response surfaces from deterministic models. *Reliab Eng Syst Saf* 94(7):1173–1182
- Shan S, Wang GG (2010a) Metamodeling for high dimensional simulation-based design problems. *J Mech Des* 132(5):051009
- Shan S, Wang GG (2010b) Survey of modeling and optimization strategies to solve high-dimensional design problems with computationally-expensive black-box functions. *Struct Multidiscip Optim* 41(2):219–241



- Shan S, Wang GG (2011) Turning black-box functions into white functions. *J Mech Des* 133(3):031003
- Shewry MC, Wynn HP (1987) Maximum entropy sampling. *J Appl Stat* 14(2):165–170
- Singh P, Deschrijver D, Dhaene T (2013) A balanced sequential design strategy for global surrogate modeling. In: 2013 Winter simulations conference, Georgia. IEEE, pp 2172–2179
- Sóbestor A, Leary SJ, Keane AJ (2005) On the design of optimization strategies based on global response surface approximation models. *J Glob Optim* 33(1):31–59
- Sobol IM (1993) Sensitivity estimates for nonlinear mathematical models. *Math Model Comput Exp* 1(4):407–414
- Sobol' I (1979) On the systematic search in a hypercube. *SIAM J Numer Anal* 16(5):790–793
- Spyromitros-Xioufis E, Tsoumakas G, Groves W, Vlahavas I (2016) Multi-target regression via input space expansion: treating targets as inputs. *Mach Learn* 104(1):55–98
- Srinivas N, Krause A, Kakade SM, Seeger M (2010) Gaussian process optimization in the bandit setting: no regret and experimental design. In: Proceedings of the 27th International conference on machine learning. ACM, pp 1015–1022
- Steinberg D, Colla P (2009) CART: classification and regression trees. *The Top Ten Algorithms in Data Mining* 9:179
- Stinstra E, den Hertog D, Stehouwer P, Vestjens A (2003) Constrained maximin designs for computer experiments. *Technometrics* 45(4):340–346
- Stocki R (2005) A method to improve design reliability using optimal Latin hypercube sampling. *Comput Assist Mech Eng Sci* 12(4):393
- Sundararajan S, Keerthi SS (2001) Predictive approaches for choosing hyperparameters in Gaussian processes. *Neural Comput* 13(5):1103–1118
- Tian W (2013) A review of sensitivity analysis methods in building energy analysis. *Renew Sust Energ Rev* 20:411–419
- Toal DJ (2015) Some considerations regarding the use of multi-fidelity kriging in the construction of surrogate models. *Struct Multidiscip Optim* 51(6):1223–1245
- Turner CJ, Crawford RH, Campbell MI (2007) Multidimensional sequential sampling for NURBs-based metamodel development. *Eng Comput* 23(3):155–174
- van Dam ER, Husslage B, den Hertog D, Melissen H (2007) Maximin Latin hypercube designs in two dimensions. *Oper Res* 55(1):158–169
- van Dam ER, Rennen G, Husslage B (2009) Bounds for maximin Latin hypercube designs. *Oper Res* 57(3):595–608
- van Dam ER, Husslage B, Den Hertog D (2010) One-dimensional nested maximin designs. *J Glob Optim* 46(2):287–306
- van der Herten J, Couckuyt I, Deschrijver D, Dhaene T (2015) A fuzzy hybrid sequential design strategy for global surrogate modeling of high-dimensional computer experiments. *SIAM J Sci Comput* 37(2):A1020–A1039
- Vasile M, Minisci E, Quagliarella D, Guénot M, Lepot I, Sainvitu C, Goblet J, Filomeno Coelho R (2013) Adaptive sampling strategies for non-intrusive POD-based surrogates. *Eng Comput* 30(4):521–547
- Viana FA (2013) Things you wanted to know about the Latin hypercube design and were afraid to ask. In: 10th world Congress on Structural and multidisciplinary optimization, Orlando, pp 1–9
