INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN ENGINEERING
*Int. J. Numer. Meth. Engng* 2007; **71**:733–756
Published online 29 December 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/nme.1969

# A non-stationary covariance-based Kriging method for metamodelling in engineering design

Ying Xiong[1], Wei Chen[1,*,†,‡], Daniel Apley[2] and Xuru Ding[3]

[1]*Department of Mechanical Engineering, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208-3111, U.S.A.*
[2]*Department of Industrial Engineering and Management Sciences, Northwestern University, 2145 Sheridan Road, Evanston, IL 60208-3111, U.S.A.*
[3]*General Motors Corp., Warren, MI, U.S.A.*

SUMMARY

Metamodels are widely used to facilitate the analysis and optimization of engineering systems that involve computationally expensive simulations. Kriging is a metamodelling technique that is well known for its ability to build surrogate models of responses with non-linear behaviour. 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 non-stationary 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 non-stationary covariance structure is incorporated into Kriging modelling for computer simulations. To represent the non-stationary covariance structure, we adopt a non-linear mapping approach based on parameterized density functions. To avoid over-parameterizing for the high dimension problems typical of engineering design, we propose a modified version of the non-linear 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.


Received 14 March 2006; Revised 13 November 2006; Accepted 13 November 2006

KEY WORDS: Kriging; Gaussian process model; metamodelling; non-stationary covariance; computer experiments; engineering design; prediction uncertainty

*Correspondence to: Wei Chen, Department of Mechanical Engineering, Northwestern University, Evanston, IL 60208-3111, U.S.A.
†E-mail: weichen@northwestern.edu
‡Associate Professor.

Contract/grant sponsor: National Science Foundation; contract/grant number: DMI-0522662

## 1. 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 metamodelling techniques [1–8], Kriging is considered powerful and flexible for building surrogate models (or metamodels) of simulated response surfaces with different functional forms [9, 10]. 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. Currin *et al.* [11] 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 [12]. This is a convenient assumption that simplifies the analysis and lessens the amount of prior information required [11]. 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 behaviour 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 [13, 14]. In those cases, the assumption of the uniformity of smoothness is not well satisfied. One such function is illustrated in Figure 1, in which the roughness in the 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 non-stationary one. The idea of using non-stationary covariance in metamodelling can be found in the engineering design literature [15, 16]. However, all existing works consider the irregular performance behaviour only in sampling but not in fitting

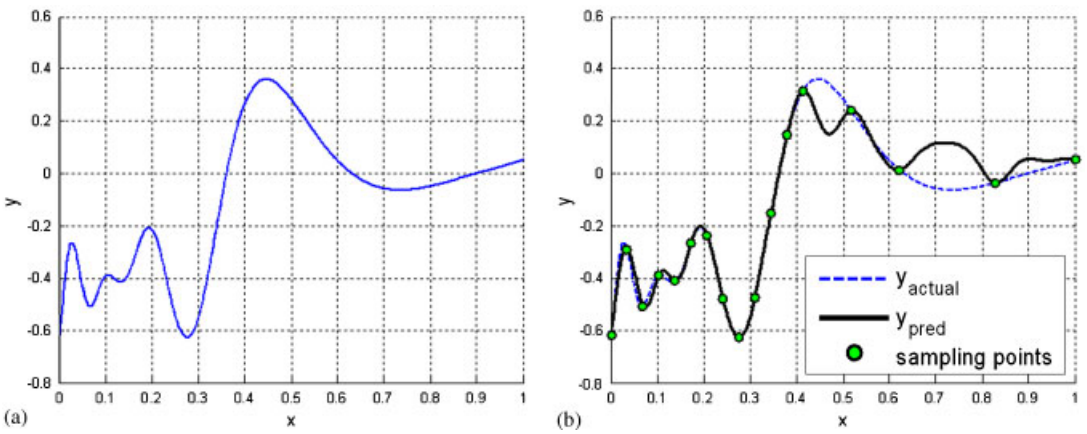


Figure 1. Example of Kriging prediction with stationary covariance: (a) true function; and (b) Kriging prediction (stationary covariance).



DOI: 10.1002/nme

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 non-stationary covariance for choosing sample sites. While samples generated by these methods tend to be non-uniform in the input space, all these existing work use 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.

Non-stationary 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 [17]. Various approaches have been proposed to formulate the non-stationary covariance structure. Sampson and Guttorp [18] developed a spatial deformation approach to reallocate all the points. Gibbs [19] and MacKay [20] proposed two methods of representing the non-stationary behaviour: the first is to directly formulate a non-stationary covariance function, and the second is the so-called non-linear map approach, in which the original Euclidean space is mapped to a new one for which the covariance can be approximated as stationary. Paciorek [12] used a process convolution approach extended from Higdon's [21] method and Gibbs' first method. Other related works could be found in Schmidt and O'Hagan [13], Pintore and Holmes [22], and Stein [23]. Gramacy *et al.* [24] presented an approach that utilizes Gaussian process trees to implement the non-stationary Gaussian process. However, discontinuity of the response across subregions cannot be avoided.

