Structural and Multidisciplinary Optimization manuscript No.
(will be inserted by the editor)

Surrogate modeling: tricks that endured the test of time and some recent developments

Felipe A. C. Viana · Christian Gogu · Tushar Goel

Received: date / Accepted: date

Abstract Tasks such as analysis, design optimization, and uncertainty quantification can be computationally expensive. Surrogate modeling is often the tool of choice for reducing the burden associated with such data-intensive tasks. However, even after years of intensive research, surrogate modeling still involves a struggle to achieve maximum accuracy within limited resources. This work summarizes various advanced, yet often straightforward, statistical tools that help. We focus on four techniques with increasing popularity in the surrogate modeling community: (i) variable screening and dimensionality reduction in both the input and the output spaces, (ii) data sampling techniques or design of experiments, (iii) simultaneous use of multiple surrogates, and (iv) sequential sampling. We close the paper with some suggestions for future research.

Keywords Dimensionality reduction · Variable screening · design of experiments · surrogate modeling · sequential sampling

Felipe A. C. Viana
Department of Mechanical and Aerospace Engineering
University of Central Florida
Orlando, FL 32816, USA
ORCID=0000-0002-2196-7603
E-mail: viana@ucf.edu

Christian Gogu
Université Toulouse III
Toulouse, F-31400, France
E-mail: christian.gogu@gmail.com

Tushar Goel
Noida, India
E-mail: tushar.goel@gmail.com

1 Introduction

Statistical modeling of computer experiments embraces the set of methodologies for generating a surrogate model (also known as metamodel or response surface approximation) used to replace an expensive simulation code [30, 56, 118, 156, 171, 209]. The goal is constructing an approximation of the response of interest based on a limited number of expensive simulations. Although it is possible to improve the surrogate accuracy by using more simulations, limited computational resources often makes us face at least one of the following problems:

- Desired accuracy of a surrogate requires more simulations than we can afford.
- The output that we want to fit is often not a scalar (scalar field) but a high-dimensional vector (vector field with several thousand components), which can be prohibitive or impractical to handle.
- We use the surrogate for global optimization and we do not know how to simultaneously obtain good accuracy near all possible optimal solutions.
- We use the surrogate for optimization, and when we do an exact analysis we find that the solution is infeasible.

This paper discusses diverse techniques that have been extensively used to address these four issues. We focus on (i) variable screening and dimensionality reduction [55, 121, 186, 213, 217], (ii) design of experiments [113, 144, 147], (iii) use of multiple surrogates [58, 178, 214], and (iv) sequential sampling and optimization [102, 108]. Although we will focus our discussion on computer models, several of these techniques can be either applied or extended to the cases in which data comes from physical experiments.

The remaining of the paper is organized as follows. Section 2 reviews the variable screening and dimension reduction techniques. Section 3 discusses issues relevant to effective methods of sampling design points. Section 4 presents the use of multiple surrogates. Section 5 focuses on sequential sampling techniques. Finally, section 6 closes the paper recapitulating salient points and concluding remarks.

2 Reducing input and output dimensionality

2.1 Variables reduction in input space

As the number of variables in the surrogate increases, the number of simulations required for surrogate construction rises exponentially (curse of dimensionality). A question at this point is then the following: is it necessary to construct the response surface approximation in terms of all the variables? Some of the variables may have only a negligible effect on the response. Several techniques have thus been proposed for evaluating the importance of the variables economically. In the next few paragraphs, we first provide a brief historical overview of methods that have been proposed in this context and which withstood the test of time. Then we focus on a few recent techniques of particular interest in more detail.

2.1.1 Variable screening techniques

A wide category of dimensionality reduction techniques in input space is commonly referred to as variables screening. Among the simplest screening techniques are so called one-at-a-time (OAT) plans [47], which evaluate in turns the effect of changing one variable at a time. It is a very inexpensive approach, but it does not estimate interaction effects between variables. Variations of OAT screening that account for interactions have been proposed by Morris [145] and Cotter [42].

Various statistical techniques such as Chi-square test, t-test, F-test, correlation coefficients (e.g. Pearson's, Spearman's, Kendall's) are also among the simplest techniques to carry out variable screening.

A very popular category of variable screening is stepwise regression, first proposed by Efromyson in 1960 [54] and refined further [10, 66, 83, 84]. Stepwise regression is achieved in one of the following methods:

1. **Forward stepwise** regression starts with no variable and tests adding one independent variable at a time and adding that variable if it is statistically significant to include that variable.
2. **Backward stepwise** regression starts with all variables, tests removing one variable at a time, and removing the variable which leads to minimal deterioration in the quality of fit.
3. **Bidirectional stepwise** regression is a combination of the above-mentioned two methods and tests which variables should be included or excluded.

Agostinelli [3] presented a method to improve robustness of the stepwise regression. Mitzi [130] presented an overview of issues associated with this technique.

Such screening techniques are relatively simple but may quickly become computationally intractable and are not always well suited for subsequent use for surrogate model construction due to the specific design of experiments they involve. To address these issues neighborhood component feature selection (NCFS) has been proposed to select relevant input variables [221]. NCFS solves an unconstrained multi-objective optimization problem that minimizes the mean loss of a neighborhood component analysis regression model while preventing over-fitting by using a regularization term. The optimum weights that are the results of the optimization problem provide information regarding the significance of the input variables, thus allowing variable screening. The NCFS technique has been recently further enhanced to reduce its computational cost and improve its robustness in [114]. The resulting normalized neighborhood component feature selection (nNCFS) method has been applied to the design of the body-in-white of a vehicle allowing to reduce the number of design variables from 31 to only 16.

2.1.2 Variance based techniques

Another category of screening techniques are variance based. A simple, commonly used approach uses a k-level factorial or fractional factorial design followed by an analysis of variance procedure (ANOVA) [191]. The procedure allows identifying the main and interaction effects that account for most of the variance in the response. ANOVA can be carried out either using a reduced fidelity, computationally inexpensive model as has been recommended in [73, 99, 179, 187, 193] and more rarely using the high fidelity model as in [48]. Note that using a reduced fidelity model is a common technique in variable screening, which will be illustrated again in the next paragraphs with global sensitivity analysis.

The iterated fractional factorial design (IFFD) method [4, 173] is another screening approach, designed to be economical for large number of variables (e.g. several thousands). The method assumes that only very few

variables account for most of the variance in the response and calculates the main effects, quadratic effects and cross term effects for these significant variables based on a clustered sampling technique.

For additional details and applications on these and additional sensitivity and screening methods, such as Bettonvil's sequential bifurcation method [15], the reader can refer to [34, 116, 172] and the references therein. In the remainder of this subsection we will develop in more detail just a few techniques that we find of particular interest.

An approach gaining popularity for finding and eliminating variables that have negligible effects is Sobol's global sensitivity analysis (GSA), initially introduced by Sobol [186]. GSA is a variance-based approach that provides the total effect (i.e. main effect and interaction effects) of each variable, and thus indicates which ones could be ignored. In the basic case, the variances are usually computed using Monte Carlo simulation based on an initial crude surrogate (assumed to be accurate enough for screening). Numerous improvements have been proposed to more efficiently compute Sobol's sensitivity indices and we refer the reader to the review in [98] for an overview. Sobol indices being variance based will also have issues for correctly assessing sensitivities that are not reflected in variances but only in higher order moments. Several methods have been proposed to address this issue and we refer the reader to [19] for a review. GSA has been successfully applied to many engineering problems and as just a few examples we can cite piston shape design [104], liquid rocket injector shape design [196], bluff body shape optimization [140], optimization of an integrated thermal protection system [79], design of high lift airfoil [143], the identification of material properties based on full field measurements [138], stream flow modeling [36], performance characterization of plasma actuators [33], uncertainty quantification in multi-phase detonations [65].

2.1.3 Variable transformation techniques

Variable transformation techniques are another major approach for reducing the number of variables and improving the quality of fit in a surrogate. It is often driven by exploiting the domain knowledge of the user. Box and Tidwell [22] suggested that it is sometimes possible to build a good quality approximation using the transformed independent variables. In these methods, the non-linear relationship between dependent and independent variable is addressed by an upfront transformation such that the relationship between transformed independent variable and the dependent variable is more amenable to surrogate modeling. A few effective transformation techniques are enumerated as follows:

1. **Mathematical operators** - functions such as log(), ln(), tanh(), power(), inverse(), abs() etc. are useful to transform variables into more meaningful variables [147] for certain problems.
2. **Variable scaling** - such as scaling of independent and/or dependent variables using uniform or normal distributions is another effective transformation technique that helps improve the approximation. This is particularly helpful when different variables have different scales.
3. **Filters** - are used to handle known bifurcations in the independent or dependent variables relationships which are hard to model using a continuous independent variable approach. In this method, values of the independent variables above (or below) are filtered out such that the same variable is applied only under certain conditions. Some examples of common filters are simple mean or moving-average based filtering, threshold-based filter functions where a variable is only considered on one side of the threshold, and auto-correlations. These filters are quite commonly used in time-series analysis and control applications [69, 135, 167, 212].

One concept related to variables transformation but relevant to physical problems consists in grouping the variables into a smaller number by non-dimensionalization (i.e. transforming variables of a physical problem into variables that have no units). The Vaschy-Buckingham theorem [25, 188, 199] provides a systematic construction procedure of non-dimensional parameters and also guarantees that the parameters found are the minimum number of parameters (even though not necessarily unique) required for an exact representation of a given partial differential equations based problem.

Following the initial works by Li and Lee [132] and Dovi et al. [49] several authors have further shown that improved accuracy polynomial response surface approximations can be obtained by using non-dimensional variables [38, 61, 111]. This is mainly because, for the same number of numerical simulations, the generally much fewer non-dimensional variables allow a fit with a higher order polynomial. While this technique is not as popular as some of the screening or sensitivity approaches, multiple works showed successful applications.

In [210] the approach was illustrated on statistical data from an automobile survey while in [126] the method was applied to several engineering case studies including a finite element (FE) based example.

In [80] the authors managed to reduce the number of variables from eight to four by constructing the

surrogate in terms of non-dimensional variables for a vibration problem of a free plate. In [201] a reduction from nine to seven variables was achieved by using non-dimensional parameters for FE analyses, modeling a mechanical problem of a plate with an abrupt change in thickness.

An even greater reduction in the number of variables is possible if non-dimensionalization is combined with other techniques. In [81] the authors achieved a reduction from fifteen to only two variables using a combination of physical reasoning, non-dimensionalization and global sensitivity analysis for a thermal design problem for an integrated thermal protection system. Physical reasoning allowed formulating simplifying assumptions that reduced the number of variables from fifteen to ten. Non-dimensionalization was then applied on the equations of the simplified problem reducing the number of variables to three non-dimensional variables. Finally, global sensitivity analysis showed that one of the non-dimensional variables had an insignificant effect thus leaving only two variables for the surrogate model.

While non-dimensionalization is by itself a variable transformation technique, it can further benefit by a combination with other techniques. Non-dimensionalization combined with exponential scaling transformations has been, for example, shown to be extremely effective at both reducing the dimensionality and improving the accuracy of surrogate models of mechatronic components over a large domain of variation of the input variables (multiple orders of magnitude) [26, 92, 142, 175].

2.1.4 Dimensionality reduction by subspace construction

Another major variable reduction approach we would like to comment on concerns problems in which the designs of interest are confined into a reduced dimension (e.g. a plane in three dimensional space). If all possible input vectors (of dimension n) lie in a lower-dimensional subspace (of dimension k, with $k < n$), then using the initial input space for surrogate construction will lead to poor results due to numerical ill-conditioning. A reduced dimensional representation of the variables should then be used by expressing the initially n-dimensional input vectors in a basis of the corresponding k-dimensional subspace.

Note that the fact that the input variables are not independent but correlated, such that they can be confined in a certain subspace, may appear as a wrong problem formulation, and one should directly express the problem in terms of independent variables. In many problems though, the choice of the input variables is majorly driven by controllable variables and often, defining a priori a subspace in which variables lie is hard if not impossible. Let's take the example of aeroelastic coupling where the lift created by a wing is affected by the deformed shape of the wing. To define the deformed shape the simplest, physically meaningful description is the deformed position of each finite element node in the mesh used for the structural finite element solver. It is obvious that the deformed positions at each node are not independent, yet defining the subspace in which these positions lie is far from trivial. Multiple techniques have thus been developed in order to uncover the reduced dimensional subspace in which the input variables lie.

Principal components regression (PCR) is a technique developed to fit an approximation (e.g. polynomial) to the data in the appropriate sub-dimensional subspace ([105] and [13, Chapter 3]). Mandel [141] has shown that the technique can be advantageous also when the variables-data is not strictly confined to a subspace, but the components outside of the subspace are relatively small. Rocha et al. [168] have shown that for the problem of fitting wing weight data of subsonic aircraft, PCR provides more accurate results compared to other fitting techniques (polynomial interpolation, kriging, radial basis function interpolation) due to its ability to account for the physical and historical trends buried within the input data.

The partial least squares [218] and sparse partial least squares [35] techniques have also been proposed as alternatives to principal component regression for reducing very high-dimensional data. Recent techniques also seek to combine both reduced dimensionality subspace learning and variance based sensitivity analysis [5, 7, 123, 226] in order to more efficiently compute accurate sensitivity indices for expensive computer models.

In the past decade multiple researchers sought to develop specific covariance kernels for Gaussian process surrogate models that intrinsically integrate dimensionality reduction. Several such techniques are based on seeking a low dimensional linear subspace on which to project the gaussian process input variables [20, 64, 184]. Various other methods to construct reduced dimensional manifolds have also been explored [27, 46, 51, 149, 185]. Many of these Gaussian process surrogates have been developed or applied in the context of sequential sampling (see section 5).

2.2 Dimensionality reduction in the output space

The question of a dimensionality reduction requirement in the output space poses itself less often than in the input space (that is, usually we want to have the minimal number of variables that controls a particular scalar field). For vector-based response, techniques that take

into account correlation between components are available and can in some cases be more accurate than considering the components independently [113]. For a relatively small dimension of the output vector, methods such as co-kriging [146] or vector splines [223] are available. However, these are not practical for approximating the high dimensional pressure field around the wing of an aircraft or approximating heterogeneous displacements fields on a complex specimen. These fields are usually described by a vector with thousands to millions of components. Fitting a surrogate for each component is then time and resource intensive and might not take advantage of the correlation between neighboring points.

Principal components analysis (PCA), which we already touched upon in the previous section addresses this problem. PCA is also known as proper orthogonal decomposition (POD) or Karhunen-Loeve expansion, and it can be applied to achieve drastic dimensionality reduction in the output space (reductions from many thousands to less than a dozen coefficients are frequent).

POD finds a low dimensional basis from a given set of N simulation samples. A field $\boldsymbol{u}$ depending of variables of interest $\boldsymbol{x}$ sampled within a sampling domain D is then approximated by

$$\tilde{\boldsymbol{u}}(\boldsymbol{x}) = \boldsymbol{\Phi}\boldsymbol{\alpha}(\boldsymbol{x}) = \sum_{i=i}^{n_{RB}} \langle \boldsymbol{u}, \boldsymbol{\Phi_i} \rangle \boldsymbol{\Phi_i}\,, \tag{1}$$

where $\boldsymbol{u} \in \mathbb{R}^n$ is the vector representation of the field (e.g. pressure or displacement fields), $\boldsymbol{\Phi} = (\boldsymbol{\Phi}_1, \ldots, \boldsymbol{\Phi}_{n_{RB}})$ is the matrix composed of the basis vectors $\boldsymbol{\Phi}_i,\ i = 1, \ldots, n_{RB} \in \mathbb{R}^n$ of the reduced-dimensional, orthogonal basis, and $\boldsymbol{\alpha}$ the vector of the coefficients of the field in this basis (i.e. the orthogonal projection of the field onto the basis vectors).

This approach allows approximating the fields in terms of n_{RB} POD coefficients α_i instead of n components of the vector $\boldsymbol{u}$. For additional details on the theoretical foundations of POD, the reader can refer to [24, Chapter 3] (Nota Bene: the book's foreword was written by Raphael Haftka, to whom this special issue is dedicated).

