A Nonstationary Covariance Based Kriging Method for Metamodeling in Engineering Design

Ying Xiong* and Weichen†
Integrated DEsign Automation Laboratory (IDEAL), Northwestern University, Evanston, IL 60208, USA

Daniel Apley‡
Northwestern University, Evanston, IL 60208, USA

and

Xuru Ding§
General Motors Corp, Warren, MI 48090, USA

Metamodels are widely used to facilitate the analysis and optimization of engineering systems that involve computationally expensive simulations. Kriging is a metamodeling technique that is well known for its ability to build surrogate models of responses with nonlinear behavior. However, the assumption of a stationary covariance structure underlying Kriging does not hold in situations where the level of smoothness of a response varies significantly. Although nonstationary Gaussian process models have been studied for years in statistics and geostatistics communities, this has largely been for physical experimental data in relatively low dimensions. In this paper, the nonstationary covariance structure is incorporated into Kriging modeling for computer simulations. To represent the nonstationary covariance structure, we adopt a nonlinear mapping approach based on a parameterized density functions. To avoid over-parameterizing for the high dimension problems typical of engineering design, we propose a modified version of the nonlinear map approach, with a sparser, yet flexible, parameterization. The effectiveness of the proposed method is demonstrated through both mathematical and engineering examples. The robustness of the method is verified by testing multiple functions under various sampling settings. We also demonstrate that our method is effective in quantifying prediction uncertainty associated with the use of metamodels.

Nomenclature

$Z(\mathbf{x})$	=	the Gaussian process indexed by $\mathbf{x}$
$\boldsymbol{\beta}^T\mathbf{h}(\mathbf{x})$	=	cylinder diameter
C_{stat}	=	stationary covariance
ρ_{stat}	=	stationary correlation
$C_{nonstat}$	=	nonstationary covariance
$\rho_{nonstat}$	=	nonstationary correlation
Θ	=	hyperparameter set

* Associate Professor, Department of Mechanical Engineering, 2145 Sheridan Rd., Tech B224, Evanston, IL 60208, weichen@northwestern.edu, Associate Fellow of AIAA.
† Graduate Student
‡ Associate Professor
§ Engineer

J = number of function basis centers
L = number of input variables
l = index of input variable (i.e. dimension)
$\theta^{(l)}$ = correlation parameter for input variable $x^{(l)}$
$f^{(l)}(\mathbf{x})$ = general mapping function
$f^{(l)}(x^{(l)})$ = mapping function (univariate)
$g^{(l)}(\mathbf{x})$ = general density function
$g^{(l)}(x^{(l)})$ = mapping function (univariate)
$\eta_k^{(l)}$ = hyperparameter in the piecewise density function
ξ_k = knot in the piecewise density function
K = number of pieces in the piecewise density function
k = index of function piece (k=1,2,…,K); index of knots or hyperparameters (k=0,1,…,K)

I. Introduction

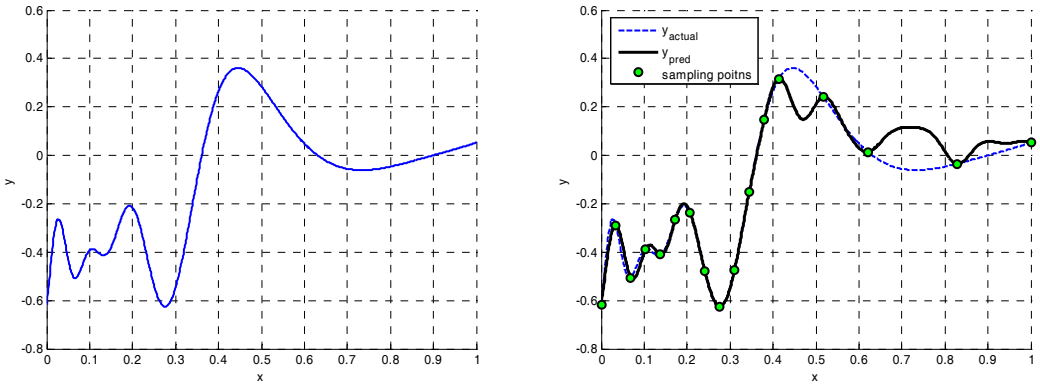
Metamodels are widely used in engineering design to facilitate the analysis and optimization of complex systems based on computationally expensive simulations. Among the widely used metamodeling techniques[1-8], Kriging is considered powerful and flexible for building surrogate models (or metamodels) of simulated response surfaces with different functional forms[6,9]. One of the distinctive advantages of Kriging is that it provides not only the prediction of the response at any site, but also the Mean Square Error (or the uncertainty) associated with the prediction. Ref. 10 provided a comprehensive review of Kriging with a Bayesian interpretation.

In a conventional Kriging model, a response is assumed to be a spatial random process with stationary covariance function. The stationary covariance implies that the smoothness of a response is fairly uniform in each region of the input space.[11] This is a convenient assumption that simplifies the analysis and lessens the amount of prior information required.[10] However, cases are common where the level of smoothness of a response could change dramatically throughout the whole design region. For example, in engineering design, when subsystem models with distinctive underlying physics are integrated, the system response behavior can differ greatly from one design region to another. Similar phenomena have been observed in geostatistics and environmental problems, where the geology of spatial locations greatly influences the correlation between responses.[12, 13] In those cases, the assumption of the uniformity of smoothness is not well satisfied. One such function is illustrated in Fig.1, in which the roughness in region $x \in [0, 0.3]$ is larger than in the region [0.3, 1]. Assuming a stationary covariance structure forces a trade-off in which the estimated stationary covariance reflects the average smoothness over the entire domain, but fails to reflect the true local smoothness in each region.

One natural approach to solving this problem is to consider relaxing the stationary covariance assumption and allowing a nonstationary one. The idea of using nonstationary covariance in metamodeling can be found in the engineering design literature.[14, 15] However, all existing works consider the irregular performance behavior only in sampling but not in fitting metamodels. Based on the prediction error and the irregularity of a response surface, the entries of the covariance matrix are adjusted, which essentially leads to a nonstationary covariance for choosing sample sites. While samples generated by these methods tend to be non-uniform in the input space, all these existing work used ordinary stationary Kriging for building the metamodels. Moreover, they use heuristic methods of adjusting the covariance matrix that do not guarantee that the covariance matrix is positive definite, a crucial property for the stability of a fitted Kriging model.

Nonstationary covariance Gaussian process models have been used for fitting response surface models in the fields of statistics, geostatistics and machine learning. However, this has primarily been under the scenario of physical experiments in relatively low dimensions.[16] Various approaches have been proposed to formulate the nonstationary covariance structure. Ref. 17 developed a spatial deformation approach to reallocate all the points. Ref. 18 and Ref. 19 proposed two methods of representing the nonstationary behavior: the first is to directly formulate a nonstationary covariance function, and the second is the so-called nonlinear map approach, in which the original Euclidean space is mapped to a new one for which the covariance can be approximated as stationary. Ref. 11 used a process convolution approach extended from Ref. 20's method and Ref. 18 (Gibbs)' first method. Other related work could be found in Ref. 12, Ref. 21, and Ref. 22. Ref. 23 presented an approach that utilizes Gaussian

process trees to implement the nonstationary Gaussian process. However, discontinuity of the response across subregions cannot be avoided.



(a) True function (b) Kriging prediction (stationary covariance)

Figure 1. Example of Kriging prediction with stationary covariance

Little prior work has been conducted on nonstationary covariance modeling for complex system design based on computer experiments. This is most likely because complex design problems are often high dimensional, and nonstationary covariance functions tend to be overparameterized in high dimensions. In this paper, we develop an efficient method that allows nonstationary covariance in Kriging metamodeling for high-dimensional engineering applications with computer experiments. We use a nonlinear mapping approach to represent the nonstationary covariance structure, in which a parameterized density function is used to map the original space to one in which the covariance becomes approximately stationary. Although high-dimensionality is undoubtedly a prohibitive problem with physical experimental data, on which most of the prior work in nonstationary covariance modeling has focused, we argue that the approach can be made quite robust for high dimensional computer experimental data. This is due to two factors: First, unlike physical experiments, computer experiments are usually perfectly repeatable, which drastically reduces the amount of data required to accurately fit nonstationary covariance functions in high dimensions. Second, we propose a modified version of Gibb's nonlinear map[18] approach, with a sparse, yet flexible, parameterization that is well suited for high-dimensional computer experimental data.

The organization of this paper is as follows. In Section II, a review of Kriging modeling with a stationary covariance function is first provided; the Gibbs' nonlinear map18 approach is then introduced. Details of our proposed approach are provided in Section III. Computational issues related to the optimization strategies for estimating the hyperparameters is also addressed. In Section IV, mathematical and engineering examples are used to illustrate the effectiveness of the proposed approach. We demonstrate that the proposed method not only improves the accuracy of metemodels for functions with changing irregularity, but also effectively quantifies the prediction uncertainty associated with the use of metamodels in engineering applications. Concluding remarks are given in Section V.