Little prior work has been conducted on non-stationary covariance modelling for complex system design based on computer experiments. This is most likely because complex design problems are often high-dimensional, and non-stationary covariance functions tend to be overparameterized in high dimensions. In this paper, we develop an efficient method that allows non-stationary covariance in Kriging metamodelling for high-dimensional engineering applications with computer experiments. We use a non-linear mapping approach to represent the non-stationary 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 non-stationary covariance modelling 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 non-stationary covariance functions in high dimensions. Second, we propose a modified version of Gibbs [19] non-linear map 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 2, a review of Kriging modelling with a stationary covariance function is first provided; the Gibbs' non-linear map approach is then introduced. Details of our proposed approach are provided in Section 3. Computational issues related to the optimization strategies for estimating the hyper parameters is also addressed. In Section 4, 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 metamodels 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 5.



DOI: 10.1002/nme

## 2. TECHNOLOGICAL BASE

### *2.1. Kriging metamodelling with a stationary covariance function*

In the conventional Kriging model [4] the performance $y(x)$ is modelled as

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

where $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{h}(\mathbf{x})$ 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 which captures local variations. From a Bayesian perspective [11, 25], 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:

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

where $\rho_{\mathrm{stat}}$ is the correlation function. The hyperparameter set $\Theta$ is composed of $\{\sigma^2; \boldsymbol{\theta}\}$. A frequently used Gaussian correlation function is

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

The variance $\sigma^2$ provides the overall vertical scale relative to the mean of Gaussian process in the output space, $\boldsymbol{\theta} = \{\theta^{(l)}\ (l = 1, 2, \ldots, L)\}$ 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 $\rho_{\mathrm{stat}}(\mathbf{x}_m, \mathbf{x}_n; \boldsymbol{\theta})$ 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 Equations (2) and (3), the subscript 'stat' means 'stationary'.

### *2.2. Existing methods using non-stationary covariance: the non-linear map approach*

Various approaches exist in literature to represent non-stationary covariance structures. The non-linear map method [19] is attractive among others because it is intuitively interpretable with a notion of space mapping. In this section, the non-linear map approach in the existing literature is briefly reviewed to set the background of our proposed method. To illustrate the idea of a non-linear map, a simple one-dimensional example is provided in Figure 2. A mapping function $f(\mathbf{x})$ is defined by integrating a density function $g(\mathbf{x})$ (see Equation (4) for details). It can be seen in the original space (Figure 2(a)) that the true function, denoted as $y(\mathbf{x})$, is hard to be modelled 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 $|\tilde{\mathbf{x}}_{\mathrm{D}} - \tilde{\mathbf{x}}_{\mathrm{C}}|$ 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



DOI: 10.1002/nme

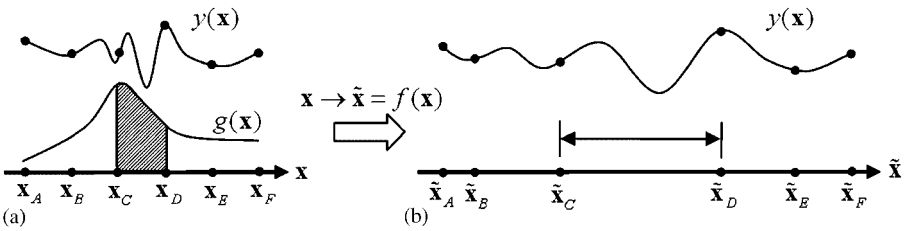


Figure 2. A conceptual illustration of the non-linear map: (a) original space; and (b) new space.

words, the relation between the density function $g(\mathbf{x})$ and the smoothness of the real response can be established.

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 non-stationary covariance can be strictly guaranteed. This is one significant characteristic of the approaches based on map and space deformation. In Gibbs' approach (refer to Paciorek [12] 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 $\tilde{\mathbf{x}} = \mathbf{f}(\mathbf{x}) = (f^{(1)}(\mathbf{x}), f^{(2)}(\mathbf{x}), \ldots, f^{(L)}(\mathbf{x}))$, where $f^{(l)}(\mathbf{x})$ defines the $l$th co-ordinate 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 $g^{(l)}(\mathbf{x})$

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

where $L$ is the number of input variables; $\mathbf{x}_0 = (x_0^{(1)}, \ldots, x_0^{(L)})$ is a reference vector, which is often chosen to be somewhere in the centre 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 $\omega_j = \mathrm{e}^{\alpha_j^{(l)}}$. The integral in Equation (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 centre 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.



DOI: 10.1002/nme

Applying the mapping function Equation (4) to the stationary covariance function in Equation (1), the non-stationary version of the covariance/correlation function in Equation (2) is obtained by

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

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

## 3. A PROPOSED NON-STATIONARY COVARIANCE STRUCTURE

Drawbacks of Gibbs' non-linear map method are immediately observable. From Equations (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 non-linear map method, the non-stationary 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 centres, and it should be large enough so that the non-stationary structure is able to cover the design space. This formulation is affordable in low-dimensional situations with small $J$ 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 non-stationary 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.