Once the reduced dimensional basis determined, one can construct surrogate models for each $\alpha_i\ i = 1, \ldots, n_{RB}$ based on the same N samples that were used for POD basis construction. It might seem surprising that, considering that n is equal to several thousands, n_{RB} can be low enough to easily allow the construction of a surrogate model per POD coefficient, but successful applications have proven the applicability of the approach to a large variety of problems. This is due to the fact that variations even in complex fields can be controlled by physical phenomena exhibiting effects characterized by relatively low dimensionality.

For example this approach has been successfully applied to the multidisciplinary design optimization of an aircraft wing [39, 40]. The method enabled cost-efficient fluid-structure interaction required for multi-disciplinary design optimization. The initial application [39] was carried out on a two-dimensional wing model, where the variations in vectors of size $n = 70$ were reduced to only two POD coefficients. A subsequent study [40] applied the same approach to more realistic 3D wing models where the pressure fields around the wing (aerodynamics model) as well as the wing displacement field (structural model) were reduced and approximated via this approach (reduction from many thousands to less than a dozen components).

Using both surrogate modeling and dimensionality reduction to model strongly coupled multidiscipliary systems such as the fluid structure interaction on an airfoil remains a challenge, in particular when seeking to construct both the reduced basis and the surrogate model as efficiently as possible. In this context sequential sampling would appear particularly promising with first attempts underway in this direction [14].

Another application of POD reduction combined with surrogate modeling concerned Bayesian identification of orthotropic elastic constants based on displacement fields on a plate with a hole [82]. Variations in the displacement fields (modelled by vectors with about 5,000 coefficients) were reduced to only four POD coefficients, containing enough information to perform the identification. Surrogate models of these POD coefficients were constructed, enabling a sufficient cost reduction for the Bayesian identification to be carried out, which requires expensive correlation information.

Other applications of these combined techniques include the heartbeat modeling of a whole heart [160, 161] and efficient reparametrization of CAD models [158].

PCA/POD are linear techniques, in that the output is expressed as a linear combination of the basis vectors. Non-linear dimensionality reduction approaches have been developed for cases where linearity will not provide a sufficiently accurate approximation. Kernel PCA [177] solves a PCA eigenproblem in a new feature space constructed through kernel methods. Local PCA subdivides the initial design space into clusters and applies PCA to each one of them. K-means [109] or spectral approaches [139] can be used as clustering methods. More recently methods based on neural networks [95], and in particular deep neural networks [97, 128], have become extremely popular and allow large dimensionality reductions for very high dimensional, highly non linear problems. In this context autoencoders [215] and in particular variational autoencoders [115, 166] have shown

to be quite effective in a large variety of domains. The main challenge in applying deep neural networks based approaches to problems involving physics based computational models resides in the requirement of large design of experiments (i.e. training dataset) for appropriate training.

To summarize, when seeking a surrogate of a high dimensional vector field it is often worth exploring first the accuracy of linear approximations by PCA/POD. Even for non-linear models these techniques have often been shown to provide sufficient accuracy. Only when this is not the case it is worth exploring the more advanced non-linear dimensionality reduction techniques.

3 Design of experiments

Surrogate models are used to reduce the cost of design and optimization by fitting "approximate" models for complex problems which are expensive to solve. These surrogate models are useful for design and optimization only if they can predict the underlying behaviour accurately. The accuracy of any surrogate model is primarily affected by two factors: (i) noise in the data and (ii) inadequacy of the fitting model (called modeling error or bias error). While errors in the approximation due to both reasons can be reduced by increasing the data available for approximation, practically the feasible number of data points is limited by budgetary constraints. So, we need to carefully plan the experiments to maximize the accuracy of surrogates, and this exercise is known as design of experiments (DOE). There have been a few studies on the influence of different experimental designs on predictions [75, 156, 180, 181]. We discuss the problem characteristics that influence the construction of DOE and different types of DOEs in the next two subsections. We follow up that by some practical tips by drawing on our experiences in this area.

3.1 Factors influencing the choice of DOE

We enumerate the following main considerations to be undertaken while constructing a design of experiment:

1. **Number of variables and order of surrogate** – This is a primary consideration in the surrogate construction. The requirement for the number of experiments increases rapidly as the number of variables increase. In Fig. 1, we notice that the number of coefficients required for quadratic and cubic polynomials increases rapidly as the number of variables increase. This means that the number of experiments required to build the model also increases rapidly.
2. **Size of the design space** – Similar to the number of variables, the requirements on the number of data points increase for problems with large design spaces. Depending on the placement of the data points, we could have large extrapolation region or large spaces between the data points. Both of these can result in poorer quality of fit.
3. **Shape of the design space** – We usually work with regular design spaces which are easier to handle as there are no restrictions on placing the experimental points. Occasionally, we get into cases where a part of the design space is not feasible due to geometric constraints or other reasons. In such cases, we need to carefully plan the DOE to account for the irregularities.
4. **Nature of the underlying function**– The number of data points required to build a surrogate model is highly dependent on the underlying function that we need to fit. When the underlying function is well-behaved, lower order models can give a good fit however when the underlying function is highly non-linear, we need a higher order model and consequently higher number of design points.
5. **Noise in experiments** – As discussed above, we can have noise or bias error as dominant sources. Experimental noise due to measurement error or experimental errors is a dominant factor for physical experiments. For computer simulations, noise could be due to numerical noise, which is usually small unless there are issues with convergence as observed in computational fluid dynamics or non-linear finite element model simulations. If noise is the dominant factor or error, we should use the DOEs that minimize the influence of noise errors such as minimum-variance designs [144].
6. **Bias errors** – The true model representing the data is rarely known. This is a possible source of error when we model a higher order true function with a lower order surrogate. To truly address this, one should try as large order surrogate as possible however, as shown in Fig. 1, the requirements of data required to build a higher order surrogate increases rapidly as we go to higher order models. There are some DOE techniques that cater to reducing the influence bias errors, such as minimum bias designs [113, 147].
7. **Cost of construction of the DOE** – Finally, we also consider the cost involved in building an efficient experimental design. For most of the problems, the cost of constructing an experimental design is much smaller than the cost of running a single experiment. However, the cost of constructing an experimental design can be significant when we have a large num-

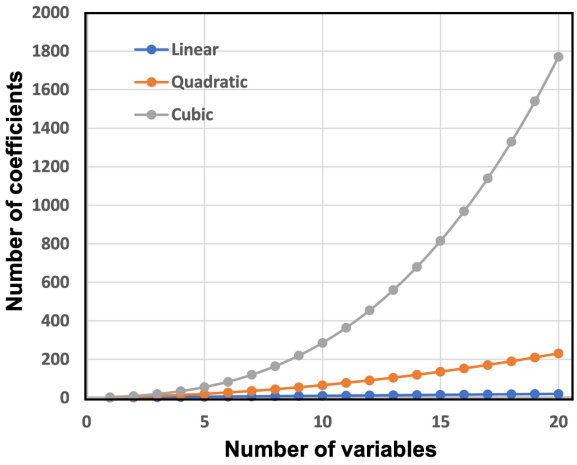


Fig. 1: Increase in the number of coefficients (y-axis) with the dimensionality of the problem (x-axis shows number of variables).

ber of variables and experimental points to select. In those cases, the selection of DOE through optimality criterion based EDs may require substantial computational effort in picking the optimal designs.

Generally, the design of experiment techniques seeks to reduce the effect of noise and reduce bias errors simultaneously. However, these objectives (noise and bias error) often conflict. For example, noise rejection criteria, such as D-optimality, usually produce designs with more points near the boundary, whereas the bias error criteria tend to distribute points more evenly in design space. Thus, the problem of selecting a DOE is a multi-objective problem with conflicting objectives (noise and bias error). The solution to this problem would be a Pareto optimal front of DOEs that yields different trade-offs between noise and bias errors [78].

3.2 Types of DOEs

In its simplest form, a DOE could consider "one-factor-at-a-time" which runs through a grid search by considering only one variable keeping others constant [47]. This is not effective as we cannot evaluate the impact of interactions among variables. Myers et al. [147] and Montgomery [144] have discussed various DOE techniques in detail. Yondo et al. [224] has presented an excellent overview of different DOE techniques. While they classify various DOEs into classical and modern DOE groups, we expand that to the following four groups:

1. **Classical DOE** –Generally focus on minimizing the influence of noise error in the approximation. In such cases, the design points are located towards the boundary of the space and there can be gaps in the interior. Some examples are, full or fractional factorial designs, central composite designs, optimal designs that minimize some variance criterion, and orthogonal arrays [147, 192].
2. **Modern DOE** – These DOEs are utilized when bias errors are dominant or the design space is huge. These techniques tend to place data points in the interior of the design space such that a part of the original design space is not covered by the largest hypercube constructed by the selected data points and that region would be subjected to extrapolation by the fitted models. If responses in this region are considered important, care must be taken to choose the surrogate models which are less-susceptible to the issues related to extrapolation. Some examples are, space filling designs, random designs, Monte-carlo sampling, quasi-random designs, Latin hypercube sampling (LHS) [203] and uniform designs,
3. **Hybrid Designs** – These DOEs are used to find balance between bias and noise errors. Some examples are, optimal LHS designs [206], orthogonal array based LHS [127], multiple criteria-based designs [78].
4. **Sequential DOE** – While most studies are focused on one-shot DOEs, the sequential DOEs are quite popular in applications involving expensive simulations. They start with a small subset of design points and keep adding points in the region of interest balancing the exploration and exploitation [44]. Section 5 will expand on such sequential sampling techniques.

Recommendations are difficult for high-dimensional problems as the curse of dimensionality will eventually make the problems hard. Successful approaches are likely to leverage a combination of strategies that include efforts in reducing the search space, either by reducing the bounds or by reducing the number of variables. In addition, it is also very common to proceed with data gathering sequentially. Strategies that rely on sequential sampling can be used, but are likely to be augmented by expert domain-knowledge.

3.3 Practical Recommendations

Based on our experience, we propose choosing the DOEs based on the dimensionality and dominant source of error as shown in Table 1 .

4 Surrogate modeling and ensembles

The simultaneous use of multiple surrogates addresses two of the problems mentioned in the introduction:

Table 1: Recommended DOE based on the dimensionality and dominant source of error

Number of variables	Type of error	Recommended DOE
≤ 5	Any	Full factorial designs, composite designs (e.g., face-centered or central composite designs), and space-filling DOEs (e.g., LHS)
> 5 and ≤ 10	Noise	Classical DOEs that minimize variance such as D-optimal DOE and space-filling DOEs (e.g., LHS)
> 5 and ≤ 10	Bias	Modern DOE (minimize bias), space-filling DOE, LHS
> 5 and ≤ 10	Noise and bias	Hybrid DOEs
> 10	Any	Sequential DOEs in combination with dimensionality and search space reduction

-- Accurate approximation requires more simulations than we can afford by offering an insurance against poorly fitted models.
-- Surrogate models for global optimization: since different surrogates might point to different regions of the design space this constitutes at least a cheap and direct approach for global optimization.

4.1 How to generate different surrogates

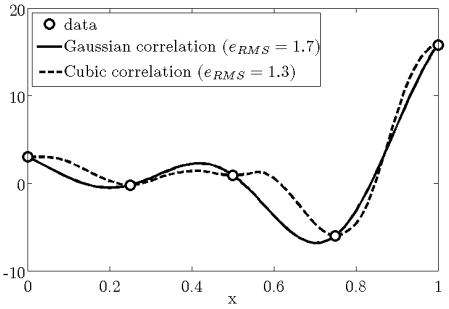
Most practitioners in the optimization community are familiar at least with the traditional polynomial response surface [23, 147], some with more sophisticated models such as the Gaussian process and kriging [164, 190], neural networks [31, 85], or support vector regression [176, 183], and few with the use of weighted average surrogates [1, 77, 205]. The diversity of surrogate models might be explained by three basic components:

1. Statistical modeling: for example, response surface techniques frequently assume that the data is noisy and the obtained model is exact. On the other hand, kriging usually assumes that the data is exact and is a realization of a Gaussian process.
2. Basis functions: response surfaces frequently use monomials. Support vector regression specifies the basis in terms of a kernel (many different functions can be used).
3. Loss function: the minimization of the mean square error is the most popular criteria for fitting the surrogate. Nevertheless, there are alternative measures such as the average absolute error (i.e., the L1 norm).

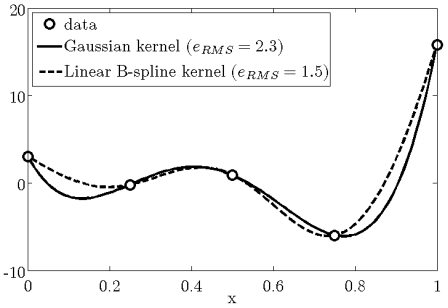
It is also possible to create different instances of the same surrogate technique. For example, we could create polynomials with different choice of monomials, kriging models with different correlation functions, and support vector regression models with different kernel and loss functions. Figure 2 illustrates this idea showing different instances of kriging and support vector regression fitted to the same set of points.

In many applications, we might have to construct surrogate models for multiple responses based on a single set of pre-sampled points. Obviously we can always build individual surrogate models for each surrogate model separately. Nevertheless, some surrogate techniques allow for the simultaneous modeling of multiple responses. As illustrated in Fig. 3, multi-layer perceptrons, one of the most popular neural network architectures, can be easily designed to accommodate multiple outputs. In such case, the hidden layers are responsible to capture shared features and inter-dependencies, while the very last layer performs final differentiation of the multiple responses. The Gaussian process can also be used to model multiple and correlated responses. With a single output, the Gaussian process is formulated as

$$f(\mathbf{x}) \sim GP(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')) \,, \tag{2}$$



(a) Kriging surrogates with different correlation functions.



(b) Support vector regression models with different kernel functions.

Fig. 2: Different surrogates fitted to five data points of the $y = (6x-2)^2 \sin(2(6x-2))$ function (adapted from [209]).

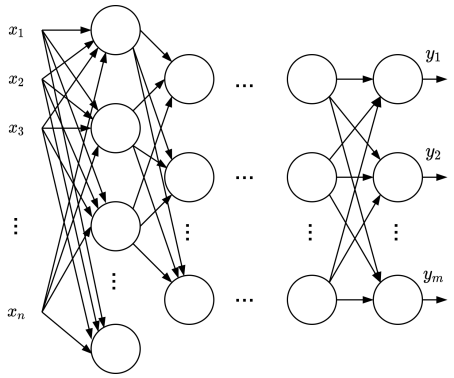


Fig. 3: Multi-layer perceptron neural network design for multiple outputs.

where $\mathbf{x}$ are prediction points, $\mathbf{x}'$ are observed points, and $m(\mathbf{x})$ and $k(\mathbf{x}, \mathbf{x}')$ define the mean and covariance, respectively. One way to extend the formulation to multiple outputs is to expand the mean and covariance functions such that

$$\mathbf{f}(\mathbf{x}) \sim GP(\mathbf{m}(\mathbf{x}), \mathbf{K}(\mathbf{x}, \mathbf{x}'))\,, \tag{3}$$

where $\mathbf{m}(\mathbf{x})$ and $\mathbf{K}(\mathbf{x}, \mathbf{x}')$ define the multiple-output mean and covariance. The interested reader is referred to [41, 60, 134] for further details.

4.2 Comparison of surrogates

The vast diversity of surrogates has motivated many papers comparing the performance among techniques. For example, Giunta and Watson [74] compared polynomial response surface approximations and kriging on analytical example problems of varying dimensions. They concluded that quadratic polynomial response surfaces were more accurate. However, they hedged that the investigation was not intended to serve as an exhaustive comparison between the two modeling methods. Jin et al. [101] compared different surrogate models based on multiple performance criteria such as accuracy, robustness, efficiency, transparency, and conceptual simplicity. They concluded that the performance of different surrogates has to do with the degree of nonlinearity of the actual function and the design of experiment (sampled points). Stander et al. [189] compared polynomial response surface approximation, kriging, and neural networks. They concluded that although neural nets and kriging seem to require a larger number of initial points, the three meta-modeling methods have comparable efficiency when attempting to achieve a converged result. Overall, the literature leads us to no clear conclusion. Instead, it confirms that the surrogate performance depends on both the nature of the problem and the sampled points.