II. Technological Base

A. Kriging Metamodeling with a Stationary Covariance Function

In the conventional Kriging model[4], the performance y(x) is modeled as

$$Y(\mathbf{x}) = \boldsymbol{\beta}^T \mathbf{h}(\mathbf{x}) + Z(\mathbf{x}), \tag{1}$$

where $\boldsymbol{\beta}^T \mathbf{h}(\mathbf{x})$ is the regression component (e.g., a polynomial) which captures global trends; $Z(\mathbf{x})$ is assumed a Gaussian process indexed by input variables $\mathbf{x}$, with zero mean and stationary covariance. From a Bayesian perspective[10], the prior knowledge of the performance y(x) is specified by a Gaussian process, which is characterized by the prior mean (i.e. the global trend) and prior covariance. Given the observations, the posterior process is also a Gaussian process (treating the covariance parameters as known and assuming a Gaussian prior distribution for $\boldsymbol{\beta}$). The prediction of y(x) is usually taken to be the posterior mean, and the prediction uncertainty is quantified by the posterior covariance.

The conventional Kriging model assumes that the Gaussian process has a stationary covariance, with the covariance function defined as follows:

$$C_{st}\left(\mathbf{x}_m, \mathbf{x}_n; \Theta\right) = \sigma^2 \rho_{st}\left(\mathbf{x}_m, \mathbf{x}_n; \boldsymbol{\theta}\right), \tag{2}$$

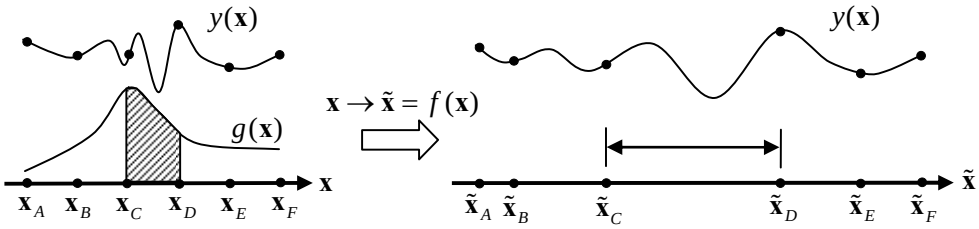
where ρ_{st} is the correlation function. The hyperparameter set Θ is composed of $\{\sigma^2; \boldsymbol{\theta}\}$. A frequently used Gaussian correlation function is:

$$\rho_{st}(\mathbf{x}_m, \mathbf{x}_n; \boldsymbol{\theta}) = \exp\left(-\sum_{l=1}^{L} \theta^{(l)} (\mathbf{x}_m^{(l)} - \mathbf{x}_n^{(l)})^2 \right). \tag{3}$$

The variance σ^2 provides the overall vertical scale relative to the mean of Gaussian process in the output space, $\boldsymbol{\theta} = \{\theta^{(l)} (l = 1, 2, ..., L)\}$ are the correlation parameters (scaling factors) associated with each input variable $\mathbf{x}^{(l)}$, which reflects the smoothness of the true performance. The stationary covariance indicates that the correlation function $\rho_{st}(\mathbf{x}_m, \mathbf{x}_n; \boldsymbol{\theta})$ between any two sites $\mathbf{x}_m$ and $\mathbf{x}_n$ depends on only the distance (scaled by $\boldsymbol{\theta}$) between $\mathbf{x}_m$ and $\mathbf{x}_n$. In Eqs. (2) and (3), the subscript '*st*' means 'stationary'.

B. Representing Nonstationary Covariance: the Nonlinear Map Approach

Various approaches exist in literature to represent nonstationary covariance structures. The nonlinear map method[18] is attractive among others because it is intuitively interpretable with a notion of space mapping. A simple one-dimensional illustration of the map is provided in Fig. 2. A mapping function $f(\mathbf{x})$ is defined by integrating a density function $g(\mathbf{x})$ (see Eq. (4) for details). It can be seen in the original space (Fig. 2 (a)) that the true function, denoted as $y(\mathbf{x})$, is hard to be modeled with a stationary covariance due to abruptly changing smoothness of $y(\mathbf{x})$. However, through mapping to the new space, the new response exhibits an improved uniformity of smoothness across the whole region. Hence the stationary covariance can be employed in the new space. Following the definition of the mapping function $f(\mathbf{x})$ as the integration of $g(\mathbf{x})$, the distance between point C and D in the new space $|\tilde{\mathbf{x}}_D - \tilde{\mathbf{x}}_C|$ corresponds to the shaded area in panel (a). Obviously, the higher density function around point C accounts for the higher abruptness of the real response. In other words, the relation between the density function $g(\mathbf{x})$ and the smoothness of the real response can be established.



(a) Original space (b) New space

Figure 2. A conceptual illustration of the nonlinear map

The mapping function becomes more complicated in high-dimensional situations. It is noteworthy that as long as the density function is positive and continuous, the positive-definiteness of the resulting nonstationary covariance can be strictly guaranteed. This is one significant characteristic of the approaches based on map and space deformation. In Gibbs' approach (refer to Ref. 11 for further details), the multidimensional mapping functions must be one-to-one and continuous, which ensures the order of points on any non-intersecting line in original $\mathbf{x}$ space is preserved in the new $\tilde{\mathbf{x}}$ space and the line in the $\tilde{\mathbf{x}}$ space is non-intersecting. Denote the mapping from $\mathbf{x}$ to $\tilde{\mathbf{x}}$ as $\tilde{\mathbf{x}} = \mathbf{f}(\mathbf{x}) = (f^{(1)}(\mathbf{x}), f^{(2)}(\mathbf{x}), \cdots, f^{(L)}(\mathbf{x}))$, where $f^{(l)}(\mathbf{x})$ defines the l^{th} coordinate of $\tilde{\mathbf{x}}$, i.e., $\tilde{x}^{(l)}$. To

achieve the aforementioned mapping, the l^{th} mapping function $f^{(l)}(\mathbf{x})$ is defined as an integral over a density function $g^{(l)}(\mathbf{x})$.

$$f^{(l)}(\mathbf{x}) = x_0^{(l)} + \int_{x_0^{(1)}}^{x^{(l)}} \cdots \int_{x_0^{(L)}}^{x^{(L)}} g^{(l)}(\mathbf{x}') dx'^{(1)} \cdots dx'^{(L)}, \tag{4}$$

where L is the number of input variables; $\mathbf{x}_0 = (x_0^{(1)}, \cdots x_0^{(L)})$ is a reference vector, which is often chosen to be somewhere in the center of the data; The density function $g^{(l)}(\mathbf{x})$ is further defined as a weighted sum of positive radial basis functions:

$$g^{(l)}(\mathbf{x}) = \sum_{j=1}^{J} \omega_j^{(l)} \psi_j(\mathbf{x}), \tag{5}$$

where $\{\psi_j(\mathbf{x})\}$ are a set of positive basis functions, common to all density functions $g^{(l)}(\mathbf{x})$; J is the number of basis functions. To ensure positive weights, $\omega_j^{(l)}$ can be taken as $\omega_j = e^{\alpha_j^{(l)}}$. The integral in Eq. (4) should be easily evaluated. To this end, uncorrelated Gaussian basis is utilized to form $\psi_j(\mathbf{x})$, i.e.

$$\psi_j(\mathbf{x}) = \exp\left[-\frac{1}{2} \sum_{l=1}^{L} \frac{(x'^{(l)} - c_j^{(l)})^2}{\sigma^{(l)2}} \right], \tag{6}$$

where $c_j^{(l)}$ and $\sigma^{(l)}$ quantify respectively the center and width for l^{th} dimension of the j^{th} basis function. $c_j^{(l)}$ and $\sigma^{(l)}$ are predetermined and viewed as known and fixed. Applying the mapping function Eq. (4) to the stationary covariance function in Eq. (1), the nonstationary version of the covariance/correlation function in Eq. (2) is obtained by

$$C_{nonst}(\mathbf{x}_m, \mathbf{x}_n; \Theta) = \sigma^2 \rho_{nonst}(\mathbf{x}_m, \mathbf{x}_n; \Theta) = \sigma^2 \exp\left(-\sum_{l=1}^{L} \left(f^{(l)}(\mathbf{x}_m) - f^{(l)}(\mathbf{x}_n) \right)^2 \right), \tag{7}$$

where the subscript '*nonst*' means 'nonstationary'. Note the original $\mathbf{x}^{(l)}$ in Eq. (4) is replaced by $f^{(l)}(\mathbf{x})$ and the original $\theta^{(l)}$ disappears. As a result, the hyperparameter set Θ becomes $\{\sigma^2; \alpha_j^{(l)} (j = 1, 2, ..., J; l = 1, 2, ..., L)\}$, as opposed to its stationary counterpart $\{\sigma^2; \theta^{(l)} (l = 1, 2, ..., L)\}$.

III. A Proposed Nonstationary Covariance Structure

Drawbacks of Gibbs' nonlinear map method are immediately observable. From Eqs. (4-6), the mapping function $f^{(l)}(\mathbf{x})$ and the density function $g^{(l)}(\mathbf{x})$ are multivariate functions of $\mathbf{x}$. The unknown hyperparameters $\alpha_j^{(l)}$ in $g^{(l)}(\mathbf{x})$ for different dimensions are indeed independent. In other words, with the nonlinear map method, the nonstationary covariance structure relies on $J \times L$ unknown hyperparameters $\alpha_j^{(l)}$ s' in total. It should be pointed out that J is the number of function basis centers, and should be large enough so that the nonstationary structure is able to cover the design space. This formulation is affordable in low dimensional situations with small and L, but can yield a large number of hyperparameters $\alpha_j^{(l)}$ in high dimensional cases because J increases with L. In other words, the nonstationary covariance will be over-parameterized in high dimensional situations, which has undesirable

consequences on the stability and robustness of the model fitting. It is our goal in this section to develop certain forms of density functions using as few hyperparameters as possible to address the aforementioned difficulties.