### *3.1. Proposed density function*

The non-stationary structure is simplified by assuming that the varying smoothness behaviour 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 Equation (4) becomes

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

where $x_0^{(l)}$ is the reference point. Instead of using the non-linear form in Gibbs' 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 2.2. For a selected number of pieces $K$, $g^{(l)}(x^{(l)})$ is defined as a summation of $K$ linear components

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



DOI: 10.1002/nme

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

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

where $\{a_k^{(l)}, b_k^{(l)}\}$ are the linear parameters for the $k$th linear component; $\{\xi_0, \xi_1, \ldots, \xi_K\}$ are a series of knots placed along input variable $x^{(l)}$. Normally, the knots $\zeta_k$ $(k = 1, \ldots, K)$ are placed evenly along each dimension, unless prior knowledge or experience is available. Imposing the continuity constraints to $g^{(l)}(x^{(l)})$, Equation (9) can be reformulated using the following linear substitutions:

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

The original linear parameters $\{a_k^{(l)}, b_k^{(l)}\}$ $(k = 1, 2, \ldots, K)$ are now replaced by $K + 1$ new parameters $\{\eta_0^{(l)}, \eta_1^{(l)}, \ldots, \eta_K^{(l)}\}$. It can be verified that $\{\eta_0^{(l)}, \eta_1^{(l)}, \ldots, \eta_K^{(l)}\}$ are equal to the density values at each knot, respectively, i.e. $g^{(l)}(\xi_k) = \eta_k^{(l)}$, $(k = 0, 1, 2, \ldots, K)$. According to Equation (8), if the reference point is placed at the knot $\xi_0$, the mapping function $f^{(l)}(x^{(l)})$ can be formulated as

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

where $M$ $(0 \leqslant M < K)$ is the index of the knot left-neighbouring $x^{(l)}$, i.e. $\xi_M < x^{(l)} \leqslant \xi_{M+1}$. The number of parameters or the degrees of freedom (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 Equation (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 Figure 3.

Constraining $g^{(l)}(x^{(l)})$ is equivalent to imposing a bound to $\eta_k^{(l)}$ since $g^{(l)}(x^{(l)})$ is linear. For instance, $\mathrm{lb} < \eta_k^{(l)} < \mathrm{ub}$, for $k = 0, 1, \ldots, K$, is equivalent to $\mathrm{lb} < g^{(l)}(x^{(l)}) < \mathrm{ub}$. When $\eta_0^{(l)} = \cdots = \eta_{K-1}^{(l)} = \eta_K^{(l)}$, $g^{(l)}(x^{(l)})$ is essentially a constant, hence the non-stationary covariance reduces to the stationary one.

The number of knots $\xi_k$ (which is $K + 1$) reflects the resolution of the density function and eventually the complexity of the whole model. It is noted that the resolution of the density function along each dimension may differ: larger resolution could be assigned to variable dimension with higher sensitivity. The selection of $K$ (and/or the placement of the knots $\zeta_k$'s) could be guided by using cross-validation (CV) or more analytical criteria like the Akaike information criterion (AIC) [26]. For illustrative purposes in this paper, we treat $K$ as a specified value.



DOI: 10.1002/nme

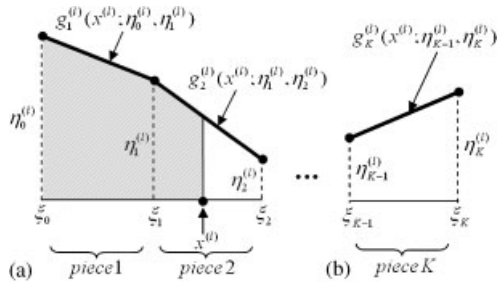


Figure 3. Continuous piecewise linear density function.

### *3.2. Determining the hyperparameters*

Various methods exist for estimating the Kriging hyperparameters. One way is to perform the integration over $\Theta$ using Monte Carlo methods [19, 27]. The method identifies the best values of the hyperparameters as the mean, median, or the mode of the posterior distribution of $\Theta$. In an alternative approach, the most probable value of $\Theta$ is identified by maximizing the maximum likelihood estimation (MLE) [20, 28]. Martin *et al.* [29] compared the use of the MLE and the CV. In our work, the MLE method is used to estimate the hyperparameters $\Theta = \{\sigma^2; \eta_k^{(l)}\ (l = 1, \ldots, L; k = 0, \ldots, K)\}$ for the non-stationary covariance. The likelihood function for the $N$ observations $\{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)\}$, given the Gaussian process parameters $\boldsymbol{\beta}$ and $\Theta$, is

$$L(\boldsymbol{\beta}, \Theta) = L(\boldsymbol{\beta}, \sigma^2, \eta_0^{(1)}, \ldots, \eta_K^{(L)})$$

$$= \frac{1}{\sqrt{(2\pi\sigma^2)^N |\mathbf{R}_{\text{non-stat}}|}} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{H}\boldsymbol{\beta})^{\mathrm{T}} \mathbf{R}_{\text{non-stat}}^{-1} (\mathbf{y} - \mathbf{H}\boldsymbol{\beta})\right\} \tag{13}$$