While different metrics can be used to compare surrogates (such as the coefficient of determination R^2, the average absolute error, the maximum absolute error [86]-[89]), here, we propose the use of the root mean square error [101, 103]. The root mean square error e_{RMS} in a design domain D of volume V is given by

$$e_{RMS} = \sqrt{\frac{1}{V}\int_D (\hat{y}(\mathbf{x}) - y(\mathbf{x}))^2 d\mathbf{x}}\,, \tag{4}$$

where $\hat{y}(\mathbf{x})$ is the surrogate model of the response of interest $y(\mathbf{x})$. The integral of Eq. (4) can be estimated using numerical integration at test points. As can be seen in Fig. 2, the e_{RMS} of a set of surrogates can greatly differ in terms of accuracy. As a result, it might be hard to point the best one for a given problem and data set.

4.3 Surrogate selection

We discussed how factors such as statistical modeling, basis functions, and loss functions determine how we can generate different surrogates. In practice, the selection of a surrogate model also involves factors such as

- Nature of the underlying function: while well-behaved functions can be handled by simple linear regression, highly nonlinear functions might call for sophisticate approaches such as support vector regression, Gaussian process, and neural networks.
- Usages of the surrogate model: most surrogate models are built for allow for certain degree of extrapolation far from observed data. Unfortunately most approaches tend to fail as prediction is pushed far away from data points. In such applications, uncertainty estimates, such as those found in Gaussian processes, are important as a reference for confidence in predictions points.
- Strategy for acquisition of data points: here we can differentiate all-at-once versus sequential sampling strategies. As we will discuss in the next section, sequential sampling strategies are mostly based on uncertainty estimates from surrogates.

If only one predictor is desired, one could apply either selection or combination of surrogates [205]. Selection is usually based on a performance index that applies to all surrogates of the set (that is, a criterion that does not depend on the assumptions of any particular surrogate technique). In this case, the use of test points is a luxury and we usually have to estimate the accuracy of the surrogate based on the sampled data only. This explains why cross validation for estimation of the prediction

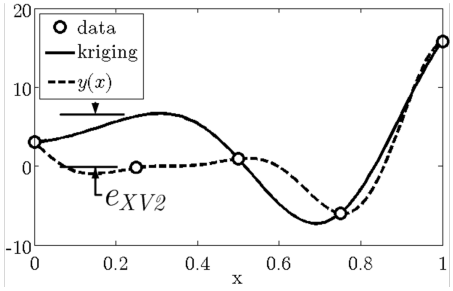


Fig. 4: Cross-validation error at the second point of the five-point experimental design e_{XV_2}. The kriging model is fitted to the remaining four points of the function $y = (6x-2)^2 \sin(2(6x-2))$.

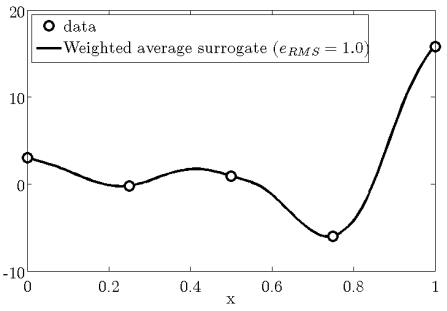


Fig. 5: Weighted average surrogate based on the models of Fig. 2.

errors and assessment of uncertainty [204] based on observed data has become so popular[1].

A cross-validation error is the error at a data point when the surrogate is fitted to a subset of the data points not including this point. When the surrogate is fitted to all the other $p-1$ points, the process has to be repeated p times (leave-one-out strategy) to obtain the vector of cross-validation errors, $\mathbf{e}_{XV}$. Figure 4 illustrates computation of the cross-validation errors for a kriging surrogate. When the leave-one-out becomes expensive, the k-fold strategy can also be used for computation of the $\mathbf{e}_{XV}$ vector. According to the classical k-fold strategy [122], after dividing the available data (p points) into p/k clusters, each fold is constructed using a point randomly selected (without replacement) from each of the clusters[2]. Of the k folds, a single fold is retained as the validation data for testing the model, and the remaining $k-1$ folds are used as training data. The cross-validation process is then repeated k times with each of the folds used once as validation data. Cross validation is a powerful way to estimate prediction errors and assess uncertainty estimates since it is based only on observed data and does not depend on the assumptions of any particular surrogate technique. Popular criteria built with cross validation include root mean square error, maximum absolute error, coefficient of determination (i.e., R^2 and adjusted R^2), prediction variance, and others.

The square root of the $PRESS$ value ($PRESS$ stands for prediction sum of squares) is the estimator of the e_{RMS}:

$$PRESS = \sqrt{\frac{1}{p}\mathbf{e}_{XV}^T \mathbf{e}_{XV}}\,. \tag{5}$$

Since $PRESS$ is an estimator of the e_{RMS}, one possible way of using multiple surrogates is to select the model with best (i.e., smallest) $PRESS$ value [195, 205, 228]. Since the quality of fit depends on the data points, the surrogate of choice may vary from one experimental design to another.

Combining surrogates is based on the hope of canceling errors in prediction through proper linear combination of models. This is shown in Fig. 5, in which the weighted average surrogate created using the four surrogates of Fig. 2 has smaller e_{RMS} than any of the basic surrogates. Cross-validation errors can be used to obtain the weights via minimization of the integrated square error [1, 205]. Alternatively, the weight computation might also involve the use of local estimator of the error in prediction. For example, Zerpa et al. [227] presented a weighting scheme that uses the prediction variance of the surrogate models (available in kriging and response surface for example).

Nevertheless, the advantages of combination over selection have never been clarified, as a manifestation of the no-free lunch theorem [219]. In addition, the success of a surrogate model is highly dependent on its computational implementation (including aspects such as optimization of hyper-parameters, numerical implementation of linear algebra, etc.). As rough guidelines, we refer to the work of Yang [222] and Viana et al. [205]. According to Yang [222], selection can be better when the errors in prediction are small and combination works better when the errors are large. Viana et al. [205] showed that, in theory, the surrogate with best e_{RMS} can be beaten via weighted average surrogate. In practice, the quality of information given by the cross-validation errors makes it very difficult. In addition, since their work used globally assigned weights, the potential gains diminish substantially in high dimensions.

[1] Nevertheless, cross validation should be used with caution, since the literature has reported problems such as bias in error estimation [152, 198].

[2] In practice, k-fold implementations are influenced by the size of datasets. With small to medium datasets, using clustering algorithms, such as k-means [110], improves robustness. When datasets are very large, clustering is very time consuming and potentially less important. Therefore, random selection is commonly used.

4.4 Multiple surrogates in optimization

A surrogate-based optimization cycle consists of choosing points in the design space (experimental design), conducting simulations at these points and fitting a surrogate (or maybe more than one) to the expensive responses. If the fitted surrogate satisfies measures of accuracy, we use it to conduct optimization. Then we verify the optimum by conducting exact simulation. If it appears that further improvements can be achieved, we update the surrogate with this new sampled points (and maybe zoom in on regions of interest) and conduct another optimization cycle. In this scenario, it seems advantageous to use multiple surrogates. After all, one surrogate may be more accurate in one region of design space while another surrogate may be more accurate in a different region. The hope is that a set of surrogates would allow exploration of different portions of the design space by pointing to different candidate solutions of the optimization problem.

Examples of this approach can be found in the literature. For instance, Mack et al. [140] employed polynomial response surfaces and radial basis neural networks to perform global sensitivity analysis and shape optimization of bluff body devices to facilitate mixing while minimizing the total pressure loss. They showed that due to small islands in the design space where mixing is very effective compared to the rest of the design space, it is difficult to use a single surrogate to capture such local but critical features. Glaz et al. [76] used polynomial response surfaces, kriging, radial basis neural networks, and weighted average surrogate for helicopter rotor blade vibration reduction. Their results indicated that multiple surrogates can be used to locate low vibration designs which would be overlooked if only a single approximation method was employed. Samad et al. [174] used polynomial response surface, kriging, radial basis neural network, and weighted average surrogate in a compressor blade shape optimization of the NASA rotor 37. It was found that the most accurate surrogate did not always lead to the best design. This demonstrated that using multiple surrogates can improve the robustness of the optimization at a minimal computational cost. The use of multiple surrogates was found to act as an insurance policy against poorly fitted models. The drawback is that the strategy generate multiple candidate points that need to be validated (potentially increasing the needed computational budget) before the optimization task is finished.

5 Sequential sampling

Surrogate-based optimization has been a standard technology for long time [8, 32]. Traditionally, the surrogate replaces the expensive simulations in the computation of the objective function (and its gradient, if that is the case). The idea of using an stochastic processes for global optimization dates back to the seminal work by Kushner [125] (already coined as ‘‘Bayesian global optimization’’ or the ‘‘random function approach’’). Yet, Jones et al. [108] added a new twist by using both prediction and prediction variance of the Gaussian process (or kriging) model to help selecting the next point to be sampled in the optimization task. Therefore, instead of exhausting the budget for data acquisition upfront, practitioners can opt for fitting a sequence of surrogates with each surrogate defining the points that need to be sampled for the next surrogate. This can improve the accuracy for a given number of points, because points may be assigned to regions where the surrogate shows sign of poor accuracy. Alternatively, this approach may focus the sampling on regions of interest (e.g., targeting optimization or limit state estimation).

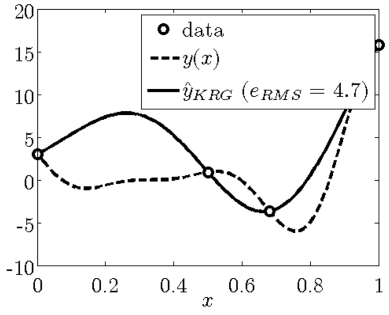
5.1 Basic idea

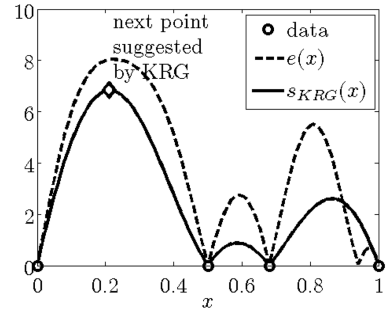
In the literature [102, 117, 197], the word ‘‘sequential’’ is sometimes substituted by ‘‘adaptive’’ or ‘‘application-driven,’’ and the word ‘‘sampling’’ is sometimes replaced by ‘‘experimental design,’’ or ‘‘design of experiment.’’ In the machine/deep learning literature, the term ‘‘Bayesian optimization’’ is also used to refer to the same idea [59, 148, 220].

The basic sequential sampling approach finds the point in the design space that maximizes the kriging prediction error (here, we use the square root of the kriging prediction variance). Fig. 6 illustrates the first cycle of the algorithm. Fig. 6a shows the initial kriging model and the corresponding prediction error. The maximization of the prediction error suggests adding $x = 0.21$ to the data set. The updated kriging model is shown in Fig. 6b. There is a substantial decrease in the root mean square error, from $e_{RMS} = 4.7$ to $e_{RMS} = 1.7$. Regions of high error estimates push exploration.

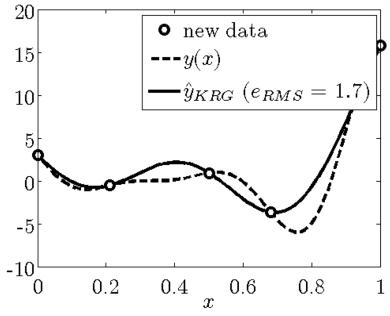
Literature on sequential sampling is vast. In this paper, we decide to discuss a sample of research works published in roughly the past 20 years.

Sample of papers published between 1998 and 2010
Papers addressed both fundamental formulation as well as presented case study compatible with the computer power available at the time. Some of the papers published in this era include:





(a) Maximizing $s_{KRG}(\mathbf{x})$.



(b) Updated KRG model.

Fig. 6: Basic sequential sampling. (a) shows that the maximization of kriging prediction standard deviation, $s_{KRG}(\mathbf{x})$, suggests adding $x = 0.21$ to the data set. (b) illustrates the updated kriging (KRG) model after $x = 0.21$ is added to the data set.

-- 1998 and 2001 - Jones et al. [108] proposed the formulation that used the uncertainty Gaussian process to guide the next set of simulations aiming at global optimization of expensive black-box functions. In a follow-up paper, Jones [106] presented an interesting taxonomy discussing advantages and disadvantages of different infill criteria.

-- 2000 - Audet et al. [6] used the expected violation, a concept similar to the expected improvement, to improve the accuracy of the surrogate along the boundaries of the feasible/unfeasible region. The approach was tested on analytical problems and a wing platform design problem.

-- 2002 - Jin et al. [102] reviewed various sequential sampling approaches (maximization of the prediction error, minimization of the integrated square error, maximization of the minimum distance, and cross validation) and compared them with simply filling of the original design (one stage approach). They found that the performance of the sequential sampling methods depended on the quality of the initial surrogate -- there is no guarantee that sequential sampling will do better than the one stage approach.

-- 2004 - Kleijnen and Van Beers [117] proposed an algorithm that, after the first model is fitted, iterates by placing a set of points using a space filling scheme and then choosing the one that maximizes the variance of the predicted output (variance of the responses taken from cross validation of the original data set). In a follow up paper, Van Beers and Kleijnen [197] improved their approach to account for noisy responses. In the works of Kleijnen and Van Beers, an improved kriging variance estimate [94] is used and that might be a reason for better results.

-- 2006 - Rai [159] introduced the qualitative and quantitative sequential sampling technique. The method combines information from multiple sources (including computer models and the designer's qualitative intuitions) through a criterion called ''confidence function.'' The capabilities of the approach were demonstrated using examples including the design of a bi-stable micro electro-mechanical system.

-- 2007 to 2009 - Ginsbourger et al. [71], Villemonteix et al. [211], and Queipo et al. [157] share the common point of proposing alternatives to the expected improvement for selection of points. Ginsbourger et al. [71] extended both the expected improvement and the probability of improvement as infill sampling criteria allowing for multiple points in each additional cycle. However, they also mention the high computational costs associated with this strategy. Villemonteix et al. [211] introduces a new criterion that they called ''conditional minimizers entropy'' with the advantage of being ready to use in noisy applications. Queipo et al. [157] focused on the assessment of the probability of being below a target value given that multiple points can be added in each optimization cycle (this strategy is more cost-efficient than the expected improvement counterpart).

-- 2007 - Turner et al. [194] proposed a heuristic scheme that samples multiple points at a time based on non uniform rational B-splines (NURBs). The candidate sites are generated by solving a multi-objective optimization problem. The effectiveness of the algorithm was demonstrated for five trial problems of engineering interest.

-- 2008 - Bichon et al. [16] proposed the efficient global reliability analysis (EGRA) algorithm. As opposed to global optimization, EGRA aims at improving the accuracy of the surrogate model near the limit state (target value). In the context of design under uncertainty, the approach is useful for estimation of the probability of failure. As the samples are added, the surrogate accuracy near the boundaries of limit state functions improves; and therefore, the accuracy of the probability of failure estimate also improves.
-- 2008 - Forrester and Keane [58] published one of the books on surrogate-based optimization that is heavily geared towards modern approaches to sequential sampling. This book is written by engineers for engineers and provides an extended and modern review, including topics such as constrained and multi-objective optimization.
-- 2009 - Gorissen et al. [86] brought multiple surrogates to adaptive sampling. The objective is to be able to select the best surrogate model by adding points iteratively. They tailored a genetic algorithm that combines automatic model type selection, automatic model parameter optimization, and sequential design exploration. They used a set of analytical functions and engineering examples to illustrate the methodology.
-- 2010 - Rennen et al. [165] proposed nested designs. The idea is that the low accuracy of a model obtained might justify the need of an extra set of function evaluations. They proposed an algorithm that expands an experimental design aiming maximization of space filling and non-collapsing points.