- Viana FA, Haftka RT, Steffen V Jr (2009) Multiple surrogates: how cross-validation errors can help us to obtain the best predictor. *Struct Multidiscip Optim* 39(4):439–457
- Viana FA, Picheny V, Haftka RT (2010a) Using cross validation to design conservative surrogates. *AIAA J* 48(10):2286–2298
- Viana FA, Venter G, Balabanov V (2010b) An algorithm for fast optimal Latin hypercube design of experiments. *Int J Numer Methods Eng* 82(2):135–156
- Viana FA, Simpson TW, Balabanov V, Toropov V (2014) Metamodeling in multidisciplinary design optimization: how far have we really come? *AIAA J* 52(4):670–690
- Villemonteix J, Vazquez E, Walter E (2009) An informational approach to the global optimization of expensive-to-evaluate functions. *J Glob Optim* 44(4):509
- Wang GG (2003) Adaptive response surface method using inherited latin hypercube design points. *J Mech Des Trans ASME* 125(2):210–220
- Wang GG, Shan S (2007) Review of metamodeling techniques in support of engineering design optimization. *J Mech Des* 129(4):370–380
- Wang L, Shan S, Wang GG (2004) Mode-pursuing sampling method for global optimization on expensive black-box functions. *Eng Optim* 36(4):419–438
- Wang H, Wu Z, Rahnamayan S (2011) Enhanced opposition-based differential evolution for solving high-dimensional continuous optimization problems. *Soft Comput* 15(11):2127–2140
- Wei X, Wu Y-Z, Chen L-P (2012) A new sequential optimal sampling method for radial basis functions. *Appl Math Comput* 218(19):9635–9646
- Williams BJ, Loeppky JL, Moore LM, Macklem MS (2011) Batch sequential design to achieve predictive maturity with calibrated computer models. *Reliab Eng Syst Saf* 96(9):1208–1219
- Xiong Y, Chen W, Apley D, Ding X (2007) A non-stationary covariance-based kriging method for metamodeling in engineering design. *Int J Numer Methods Eng* 71(6):733–756
- Xiong F, Xiong Y, Chen W, Yang S (2009) Optimizing Latin hypercube design for sequential sampling of computer experiments. *Eng Optim* 41(8):793–810
- Xu S, Liu H, Wang X, Jiang X (2014) A robust error-pursuing sequential sampling approach for global metamodeling based on voronoi diagram and cross validation. *J Mech Des* 136(7):071009
- Yang Z, Tang K, Yao X (2007) Differential evolution for high-dimensional function optimization. In: 2007 I.E. Congress on evolutionary computation, Singapore. IEEE, pp 3523–3530
- Yao W, Chen X, Luo W (2009) A gradient-based sequential radial basis function neural network modeling method. *Neural Comput & Applic* 18(5):477–484
- Younis A, Dong Z (2010) Metamodeling and search using space exploration and unimodal region elimination for design optimization. *Eng Optim* 42(6):517–533
- Zhang Y, Hoang TN, Low KH, Kankanhalli M (2016) Near-optimal active learning of multi-output Gaussian processes. In: Proceedings of the 30th AAAI conference on artificial intelligence. AAAI Press, pp 2351–2357
- Zhao L, Choi K, Lee I, Gorsich D (2009) Sequential-sampling-based kriging method with dynamic basis selection. In: 8th world Congress on Structural and multidisciplinary optimization, Lisbon
- Zhou Q, Shao X, Jiang P, Gao Z, Zhou H, Shu L (2016) An active learning variable-fidelity metamodeling approach based on ensemble of metamodels and objective-oriented sequential sampling. *J Eng Des* 27(4–6):205–231
- Zhou Q, Jiang P, Shao X, Hu J, Cao L, Wan L (2017) A variable fidelity information fusion method based on radial basis function. *Adv Eng Inform* 32:26–39