A. Proposed Density Function

The nonstationary structure is simplified by assuming that the varying smoothness behavior in any single input variable is independent with respect to the other input variables. This simplifying assumption reduces the multivariate density functions to *univariate* density functions in which $g^{(l)}(\mathbf{x})$ depends only on a particular $x^{(l)}$. By substituting $g^{(l)}(x^{(l)})$ in place of $g^{(l)}(\mathbf{x})$, the mapping function in Eq. (4) becomes

$$f^{(l)}(\mathbf{x}) = f^{(l)}(x^{(l)}) = x_0^{(l)} + \int_{x_0^{(l)}}^{x^{(l)}} g^{(l)}(x')dx', \tag{8}$$

where $x_0^{(l)}$ is the reference point. Instead of using the nonlinear form in Gibb's approach, we make further simplifications by assuming that $g^{(l)}(x^{(l)})$ is a continuous piecewise linear function in $x^{(l)}$. The continuity of $g^{(l)}(x^{(l)})$ is emphasized here because it is a critical requirement of the mapping functions as stated in Section II.B. For a selected number of pieces K, $g^{(l)}(x^{(l)})$ is defined as a summation of K linear components,

$$g^{(l)}(x^{(l)}) = \sum_{k=1}^{K} g_k^{(l)}(x^{(l)}; a_k^{(l)}, b_k^{(l)}). \tag{9}$$

Each component $g_k^{(l)}$ is a single linear function over its support interval, while being zero elsewhere:

$$g_k^{(l)}(x^{(l)}; a_k^{(l)}, b_k^{(l)}) = \begin{cases} a_k^{(l)} + b_k^{(l)} x^{(l)}, & x^{(l)} \in [\xi_{k-1}, \xi_k] \\ 0, & x^{(l)} \notin [\xi_{k-1}, \xi_k] \end{cases}, (k = 1, 2, ..., K), \tag{10}$$

where $\{a_k^{(l)}, b_k^{(l)}\}$ are the linear parameters for the k^{th} linear component; $\{\xi_0, \xi_1, ..., \xi_K\}$ are a series of knots placed along input variable $x^{(l)}$. Imposing the continuity constraints to $g^{(l)}(x^{(l)})$, Eq. (9) can be reformulated using the following linear substitutions:

$$\begin{cases} a_k^{(l)} = \left(\xi_k \eta_{k-1}^{(l)} - \xi_{k-1} \eta_k^{(l)}\right)\left(\xi_k - \xi_{k-1}\right)^{-1} \\ b_k^{(l)} = \left(\eta_k^{(l)} - \eta_{k-1}^{(l)}\right)\left(\xi_k - \xi_{k-1}\right)^{-1} \end{cases}, (k = 1, 2, ..., K). \tag{11}$$

The original linear parameters $\{a_k^{(l)}, b_k^{(l)}\}$ $(k = 1, 2, ..., K)$ are now replaced by K+1 new parameters $\{\eta_0^{(l)}, \eta_1^{(l)}, ..., \eta_K^{(l)}\}$. It can be verified that $\{\eta_0^{(l)}, \eta_1^{(l)}, ..., \eta_K^{(l)}\}$ are equal to the density values at each knot respectively, i.e., $g^{(l)}(\xi_k) = \eta_k^{(l)}, (k = 0, 1, 2, ..., K)$. According to Eq. (8), if the reference point is placed at the knot ξ_0, the mapping function $f^{(l)}(x^{(l)})$ can be formulated as

$$\begin{aligned} f^{(l)}(x^{(l)}) &= \xi_0 + \int_{\xi_0}^{x^{(l)}} g^{(l)}(x')dx' = \xi_0 + \int_{\xi_0}^{\xi_1} g^{(l)}(x')dx' + \cdots + \int_{\xi_{M-1}}^{\xi_M} g^{(l)}(x')dx' + \int_{\xi_M}^{x^{(l)}} g^{(l)}(x')dx' \\ &= \xi_0 + \sum_{k=1}^{M} \int_{\xi_{k-1}}^{\xi_k} g_k^{(l)}(x'; \eta_{k-1}^{(l)}, \eta_k^{(l)})dx' + \int_{\xi_M}^{x^{(l)}} g^{(l)}(x'; \eta_{k-1}^{(l)}, \eta_k^{(l)})dx' \end{aligned}, \tag{12}$$

where M ($0 \le M < K$) is the index of the knot left-neighboring $x^{(l)}$, i.e., $\xi_M < x^{(l)} \le \xi_{M+1}$. The number of parameters or the d.o.f. of $g^{(l)}(x^{(l)})$ is K+1. In particular, K=1 means that $g^{(l)}(x^{(l)})$ reduces to a single linear function $g^{(l)}(x^{(l)};\eta_0^{(l)},\eta_1^{(l)})$. Because $g^{(l)}(x^{(l)})$ is linear and univariate, all integrals in Eq. (12) can be easily computed in analytical way. The sum of integrals in $f^{(l)}(x^{(l)})$ over the interval $[\xi_0, x^{(l)}]$ corresponds to the shaded area in Fig. 3.

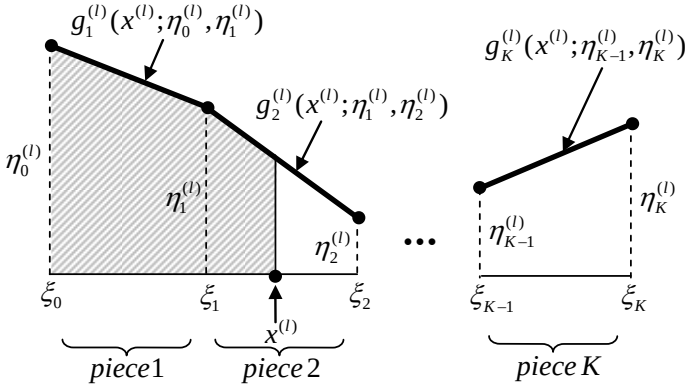


Figure 3. Continuous piecewise linear density function

Constraining $g^{(l)}(x^{(l)})$ is equivalent to imposing a bound to $\eta_k^{(l)}$ since $g^{(l)}(x^{(l)})$ is linear. For instance, $lb < \eta_k^{(l)} < ub$, for $k = 0,1,...,K$, is equivalent to $lb < g^{(l)}(x^{(l)}) < ub$. When $\eta_0^{(l)} = \cdots = \eta_{K-1}^{(l)} = \eta_K^{(l)}$, $g^{(l)}(x^{(l)})$ is essentially a constant, hence the nonstationary covariance reduces to the stationary one. The number of knots ξ_k (which is K+1) reflects the resolution and complexity of the density function.

B. Determining the Hyperparameters

Various methods exist for estimating the Kriging hyperparameters. One way is to perform the integration over Θ using Monte Carlo methods[18, 24]. The method identifies the best values of the hyperparmeters as the mean, median, or the mode of the posterior distribution of Θ. In an alternative approach, the most probable value of Θ is identified by maximizing the MLE (Maximum Likelihood Estimation) [14,19]. Ref. 25 compared the use of the AIC (Akaike Information Criterion) and the CV (Cross Validation) criterion. In our work, the MLE method is used to estimate the hyperparameters for the nonstationary covariance. The Simulated Annealing method[26] is used for optimization due to the highly nonlinear nature of the MLE function.

Compared to the other nonstationary covariance structures, the proposed nonstationary covariance structure employs very few hyperparameters, even though the total number will be larger than that of the stationary Kriging model. One way to alleviate the problem is to place fewer knots along each variable in high dimensional problems or to place more knots along critical variables. The other way is to impose reasonable bounds for each $\eta_k^{(l)}$ to expedite the search. This is easy to implement through our proposed density function. For bounding $\eta_k^{(l)}$, a multiple-stage strategy is employed, in which the density function is estimated sequentially. Fig. 4 illustrates the idea of such strategy with, for example, two stages. In the first stage, the density function is constructed with a low d.o.f., say, K=1 (i.e. d.o.f.=2). After $\eta_0^{(l)}$ and $\eta_1^{(l)}$ are estimated (the filled squares), the density function is determined (the dashed line). In the second stage, the complexity of the density function could be increased, say, with $\{\eta_0^{(l)},\eta_1^{(l)},\eta_2^{(l)}\}$. Based on the estimated density function from the first stage, reasonable bounds can be set for $\{\eta_0^{(l)},\eta_1^{(l)},\eta_2^{(l)}\}$. It is worth noting that imposing the bounds for $\eta_k^{(l)}$ not only facilitates the optimization, but also allows us to express the belief about the abruptness of changing smoothness. If the smoothness is not expected to change abruptly, narrower bounds closer to their neighbors are preferred.