Sample of papers published between 2011 and 2021
In the past ten years, researchers kept the focus on fundamental issues as well as addressed very important practical aspects of sequential sampling. Some of the papers published in this era include:

-- 2011 - Echard et al. [52] proposed a sequential sampling procedure for the estimation of limit state functions in reliability analysis. While similar in spirit to EGRA [16] it proposes a different enrichment criterion (U criterion) and has the advantage of very simple integration in a Monte Carlo sampling framework.
-- 2011 - Bichon et al. [17] extended the efficient global reliability analysis algorithm to simultaneously handle multiple limit state functions (i.e., failure modes). Authors studied the performance of three different approaches (i) a Gaussian process model for each limit state function, (ii) a single Gaussian process model to capture the ''composite'' limit state, and finally (iii) a composite expected feasibility function. These approaches were tested against analytical problems as well as the design of a liquid hydrogen tank and a vehicle side impact problem.
-- 2012 - Bect et al. [11] expanded the sequential sampling method aiming directly at the probability of failure, as opposed to the limit state function. Authors derive a stepwise uncertainty reduction strategy and apply to analytical and structural reliability problems.
-- 2012 and 2013 - Viana et al. [207, 208] proposed using multiple surrogates that optimize the expected improvement and expected feasibility as a cheap and efficient way of generating multiple points in each additional cycle. The diversity of surrogate models drives the suggestion of multiple points per cycle and keeping the implementation simple (avoiding the computationally expensive implementations of multi-point versions of the probability of improvement or expected improvement).
-- 2014 - Zaefferer et al. [225] extended the formulation of the efficient global optimization algorithm to handle combinatorial problems. Their approach involved using permutation measures such as the Hamming distance and others, as opposed to simple Euclidean distance. Authors tested their approach using a collection of analytical problems.
-- 2015 - Hu and Du [96] demonstrated the use of sequential sampling for reliability analysis with time-dependent responses. Authors used a Gaussian process that models the output of interest as a function of input variables and time and discuss issues related to the initial data set as well as uncertainty associated with the computation of the time-dependent probability of failure.
-- 2016 - Haftka et al. [91] and Li et al. [133] published almost simultaneously papers where they studied parallelization of sequential sampling for global optimization. Haftka et al. [91] presented a formal literature survey where authors discussed exploration versus exploitation, strategies based on one or multiple surrogate models, and evolutionary approaches. Li et al. [133] presented a comparative study of parallel version of sequential sampling involving multi-point sampling criteria and domain decomposition. They discussed infill criteria such as the expected improvement and mutual information. Their numerical study included analytical problems, looking into convergence and robustness, as well as an engineering application (air conditioner cover design).
-- 2017 - Rana et al. [162] proposed an algorithm to address the curse of dimensionality in sequential sampling using the standard implementation of the

Gaussian process. In high-dimensions different infill criteria can invariably show only few peaks surrounded by almost flat surface, which makes their optimization a very hard problem. The proposed method builds on two assumptions (i) when the Gaussian process length-scales (i.e. correlation length) are large enough, they can make the gradient of infill criteria useful, (ii) extrema of consecutive acquisition functions are close if the difference in the used length-scales is small. The algorithm switches between finding ideal length scales and optimizing the infill criterion. Authors tested their approach on analytical and engineering problems ranging from 6 to 50 input variables.

-- 2018 - Chaudhuri et al. [28] presented an approach for analysis of feedback-coupled systems (fire detection satellite model and aerostructural analysis of a wing) based on sequential sampling so that the computational cost is minimized. They used an information-gain-based and a residual-error-based adaptive sampling strategy in their work.

-- 2019 - Bartoli et al. [9] proposed a framework to handle nonlinear equality or inequality constraints while performing global optimization using sequential sampling. Their approach builds on mixture of experts and augmented infill criteria (using a combination of expected improvement and the estimated response given by the surrogate). Using both analytical problems and engineering examples, authors compared the results of their approach with other optimization algorithms using criteria such as percentage of converged runs, mean number of function evaluations for converged runs, and standard deviation of the number of function evaluations for converged runs.

-- 2020 - Knudde et al. [120] implemented sequential sampling using deep Gaussian process models. The claimed advantage of deep Gaussian process is their ability to model highly non-linear and nonstationary responses. Authors compared regular Gaussian process, sparse Gaussian process, deep Gaussian process, and Gaussian process models with nonstationary kernels against an analytical problem, as well as the Langley Glide-Back Booster problem by NASA [150] and an electronic amplifier design.

-- 2021 - Beek et al. [12] proposed a batch sampling scheme for optimization of expensive computer models with spatially varying noise. Authors reformulated the posterior prediction variance to account for the nonstationary noise. The approach is empirically demonstrated on a set of analytical problems as well as the design of an organic photovoltaic cell.

5.2 A practical view on sequential sampling

Understanding two important trade-offs - Surrogate models are usually employed on problems where data is computationally expensive to obtain. It is very convenient to think that given a budget, defined in terms of number of data points, one can go about acquiring data, fitting one or more surrogate models, and then build a model that will replace all the expensive simulations in tasks such as design space exploration, global sensitivity analysis, uncertainty quantification, optimization, etc. The first trade-off one needs to understand when using sequential sampling is that goal justifies the means that data is obtained. In practical terms, if the goal is to perform global optimization, one will use metrics such as the expected improvement to sequentially gather data and eventually find the global optimum. Since the surrogate model is just a mean to achieve the end goal, efficiency gains most likely come at the cost of a highly specialized final surrogate model.

The second important trade-off to understand is the compromise between exploration and exploitation. This is very well explained in [106]. Essentially, even though points are ideally added to achieve a specific goal (exploitation), some of these points are inevitably added to compensate for surrogate uncertainty (exploration).

Formulating the sequential sampling problem - While sequential sampling can provide efficiency gains in terms of number of data points and wall-clock time, it comes at the cost of having to pre-define the intended task. Table 2 summarizes some of the important aspects to consider. Obviously, the first one is the goal of the task itself, which is then used to define the sampling criteria. The number of points added per cycle can influence the sampling criteria as well and it is usually a decision that is application dependent. For example, in simulation-based design optimization, it is common to run multiple simulations at once using parallel computing. The choice of surrogate models to be used depends on the quality of the available uncertainty estimates, but can also be related to the number of points per cycle. After all, each surrogate can provide one or more multiple points. Last, but not least, the fidelity level of the data is important, as it defines the one more level of complexity in the surrogate modeling. Multi-fidelity approaches are available and detailed in [28, 70, 229].

Choosing an infill (sampling) criteria - In its standard form, the efficient global optimization (EGO) algorithm starts by fitting a Gaussian process (kriging) model for the initial set of data points. After that, the algorithm iteratively adds points to the data set in an effort to

Table 2: Points to consider in sequential sampling

Feature	Common options
End goal	Surrogate accuracy / Global optimization / Limit state estimation
Points per cycle	Single / multiple
Surrogate modeling	Single / multiple
Fidelity of data	Single / multiple (variable)

improve upon the present best sample y_{PBS}. In each cycle, the next point to be sampled is the one that maximizes the expected improvement

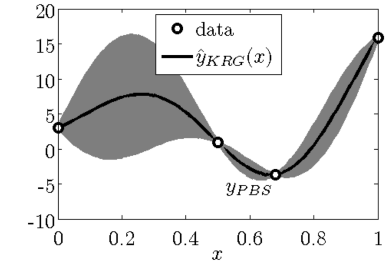
$$EI(\mathbf{x}) = \mathbb{E}\left[I(\mathbf{x})\right] = s(\mathbf{x})\left[u\Phi(u) + \phi(u)\right] \ , \\ u = \left[y_{PBS} - \hat{y}(\mathbf{x})\right]/s(\mathbf{x}) \ , \tag{6}$$

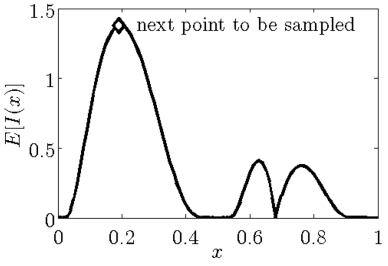
where $\Phi(.)$ and $\phi(.)$ are the cumulative density function (CDF) and probability density function (PDF) of a normal distribution, $\hat{y}(\mathbf{x})$ is the kriging prediction; and $s(\mathbf{x})$ is the prediction standard deviation -- here estimated as the square root of the prediction variance $s^2(\mathbf{x})$.

Unlike methods that only look for the optimum predicted by the surrogate, EGO will also favor points where surrogate predictions have high uncertainty. After adding the new point to the existing data set, the kriging model is updated (usually without going to the costly optimization of the correlation parameters). Fig. 7 illustrates the one cycle of the EGO algorithm. Fig. 7a shows the initial kriging model and the corresponding expected improvement. The maximization of $E[I(\mathbf{x})]$ adds $x = 0.19$ to the data set. In the next cycle, EGO uses the updated kriging model shown in Fig. 7b.

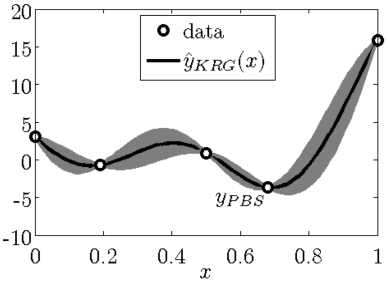
Table 3 summarizes the **basic** versions of popular infill criteria used for different goals. The interested reader can find discussion on other, and potentially more sophisticated, infill criteria in [9, 90, 91, 133, 151, 154, 169]. There are also variations for "noisy" data (coming from stochastic simulators or physical experiments), as discussed in [57, 129, 154, 200]. For the most part, implementation of sequential sampling adding one point at a time should be relatively straightforward; except for the fact that the infill criteria is often challenging to be optimized. As illustrated in Figs. 6 and 7, the infill criteria used in sequential sampling can present several local optima. Therefore, in practical terms, optimization is carried over with specialized implementations of optimization algorithms [107, 182].

Other important remarks - Research on concurrent implementations of sequential sampling has extended infill criteria, such as the probability of improvement and the expected improvement, to versions that return the infill criteria when multiple points are added to the data set (rather than a single point) [93, 155]. However, maximizing the multiple point infill criteria is computationally challenging [72]. Alternatively, there are approximations and heuristics that can be used to reduce the computational cost. For example, one can approximate the multi-point probability by neglecting the correlation (assuming that points would be far enough). In that case, the single point probability can be used as a computationally efficient approximation. A heuristic called kriging-believer [72] has been extensively used. After selecting a new sample, kriging is updated using the estimated value as if it were data. The process is repeated until one obtains the desired number of samples. Alternatively, multiple surrogates and ensembles can also be used to suggest multiple points per cycle [86, 208]. The interested reader can find survey and dedicated studies on parallelization of sequential sampling for global optimization in [91, 133].





(a) Maximizing $E[I(\mathbf{x})]$.



(b) Updated KRG model.

Fig. 7: Cycle of the efficient global optimization (EGO) algorithm. (a) shows that the maximization of the expected improvement, $E[I(\mathbf{x})]$, suggests adding $x = 0.19$ to the data set. (b) illustrates the updated kriging (KRG) model after $x = 0.19$ is added to the data set.

Table 3: Overview of a subset of initially developed infill criteria. $\Phi(.)$ and $\phi(.)$ are the cumulative density function and probability density function of a normal distribution. $\hat{y}(\mathbf{x})$ and $s(\mathbf{x})$ are the Gaussian process prediction mean and standard deviation. Here, global optimization targets finding the minimum of a function.

Goal	Criteria	Definition	Comments
Surrogate accuracy	Entropy [45]	$\lvert\mathbf{R}\rvert \lvert\mathbf{J}^T\mathbf{R}^{-1}\mathbf{J}\rvert$	$\mathbf{R}$ is the correlation matrix of all the $n+m$ points (n are points already sampled and m are points to be sampled) and $\mathbf{J}$ is a vector of 1's. This is a simple metric to compute; although it can be come expensive when m is large.
	Integrated mean squared error [171]	$-\int s^2(\mathbf{x})d\mathbf{x}$	Very elegant but computationally intractable.
	Prediction uncertainty	$s^2(\mathbf{x})$	This implementation is simplistic but can exhibit slow convergence.
Global optimization [106]	Statistical lower bound	$\hat{y}(\mathbf{x}) - \kappa s(\mathbf{x})$	Parameter κ is an arbitrary positive number. This implementation is simplistic but lacks the balance between exploration and exploitation.
	Probability of improvement	$\Phi\left(\frac{y_T-\hat{y}(\mathbf{x})}{s(\mathbf{x})}\right)$	y_T is a user-defined target value. The choice of y_T is arbitrary and can influence the performance of the criterion.
	Expected improvement	$s(\mathbf{x})\left[u\Phi(u)+\phi(u)\right]$	$u = [y_{PBS} - \hat{y}(\mathbf{x})]/s(\mathbf{x})$ and y_{PBS} is the present best solution. Improvement is defined as $I = \max[y_{PBS} - y((x)), 0]$. This criterion has become the gold standard due to its simplicity and robustness.
Contour / limit state estimation	Expected feasibility [16]	$(\hat{y}(\mathbf{x}) - \bar{y})\left[2\Phi(u) - \Phi(u^+) - \Phi(u^-)\right] - s(\mathbf{x})\left[2\phi(u) - \phi(u^+) - \phi(u^-)\right] + \epsilon\left[\Phi(u^+) - \Phi(u^-)\right]$	$\bar{y}$ is the target value for the limit state function, $u = \frac{\bar{y}-\hat{y}(\mathbf{x})}{s(\mathbf{x})}$, $u^+ = \frac{\bar{y}+\epsilon-\hat{y}(\mathbf{x})}{s(\mathbf{x})}$, $u^- = \frac{\bar{y}-\epsilon-\hat{y}(\mathbf{x})}{s(\mathbf{x})}$, $\epsilon = \alpha s(\mathbf{x})$, and usually $\alpha = 2$. This is the equivalent of expected improvement for the case of contour estimation.
	Modified expected improvement [163]	$[\alpha^2 s^2(\mathbf{x}) - (\hat{y}(\mathbf{x}) - \bar{y})^2]\left[\Phi(u+\alpha) - \Phi(u-\alpha)\right] + 2(\hat{y}(\mathbf{x}) - \bar{y})s^2(\mathbf{x})\left[\phi(u+\alpha) - \phi(u-\alpha)\right] - \int_{\bar{y}-\alpha s(\mathbf{x})}^{\bar{y}+\alpha s(\mathbf{x})}(y - \hat{y}(\mathbf{x}))^2\phi(u)dy$	This comes after redefining improvement as $I = \alpha s(\mathbf{x}) - \min\left[(y(\mathbf{x}) - \bar{y})^2, \alpha^2 s^2(\mathbf{x})\right]$.
	Weighted integrated mean squared error [153]	$-\int s^2(\mathbf{x})W(\mathbf{x})d\mathbf{x}$	$W(\mathbf{x}) = \Phi(u^+) - \Phi(u^-)$, $u^+ = \frac{\bar{y}+\epsilon-\hat{y}(\mathbf{x})}{s(\mathbf{x})}$, $u^- = \frac{\bar{y}-\epsilon-\hat{y}(\mathbf{x})}{s(\mathbf{x})}$. This is a modified version of the classical integrated mean square error criterion by weighting the prediction variance with the expected proximity to the target level of response.

Besides global optimization or contour estimation, practitioners very often deal with constrained and multi-objective optimization. [151] discussed how to use the expected improvement of the objective function and the probability of feasibility calculated from the constraint functions to handle constrained global optimization. Durantin et al. [50] presented a friendly discussion on sequential sampling for multi-objective optimization problems including adaptations for multi-objective constrained optimization. Examples of research that furthered the treatment of constrains and multiple objectives include, but are not limited to, [9, 43, 90]. Bartoli et al. [9] proposed a framework to handle nonlinear equality or inequality constraints while performing global optimization of multimodal functions. Their approach is based on a robust implementation of the Gaussian process (with conditioning of the correlation matrix to support large number of design variable), a modified expected improvement that is less multimodal than the original version, and mixture of experts (a form of ensemble). Aware of the potential challenges associated with the computational cost of using the probability of improvement and/or the expected improvement for multi-objective problems, Couckuyt et al. [43] proposed

a hypervolume-based improvement function coined for multi-objective problems. Their approach also included tracking changes in the Pareto front, and decomposing the objective space. Guo et al. [90] explicitly proposed using ensemble of surrogates for managing multiple objectives in sequential sampling. They used support vector machines, radial basis functions, and Gaussian process with the argument that using ensembles is an scalable way to achieve diversity, which is important as the complexity of the problem grows.