Satisfying the conditions $\partial L/\partial \boldsymbol{\beta} = 0$ and $\partial L/\partial \sigma^2 = 0$, the MLE of $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ are analytically expressed as functions of $\eta_0^{(1)}, \ldots, \eta_K^{(L)}$, i.e. $\hat{\boldsymbol{\beta}} = (\mathbf{H}^{\mathrm{T}} \mathbf{R}_{\text{non-stat}}^{-1} \mathbf{H})^{-1} \mathbf{H}^{\mathrm{T}} \mathbf{R}_{\text{non-stat}}^{-1} \mathbf{y}_0$ and $\hat{\sigma}^2 = (1/N)(\mathbf{y}_0 - \mathbf{H}\hat{\boldsymbol{\beta}})^{\mathrm{T}} \mathbf{R}_{\text{non-stat}}^{-1} (\mathbf{y}_0 - \mathbf{H}\hat{\boldsymbol{\beta}})$, where $\mathbf{H} = [\mathbf{h}(\mathbf{x}_1), \mathbf{h}(\mathbf{x}_2), \ldots, \mathbf{h}(\mathbf{x}_N)]^{\mathrm{T}}$, $\mathbf{y}_0 = [y_1, y_2, \ldots, y_N]^{\mathrm{T}}$. $\mathbf{R}_{\text{non-stat}}$ is the non-stationary correlation matrix of size $N \times N$, with each entry calculated by $\rho_{\text{non-stat}}(\mathbf{x}_m, \mathbf{x}_n)$ as shown in Equation (7), which constitutes the non-stationary covariance matrix $\mathbf{C}_{\text{non-stat}} = \sigma^2 \mathbf{R}_{\text{non-stat}}$. Note the likelihood function in Equation (13) is identical to that from a stationary covariance model if we simply replace $\mathbf{R}_{\text{non-stat}}$ with $\mathbf{R}_{\text{stat}}$. By plugging in $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$, the likelihood function in Equation (13) becomes a function of $\eta_0^{(1)}, \ldots, \eta_K^{(L)}$, i.e. $L(\eta_0^{(1)}, \ldots, \eta_K^{(L)})$. The MLEs of $\eta_0^{(1)}, \ldots, \eta_K^{(L)}$ are obtained by numerically solving an optimization problem, i.e. maximizing $L(\eta_0^{(1)}, \ldots, \eta_K^{(L)})$. Due to the highly non-linear nature of the MLE function, in this work, the simulated annealing method [30, 31] is used for optimization.

Compared to other non-stationary covariance structures, the proposed non-stationary 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



DOI: 10.1002/nme

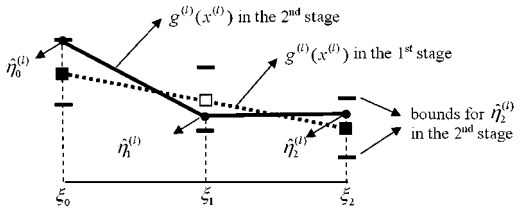


Figure 4. Setting bounds for the to-be-estimated $\eta_k^{(l)}$'s.

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. Figure 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 neighbours are preferred.

## 4. NUMERICAL EXAMPLES AND VERIFICATION

In this section, we present a few mathematical and engineering examples to illustrate and verify the effectiveness of our proposed non-stationary covariance structure for modelling the varying smoothness of different responses. We first illustrate important characteristics of the approach with simple one- and two-dimensional examples. In Section 4.4, 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 3.2, 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 non-stationary covariance are used in the second stage. In Section 4.6, we illustrate the effectiveness of our method in quantifying prediction uncertainty.

### *4.1. One-dimensional example*

First consider the same example (Figure 1) used in Section 1, 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 (Equation (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 non-stationary Kriging method proposed in this work, two Kriging models are built by using density functions of different d.o.f. Figures 5 and 6 show, respectively, the results of the fitted Kriging model and the density function used for both cases with our proposed method.



DOI: 10.1002/nme

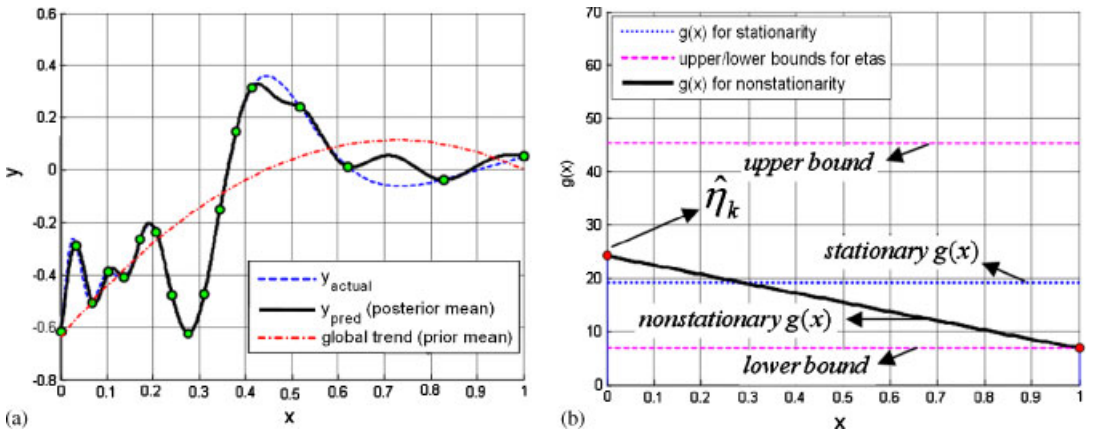


Figure 5. Results of non-stationary Kriging with single linear density function ($K = 1$): (a) Kriging prediction; and (b) estimated density function.