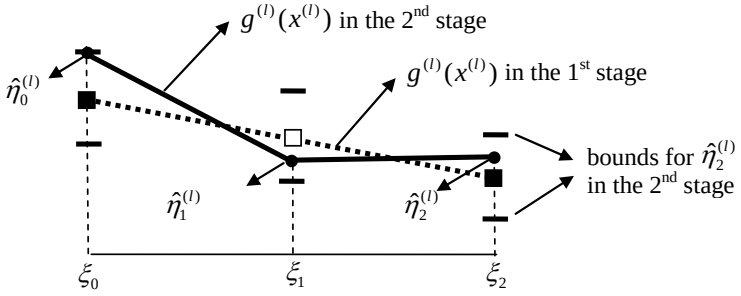


Figure 4. Setting bounds for the to-be-estimated $\eta_k^{(l)}$'s

IV. Numerical Examples and Verification

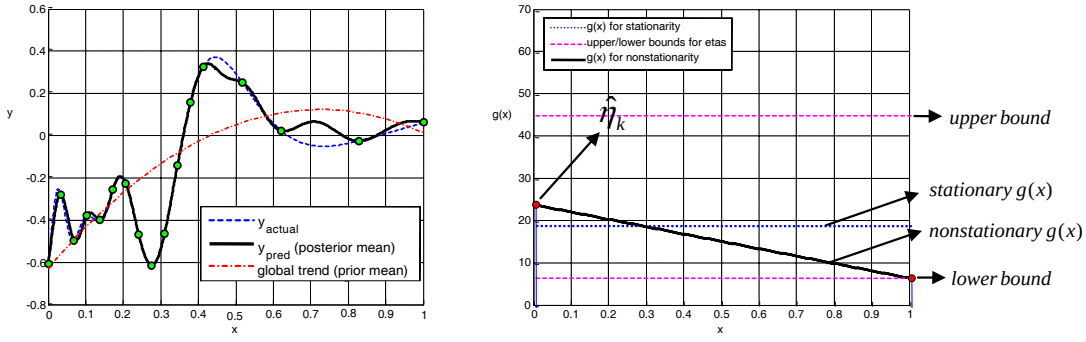
In this section, we present a few mathematical and engineering examples to illustrate and verify the effectiveness of our proposed nonstationary covariance structure for modeling the varying smoothness of different responses. We first illustrate important characteristics of the approach with simple one- and two-dimensional examples. In Section IV.D, we demonstrate the approach for a high-dimensional vehicle crash/safety design example. A quadratic polynomial is used for the prior mean function in all cases. All input variables are normalized to the range [0, 1]. In light of the computational issues brought up in Section III.B, a two-stage strategy is applied in all cases. In the first stage, the Kriging model with one d.o.f. density function on each dimension is fitted. Based on the density function derived from the first stage, reasonable bounds for the Kriging model with nonstationary covariance are used in the second stage. In Section IV.F, we illustrate the effectiveness of our method in quantifying prediction uncertainty.

A. One-Dimensional Example

First consider the same example (Fig. 1) used in Section I, in which 17 sampling points are used to fit the Kriging model with a quadratic prior mean. The mathematical form of the true function is Function 11 (Eq. A11) in Appendix A. As observed earlier, the conventional Kriging model with stationary covariance fails to capture the varying level of smoothness of the function. Using the nonstationary Kriging method proposed in this work, two Kriging models are built by using density functions of different d.o.f. Fig. 5 and Fig. 6 show respectively the results of the fitted Kriging model and the density function used for both cases with our proposed method. In Fig. 5(b) and Fig. 6(b), the stationary density function (a horizontal straight line) used for the conventional Kriging is also provided for comparison. The difference between Fig. 5 and Fig. 6 is that the former has only two d.o.f. (K=1) for the density function, while the latter has nine d.o.f. (K=8). The hyperparameters $\hat{\eta}_k^{(l)}$ (hence the density functions) are estimated for each Kriging model by using the MLE approach discussed in Section III.B.

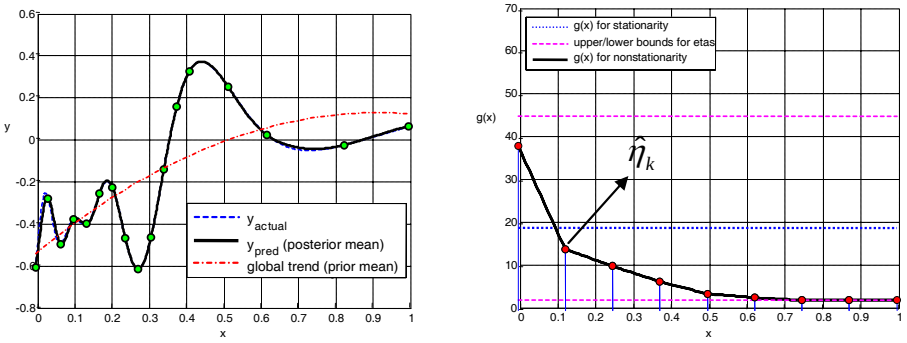
To assess the accuracy of the fitted Kriging models, the response was predicted for 1000 evenly spaced test points generated over the interval [0, 1]. Table (4.1) provides the accuracy comparison of the three Kriging models. Three well-known accuracy metrics are employed, namely the R^2 (R-square), the RMSE (Rooted Mean Square Error) and the RAME (Relative Absolute Max Error). For R^2, the larger the better; for RMSE and RAME, the smaller the better. Formulations of these metrics are provided in Appendix B.

From Fig. 5(b) it is noted that the single linear mapping function identified by maximizing the MLE has a negative slope, i.e., density function is smaller in the right half than in the left half. This properly reflects the fact that the left half region has higher roughness than the right part. Fig. 5(a) indicates that prediction gaps still exist when using the single density function. The model in Fig. 6 uses a slightly more complicated piecewise density function as well as wider bounds for $\hat{\eta}_k^{(l)}$. The density function adapts to the local behavior more closely and leads to almost perfect predictions.



(a) Kriging prediction (b) Estimated density function

Figure 5. Results of nonstationary Kriging with single linear density function (K= 1)



(a) Kriging prediction (b) Estimated density function

Figure 6. Results of nonstationary Kriging with single linear density function (*K*= 8)

The accuracy comparison across the above three metamodels is summarized in Table 1. The worst performance comes with the conventional Kriging with the stationary covariance shown in Fig. 1, where R^2 is 0.9196, RMSE is 0.0743, and RAME is 0.7523. For the nonstationary Kriging model with the single linear density function in Fig. 5(b), the accuracy is significantly improved. R^2 increases to 0.9771, while RMSE and RAME drop to 0.0396 and 0.4401 respectively. The best accuracy is offered by the Kriging model with the largest complexity (K=8), which provides enough flexibility and resolution to capture the locally changing smoothness.

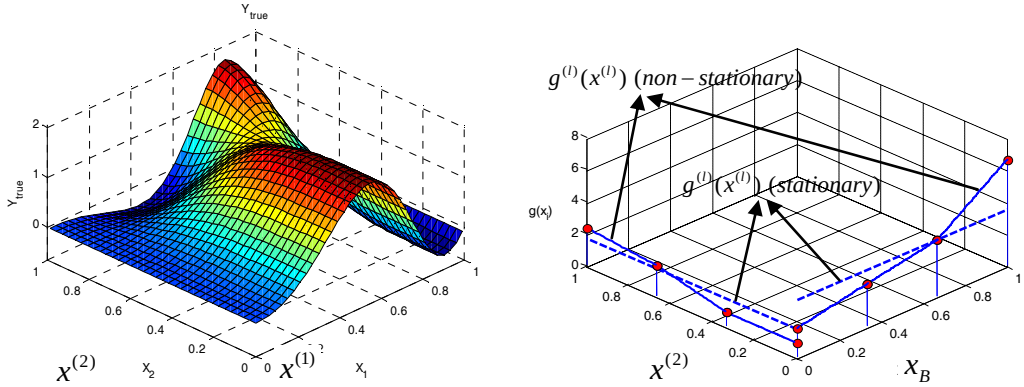
Table 1. The accuracy comparison (* the best values)

Covariance structure	*K*	*d.o.f.*	R^2	*RMSE*	*RAME*
Stationary	N/A	1×1	0.9196	0.0743	0.7523
Nonstationary	1	2×1	0.9771	0.0396	0.4401
	8	9×1	0.9991*	0.0109*	0.2157*

B. Two-Dimensional Example: Evenly Spaced Sampling Points

The second mathematical example (Function 1, Appendix A) is a two-dimensional problem. From the true surface in Fig. 7(a), it is observed that the smoothness of the real performance varies spatially. For the nonstationary Kriging, a density function with 4 d.o.f. (*K*=3) is used for each $x^{(l)}$. In this example, 18 sampling points are generated using the Optimal Latin Hypercube approach[27].

For verification, after fitting the various models based on the 18 sampling points, the response was predicted for a 300×300 grid of input values over the $[0, 1]^2$ region. The accuracy comparison between the stationary Kriging model and nonstationary Kriging models is summarized in Table 2. R^2, RMSE and RAME are all significantly improved, respectively from 0.9591 to 0.9861, from 0.1238 to 0.0723, and from 1.0779 to 0.6058.