In many real-world applications, engineers and scientists are likely to face problems involving multiple responses and/or multi-fidelity models. Liu et al. [136] proposed an approach for handling sequential sampling when dealing with multi-fidelity models that is based on Voronoi tessellation. Using a combination of Voronoi tessellation for region division, Pearson correlation coefficients and cross validation analysis, authors determine the candidate region for infilling a new sample. They demonstrated the success of their approach using well-kown set of analytical test functions as well as the design optimization of a hoist sheave (based on finite element modeling). Khatamsaz et al. [112] exploits the structure of the Gaussian process to perform model fusion and expected hypervolume improvement to perform efficient multi-objective optimization when multiple sources of information are available. After performing an study with analytical functions, authors applied their framework to a coupled aerostructural wing design optimization problem. Results were compared to ParEGO and NSGA-II, two other well-known multi-objective optimization approaches. Chaudhuri et al. [29] extended the EGRA algorithm to multifidelity problems. Authors proposed a two-stage sequential sampling approach combining expected feasibility and one-step lookahead information gain, where multi-fidelity modeling is handled by Gaussian processes. Numerical study includes analytical functions and the reliability analysis of an acoustic horn

Finally, sequential sampling techniques are very important for high-dimensional, computationally expensive, and black-box problems. However, efficient implementation often finds challenges associated with the curse of dimensionality, such as the relative distance between the points, the need for large datasets, etc. Besides, the Gaussian process itself suffers from problems such as ill-conditioning of the covariance matrix and convergence to mean prediction and saturation of uncertainty estimates. These challenges have motivated research in the field, and we refer the interested readers to [131, 162, 216] for further details.

5.3 Software

Currently, there exists a number of commercial software an dedicated packages (in many programming languages) that implement a variety of surrogate modeling techniques. Sometimes surrogate modeling is not a final goal; but rather, it is a companion to optimization and design exploration capabilities. Table 4 lists a few software and packages that have implemented, integrated, or adapted, in one way or another, sequential sampling using Gaussian process in the form discussed in this paper. We do not claim this to be a comprehensive list, although it shows the diversity in terms of commercial/industrial software and open source packages.

6 Concluding Remarks and Future Research

In this paper, we have summarized four methodologies that allow smarter use of the sampled data and surrogate modeling. We discussed (i) screening and variable reduction, (ii) design of experiments, (iii) surrogate modeling and ensembles, and (iv) sequential sampling. Based on our understanding and experience we can say that:

- **Screening and variable reduction:** is an efficient step for reducing the cost of the surrogate's construction, with drastic dimensionality reductions being possible. Some approaches such as non-dimensionalization or principal component regression can at the same time improve the accuracy of the approximations.
- **Design of experiments:** is extremely useful to optimally sample the design space such that we maximize the quality of surrogate model using the knowledge about the problem while working within the budgetary constraints in gathering the data.
- **Surrogate modeling and ensembles:** is attractive because no single surrogate works well for all problems and the cost of constructing multiple surrogates is often small compared to the cost of simulations.
- **Sequential sampling:** is an efficient way of making use of limited computational budget. Techniques make use of both the prediction and the uncertainty estimates of the surrogate models to intelligently sample the design space.

With this paper, we hope to have summarized practices that have endured the test of time. The research in each of the topics we discussed has been very active and we would like to contribute by mentioning potential topics of future research regarding:

- **Screening and variable reduction:**

Table 4: Short list of software and openly available packages for sequential sampling.

Software	**Link**	
DAKOTA [2]	`https://dakota.sandia.gov`	
OpenMDAO [89]	`https://openmdao.org/`	
TOMLAB [53]	`https://tomopt.com/tomlab`	
GEBHM [124]	Proprietary	
Package	**Language**	**Link**
GPFlow [62]	Python	`https://github.com/GPflow/GPflow`
GPyTorch [63]	Python	`https://github.com/cornellius-gp/gpytorch`
SEPIA [67]	Python	`https://github.com/lanl/SEPIA`
GPyOpt [119]	Python	`https://github.com/SheffieldML/GPyOpt`
pyGPGO [100]	Python	`https://pygpgo.readthedocs.io`
SMT [21]	Python	`https://smt.readthedocs.io/en/latest/`
GPMSA [68]	Matlab	`https://github.com/lanl/GPMSA`
DACE [137]	Matlab	`http://www.omicron.dk/dace`
SURROGATES Toolbox [202]	Matlab	`https://sites.google.com/site/srgtstoolbox`
SUMO Toolbox [87]	Matlab	`http://www.sumowiki.intec.ugent.be/Main_Page`
DiceKriging and DiceOptim [170]	R	`https://cran.r-project.org/web/packages/DiceOptim`
laGP [88]	R	`https://cran.r-project.org/web/packages/laGP`
mlrMBO [18]	R	`https://cran.r-project.org/web/packages/mlrMBO`
Metrics Optimization Engine [37]	C++	`https://github.com/Yelp/MOE`

a) automatic construction of most appropriate variable transformation,
b) efficient construction of most relevant embedded subspace for variable reduction or dimensionality reduction, and
c) combination of dimensionality reduction, sensitivity analysis, and sequential sampling techniques to enhance efficiency.

-- Design of experiments:

a) framework for the use of appropriate criteria for the construction of DOEs,
b) optimal DOEs suitable to a large class of surrogate techniques, and
c) budget allocation in sequential DOE sampling.

-- Surrogate modeling and ensembles:

a) resource allocation of simulators with tunable fidelity (unified framework as opposed to adapted ideas from multi-fidelity modeling),
b) investigation of the benefits multiple surrogates on sequential sampling and reliability-based optimization, and
c) visualization and design space exploration (since different surrogates might be more accurate in different regions of the design space).

-- Sequential sampling:

a) development of variable fidelity approaches,
b) measurement (and reduction) of the influence of the surrogate accuracy on the method, and
c) combined use of space-filling and adapted strategies for increased robustness.

Finally, complexity in terms of numerical implementation and optimization and, in some cases, the small commercial software footprint may hinder these techniques from popularity in the near term. Therefore, we believe that the investment in packages and learning tools together with ongoing scientific investigation can continue to be very beneficial.

Acknowledgements This paper is dedicated to Dr. Raphael T. Haftka. All three authors of this paper had the honor to have Dr. Haftka as their doctoral supervisor. Dr. Haftka's legacy in multidisciplinary optimization, surrogate modeling, uncertainty quantification, structural analysis, and other engineering disciplines is everlasting. His influence in the education of scientists and engineers will remain alive for many years to come.

Conflict of interest

The authors declare that they have no known competing financial interests, personal relationships, or any other conflicts of interest that could have appeared to influence the work reported in this paper.

Replication of results

This is a literature review paper. Therefore, we do not provide results that could be replicated.

References

1. Acar E, Rais-Rohani M (2009) Ensemble of metamodels with optimized weight factors. Structural

and Multidisciplinary Optimization 37(3):279--294, DOI 10.1007/s00158-008-0230-y

2. Adams BM, Bohnhoff WJ, Dalbey KR, Ebeida MS, Eddy JP, Eldred MS, Hooper RW, Hough PD, Hu KT, Jakeman JD, Khalil M, Maupin KA, Monschke JA, Ridgway EM, Rushdi AA, Seidl DT, Stephens JA, Swiler LP, Winokur JG (2020) Dakota, a multilevel parallel object-oriented framework for design optimization, parameter estimation, uncertainty quantification, and sensitivity analysis: Version 6.12 user's manual. Technical Report SAND2020-12495, Sandia National Laboratories
3. Agostinelli C (2002) Robust stepwise regression. Journal of Applied Statistics 29(6):825--840, DOI 10.1080/02664760220136168, URL https://doi.org/10.1080/02664760220136168
4. Andres TH, Hajas WC (1993) Using iterated fractional factorial design to screen parameters in sensitivity analysis of a probabilistic risk assessment model. In: Mathematical Methods and Supercomputing in Nuclear Applications, Germany, URL http://inis.iaea.org/search/search.aspx?orig_q=RN:25062495
5. Arnst M, Soize C, Bulthuis K (2020) Computation of sobol indices in global sensitivity analysis from small data sets by probabilistic learning on manifolds. International Journal for Uncertainty Quantification DOI 10.1615/int.j.uncertaintyquantification.2020032674, URL https://doi.org/10.1615/int.j.uncertaintyquantification.2020032674
6. Audet C, Dennis J, Moore D, Booker A, Frank P (2000) A surrogate-model-based method for constrained optimization. In: 8th AIAA/NASA/USAF/ISSMO Symposium on Multidisciplinary Analysis and Optimization, Long Beach, USA, DOI 10.2514/6.2000-4891
7. Ballester-Ripoll R, Paredes EG, Pajarola R (2019) Sobol tensor trains for global sensitivity analysis. Reliability Engineering & System Safety 183:311--322, DOI https://doi.org/10.1016/j.ress.2018.11.007, URL https://www.sciencedirect.com/science/article/pii/S0951832018303132
8. Barthelemy JFM, Haftka RT (1993) Approximation concepts for optimum structural design -- a review. Structural and Multidisciplinary Optimization 5(3):129--144, DOI 10.1007/BF01743349, URL https://doi.org/10.1007/BF01743349
9. Bartoli N, Lefebvre T, Dubreuil S, Olivanti R, Priem R, Bons N, Martins J, Morlier J (2019) Adaptive modeling strategy for constrained global optimization with application to aerodynamic wing design. Aerospace Science and Technology 90:85--102, DOI https://doi.org/10.1016/j.ast.2019.03.041, URL https://www.sciencedirect.com/science/article/pii/S1270963818306011
10. Beale EML, Kendall MG, Mann DW (1967) The discarding of variables in multivariate analysis. Biometrika 54(3-4):357--366, DOI 10.1093/biomet/54.3-4.357, URL https://doi.org/10.1093/biomet/54.3-4.357
11. Bect J, Ginsbourger D, Li L, Picheny V, Vazquez E (2012) Sequential design of computer experiments for the estimation of a probability of failure. Statistics and Computing 22(3):773--793, DOI 10.1007/s11222-011-9241-4, URL https://doi.org/10.1007/s11222-011-9241-4
12. van Beek A, Ghumman UF, Munshi J, Tao S, Chien T, Balasubramanian G, Plumlee M, Apley D, Chen W (2021) Scalable adaptive batch sampling in simulation-based design with heteroscedastic noise. Journal of Mechanical Design 143(3):031709 (15 pages), DOI 10.1115/1.4049134, URL https://doi.org/10.1115/1.4049134
13. Belsley D, Kuh E, Welsch R (2004) Regression diagnostics: Identifying influential data and sources of collinearity. Wiley-IEEE, New York
14. Berthelin G, Dubreuil S, Salaün M, Gogu C, Bartoli N (2020) Model order reduction coupled with efficient global optimization for multidisciplinary optimization. In: 14th World Congress on Computational Mechanics (WCCM)
15. Bettonvil B (1990) Detection of Important Factors by Sequential Bifurcation. Tilburg University Press, Tilburg
16. Bichon BJ, Eldred MS, Swiler LP, Mahadevan S, McFarland JM (2008) Efficient global reliability analysis for nonlinear implicit performance functions. AIAA Journal 46(10):2459--2468, DOI 10.2514/1.34321, URL http://dx.doi.org/10.2514/1.34321
17. Bichon BJ, McFarland JM, Mahadevan S (2011) Efficient surrogate models for reliability analysis of systems with multiple failure modes. Reliability Engineering & System Safety 96(10):1386--1395, DOI https://doi.org/10.1016/j.ress.2011.05.008, URL https://www.sciencedirect.com/science/article/pii/S0951832011001062
18. Bischl B, Richter J, Bossek J, Horn D, Thomas J, Lang M (2017) mlrMBO: A Modular Framework for Model-Based Optimization of Expensive Black-Box Functions. arXiv preprint arXiv:170303373
19. Borgonovo E, Plischke E (2016) Sensitivity analysis: A review of recent advances. European Journal of Operational Research 248(3):869-

-887, DOI https://doi.org/10.1016/j.ejor.2015.06.032, URL https://www.sciencedirect.com/science/article/pii/S0377221715005469
20. Bouhlel MA, Bartoli N, Otsmane A, Morlier J (2016) Improving kriging surrogates of high-dimensional design models by partial least squares dimension reduction. Structural and Multidisciplinary Optimization 53(5):935--952, DOI 10.1007/s00158-015-1395-9, URL https://doi.org/10.1007/s00158-015-1395-9
21. Bouhlel MA, Hwang JT, Bartoli N, Lafage R, Morlier J, Martins JRRA (2019) A python surrogate modeling framework with derivatives. Advances in Engineering Software 135:102662, DOI https://doi.org/10.1016/j.advengsoft.2019.03.005, URL https://www.sciencedirect.com/science/article/pii/S0965997818309360
22. Box GEP, Tidwell PW (1962) Transformation of the independent variables. Technometrics 4(4):531--550, DOI 10.1080/00401706.1962.10490038, URL https://www.tandfonline.com/doi/abs/10.1080/00401706.1962.10490038
23. Box GEP, Hunter JS, Hunter WG (1978) Statistics for experimenters: an introduction to design, data analysis, and model building. John Wiley & Sons, Inc, New York, USA
24. Breitkopf P, Coelho RF (2013) Multidisciplinary design optimization in computational mechanics. John Wiley & Sons
25. Buckingham E (1914) On physically similar systems; illustrations of the use of dimensional equations. Physical Review 4(4):345--376, DOI 10.1103/PhysRev.4.345, URL https://doi.org/10.1103/PhysRev.4.345
26. Budinger M, Passieux JC, Gogu C, Fraj A (2014) Scaling-law-based metamodels for the sizing of mechatronic systems. Mechatronics 24(7):775--787, DOI https://doi.org/10.1016/j.mechatronics.2013.11.012, URL https://www.sciencedirect.com/science/article/pii/S0957415813002328
27. Calandra R, Peters J, Rasmussen CE, Deisenroth MP (2016) Manifold gaussian processes for regression. In: 2016 International Joint Conference on Neural Networks (IJCNN), IEEE, Vancouver, Canada, pp 3338--3345, DOI 10.1109/IJCNN.2016.7727626
28. Chaudhuri A, Lam R, Willcox K (2018) Multifidelity uncertainty propagation via adaptive surrogates in coupled multidisciplinary systems. AIAA Journal 56(1):235--249, DOI 10.2514/1.J055678, URL https://doi.org/10.2514/1.J055678
29. Chaudhuri A, Marques AN, Willcox K (2021) mfEGRA: Multifidelity efficient global reliability analysis through active learning for failure boundary location. Structural and Multidisciplinary Optimization DOI 10.1007/s00158-021-02892-5, URL https://doi.org/10.1007/s00158-021-02892-5
30. Chen VCP, Tsui KL, Barton RR, Meckesheimer M (2006) A review on design, modeling and applications of computer experiments. IIE Transactions 38(4):273--291, DOI 10.1080/07408170500232495, URL https://doi.org/10.1080/07408170500232495
31. Cheng B, Titterington DM (1994) Neural networks: a review from a statistical perspective. Statistical Science 9(1):2--54, DOI 10.1214/ss/1177010638
32. Chernoff H (1959) Sequential design of experiments. The Annals of Mathematical Statistics 30(3):755--770, URL http://www.jstor.org/stable/2237415
33. Cho YC, Jayaraman B, Viana FAC, Haftka RT, Shyy W (2010) Surrogate modelling for characterising the performance of a dielectric barrier discharge plasma actuator. International Journal of Computational Fluid Dynamics 24(7):281--301, DOI 10.1080/10618562.2010.521129, URL http://dx.doi.org/10.1080/10618562.2010.521129
34. Christopher Frey H, Patil SR (2002) Identification and review of sensitivity analysis methods. Risk Analysis 22(3):553--578, DOI 10.1111/0272-4332.00039, URL https://onlinelibrary.wiley.com/doi/abs/10.1111/0272-4332.00039
35. Chun H, Keleş S (2010) Sparse partial least squares regression for simultaneous dimension reduction and variable selection. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 72(1):3--25, DOI https://doi.org/10.1111/j.1467-9868.2009.00723.x, URL https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2009.00723.x
36. Cibin R, Sudheer KP, Chaubey I (2010) Sensitivity and identifiability of stream flow generation parameters of the swat model. Hydrological Processes 24(9):1133--1148, DOI https://doi.org/10.1002/hyp.7568, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/hyp.7568
37. Clark SC, Liu E, Frazier PI, Wang J, Oktay D, Vesdapunt N (2014) Metrics optimization engine. URL http://yelp.github.io/MOE
38. Coatanéa E, Nonsiri S, Ritola T, Tumer IY, Jensen DC (2011) A framework for building dimensionless behavioral models to aid in function-based failure propagation analysis. Journal of Mechanical Design 133(12), DOI 10.1115/1.4005230, URL https://doi.org/10.1115/1.4005230