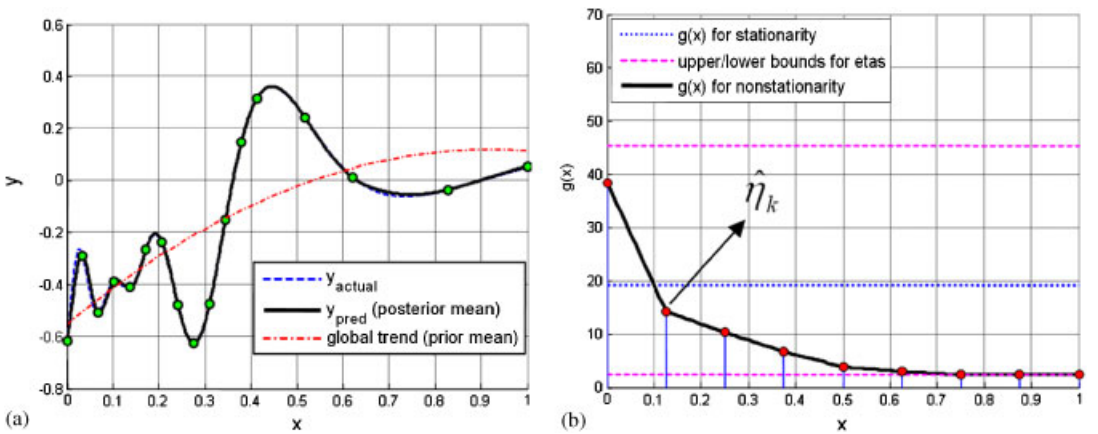


Figure 6. Results of non-stationary Kriging with piecewise linear density function ($K = 8$): (a) Kriging prediction; and (b) estimated density function.

In Figures 5(b) and 6(b), the stationary density function (a horizontal straight line) used for the conventional Kriging is also provided for comparison. The difference between Figures 5 and 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 3.2.

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 I provides the accuracy comparison of the three Kriging models. Three well-known accuracy metrics are employed, namely the $R$-square ($R^2$), the rooted mean square error (RMSE) and the relative absolute max error (RAME). For $R^2$, the larger the better; for RMSE and RAME, the smaller the better. Formulations of these metrics are provided in Appendix B.



DOI: 10.1002/nme

Table I. The accuracy comparison.

| Covariance structure | $K$ | d.o.f. | $R^2$ | RMSE | RAME |
|---|---|---|---|---|---|
| Stationary | N/A | $1 \times 1$ | 0.9196 | 0.0743 | 0.7523 |
| Non-stationary | 1 | $2 \times 1$ | 0.9771 | 0.0396 | 0.4401 |
| | 8 | $9 \times 1$ | 0.9991* | 0.0109* | 0.2157* |

*The best values.

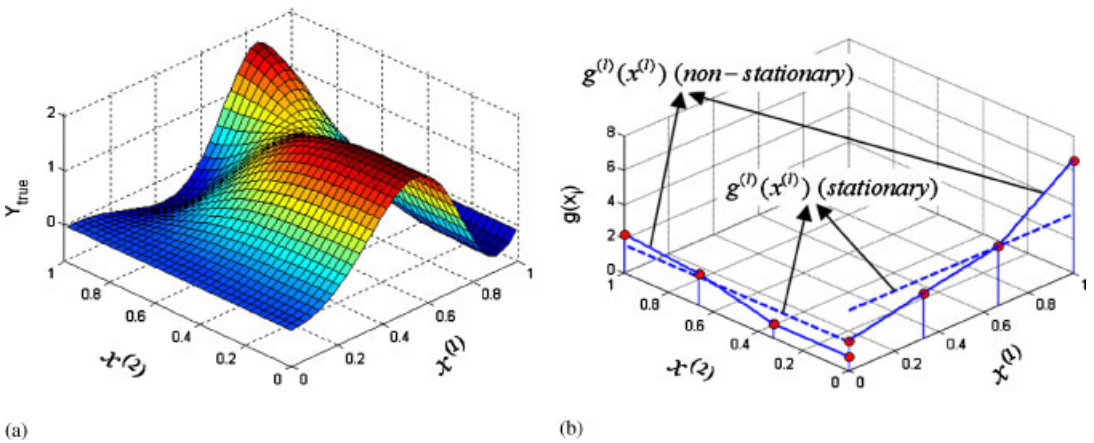


Figure 7. The true surface and the results of density functions (Function 1): (a) true surface; and (b) density functions (stationary and non-stationary).

From Figure 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. Figure 5(a) indicates that prediction gaps still exist when using the single density function. The model in Figure 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 behaviour more closely and leads to almost perfect predictions.