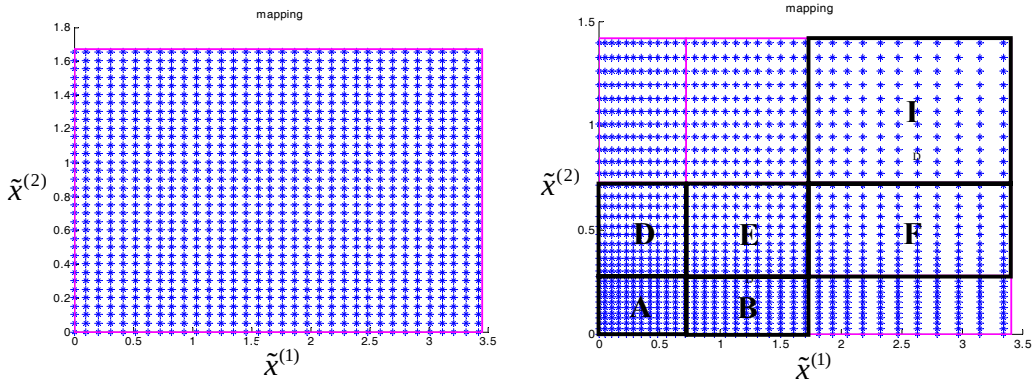
(a) True surface (b) Density functions (stationary and nonstationary)

Figure 7. The true surface and the results of density functions (Function 1)

Table 2. The accuracy comparison (* the best values)

Covariance structure	*K*	*d.o.f.*	R^2	*RMSE*	*RAME*
Stationary	N/A	1×2	0.9591	0.1238	1.0779
Nonstationary	3	4×2	0.9861*	0.0723*	0.6058*

The point mapping plots are shown in Fig. 8, which are the two-dimensional versions of the mapping plot shown in Fig. 2 (b). The point mapping plots allow one to visualize the mapping between the uniformly spaced points in the original space, which are the 34×34 grid points in the $[0, 1]^2$ region, to the new space via the mapping function $\tilde{\mathbf{x}} = \mathbf{f}(\mathbf{x})$. The pattern of mapped points in the new space reflects the interaction of $g^{(1)}(x^{(1)})$ and $g^{(2)}(x^{(2)})$ at different locations. It should be noted that the plots in Fig. 8 show the points in the new space, not the grid points in the original space. Fig. 8(a) is obtained based on the mapping function of the conventional Kriging, where the density functions $g^{(1)}(x^{(1)})$ and $g^{(2)}(x^{(2)})$ are both constant over $x^{(1)}$ and $x^{(2)}$. Since the reference point ξ_0 in Eq. (12) is set to 0, the origins in both spaces are the same point. The mapped points are uniformly spaced in the converted space within [0, 3.43] for $\tilde{x}^{(1)}$ and [0, 1.68] for $\tilde{x}^{(2)}$. Although the constant density functions reflect the average smoothness over $x^{(1)}$ and $x^{(2)}$, they cannot capture the local smoothness. In contrast, the mapped points in Fig. 8(b) via the nonstationary Kriging are non-uniform, ranging within [0, 3.40] for $\tilde{x}^{(1)}$ and [0, 1.45] for $\tilde{x}^{(2)}$.



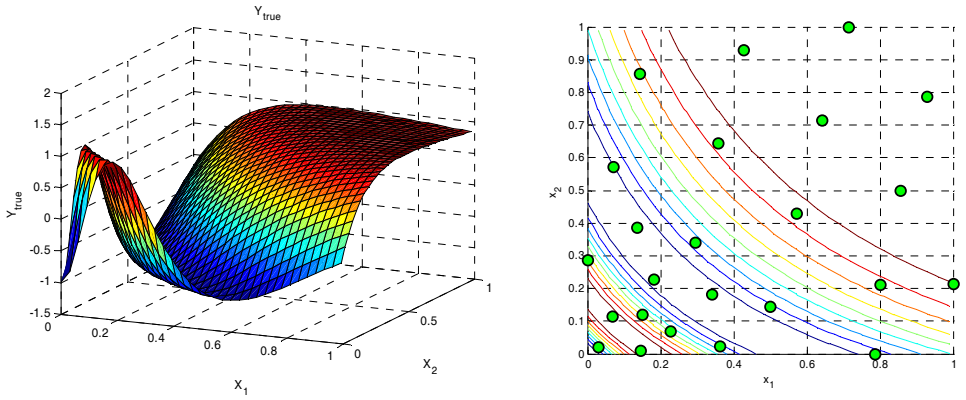
(a) Stationary covariance (b) Nonstationary covariance

Figure 8. The point mapping plot (Function 1)

Characteristics of the proposed nonstationary covariance structure can be revealed by inspecting the mapping plot in Fig. 8(b). Due to the continuity of the density function, the unit region $[0, 1]^2$ is mapped to the new region [0, 3.43] × [0, 1.68]. Moreover, because each density function $g^{(l)}(x^{(l)})$ in $x^{(1)}$ and $x^{(2)}$ is a univariate piecewise linear function, composed of three single linear pieces, the whole region can be divided into nine (3×3) rectangular patches. The point mapping within each patch can be described by two single linear components in $x^{(1)}$ and $x^{(2)}$. Because we use one Kriging model to cover all the nine regions, with those local features incorporated in the nonstationary covariance structure, the model does not suffer from discontinuity as in the tree-based approaches[23]. Relating Fig. 8(b) with Fig. 7(a), it is observed that the density functions can be used to identify areas of larger irregularity (covered by rectangles F and I) and those that are relative smooth (covered by rectangles A, B, D, and E).

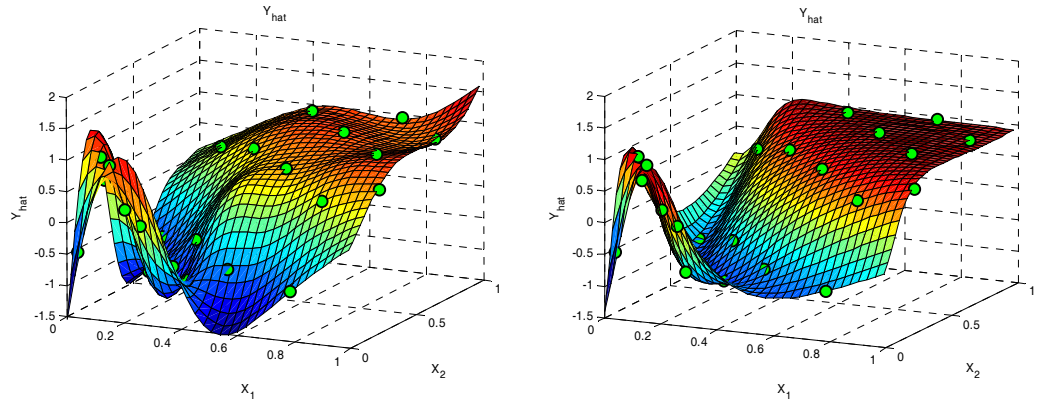
C. Two-Dimensional Example: Adaptive Sampling Points

In the above two-dimensional example, evenly spaced sampling points are generated via optimal design of computer experiments. We also test here the scenario where adaptive sampling points are used.[4,10,14,15] With adaptive sampling, additional points are sequentially placed in regions that are identified with highly nonlinear or irregular behavior. Fig. 9(a) provides the plot of the function considered here (Function 3, Appendix B). Fig. 9(b) shows the 2-D plot of 23 sampling points, which are generated in two sequential stages (16 + 7). Due to the fluctuation of the surface around the origin, more sampling points are placed in the region [0, 0.3] × [0, 0.3], while in the remaining areas, sparse sampling points are used. Based on these 23 sampling points, the stationary Kriging and the nonstationary Kriging models are created for comparison.



(a) True surface (b) Adaptive sampling points

Figure 9. The true surface and adaptive sampling points (Function 3)



(a) Stationary Kriging (b) Nonstationary Kriging

Figure 10. The predicted surface via the two Kriging models (Function 3)

The assessment of the accuracy of the two Kriging models is summarized in Table 3. The accuracy of the nonstationary Kriging model is found to be much better than the stationary Kriging in terms of R^2, RMSE, and RAME.

Table 3. The accuracy comparison (* the better values)

Covariance structure	*K*	*d.o.f.*	R^2	*RMSE*	*RAME*
Stationary	N/A	1×2	0.9346	0.1758	1.0996
Nonstationary	3	4×2	0.9777*	0.1025*	0.8610*

The predicted surfaces via the stationary Kriging and the nonstationary Kriging are shown in Fig. 10, indicating that the stationary Kriging model yields worse prediction in the smoother regions, in which fewer sampling points are placed. In contrast, the nonstationary Kriging shows superior capability in capturing the varying density of sampling points as well as the varying smoothness between regions. From the point mapping plot via the nonstationary covariance in Fig. 11, the unevenly spaced sampling points in the original space appear to be evenly spaced after mapping to the new space. This indicates that using nonstationary Kriging is even more beneficial in the scenario of adaptive sampling as the modeling is adaptive to the smoothness of a response over different regions.