39. Coelho RF, Breitkopf P, Knopf-Lenoir C (2008) Model reduction for multidisciplinary optimization - application to a 2d wing. Structural and Multidisciplinary Optimization 37(1):29--48, DOI 10.1007/s00158-007-0212-5, URL https://doi.org/10.1007/s00158-007-0212-5
40. Coelho RF, Breitkopf P, Knopf-Lenoir C, Villon P (2008) Bi-level model reduction for coupled problems. Structural and Multidisciplinary Optimization 39(4):401--418, DOI 10.1007/s00158-008-0335-3, URL https://doi.org/10.1007/s00158-008-0335-3
41. Conti S, O'Hagan A (2010) Bayesian emulation of complex multi-output and dynamic computer models. Journal of Statistical Planning and Inference 140(3):640--651, DOI https://doi.org/10.1016/j.jspi.2009.08.006, URL https://www.sciencedirect.com/science/article/pii/S0378375809002559
42. Cotter SC (1979) A screening design for factorial experiments with interactions. Biometrika 66(2):317--320, DOI 10.1093/biomet/66.2.317, URL https://doi.org/10.1093/biomet/66.2.317
43. Couckuyt I, Deschrijver D, Dhaene T (2013) Fast calculation of multiobjective probability of improvement and expected improvement criteria for pareto optimization. Journal of Global Optimization 60(3):575--594, DOI 10.1007/s10898-013-0118-2, URL https://doi.org/10.1007/s10898-013-0118-2
44. Crombecq K, Gorissen D, Deschrijver D, Dhaene T (2011) A novel hybrid sequential design strategy for global surrogate modeling of computer experiments. SIAM Journal on Scientific Computing 33(4):1948--1974, DOI 10.1137/090761811, URL https://doi.org/10.1137/090761811, https://doi.org/10.1137/090761811
45. Currin C, Mitchell T, Morris M, Ylvisaker D (1991) Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. Journal of the American Statistical Association 86(416):953--963, DOI 10.1080/01621459.1991.10475138, URL https://doi.org/10.1080/01621459.1991.10475138
46. Damianou A, Lawrence ND (2013) Deep Gaussian processes. In: Carvalho CM, Ravikumar P (eds) Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics, PMLR, Scottsdale, Arizona, USA, Proceedings of Machine Learning Research, vol 31, pp 207--215, URL http://proceedings.mlr.press/v31/damianou13a.html
47. Daniel C (1973) One-at-a-time plans. Journal of the American Statistical Association 68(342):353–360, DOI 10.1080/01621459.1973.10482433, URL https://doi.org/10.1080/01621459.1973.10482433
48. Doebling S, Schultze J, Hemez F (2002) Overview of structural dynamics model validation activities at los alamos national laboratory. In: 43rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, American Institute of Aeronautics and Astronautics, Denver, CO, DOI 10.2514/6.2002-1643, URL https://doi.org/10.2514/6.2002-1643
49. Dovi VG, Reverberi AP, Maga L, De Marchi G (1991) Improving the statistical accuracy of dimensional analysis correlations for precise coefficient estimation and optimal design of experiments. International Communications in Heat and Mass Transfer 18(4):581--590, DOI https://doi.org/10.1016/0735-1933(91)90071-B, URL https://www.sciencedirect.com/science/article/pii/073519339190071B
50. Durantin C, Marzat J, Balesdent M (2016) Analysis of multi-objective kriging-based methods for constrained global optimization. Computational Optimization and Applications 63(3):903--926, DOI 10.1007/s10589-015-9789-6, URL https://doi.org/10.1007/s10589-015-9789-6
51. Duvenaud D, Lloyd J, Grosse R, Tenenbaum J, Zoubin G (2013) Structure discovery in nonparametric regression through compositional kernel search. In: Dasgupta S, McAllester D (eds) Proceedings of the 30th International Conference on Machine Learning, PMLR, Atlanta, Georgia, USA, Proceedings of Machine Learning Research, vol 28, pp 1166--1174, URL http://proceedings.mlr.press/v28/duvenaud13.html
52. Echard B, Gayton N, Lemaire M (2011) Ak-mcs: an active learning reliability method combining kriging and monte carlo simulation. Structural Safety 33(2):145--154
53. Edvall1 MM, Göran A, Holmström K (2020) TOMLAB - Version 8.7. Vallentuna, Sweden, URL https://tomopt.com/docs/tomlab/tomlab.php
54. Efroymson MA (1960) Multiple regression analysis. In: Ralston A, Wilf H (eds) In Mathematical Methods for Digital Computers, John Wiley, New York, pp 191--203
55. Flury B (1988) Common principal components and related multivariate models. Wiley, New York
56. Forrester AI, Keane AJ (2009) Recent advances in surrogate-based optimization. Progress in Aerospace Sciences 45(1):50--79,

DOI 10.1016/j.paerosci.2008.11.001, URL https://www.sciencedirect.com/science/article/pii/S0376042108000766
57. Forrester AIJ, Keane AJ, Bressloff NW (2006) Design and analysis of "noisy" computer experiments. AIAA Journal 44(10):2331--2339, DOI 10.2514/1.20068, URL https://doi.org/10.2514/1.20068
58. Forrester AIJ, Sobester A, Keane AJ (2008) Engineering design via surrogate modelling: a practical guide. John Wiley & Sons, Ltd, Chichester, UK
59. Frazier PI (2018) Bayesian optimization. In: Recent Advances in Optimization and Modeling of Contemporary Problems, INFORMS, pp 255--278, DOI 10.1287/educ.2018.0188, URL https://doi.org/10.1287/educ.2018.0188
60. Fricker TE, Oakley JE, Urban NM (2012) Multivariate Gaussian process emulators with nonseparable covariance structures. Technometrics 55(1):47--56, DOI 10.1080/00401706.2012.715835, URL https://doi.org/10.1080%2F00401706.2012.715835
61. Frits A, Reynolds K, Weston N, Mavris D (2004) Benefits of non-dimensionalization in creation of designs of experiments for sizing torpedo systems. In: 10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, American Institute of Aeronautics and Astronautics, Albany, USA, pp AIAA 2004--4491, DOI 10.2514/6.2004-4491, URL https://doi.org/10.2514/6.2004-4491
62. de G Matthews AG, van der Wilk M, Nickson T, Fujii K, Boukouvalas A, León-Villagrá P, Ghahramani Z, Hensman J (2017) Gpflow: A gaussian process library using tensorflow. Journal of Machine Learning Research 18(40):1--6, URL http://jmlr.org/papers/v18/16-537.html
63. Gardner JR, Pleiss G, Bindel D, Weinberger KQ, Wilson AG (2018) Gpytorch: Blackbox matrix-matrix Gaussian process inference with GPU acceleration. arXiv preprint arXiv:180911165
64. Garnett R, Osborne MA, Hennig P (2014) Active learning of linear embeddings for gaussian processes. In: Proceedings of the Thirtieth Conference on Uncertainty in Artificial Intelligence, AUAI Press, Arlington, Virginia, USA, UAI'14, p 230–239
65. Garno J, Ouellet F, Bae S, Jackson TL, Kim NH, Haftka RT, Hughes KT, Balachandar S (2020) Calibration of reactive burn and jones-wilkins-lee parameters for simulations of a detonation-driven flow experiment with uncertainty quantification. Physical Review Fluids 5(12):123201, DOI 10.1103/PhysRevFluids.5.123201, URL https://doi.org/10.1103/PhysRevFluids.5.123201
66. Garside MJ (1965) The best sub-set in multiple regression analysis. Journal of the Royal Statistical Society: Series C (Applied Statistics) 14(2-3):196--200, DOI https://doi.org/10.2307/2985341, URL https://rss.onlinelibrary.wiley.com/doi/abs/10.2307/2985341
67. Gattiker J, Klein N, Lawrence E, Hutchings G (2014) lanl/SEPIA. Los Alamos, New Mexico, USA, DOI 10.5281/zenodo.4048801, URL https://github.com/lanl/SEPIA
68. Gattiker JR (2008) Gaussian Process Models for Simulation Analysis (GPM/SA) Command, Function, and Data Structure Reference. Technical Report LA-UR-08-08057, Los Alamos National Laboratory, Los Alamos, New Mexico, USA
69. Getis A, Griffith DA (2002) Comparative spatial filtering in regression analysis. Geographical Analysis 34(2):130--140, DOI https://doi.org/10.1111/j.1538-4632.2002.tb01080.x, URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1538-4632.2002.tb01080.x
70. Ghoreishi SF, Allaire D (2018) Multi-information source constrained Bayesian optimization. Structural and Multidisciplinary Optimization 59(3):977--991, DOI 10.1007/s00158-018-2115-z, URL https://doi.org/10.1007/s00158-018-2115-z
71. Ginsbourger D, Le Riche R, Carraro L (2007) Multi-points criterion for deterministic parallel global optimization based on kriging. In: International Conference on Nonconvex Programming, Rouen, France
72. Ginsbourger D, Le Riche R, Carraro L (2010) Kriging Is Well-Suited to Parallelize Optimization, Springer Berlin Heidelberg, Berlin, Heidelberg, pp 131--162. DOI 10.1007/978-3-642-10701-6_6, URL https://doi.org/10.1007/978-3-642-10701-6_6
73. Giunta A (2002) Use of data sampling, surrogate models, and numerical optimization in engineering design. In: 40th AIAA Aerospace Sciences Meeting & Exhibit, American Institute of Aeronautics and Astronautics, Reno, NV, pp AIAA 2002--0538, DOI 10.2514/6.2002-538, URL https://doi.org/10.2514/6.2002-538
74. Giunta A, Watson LT (1998) A comparison of approximation modeling techniques: polynomial versus interpolating models. In: 7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, St, St. Louis, USA, p –98–4758, DOI 10.2514/6.1998-4758
75. Giunta AA, Wojtkiewicz S, Eldred M (2003) Overview of modern design of experiments methods

for computational simulations. In: 41st Aerospace Sciences Meeting and Exhibit, Reno, NV, pp AIAA--2003--0649, DOI 10.2514/6.2003-649, URL https://doi.org/10.2514/6.2003-649
76. Glaz B, Goel T, Liu L, Friedmann PP, Haftka RT (2009) Multiple-surrogate approach to helicopter rotor blade vibration reduction. AIAA Journal 47(1):271--282, DOI 10.2514/1.40291, URL https://doi.org/10.2514/1.40291
77. Goel T, Haftka RT, Shyy W, Queipo NV (2007) Ensemble of surrogates. Structural and Multidisciplinary Optimization 32(3):199--216, DOI 10.1007/s00158-006-0051-9
78. Goel T, Haftka RT, Shyy W, Watson LT (2008) Pitfalls of using a single criterion for selecting experimental designs. International Journal for Numerical Methods in Engineering 75(2):127--155, DOI https://doi.org/10.1002/nme.2242, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/nme.2242
79. Gogu C, Bapanapalli SK, Haftka RT, Sankar BV (2009) Comparison of materials for an integrated thermal protection system for spacecraft reentry. Journal of Spacecraft and Rockets 46(3):501--513, DOI 10.2514/1.35669, URL https://doi.org/10.2514/1.35669
80. Gogu C, Haftka R, Le Riche R, Molimard J (2009) Dimensional analysis aided surrogate modelling for natural frequencies of free plates. In: 8th World Congress on Structural and Multidisciplinary Optimization, Lisbon, Portugal
81. Gogu C, Haftka RT, Bapanapalli SK, Sankar BV (2009) Dimensionality reduction approach for response surface approximations: Application to thermal design. AIAA Journal 47(7):1700--1708, DOI 10.2514/1.41414, URL https://doi.org/10.2514/1.41414
82. Gogu C, Yin W, Haftka R, Ifju P, Molimard J, Le Riche R, Vautrin A (2012) Bayesian identification of elastic constants in multi-directional laminate from moiré interferometry displacement fields. Experimental Mechanics 53(4):635--648, DOI 10.1007/s11340-012-9671-8, URL https://doi.org/10.1007/s11340-012-9671-8
83. Goldberger AS (1961) Stepwise least squares: Residual analysis and specification error. Journal of the American Statistical Association 56(296):998--1000, DOI 10.1080/01621459.1961.10482142, URL https://www.tandfonline.com/doi/abs/10.1080/01621459.1961.10482142
84. Goldberger AS, Jochems DB (1961) Note on stepwise least squares. Journal of the American Statistical Association 56(293):105--110, DOI 10.1080/01621459.1961.10482095, URL https://www.tandfonline.com/doi/abs/10.1080/01621459.1961.10482095
85. Goodfellow I, Bengio Y, Courville A (2016) Deep Learning. MIT Press, URL http://www.deeplearningbook.org
86. Gorissen D, Dhaene T, Turck FD (2009) Evolutionary model type selection for global surrogate modeling. Journal of Machine Learning Research 10(71):2039--2078, URL http://jmlr.org/papers/v10/gorissen09a.html
87. Gorissen D, Couckuyt I, Demeester P, Dhaene T, Crombecq K (2010) A surrogate modeling and adaptive sampling toolbox for computer based design. Journal of Machine Learning Research 11(68):2051--2055, URL http://jmlr.org/papers/v11/gorissen10a.html
88. Gramacy RB (2016) lagp: Large-scale spatial modeling via local approximate gaussian processes in r. Journal of Statistical Software 72(1):1--46, DOI 10.18637/jss.v072.i01, URL https://www.jstatsoft.org/v072/i01
89. Gray JS, Hwang JT, Martins JRRA, Moore KT, Naylor BA (2019) OpenMDAO: An open-source framework for multidisciplinary design, analysis, and optimization. Structural and Multidisciplinary Optimization 59(4):1075--1104, DOI 10.1007/s00158-019-02211-z
90. Guo D, Jin Y, Ding J, Chai T (2019) Heterogeneous ensemble-based infill criterion for evolutionary multiobjective optimization of expensive problems. IEEE Transactions on Cybernetics 49(3):1012--1025, DOI 10.1109/TCYB.2018.2794503
91. Haftka RT, Villanueva D, Chaudhuri A (2016) Parallel surrogate-assisted global optimization with expensive functions – a survey. Structural and Multidisciplinary Optimization 54(1):3--13, DOI 10.1007/s00158-016-1432-3, URL https://doi.org/10.1007/s00158-016-1432-3
92. Hazyuk I, Budinger M, Sanchez F, Gogu C (2017) Optimal design of computer experiments for surrogate models with dimensionless variables. Structural and Multidisciplinary Optimization 56(3):663--679, DOI 10.1007/s00158-017-1683-7, URL https://doi.org/10.1007/s00158-017-1683-7
93. Henkenjohann N, Kunert J (2007) An efficient sequential optimization approach based on the multivariate expected improvement criterion. Quality Engineering 19(4):267--280, DOI 10.1080/08982110701621312, URL https://doi.org/10.1080/08982110701621312