The accuracy comparison across the above three metamodels is summarized in Table I. The worst performance comes with the conventional Kriging with the stationary covariance shown in Figure 1, where $R^2$ is 0.9196, RMSE is 0.0743 and RAME is 0.7523. For the non-stationary Kriging model with the single linear density function in Figure 5(b), the accuracy is significantly improved. The $R^2$ increases to 0.9771, while the 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.

### 4.2. *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 Figure 7(a), it is observed that the smoothness of the real performance varies spatially. For the non-stationary Kriging, a density function with 4 d.o.f. ($K = 3$) is used for



DOI: 10.1002/nme

Table II. The accuracy comparison.

| Covariance structure | $K$ | d.o.f. | $R^2$ | RMSE | RAME |
|---|---|---|---|---|---|
| Stationary | N/A | $1 \times 2$ | 0.9591 | 0.1238 | 1.0779 |
| Non-stationary | 3 | $4 \times 2$ | 0.9861* | 0.0723* | 0.6058* |

*The better values.

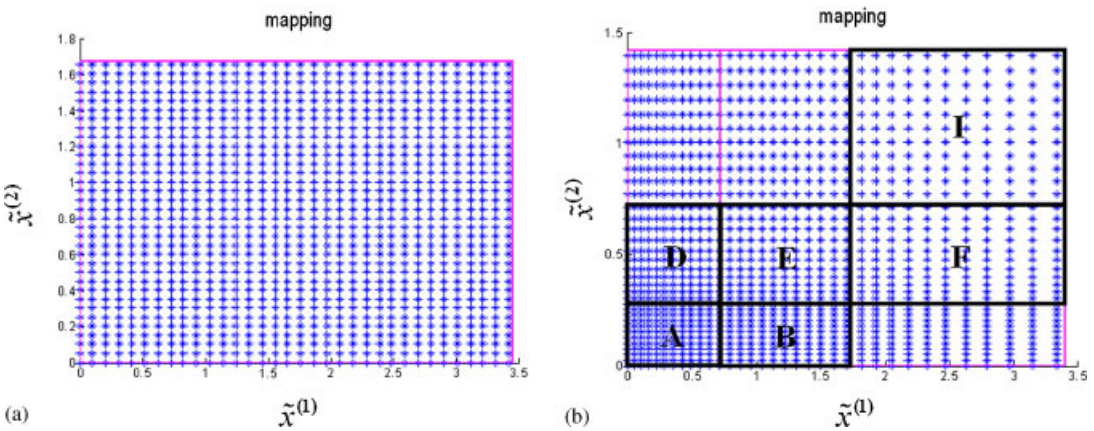


Figure 8. The point mapping plot (Function 1): (a) stationary covariance; and (b) non-stationary covariance.

each $x^{(l)}$. In this example, 18 sampling points are generated using the optimal Latin hypercube approach [32].

For verification, after fitting the various models based on the 18 sampling points, the response was predicted for a $300 \times 300$ grid of input values over the $[0, 1]^2$ region. The accuracy comparison between the stationary Kriging model and non-stationary Kriging models is summarized in Table II. $R^2$, 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.

The point mapping plots are shown in Figure 8, which are the two-dimensional versions of the mapping plot shown in Figure 2(b). The point mapping plots allow one to visualize the mapping between the uniformly spaced points in the original space, which are the $34 \times 34$ 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 Figure 8 show the points in the new space, not the grid points in the original space. Figure 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 $\xi_0$ in Equation (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 Figure 8(b) *via* the non-stationary Kriging are non-uniform, ranging within [0, 3.40] for $\tilde{x}^{(1)}$ and [0, 1.45] for $\tilde{x}^{(2)}$.



DOI: 10.1002/nme

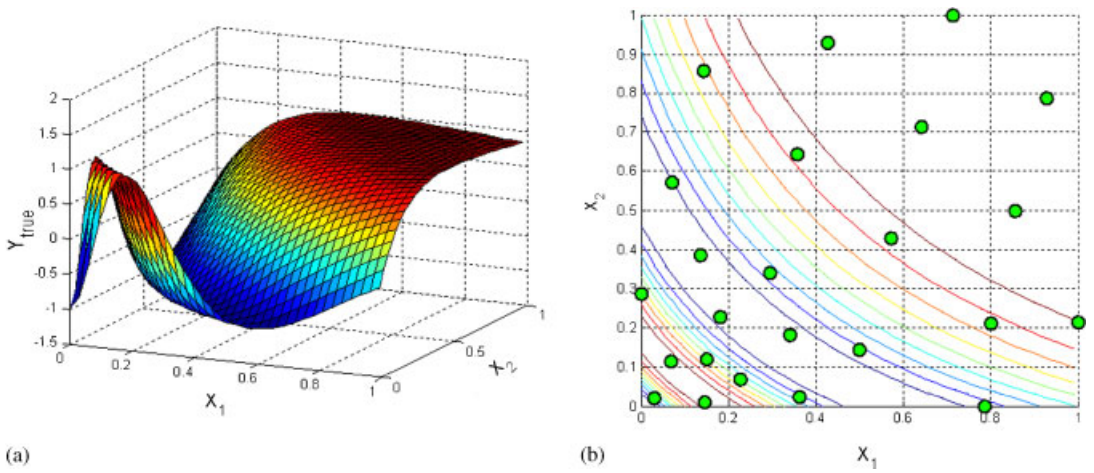


Figure 9. The true surface and adaptive sampling points (Function 3): (a) true surface; and (b) adaptive sampling points.