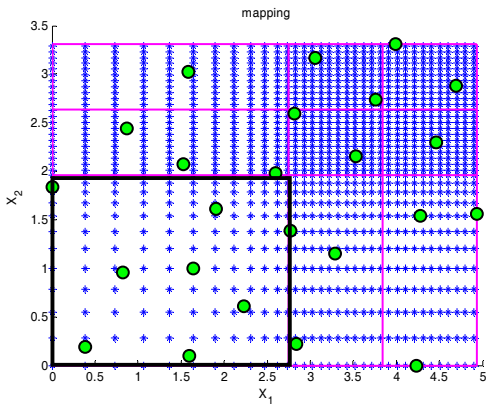


Figure 11. The point mapping plot (circles: relocated sampling points in the new space)

D. A High-Dimensional Example: Vehicle Crash/Safety Models

This example is based on computer data from a vehicle crash simulation model used at General Motors. Due to the destructive nature of physical tests, computer-based approach is playing an increasingly important role in safety engineering. Because the crash simulations are computationally expensive (taking hours to days to run a single simulation), metamodeling is widely used for building surrogate models of crash/safety simulations by automakers. Two responses are considered in our example, namely the HIC (Head Injury Criterion) and the Chest G for the chest acceleration. There are 19 input variables in total for the simulation model. Since each simulation is run at high computational cost, only 101 simulation points are available. With such a limited amount of samplings, the accuracy assessment is conducted through cross validation. Specifically, five-fold cross-validation (CV) is employed in which 80% of the data is used to fit the model, and the remaining 20% is used to verify the model. Due to the high dimension (19), a single linear density function (i.e. K=1) is used for each input variable to limit the number of hyperparamters. As a result, there are $2\times19=38$ unknown hyperparameters in the nonstationary Kriging model. The cross-validation estimate of prediction errors (represented by the average of squared errors) of the stationary Kriging and nonstatioanry Kriging for modeling HIC and CHEST G are summarized in Table 4. It is noted that nonstationary Kriging leads to more accurate prediction (i.e., smaller CV errors) for both responses.

Table 4. The accuracy comparison for modeling HIC and Chest G (* the better values)

Covariance structure	*d.o.f.*	CV_{HIC}	$CV_{Chest\ G}$
Stationary	1×19	679717.80	18.92
Nonstationary	2×19	636628.45*	16.81*

E. Tests via Multiple Functions and Various Sampling Sizes

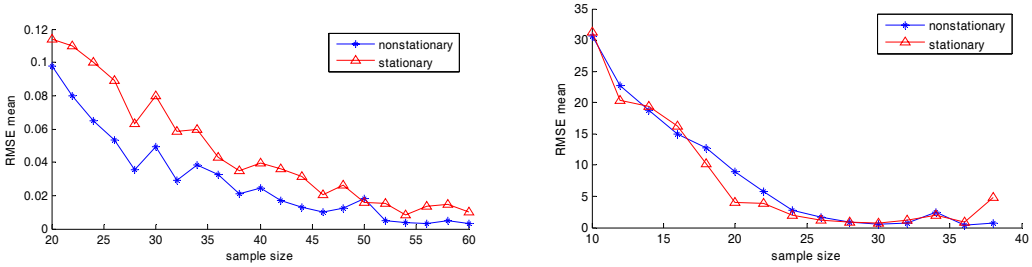
Many factors contribute to the accuracy of a metamodel, e.g., the true response behavior, problem dimension, and the sampling size. It is our interest to test the robustness of the proposed method against various conditions on a set of functions (formulations and 3-D plots of ten testing functions are provided in Appendix A). We note that Functions 1-5 exhibit large changing smoothness; Functions 6-10 appear to be of less variability. We roughly categorize these into two groups, i.e., a nonstationary group and a stationary group.

To compare the robustness with respect to different sampling sizes, ten tests are conducted for each of the selected functions. In each test, we compare the results from using stationary Kriging and nonstationary Kriging. The number of pieces (i.e. K) is chosen at 3 for the nonstationary Kriging. For Functions 1-5, {30, 33, 36, 39, 42, 45, 48, 51, 55, 60} are used as the sampling size for the ten tests; for Functions 6-10, the sampling size follows {15, 17, 19, 21, 23, 25, 27, 29, 31, 33}. Note the sampling size in the latter set is smaller considering that Functions 6-10 are smoother than Functions 1-5. The Optimal Latin Hypercube is used to produce the sampling points throughout all tests; 300×300 grid points are used for accuracy assessment. The results for the tests on Functions 1-5 and Functions 6-10 are summarized in Table 5.

Table 5. The accuracy comparison for Functions 1-10

Fun. No.	1	2	3	4	5	Total
N_{Stat}	1	3	2	2	2	10
$N_{Nonstat}$	9	7	8	8	8	40
Fun. No.	6	7	8	9	10	
N_{Stat}	3	3	3	5	4	18
$N_{Nonstat}$	7	7	7	5	6	32

In Table 5, N_{stat} is used to count the tests in which the stationary Kriging outperforms the nonstationary one; $N_{Nonstat}$ is used to count the tests in which the nonstationary Kriging outperforms the stationary one. For Functions 1-5, the N_{stat} / $N_{Nonstat}$ ratio is 10/40, which means that about 80% of the tests favor the nonstationary Kriging models when the function behavior is nonstationary. For Functions 6-10, the N_{stat} / $N_{Nonstat}$ ratio is 18/32, indicating that the nonstationary Kriging slightly outperforms the stationary Kriging. Results imply that using nonstationary Kriging at least will not deteriorate the prediction.



(a) Function 1 (b) Function 9

Figure 12. Average RMSE vs. sampling size

Fig. 12 illustrates the average RMSE vs. sampling size, for Functions 1 and 9, which are selected to represent the typical nonstationary and stationary functions, respectively. Considering the random nature of Optimal Latin Hypercube, for each sampling size, ten experiment designs are conducted. From Fig. 12(a), the nonstationary Kriging consistently achieves the lower average RMSE across over all the tested sampling sizes. In Fig. 12(b), when the sampling size is not sufficient, the nonstationary Kriging is slightly less accurate than the stationary one; as the sampling size increases, they achieve the same level of accuracy. This indicates that when the sampling size is not sufficient, the proposed method might be misled by the sampling data.

We further test the average performance of our proposed nonstationary Kriging. In these tests, the sampling size (i.e. N_s in Table 6) is chosen as N_s = 35 for Functions 1-5 and N_s = 25 for Functions 6-10. Three Kriging models (one stationary Kriging model; two nonstationary Krigign models with K=1 and 3) are created in each test. Each test is repeated for ten times with the same sampling size (N_s) but different designs of Optimal Latin Hypercube sampling.

Table 6. The RMSE average over ten tests for each function
(the best values; * the second best values)**

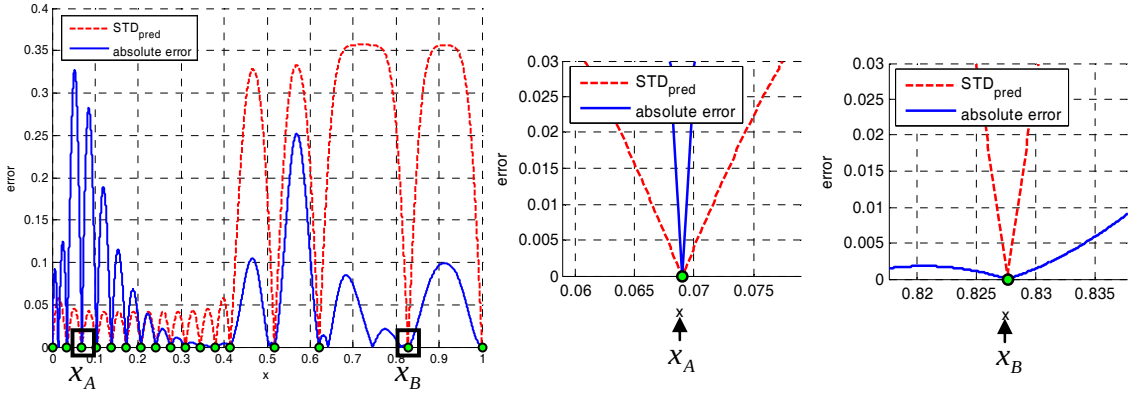
Fun. No.	N_s	average RMSE		
		Stat	Nonstat (K=1)	Nonstat (K=3)
1	35	0.0029	0.0009**	0.0012*
2	35	0.0213**	0.0263*	0.0304
3	35	0.0635	0.0592*	0.0371**
4	35	0.0001*	0.0001*	0.0001**
5	35	0.0029	0.0028*	0.0020**
6	25	0.0106*	0.0321	0.0041**
7	25	0.2553*	0.2375**	0.2782
8	25	0.2882	0.2560*	0.1606**
9	25	3.8846	2.6053*	1.8511**
10	25	1.94E9*	2.13E9	1.93E9**

From Table 6, it is observed that most of the best and second best values of average RMSE result from the nonstationary Kriging model. When the sampling sizes are sufficient for the respective function behaviors, the nonstationary Kriging outperforms the stationary Kriging in majority of the tested functions. It is also observed that the more complicated (K=3) nonstationary Kriging models are generally more accurate than the less complicated (K=1) counterparts.