94. den Hertog D, Kleijnen JPC, Siem AYD (2006) The correct kriging variance estimated by bootstrapping. Journal of the Operational Research Society 57(4):400--409, DOI 10.1057/palgrave.jors.2601997, URL https://doi.org/10.1057/palgrave.jors.2601997
95. Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. Science 313(5786):504--507, DOI 10.1126/science.1127647, URL https://science.sciencemag.org/content/313/5786/504
96. Hu Z, Du X (2015) Mixed efficient global optimization for time-dependent reliability analysis. Journal of Mechanical Design 137(5), DOI 10.1115/1.4029520, URL https://doi.org/10.1115/1.4029520
97. Huang H (2018) Mechanisms of dimensionality reduction and decorrelation in deep neural networks. Physical Review E 98(6):062313, DOI 10.1103/PhysRevE.98.062313, URL https://doi.org/10.1103/PhysRevE.98.062313
98. Iooss B, Lemaître P (2015) A review on global sensitivity analysis methods. In: Uncertainty Management in Simulation-Optimization of Complex Systems, Springer US, pp 101--122, DOI 10.1007/978-1-4899-7547-8_5, URL https://doi.org/10.1007/978-1-4899-7547-8_5
99. Jilla CD, Miller DW (2004) Multi-objective, multidisciplinary design optimization methodology for distributed satellite systems. Journal of Spacecraft and Rockets 41(1):39--50, DOI 10.2514/1.9206, URL https://doi.org/10.2514/1.9206
100. Jiménez J, Ginebra J (2017) pygpgo: Bayesian optimization for python. Journal of Open Source Software 2(19):431, DOI 10.21105/joss.00431, URL https://doi.org/10.21105/joss.00431
101. Jin R, Chen W, Simpson T (2001) Comparative studies of metamodelling techniques under multiple modelling criteria. Structural and Multidisciplinary Optimization 23(1):1--13, DOI 10.1007/s00158-001-0160-4, URL https://doi.org/10.1007/s00158-001-0160-4
102. Jin R, Chen W, Sudjianto A (2002) On sequential sampling for global metamodeling in engineering design. In: ASME 2002 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference, ASME, Montreal, Canada, pp 539--548, DOI 10.1115/detc2002/dac-34092, URL https://doi.org/10.1115/detc2002/dac-34092
103. Jin R, Du X, Chen W (2003) The use of metamodeling techniques for optimization under uncertainty. Structural and Multidisciplinary Optimization 25(2):99--16, DOI 10.1007/s00158-002-0277-0
104. Jin R, Chen W, Sudjianto A (2004) Analytical metamodel-based global sensitivity analysis and uncertainty propagation for robust design. SAE Transactions 113:121--128, URL http://www.jstor.org/stable/44699914
105. Jolliffe I (2002) Principal component analysis, 2nd edn. Springer, New York
106. Jones DR (2001) A taxonomy of global optimization methods based on response surfaces. Journal of Global Optimization 21(4):345--383, DOI 10.1023/a:1012771025575, URL http://dx.doi.org/10.1023/A:1012771025575
107. Jones DR, Martins JRRA (2020) The DIRECT algorithm: 25 years later. Journal of Global Optimization DOI 10.1007/s10898-020-00952-6, URL https://doi.org/10.1007/s10898-020-00952-6
108. Jones DR, Schonlau M, Welch WJ (1998) Efficient global optimization of expensive black-box functions. Journal of Global Optimization 13(4):455--492, DOI 10.1023/a:1008306431147, URL https://doi.org/10.1023/a:1008306431147
109. Kambhatla N, Leen TK (1997) Dimension reduction by local principal component analysis. Neural computation 9(7):1493--1516, DOI 10.1162/neco.1997.9.7.1493, URL https://doi.org/10.1162/neco.1997.9.7.1493
110. Kanungo T, Mount D, Netanyahu N, Piatko C, Silverman R, Wu A (2002) An efficient k-means clustering algorithm: analysis and implementation. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(7):881--892, DOI 10.1109/TPAMI.2002.1017616
111. Kaufman M, Balabanov V, Grossman B, Mason WH, Watson LT, Haftka RT (1996) Multidisciplinary optimization via response surface techniques. In: 36th Israel Conference on Aerospace Sciences, Omanuth, Israel, p 57—67
112. Khatamsaz D, Peddareddygari L, Friedman S, Allaire D (2021) Bayesian optimization of multiobjective functions using multiple information sources. AIAA Journal 59(6):1964--1974, DOI 10.2514/1.j059803, URL https://doi.org/10.2514/1.j059803
113. Khuri AI, Cornell JA (1996) Response Surfaces: Designs and Analyses, 2nd edn. Dekker
114. Kim H, Lee TH, Kwon T (2021) Normalized neighborhood component feature selection and feasible-improved weight allocation for input variable selection. Knowledge-Based Systems 218:106855
115. Kingma DP, Welling M (2013) Auto-encoding variational bayes. arXiv preprint arXiv:13126114

116. Kleijnen JP (1997) Sensitivity analysis and related analyses: A review of some statistical techniques. Journal of Statistical Computation and Simulation 57(1-4):111--142, DOI 10.1080/00949659708811805, URL https://doi.org/10.1080/00949659708811805
117. Kleijnen JPC, van Beers WCM (2004) Application-driven sequential designs for simulation experiments: kriging metamodelling. Journal of the Operational Research Society 55(8):876--883, DOI 10.1057/palgrave.jors.2601747, URL https://doi.org/10.1057/palgrave.jors.2601747
118. Kleijnen JPC, Sanchez SM, Lucas TW, Cioppa TM (2005) State-of-the-art review: A user's guide to the brave new world of designing simulation experiments. INFORMS Journal on Computing 17(3):263--289, DOI 10.1287/ijoc.1050.0136, URL https://doi.org/10.1287/ijoc.1050.0136
119. Knudde N, van der Herten J, Dhaene T, Couckuyt I (2017) GPflowOpt: A Bayesian optimization library using TensorFlow. arXiv preprint arXiv:171103845
120. Knudde N, Dutordoir V, Herten JVD, Couckuyt I, Dhaene T (2020) Hierarchical gaussian process models for improved metamodeling. ACM Transactions on Modeling and Computer Simulation 30(4):1--17, DOI 10.1145/3384470, URL https://doi.org/10.1145/3384470
121. Koch PN, Simpson TW, Allen JK, Mistree F (1999) Statistical approximations for multidisciplinary design optimization: The problem of size. Journal of Aircraft 36(1):275--286, DOI 10.2514/2.2435, URL https://doi.org/10.2514/2.2435
122. Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Fourteenth International Joint Conference on Artificial Intelligence, IJCAII, Montreal, QC, CA, pp 1137--1143
123. Konakli K, Sudret B (2016) Global sensitivity analysis using low-rank tensor approximations. Reliability Engineering & System Safety 156:64--83, DOI https://doi.org/10.1016/j.ress.2016.07.012, URL https://www.sciencedirect.com/science/article/pii/S0951832016302423
124. Kristensen J, Subber W, Zhang Y, Ghosh S, Kumar NC, Khan G, Wang L (2020) Industrial applications of intelligent adaptive sampling methods for multi-objective optimization. In: Design and Manufacturing, IntechOpen, DOI 10.5772/intechopen.88213, URL https://doi.org/10.5772/intechopen.88213
125. Kushner HJ (1964) A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise. Journal of Basic Engineering 86(1):97--106, DOI 10.1115/1.3653121, URL https://doi.org/10.1115/1.3653121
126. Lacey D, Steele C (2006) The use of dimensional analysis to augment design of experiments for optimization and robustification. Journal of Engineering Design 17(1):55--73, DOI 10.1080/09544820500275594, URL https://doi.org/10.1080/09544820500275594
127. Leary S, Bhaskar A, Keane A (2003) Optimal orthogonal-array-based latin hypercubes. Journal of Applied Statistics 30(5):585--598, DOI 10.1080/0266476032000053691, URL https://doi.org/10.1080/0266476032000053691
128. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436--444, DOI 10.1038/nature14539, URL https://doi.org/10.1038/nature14539
129. Letham B, Karrer B, Ottoni G, Bakshy E (2019) Constrained bayesian optimization with noisy experiments. Bayesian Analysis 14(2), DOI 10.1214/18-ba1110, URL https://doi.org/10.1214/18-ba1110
130. Lewis M (2007) Stepwise versus hierarchical regression: Pros and cons. In: Proceedings of the Annual Meeting of the Southwest Educational Research Association, San Antonio
131. Li C, Gupta S, Rana S, Nguyen V, Venkatesh S, Shilton A (2017) High dimensional bayesian optimization using dropout. In: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17, pp 2096--2102, DOI 10.24963/ijcai.2017/291, URL https://doi.org/10.24963/ijcai.2017/291
132. Li CC, Lee YC (1990) A statistical procedure for model building in dimensional analysis. International journal of heat and mass transfer; A statistical procedure for model building in dimensional analysis 33(7):1566--1567
133. Li Z, Ruan S, Gu J, Wang X, Shen C (2016) Investigation on parallel algorithms in efficient global optimization based on multiple points infill criterion and domain decomposition. Structural and Multidisciplinary Optimization 54(4):747--773, DOI 10.1007/s00158-016-1441-2, URL https://doi.org/10.1007/s00158-016-1441-2
134. Liu H, Cai J, Ong YS (2018) Remarks on multi-output Gaussian process regression. Knowledge-Based Systems 144:102--121, DOI https://doi.org/10.1016/j.knosys.2017.12.034, URL https://www.sciencedirect.com/science/article/pii/S0950705117306123

135. Liu Q, Ding F (2018) Auxiliary model-based recursive generalized least squares algorithm for multivariate output-error autoregressive systems using the data filtering. Circuits, Systems, and Signal Processing 38(2):590--610, DOI 10.1007/s00034-018-0871-z, URL https://doi.org/10.1007/s00034-018-0871-z
136. Liu Y, Li K, Wang S, Cui P, Song X, Sun W (2021) A sequential sampling generation method for multi-fidelity model based on Voronoi region and sample density. Journal of Mechanical Design pp 1--17, DOI 10.1115/1.4051014, URL https://doi.org/10.1115/1.4051014
137. Lophaven SN, Nielsen HB, Søndergaard J (2002) Aspects of the Matlab toolbox DACE. Roskilde, Denmark, URL http://www.omicron.dk/dace/dace-aspects.pdf
138. Lubineau G (2009) A goal-oriented field measurement filtering technique for the identification of material model parameters. Computational Mechanics 44(5):591--603, DOI 10.1007/s00466-009-0392-5, URL https://doi.org/10.1007/s00466-009-0392-5
139. von Luxburg U (2007) A tutorial on spectral clustering. Statistics and Computing 17(4):395--416, DOI 10.1007/s11222-007-9033-z, URL https://doi.org/10.1007/s11222-007-9033-z
140. Mack Y, Goel T, Shyy W, Haftka R, Queipo N (2005) Multiple surrogates for the shape optimization of bluff body-facilitated mixing. In: 43rd AIAA aerospace sciences meeting and exhibit, AIAA, Reno, NV, USA, pp AIAA--2005--0333, DOI 10.2514/6.2005-333, URL https://doi.org/10.2514/6.2005-333
141. Mandel J (1982) Use of the singular value decomposition in regression analysis. The American Statistician 36(1):15--24, DOI 10.1080/00031305.1982.10482771, URL https://amstat.tandfonline.com/doi/abs/10.1080/00031305.1982.10482771
142. Mendez PF, Ordóñez F (2005) Scaling laws from statistical data and dimensional analysis. Journal of Applied Mechanics 72(5):648--657, DOI 10.1115/1.1943434, URL https://doi.org/10.1115/1.1943434
143. Meunier M (2009) Simulation and optimization of flow control strategies for novel high-lift configurations. AIAA Journal 47(5):1145--1157, DOI 10.2514/1.38245, URL https://doi.org/10.2514/1.38245
144. Montgomery DC (2012) Design and Analysis of Experiments. John Wiley & Sons, Inc, New York, USA
145. Morris MD (1991) Factorial sampling plans for preliminary computational experiments. Technometrics 33(2):161--174, DOI 10.1080/00401706.1991.10484804, URL https://amstat.tandfonline.com/doi/abs/10.1080/00401706.1991.10484804, https://amstat.tandfonline.com/doi/pdf/10.1080/00401706.1991.10484804
146. Myers DE (1982) Matrix formulation of co-kriging. Journal of the International Association for Mathematical Geology 14(3):249--257, DOI 10.1007/bf01032887, URL https://doi.org/10.1007/bf01032887
147. Myers RH, Montgomery DC (1995) Response Surface Methodology: Process and Product Optimization Using Designed Experiments. John Wiley & Sons, Inc, New York, USA
148. Nayebi A, Munteanu A, Poloczek M (2019) A framework for Bayesian optimization in embedded subspaces. In: Chaudhuri K, Salakhutdinov R (eds) Proceedings of the 36th International Conference on Machine Learning, PMLR, Proceedings of Machine Learning Research, vol 97, pp 4752--4761, URL http://proceedings.mlr.press/v97/nayebi19a.html
149. Oh C, Gavves E, Welling M (2018) BOCK : Bayesian optimization with cylindrical kernels. In: Dy J, Krause A (eds) Proceedings of the 35th International Conference on Machine Learning, PMLR, Stockholmsmässan, Stockholm Sweden, Proceedings of Machine Learning Research, vol 80, pp 3868--3877, URL http://proceedings.mlr.press/v80/oh18a.html
150. Pamadi B, Covell P, Tartabini P, Murphy K (2004) Aerodynamic characteristics and glide-back performance of langley glide-back booster. In: 22nd Applied Aerodynamics Conference and Exhibit, AIAA, Providence, USA, DOI 10.2514/6.2004-5382, URL https://arc.aiaa.org/doi/abs/10.2514/6.2004-5382
151. Parr JM, Keane AJ, Forrester AI, Holden CM (2012) Infill sampling criteria for surrogate-based optimization with constraint handling. Engineering Optimization 44(10):1147--1166, DOI 10.1080/0305215x.2011.637556, URL https://doi.org/10.1080/0305215x.2011.637556
152. Picard RR, Cook RD (1984) Cross-validation of regression models. Journal of the American Statistical Association 79(387):575--583
153. Picheny V, Ginsbourger D, Roustant O, Haftka RT, Kim NH (2010) Adaptive designs of experiments for accurate approximation of a target region. Journal of Mechanical Design 132(7),

DOI 10.1115/1.4001873, URL https://doi.org/10.1115/F1.4001873
154. Picheny V, Wagner T, Ginsbourger D (2013) A benchmark of kriging-based infill criteria for noisy optimization. Structural and Multidisciplinary Optimization 48(3):607--626, DOI 10.1007/s00158-013-0919-4, URL https://doi.org/10.1007/s00158-013-0919-4
155. Ponweiser W, Wagner T, Vincze M (2008) Clustered multiple generalized expected improvement: A novel infill sampling criterion for surrogate models. In: 2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence), Hong Kong, China, pp 3515--3522, DOI 10.1109/CEC.2008.4631273
156. Queipo NV, Haftka RT, Shyy W, Goel T, Vaidyanathan R, Kevin Tucker P (2005) Surrogate-based analysis and optimization. Progress in Aerospace Sciences 41(1):1--28, DOI 10.1016/j.paerosci.2005.02.001, URL https://www.sciencedirect.com/science/article/pii/S0376042105000102
157. Queipo NV, Verde A, Pintos S, Haftka RT (2009) Assessing the value of another cycle in gaussian process surrogate-based optimization. Structural and Multidisciplinary Optimization 39(5):459--475, DOI 10.1007/s00158-008-0346-0, URL https://doi.org/10.1007/s00158-008-0346-0
158. Raghavan B, Xia L, Breitkopf P, Rassineux A, Villon P (2013) Towards simultaneous reduction of both input and output spaces for interactive simulation-based structural design. Computer Methods in Applied Mechanics and Engineering 265:174--185, DOI https://doi.org/10.1016/j.cma.2013.06.010, URL https://www.sciencedirect.com/science/article/pii/S0045782513001710
159. Rai R (2006) Qualitative and quantitative sequential sampling. PhD thesis, University of Texas, Austin, USA
160. Rama RR, Skatulla S (2020) Towards real-time modelling of passive and active behaviour of the human heart using podi-based model reduction. Computers & Structures 232:105897, DOI https://doi.org/10.1016/j.compstruc.2018.01.002, URL https://www.sciencedirect.com/science/article/pii/S0045794917306727
161. Rama RR, Skatulla S, Sansour C (2016) Real-time modelling of diastolic filling of the heart using the proper orthogonal decomposition with interpolation. International Journal of Solids and Structures 96:409--422, DOI https://doi.org/10.1016/j.ijsolstr.2016.04.003, URL https://www.sciencedirect.com/science/article/pii/S002076831630021X
162. Rana S, Li C, Gupta S, Nguyen V, Venkatesh S (2017) High dimensional Bayesian optimization with elastic Gaussian process. In: Precup D, Teh YW (eds) Proceedings of the 34th International Conference on Machine Learning, PMLR, International Convention Centre, Sydney, Australia, Proceedings of Machine Learning Research, vol 70, pp 2883--2891, URL http://proceedings.mlr.press/v70/rana17a.html
163. Ranjan P, Bingham D, Michailidis G (2008) Sequential experiment design for contour estimation from complex computer codes. Technometrics 50(4):527--541, DOI 10.1198/004017008000000541, URL https://doi.org/10.1198/004017008000000541
164. Rasmussen CE, Williams CKI (2006) Gaussian Processes for Machine Learning. The MIT Press
165. Rennen G, Husslage B, Van Dam ER, Den Hertog D (2010) Nested maximin latin hypercube designs. Structural and Multidisciplinary Optimization 41(3):371--395, DOI 10.1007/s00158-009-0432-y, URL https://doi.org/10.1007/s00158-009-0432-y
166. Rezende DJ, Mohamed S, Wierstra D (2014) Stochastic backpropagation and approximate inference in deep generative models. In: Xing EP, Jebara T (eds) Proceedings of the 31st International Conference on Machine Learning, PMLR, Bejing, China, Proceedings of Machine Learning Research, vol 32, pp 1278--1286, URL http://proceedings.mlr.press/v32/rezende14.html
167. RH Shumway DS (2017) Time Series Regression and Exploratory Data Analysis. Springer, New York, New York, NY, DOI 10.1007/0-387-36276-2_2, URL https://doi.org/10.1007/978-3-319-52452-8_2
168. Rocha H, Li W, Hahn A (2006) Principal component regression for fitting wing weight data of subsonic transports. Journal of Aircraft 43(6):1925--1936, DOI 10.2514/1.21934, URL https://doi.org/10.2514/1.21934
169. Rojas-Gonzalez S, Van Nieuwenhuyse I (2020) A survey on kriging-based infill algorithms for multiobjective simulation optimization. Computers & Operations Research 116:104869, DOI https://doi.org/10.1016/j.cor.2019.104869, URL https://www.sciencedirect.com/science/article/pii/S0305054819303119
170. Roustant O, Ginsbourger D, Deville Y (2012) DiceKriging, DiceOptim: Two R packages for the analysis of computer experiments by kriging-based metamodeling and optimization. HAL preprint: 00495766 URL https://hal.