Characteristics of the proposed non-stationary covariance structure can be revealed by inspecting the mapping plot in Figure 8(b). Due to the continuity of the density function, the unit region $[0, 1]^2$ is mapped to the new region $[0, 3.43] \times [0, 1.68]$. 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 is divided into nine $(3 \times 3)$ rectangular patches. The point mapping within each patch is 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 non-stationary covariance structure, the model does not suffer from discontinuity as in the tree-based approaches (e.g. [24]). Comparing Figure 8(b) with Figure 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).

### *4.3. Two-dimensional example*: *adaptive sampling points*

In the above two-dimensional example, evenly spaced sampling points are generated *via* optimization of computer experiments. We also test here the scenario when adaptive sampling points are used (e.g. [4, 11, 15, 16]). With adaptive sampling, additional points are sequentially placed in regions that are identified with highly non-linear or irregular behaviour. Figure 9(a) provides the plot of the function considered here (Function 3, Appendix B). Figure 9(b) shows the 2-D plot of 23 sampling points, which are generated in two sequential stages $(16 + 7)$. To mimic the adaptive sampling, in the second stage, more sampling points are placed manually in the region $[0, 0.3] \times [0, 0.3]$ where the function fluctuates the most; while in the remaining areas, sparse sampling points are used. Based on these 23 sampling points, the stationary Kriging and the non-stationary Kriging are built for comparison.

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



DOI: 10.1002/nme

Table III. The accuracy comparison.

| Covariance structure | $K$ | d.o.f. | $R^2$ | RMSE | RAME |
|---|---|---|---|---|---|
| Stationary | N/A | $1 \times 2$ | 0.9346 | 0.1758 | 1.0996 |
| Non-stationary | 3 | $4 \times 2$ | 0.9777* | 0.1025* | 0.8610* |

*The better values.

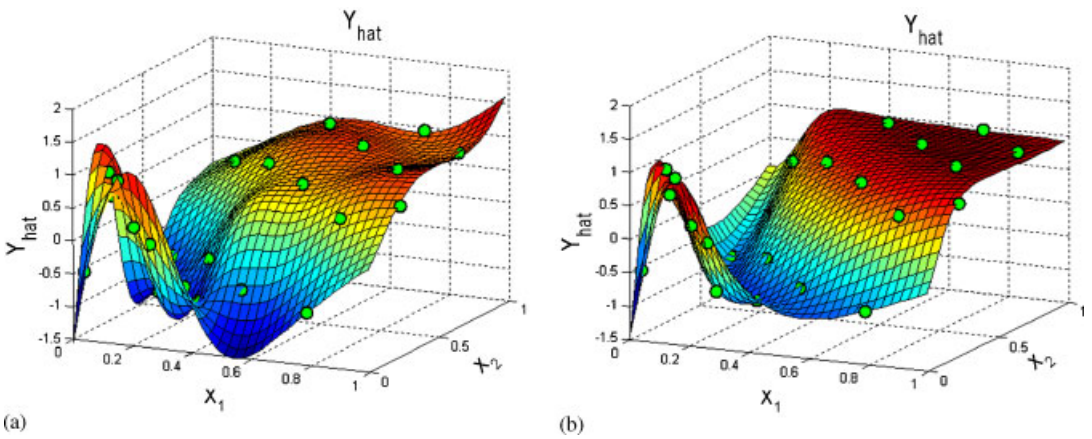


Figure 10. The predicted surface *via* the two Kriging models (Function 3): (a) stationary Kriging; and (b) non-stationary Kriging.

The predicted surfaces *via* the stationary Kriging and the non-stationary Kriging are shown in Figure 10, indicating that the stationary Kriging model yields worse prediction in the smoother regions, in which fewer sampling points are placed. In contrast, the non-stationary 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 non-stationary covariance in Figure 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 non-stationary Kriging is even more beneficial in the scenario of adaptive sampling as the modelling is adaptive to the smoothness of a response over different regions.

### *4.4. A high-dimensional example*: *vehicle crash/safety models*

This example is based on the computer simulation data from a vehicle crash model used in safety optimization 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), metamodelling is widely used for building surrogate models of crash/safety simulations by automakers. Two responses are considered in our example, namely the head injury criterion (HIC) 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 CV. Specifically,



DOI: 10.1002/nme

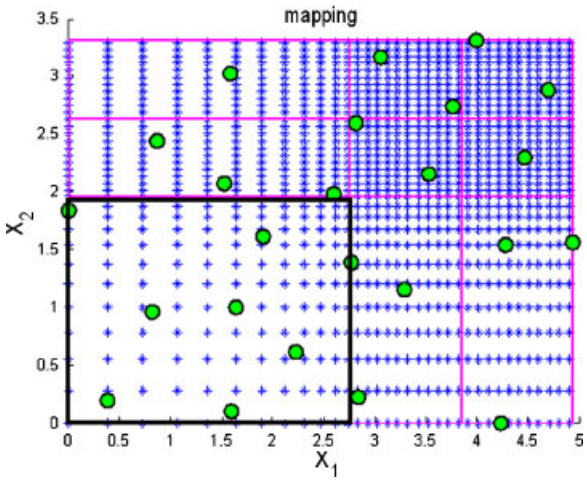


Figure 11. The point mapping plot (circles: relocated sampling points in the new space).