F. Improvement on Quantifying Prediction Uncertainty

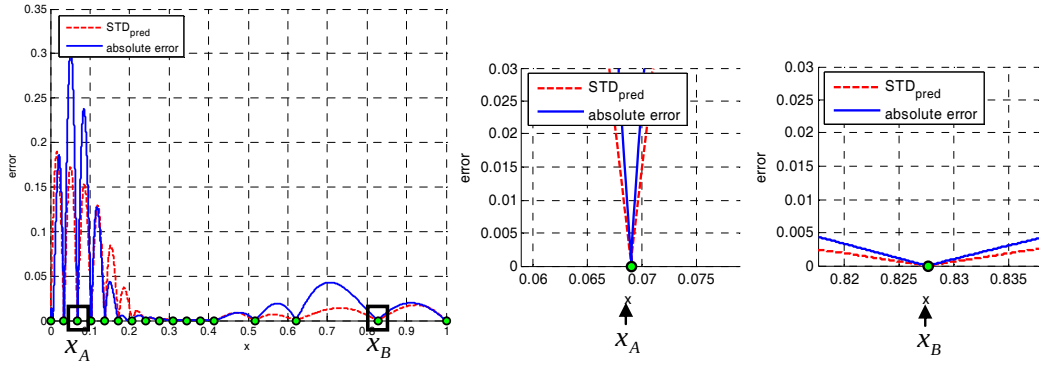
Kriging is considered a very attractive metamodeling technique not only because of its ability to fit a wide variety of functional behavior, but also because of its capability to quantify the prediction uncertainty due to having to fit the metamodel using a limited number of computer simulations. With Kriging, the prediction uncertainty can be analytically derived based on the formula of the *prediction error variance* (denoted as $MSE(\hat{y}(x))$ [4,28]. Ref. 29 used the prediction error variance (and covariance) to quantify metamodel interpolation uncertainty in robust design. In Ref. 30, the prediction error variance is used as one optimality criterion in sequential sampling. In this work, the formula of the prediction error variance from Ref. 4 is used. We are interested in examining whether the use of nonstatioanary covariance could improve the quantification of prediction uncertainty by Kriging model.

The one-dimensional example in Section IV.A (Fig. 6) is used here for demonstrative purpose. To verify whether the theoretical prediction error variance provided by the Kriging modeling approach accurately quantifies the actual prediction error, plots of the actual absolute prediction error (i.e., $|e(x)|$, where $e(x) = y(x) - \hat{y}(x)$) and the theoretical prediction error standard deviation (i.e., $[MSE(\hat{y}(x))]^{1/2}$, denoted as $STD_{pred}(x)$) are provided in Fig. 13 and Fig. 14, respectively for the stationary Kriging and the nonstationary Kriging. In both Fig. 13 and Fig 14, panels (b) and (c) are two zoom-in plots of panel (a) around two selected sampling points x_A(= 0.0690) and x_B(= 0.8276). It is observed from panel (a) that, overall, the actual and theoretical prediction error quantification is in much better agreement for the nonstationary Kriging model than for the stationary Kriging model.



(a) (b) (c)

Figure 13. The theoretical standard deviation of prediction error vs. the actual absolute prediction error yielded by stationary Kriging



(a) (b) (c)

Figure 14. The theoretical standard deviation of prediction error vs. the actual absolute prediction error yielded by nonstationary Kriging

Moreover, from the zoom-in plots in panels (b) and (c) of Fig. 14, we see that the theoretical prediction error standard deviation from the nonstationary Kriging model has different behavior in different regions of the input space, depending on the local level of smoothness in the function. The stationary Kriging model does not possess this desirable characteristic. Specifically, Fig. 13(b and c) show that for the stationary covariance model, when x moves away from sampling points x_A (plot (b)) and x_B (plot (c)), the theoretical $STD_{pred}(x)$ increases at nearly identical rates (the slight slower increase of $STD_{pred}(x)$ around x_A is caused by the fact that more sampling points were placed near x_A than near x_B). This is in disagreement with the fact that the response is known to be smoother in the vicinity of x_B, and hence we would expect the prediction uncertainty to increase at a slower rate as

we move away from x_B. In contrast, the nonstationary Kriging model is able to capture this phenomenon. Fig. 14(b and c) shows that the theoretical prediction uncertainty for the nonstationary Kriging model increases at much higher rates in regions in which the response is rougher than in regions in which the response is smoother.

V. Conclusion

In this paper, we investigate the development of Kriging models with nonstationary covariance structure for metamodeling in the design of engineering systems based on data from computer experiments. Nonstationary covariance based methods in the existing literature suffer from over-parameterization difficulties, which are compounded in high-dimensional problems typical in complex system design optimization. To this end, a modified version of Gibb's nonlinear map approach is proposed, with a sparser, yet flexible, parameterization. Through both mathematical and engineering examples we demonstrate that Kriging modeling based on the proposed nonstationary covariance representation is flexible enough to capture the changing smoothness behavior of the response.

The robustness of the proposed nonstationary Kriging method was also demonstrated via testing multiple functions under different sampling situations. When the performance exhibits obvious varying smoothness levels, the nonstationary Kriging model is able to effectively capture local features and significantly enhance the prediction accuracy. We also demonstrated that for general functions that do not exhibit strong changing smoothness behavior, the nonstationary Kriging model performs no worse than the nonstationary covariance structure. With moderately selected model complexity and sufficient sampling, the robustness of the nonstationary Kriging is reasonably assured.

The nonstatioanry Kriging appears to be even more effective with data obtained from adaptive sequential sampling in which the density of sampling points varies over the regions with different irregularity. We also investigated the benefit of the nonstationary covariance to prediction uncertainty quantification. It is demonstrated that, when the response surface has nonstationary behavior, the nonstationary Kriging yields more effective quantification of prediction error variance than the stationary Kriging,

Despite the argument that our proposed nonstationary covariance function employs as few hyperparameters as possible, the total size of hyperparameters could still be large when the problem dimension is high and larger d.o.f. is desired. In our example problems, density functions with smaller d.o.f. have been used to limit the total number of hyperparameters. Further efforts are needed in identifying the critical dimension where larger d.o.f. is desired and developing more efficient optimization strategies for estimating the hyperparameters to allow for larger d.o.f.

Appendix A

Function 1

$$f(x_1,x_2)=\sin(30(x-0.9)^4)\cos(2(x-0.9))+(x-0.9)/2 \qquad x_1,x_2\in[0,1] \tag{A1}$$

Function 2 ('Mystery Function', Ref. 1)

$$f(x_1,x_2)=2+0.01(x_2-x_1^{\,2})^2+(1-x_1)+2(2-x_2)^2+7\sin(0.5x_1)\sin(0.7x_1x_2) \qquad x_1,x_2\in[0,1] \tag{A2}$$

Function 3 (Ref. 11)

$$f(x_1,x_2)=\sin(1/(x_1*x_2)) \qquad x_1,x_2\in[0.3,1] \tag{A3}$$

Function 4

$$f(x_1,x_2)=x_1\exp(-x_1^{\,2}-x_2^{\,2}) \qquad x_1,x_2\in[-2.5,2.5] \tag{A4}$$

Function 5 (Ref. 30)

$$f(x_1,x_2)=\cos(6(x_1-0.5))+3.1|x_1-0.7|+2(x_1-0.5) +7\sin(1/(|x_1-0.5|+0.31))+0.5x_2 \qquad x_1,x_2\in[0,1] \tag{A5}$$

Function 6

$$f(x_1,x_2)=\cos(5(x_1-0.5))+3.1|x_2-0.7|+2(x_1-0.5)+7\sin(1/|0.5x_2+0.31|) \qquad x_1,x_2\in[0,1] \tag{A6}$$

Function 7 (Ref. 11)

$$f(x_1,x_2)=1.9(1.35+\exp(x_1)\sin(13(x_1-0.6)^2)\exp(-x_2)\sin(7x_2)) \qquad x_1,x_2\in[0,1] \tag{A7}$$

Function 8 ('Six-hump Function', Ref. 31)

$$f(x_1,x_2)=(4-2.1x_1^{\,2}+x_1^{\,4}/3)x_1^{\,2}+x_1x_2+(-1+4x_2^{\,2})x_2^{\,2} \quad x_1\in[-2,2] \quad x_2\in[-1,1] \tag{A8}$$

Function 9 ('Branin Function', Ref. 30)

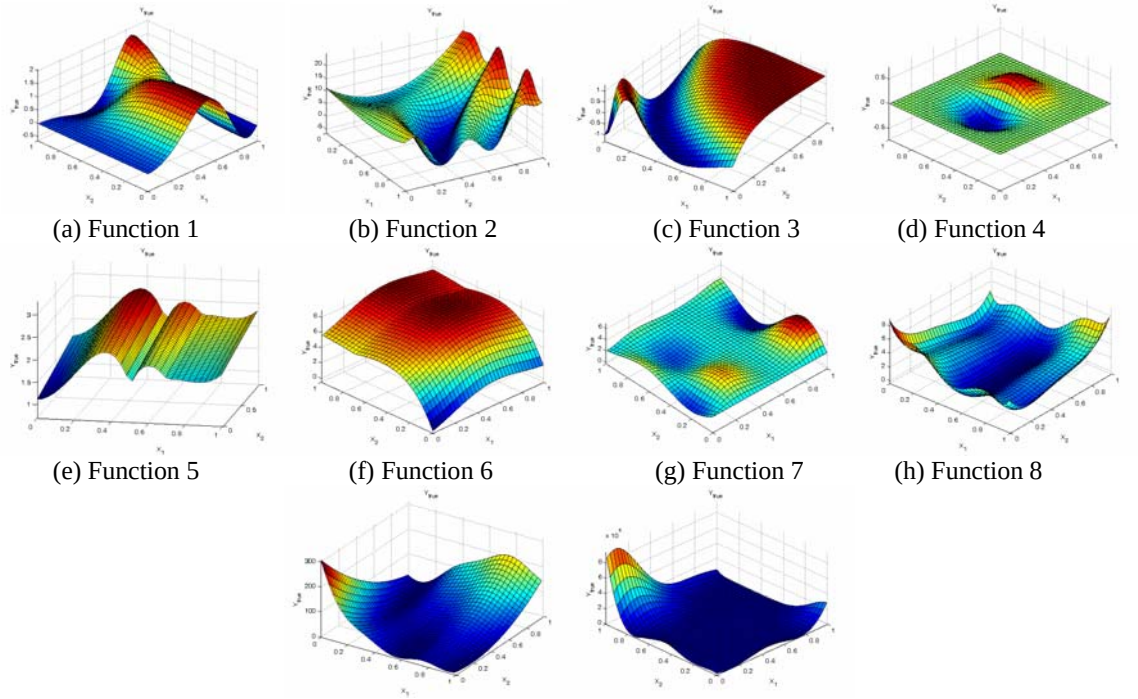
$$f(x_1,x_2)=(x2-5.1x_1^{\,2}/(4\pi^2)+5x_1/\pi-6)^2+10(1-1/(8\pi))\cos(x_1)+10 \quad x_1\in[-5,10] \quad x_2\in[0,15] \tag{A9}$$