archives-ouvertes.fr/hal-00495766
171. Sacks J, Welch WJ, Mitchell TJ, Wynn HP (1989) Design and analysis of computer experiments. Statistical Science 4(4):409--423, URL http://www.jstor.org/stable/2245858
172. Saltelli A (2009) Sensitivity Analysis. John Wiley & Sons, New York
173. Saltelli A, Andres T, Homma T (1995) Sensitivity analysis of model output. performance of the iterated fractional factorial design method. Computational Statistics & Data Analysis 20(4):387--407, DOI https://doi.org/10.1016/0167-9473(95)92843-M, URL https://www.sciencedirect.com/science/article/pii/016794739592843M
174. Samad A, Kim KY, Goel T, Haftka RT, Shyy W (2008) Multiple surrogate modeling for axial compressor blade shape optimization. Journal of Propulsion and Power 24(2):301--310, DOI 10.2514/1.28999, URL https://doi.org/10.2514/1.28999
175. Sanchez F, Budinger M, Hazyuk I (2017) Dimensional analysis and surrogate models for the thermal modeling of multiphysics systems. Applied Thermal Engineering 110:758--771, DOI https://doi.org/10.1016/j.applthermaleng.2016.08.117, URL https://www.sciencedirect.com/science/article/pii/S1359431116314776
176. Schölkopf B, Smola AJ (2002) Learning with Kernels. The MIT Press, Cambridge, USA
177. Schölkopf B, Smola A, Müller KR (1998) Nonlinear component analysis as a kernel eigenvalue problem. Neural computation 10(5):1299--1319, DOI 10.1162/089976698300017467, URL https://doi.org/10.1162/089976698300017467
178. Simpson T, Poplinski J, Koch PN, Allen J (2001) Metamodels for computer-based engineering design: Survey and recommendations. Engineering with Computers 17(2):129--150, DOI 10.1007/pl00007198, URL https://doi.org/10.1007/pl00007198
179. Simpson T, Poplinski J, Koch PN, Allen J (2001) Metamodels for computer-based engineering design: Survey and recommendations. Engineering with Computers 17(2):129--150, DOI 10.1007/pl00007198, URL https://doi.org/10.1007/pl00007198
180. Simpson T, Booker A, Ghosh D, Giunta A, Koch P, Yang RJ (2004) Approximation methods in multidisciplinary analysis and optimization: a panel discussion. Structural and Multidisciplinary Optimization 27(5), DOI 10.1007/s00158-004-0389-9, URL https://doi.org/10.1007/s00158-004-0389-9
181. Simpson TW, Lin DKJ, Chen W (2001) Sampling strategies for computer experiments: design and analysis. International Journal of Reliability and Applications 2(3):209--240
182. Singh G, Viana FAC, Subramaniyan AK, Wang L, DeCesare D, Khan G, Wiggs G (2012) Multimodal particle swarm optimization: Enhancements and applications. In: 53rd AIAA / ASME / ASCE / AHS / ASC Structures, Structural Dynamics and Materials Conference, AIAA, Honolulu, USA, DOI 10.2514/6.2012-5570, URL http://dx.doi.org/10.2514/6.2012-5570
183. Smola AJ, Schölkopf B (2004) A tutorial on support vector regression. Statistics and Computing 14(3):199--222, DOI 10.1023/B:STCO.0000035301.49549.88
184. Snelson E, Ghahramani Z (2006) Variable noise and dimensionality reduction for sparse gaussian processes. In: Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence, AUAI Press, Arlington, Virginia, USA, p 461–468
185. Snoek J, Swersky K, Zemel R, Adams R (2014) Input warping for bayesian optimization of non-stationary functions. In: Xing EP, Jebara T (eds) Proceedings of the 31st International Conference on Machine Learning, PMLR, Bejing, China, Proceedings of Machine Learning Research, vol 32, pp 1674--1682, URL http://proceedings.mlr.press/v32/snoek14.html
186. Sobol I (1993) Sensitivity estimates for non-linear mathematical models. Mathematical Modelling and Computational Experiment 4:407–--414
187. Song W, Keane A (2006) Parameter screening using impact factors and surrogate-based ANOVA techniques. In: 11th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, American Institute of Aeronautics and Astronautics, Portsmouth, Virginia, pp AIAA 2006--7088, DOI 10.2514/6.2006-7088, URL https://doi.org/10.2514/6.2006-7088
188. Sonin AA (2001) The Physical Basis of Dimensional Analysis, 2nd edn. Massachusetts Institute of Technology, Cambridge, MA
189. Stander N, Roux W, Giger M, Redhe M, Fedorova N, Haarhoff J (2004) A comparison of metamodeling techniques for crashworthiness optimization. In: 10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, American Institute of Aeronautics and Astronautics, Albany, NY, DOI 10.2514/6.2004-4489, URL https://doi.org/10.2514/6.2004-4489
190. Stein ML (1999) Interpolation of Spatial Data: Some Theory for Kriging. Springer

191. Tabachnick BG, Fidell LS (2007) Experimental designs using ANOVA. Thomson/Brooks/Cole Belmont, CA
192. Tang B (1993) Orthogonal array-based latin hypercubes. Journal of the American Statistical Association 88(424):1392--1397, DOI 10.1080/01621459.1993.10476423, URL https://www.tandfonline.com/doi/abs/10.1080/01621459.1993.10476423, https://www.tandfonline.com/doi/pdf/10.1080/01621459.1993.10476423
193. Tenne Y, Izui K, Nishiwaki S (2010) Dimensionality-reduction frameworks for computationally expensive problems. In: IEEE Congress on Evolutionary Computation, IEEE, Barcelona, Spain, pp 1--8, DOI 10.1109/CEC.2010.5586251
194. Turner CJ, Crawford RH, Campbell MI (2007) Multidimensional sequential sampling for nurbs-based metamodel development. Engineering with Computers 23(3):155--174, DOI 10.1007/s00366-006-0051-9, URL https://doi.org/10.1007/s00366-006-0051-9
195. Utans J, Moody J (1991) Selecting neural network architectures via the prediction risk: application to corporate bond rating prediction. In: IEEE 1st International Conference on AI Applications on Wall Street, IEEE, New York, NY, USA, pp 35--41
196. Vaidyanathan R, Goel T, Haftka R, Quiepo N, Shyy W, Tucker K (2004) Global sensitivity and trade-off analyses for multi-objective liquid rocket injector design. In: 40th AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit, Fort Lauderdale, FL, pp AIAA 2004--4007, DOI 10.2514/6.2004-4007, URL https://doi.org/10.2514/6.2004-4007
197. van Beers WC, Kleijnen JP (2008) Customized sequential designs for random simulation experiments: kriging metamodeling and bootstrapping. European Journal of Operational Research 186(3):1099--1113, DOI https://doi.org/10.1016/j.ejor.2007.02.035, URL https://www.sciencedirect.com/science/article/pii/S0377221707002895
198. Varma S, Simon R (2006) Bias in error estimation when using cross-validation for model selection. BMC Bioinformatics 7(91):1290--1300, DOI 10.1186/1471-2105-7-91
199. Vaschy A (1882) Sur les lois de similitude en physique. Annales Télégraphiques 19:25
200. Vazquez E, Villemonteix J, Sidorkiewicz M, Walter É (2008) Global optimization based on noisy evaluations: An empirical study of two statistical approaches. Journal of Physics: Conference Series 135:012100, DOI 10.1088/1742-6596/135/1/012100, URL https://doi.org/10.1088%2F1742-6596%2F135%2F1%2F012100
201. Venter G, Haftka RT, Starnes JH (1998) Construction of response surface approximations for design optimization. AIAA Journal 36(12):2242--2249, DOI 10.2514/2.333, URL https://doi.org/10.2514/2.333
202. Viana FAC (2011) SURROGATES Toolbox User's Guide. Gainesville, FL, USA, version 3.0 edn, URL https://sites.google.com/site/srgtstoolbox
203. Viana FAC (2016) A tutorial on Latin hypercube design of experiments. Quality and Reliability Engineering International 32(5):1975--1985, DOI 10.1002/qre.1924, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/qre.1924, https://onlinelibrary.wiley.com/doi/pdf/10.1002/qre.1924
204. Viana FAC, Haftka RT (2009) Cross validation can estimate how well prediction variance correlates with error. AIAA Journal 47(9):2266--2270, DOI 10.2514/1.42162, URL http://dx.doi.org/10.2514/1.42162
205. Viana FAC, Haftka RT, Steffen V (2009) Multiple surrogates: how cross-validation errors can help us to obtain the best predictor. Structural and Multidisciplinary Optimization 39(4):439--457, DOI 10.1007/s00158-008-0338-0, URL http://dx.doi.org/10.1007/s00158-008-0338-0
206. Viana FAC, Venter G, Balabanov V (2010) An algorithm for fast optimal latin hypercube design of experiments. International Journal for Numerical Methods in Engineering 82(2):135--156, DOI 10.1002/nme.2750, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/nme.2750
207. Viana FAC, Haftka RT, Watson LT (2012) Sequential sampling for contour estimation with concurrent function evaluations. Structural and Multidisciplinary Optimization 45(4):615--618, DOI 10.1007/s00158-011-0733-9, URL http://dx.doi.org/10.1007/s00158-011-0733-9
208. Viana FAC, Haftka RT, Watson LT (2013) Efficient global optimization algorithm assisted by multiple surrogate techniques. Journal of Global Optimization 56(2):669--689, DOI 10.1007/s10898-012-9892-5, URL https://doi.org/10.1007/s10898-012-9892-5
209. Viana FAC, Simpson TW, Balabanov V, Toropov V (2014) Metamodeling in multidisciplinary design optimization: how far have we really come? AIAA Journal 52(4):670--690, DOI 10.2514/1.J052375, URL https://arc.aiaa.org/doi/abs/10.2514/

1.J052375
210. Vignaux GA, Scott JL (1999) Simplifying regression models using dimensional analysis. Austalia & New Zealand Journal of Statistics 41(2):31–41
211. Villemonteix J, Vazquez E, Walter E (2008) An informational approach to the global optimization of expensive-to-evaluate functions. Journal of Global Optimization 44(4):509--534, DOI 10.1007/s10898-008-9354-2, URL https://doi.org/10.1007/s10898-008-9354-2
212. Wang DQ (2011) Least squares-based recursive and iterative estimation for output error moving average systems using data filtering. IET Control Theory & Applications 5:1648--1657(9), DOI 10.1049/iet-cta.2010.0416, URL https://digital-library.theiet.org/content/journals/10.1049/iet-cta.2010.0416
213. Wang GG, Shan S (2004) Design space reduction for multi-objective optimization and robust design optimization problems. SAE Transactions 113:101--110, URL http://www.jstor.org/stable/44699911
214. Wang GG, Shan S (2006) Review of metamodeling techniques in support of engineering design optimization. Journal of Mechanical Design 129(4):370--380, DOI 10.1115/1.2429697, URL https://doi.org/10.1115/1.2429697
215. Wang W, Huang Y, Wang Y, Wang L (2014) Generalized autoencoder: A neural network framework for dimensionality reduction. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 490--497
216. Wang Z, Li C, Jegelka S, Kohli P (2017) Batched high-dimensional Bayesian optimization via structural kernel learning. In: Precup D, Teh YW (eds) Proceedings of the 34th International Conference on Machine Learning, PMLR, Proceedings of Machine Learning Research, vol 70, pp 3656--3664, URL http://proceedings.mlr.press/v70/wang17h.html
217. Welch WJ, Buck RJ, Sacks J, Wynn HP, Mitchell TJ, Morris MD (1992) Screening, predicting, and computer experiments. Technometrics 34(1):15--25, DOI 10.1080/00401706.1992.10485229, URL https://amstat.tandfonline.com/doi/abs/10.1080/00401706.1992.10485229
218. Wold H (1983) Systems analysis by partial least squares. Iiasa collaborative paper, IIASA, IIASA, Laxenburg, Austria, URL http://pure.iiasa.ac.at/id/eprint/2336/
219. Wolpert DH (1996) The lack of a priori distinctions between learning algorithms. Neural Computation 8(7):1341--1390, DOI 10.1162/neco.1996.8.7.1341, URL https://doi.org/10.1162/neco.1996.8.7.1341
220. Wu J, Toscano-Palmerin S, Frazier PI, Wilson AG (2020) Practical multi-fidelity Bayesian optimization for hyperparameter tuning. In: Adams RP, Gogate V (eds) Proceedings of The 35th Uncertainty in Artificial Intelligence Conference, PMLR, Tel Aviv, Israel, Proceedings of Machine Learning Research, vol 115, pp 788--798, URL http://proceedings.mlr.press/v115/wu20a.html
221. Yang W, Wang K, Zuo W (2012) Neighborhood component feature selection for high-dimensional data. JCP 7(1):161--168
222. Yang Y (2003) Regression with multiple candidate models: selecting or mixing? Statistica Sinica 13(3):783--809, DOI 10.1.1.17.753
223. Yee TW (2000) Vector splines and other vector smoothers. In: eds (ed) Proc. Computational Statistics COMPSTAT 2000, Physica-Verlag HD, Bethlehem, J.G., Van der Heijden, P.G.M, pp 529--534, DOI 10.1007/978-3-642-57678-2_75, URL https://doi.org/10.1007/978-3-642-57678-2_75
224. Yondo R, Andrés E, Valero E (2018) A review on design of experiments and surrogate models in aircraft real-time and many-query aerodynamic analyses. Progress in Aerospace Sciences 96:23--61, DOI https://doi.org/10.1016/j.paerosci.2017.11.003, URL https://www.sciencedirect.com/science/article/pii/S0376042117300611
225. Zaefferer M, Stork J, Friese M, Fischbach A, Naujoks B, Bartz-Beielstein T (2014) Efficient global optimization for combinatorial problems. In: Proceedings of the 2014 Annual Conference on Genetic and Evolutionary Computation, ACM, DOI 10.1145/2576768.2598282, URL https://doi.org/10.1145/2576768.2598282
226. Zahm O, Constantine PG, Prieur C, Marzouk YM (2020) Gradient-based dimension reduction of multivariate vector-valued functions. SIAM Journal on Scientific Computing 42(1):A534--A558, DOI 10.1137/18M1221837, URL https://doi.org/10.1137/18M1221837
227. Zerpa LE, Queipo NV, Pintos S, Salager JL (2005) An optimization methodology of alkaline-surfactant-polymer flooding processes using field scale numerical simulation and multiple surrogates. Journal of Petroleum Science and Engineering 7(3-4):197--208, DOI 10.1016/j.petrol.2005.03.002
228. Zhang P (1993) Model selection via multifold cross validation. The Annals of Statistics 21(1):299--313, DOI 10.1214/aos/1176349027
229. Zhou Q, Wang Y, Choi SK, Jiang P, Shao X, Hu J (2017) A sequential multi-fidelity

metamodeling approach for data regression. Knowledge-Based Systems 134:199--212, DOI https://doi.org/10.1016/j.knosys.2017.07.033, URL https://www.sciencedirect.com/science/article/pii/S0950705117303556