Function 10 ('Goldstein-Price Function', Ref. 30)

$$f(x_1,x_2)=[1+(x_1+x_2+1)^2(19-14x_1+3x_1^2-14x_2+6x_1x_2+3x_2^2)]\cdot [30+(2x_1-3x_2)^2(18-32x_1+12x_1^2+48x_2-36x_1x_2+27x_2^2)] \qquad x_1,x_2\in[-2,2] \tag{A10}$$

Function 11

$$f(x)=\sin(30(x-0.9)^4)\cos(2(x-0.9))+(x-0.9)/2 \qquad x\in[0,1] \tag{A11}$$



(a) Function 1 (b) Function 2 (c) Function 3 (d) Function 4

(e) Function 5 (f) Function 6 (g) Function 7 (h) Function 8

(i) Function 9 (j) Function 10

Figure 15. The 3-D plots of Function 1 ~ Function 10
(a~e: there are obvious nonstationary behavior; f~j: there are no obvious nonstationary behavior)

Appendix B

The formulations of R^2, RMSE, and RAME are provided as follows, where N represents the total number of validation points; y_i and $\hat{y}_i$ represent the real and predicted value at validation point, respectively; $\bar{y}$ represents the mean of y_i.

$$R^2=1-\sum_{i=1}^{N}(y_i-\hat{y}_i)^2\Big/\sum_{i=1}^{N}(y_i-\bar{y})^2 \tag{B1}$$

$$RMSE=\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i-\hat{y}_i\right)^2} \tag{B2}$$

$$RAME=\max_{i=1,..,N}\left|y_i-\hat{y}_i\right|\Big/\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i-\bar{y})^2} \tag{B3}$$

Acknowledgement

Grant support form National Science Foundation (DMI-0522662) is greatly appreciated.

References

[1]Friedman, J. H., "Multivariate Adaptive Regression Splines", The Annals of Statistics, Vol. 19, No. 1, 1991, pp. 1-141.

[2]Box, G. E. P., Hunter, W. G., and Hunter, J. S., 1978, Statistics for Experimenters, John Wiley & Sons, NY.

[3]Hardy, R.L., "Multiquadratic equations of topography and other irregular surfaces," J. Geophys. Res., Vol. 76, 1971, pp. 1905-1915.

[4]Sacks, J., Welch, W. J., Mitchell, T. J. and Wynn, H. P., "Design and analysis of computer experiments," Statistical Science, Vol. 4, No. 4, 1989, pp. 409-435.

[5]Jin, R, Chen, W., and Simpson T., "Comparative Studies of Metamodeling Techniques under Multiple Modeling Criteria," Journal of Structural Optimization, Vol. 23, No. 1, 2001, pp. 1-13.

[6]Wang, G., Dong, Z., and Aitchison, P., "Adaptive Response Surface Method — A Global Optimization Scheme for Computation-intensive Design Problems," Journal of Engineering Optimization, Vol. 33, No. 6, 2001, pp. 707-734.

[7]Booker, A. J., Dennis, J. E., Frank, P. D., Serafini, D.B., Torczon, V., and Trosset, M.W., "A rigorous framework for optimization of expensive functions by surrogates", Structural Optimization, Vol. 17, 1999, pp.1-13.

[8]Barthelemy, J.-F. M. and Haftka, R. T., "Approximation Concepts for Optimum Structural Design - A Review," Structural Optimization, Vol. 5, 1993, pp. 129-144.

[9]Simpson, T. W., Peplinski, J., Koch, P. N., Allen, J. K., "Metamodels for Computer-Based Engineering Design: Survey and Recommendations," Engineering with Computers, Vol. 17, No. 2, 2001, pp.129-150.

[10]Currin, C., Mitchell, T., Morris, M., Ylvisaker, D., "Bayesian Prediction of Deterministic Functions, with Applications to the Design and Analysis of Computer Experiments," Journal of the American Statistical Association, Vol. 86, 1991, pp. 953-963.

[11]Paciorek, C.J., "Nonstationary Gaussian Processes for Regression and Spatial Modelling," PhD dissertation, Carnegie Mellon University, Pittsburgh, PA, USA, 2003.

[12]Schmidt, A. M. and O'Hagan, A., "Bayesian Inference for Nonstationary Spatial Covariance Structure via Spatial Deformations", Journal of the Royal Statistical Society, Series B, Vol. 65, 2003, pp. 745-758.

[13]Kim, H. M., Mallick B. K, Holmes C. C., "Analyzing Nonstationary Spatial Data Using Piecewise Gaussian Processes," Journal of the American Statistical Association, Vol. 100, No. 470, 2005, pp. 653-668.

[14]Lin, Y., F. Mistree, J.K. Allen, Tsui K-L, Chen V., "Sequential Metamodeling in Engineering Design," AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, Albany, NY. USA, 2004, Paper Number AIAA-2004-4304.

[15]Farhang Mehr, A., Azarm, S., "Bayesian Meta-Modeling of Engineering Design Simulations: A Sequential Approach with Adaptation to Irregularities in the Response Behavior," International Journal for Numerical Methods in Engineering, John Wiley InterScience Publishing, Vol. 62, No. 15, 2005, pp. 2104-2126.

[16]Santner, T. J., Williams B. J., and Notz W. I., The Design and Analysis of Computer Experiments, Springer-Verlag, New York, 2003.

[17]Sampson, P., Guttorp, P. D., "Nonparametric Estimation of Nonstationary Spatial Covariance Structure," Journal of the American Statistical Association, Vol. 87, 1992, pp. 108-119.

[18]Gibbs, M. N., "Bayesian Gaussian Processes for Regression and Classification," PhD dissertation, University of Cambridge, UK, 1997.

[19]MacKay, D. J. C., "Introduction to Gaussian processes," Neural Networks and Machine Learning, Springer-Verlag, Berlin, Vol. 168, 1998, pp. 133-165.

[20]Higdon, D., "A process-convolution approach to modeling temperatures in the North Atlantic Ocean," Journal of Environmental and Ecological Statistics, Vol. 5, 1998, pp. 173-190.

[21]Pintore, A., Holmes, C. C., "Non-stationary covariance functions via spatially adaptive spectra," Technical Report, University of Oxford, UK, 2004.

[22]Stein, M., "Nonstationary Spatial Covariance Functions," Technical Report No. 21, Center for Integrating Statistical and Environmental Science, University of Chicago, IL, USA, 2005.

[23]Gramacy, R. B., Lee, H. K. H., Macready, W. G., "Parameter Space Exploration with Gaussian Process Trees." Proceedings of the 21st International Conference on Machine Learning, Banff, Canada, 2004, pp. 353-360.

[24]Handcock, M.S., Stein, M.L., "A Bayesian Analysis of Kriging," Technometrics, Vol. 35, No. 4, 1993, pp. 403-410.

[25]Martin, J. D., Simpson, T. W., "On the Use of Kriging Models to Approximate Deterministic Computer Models," AIAA Journal, Vol. 43, No. 4, 2004, pp. 853-863.

[26]Goffe, W.L., Ferrier, G.D., Rodger, J., "Global Optimization of Statistical Functions with Simulated Annealing", Journal of Econometrics, Vol. 60, No. 1/2, 1994, pp. 65-99.

[27]Jin, R., Chen, W., and Sudjianto, A., "An Efficient Algorithm for Constructing Optimal Design of Computer Experiments," Journal of Statistical Planning and Inference, Vol. 134, No. 1, 2005, pp. 268-287.

[28]O'Hagan, A., "Some Bayesian numerical analysis," Bayesian Statistics 4, J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (eds.), Oxford University Press, 1992, pp. 345-363.

[29]Apley, D. W., Liu, J., Chen, W., "Understanding the Effects of Model Uncertainty in Robust Design with Computer Experiments," *ASME Journal of Mechanical Design*, Vol. 128, 2006, pp. 745-958.

[30]Jin, R., Chen, W., Sudjianto, A., "On Sequential Sampling for Global Metamodeling in Engineering Design", Proceedings of ASME Design Automation Conference, Montreal, Canada, Sept.- Oct. 2002, Paper Number DETC2002/DAC-34092.

[31]Sasena, M., "Efficiency and Flexibility Enhancements for Constrained Global Design Optimization with Kriging Approximations," PhD Dissertation, University of Michigan, Ann Arbor, MI, USA, 2002.