Table IV. The accuracy comparison for modelling HIC and Chest G.

| Covariance structure | d.o.f. | $CV_{HIC}$ | $CV_{Chest\ G}$ |
|---|---|---|---|
| Stationary | $1 \times 19$ | 679717.80 | 18.92 |
| Non-stationary | $2 \times 19$ | 636628.45* | 16.81* |

*The better values.

five-fold 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 [33]. 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 hyperparameters. As a result, there are $2 \times 19 = 38$ unknown hyperparameters in the non-stationary Kriging model, which is twice as many as that in the stationary Kriging model, indicated by the d.o.f. in Table IV. The cross-validation estimate of prediction errors (represented by the average of squared errors) of the stationary Kriging and non-stationary Kriging for modelling HIC and CHEST G are summarized in Table IV. It is noted that non-stationary Kriging leads to more accurate prediction (i.e. smaller CV errors) for both responses.

### *4.5. Tests via multiple functions and various sampling sizes*

Many factors contribute to the accuracy of a metamodel, e.g. the true response behaviour, 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 10 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 non-stationary group and a stationary group.

To compare the robustness with respect to different sampling sizes, 10 tests are conducted for each of the selected functions. In each test, we compare the results from using stationary Kriging and non-stationary Kriging. The number of pieces (i.e. $K$) is chosen at 3 for the non-stationary



DOI: 10.1002/nme

Table V. The accuracy comparison for Functions 1–10.

| Fun. no. | 1 | 2 | 3 | 4 | 5 | Total |
|---|---|---|---|---|---|---|
| $N_{\text{stat}}$ | 1 | 3 | 2 | 2 | 2 | 10 |
| $N_{\text{non-stat}}$ | 9 | 7 | 8 | 8 | 8 | 40 |
| Fun. no. | 6 | 7 | 8 | 9 | 10 | |
| $N_{\text{stat}}$ | 3 | 3 | 3 | 5 | 4 | 18 |
| $N_{\text{non-stat}}$ | 7 | 7 | 7 | 5 | 6 | 32 |

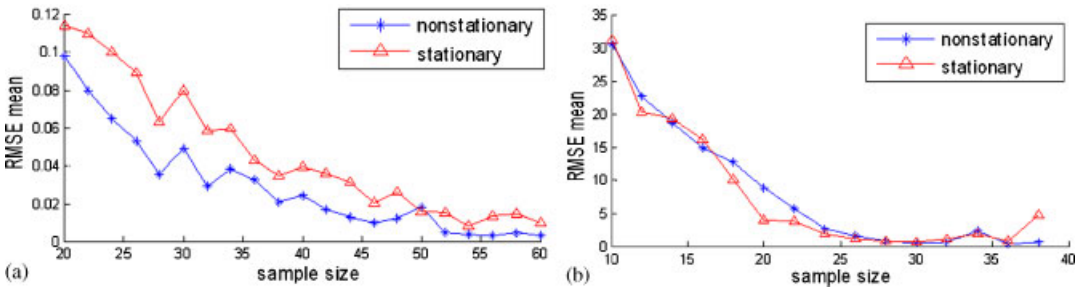


Figure 12. Average RMSE *vs* sampling size: (a) Function 1; and (b) Function 9.

Kriging. For Functions 1–5, {30, 33, 36, 39, 42, 45, 48, 51, 55, 60} are used as the sampling size for the 10 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; $300 \times 300$ grid points are used for accuracy assessment. The results for the tests on Functions 1–5 and Functions 6–10 are summarized in Table V.

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

Figure 12 illustrates the average RMSE *vs* sampling size, for Functions 1 and 9, which are selected to represent the typical non-stationary and stationary functions, respectively. Considering the random nature of optimal Latin hypercube, for each sampling size, 10 experiment designs are conducted. From Figure 12(a), the non-stationary Kriging consistently achieves the lower average RMSE across over all the tested sampling sizes. In Figure 12(b), when the sampling size is not sufficient, the non-stationary 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 non-stationary Kriging. In these tests, the sampling size (i.e. $N_s$ in Table VI) is chosen as $N_s = 35$ for Functions 1–5 and $N_s = 25$ for



DOI: 10.1002/nme

Table VI. The RMSE average over 10 tests for each function.

| | | Average RMSE | | |
|---|---|---|---|---|
| Fun. no. | $N_s$ | Stat | Non-stat ($K=1$) | Non-stat ($K=3$) |
| 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* |

*The best values.
†The second best values.

Functions 6–10. Three Kriging models (one stationary Kriging model; two non-stationary Kriging models with $K=1$ and 3) are created in each test. Each test is repeated for 10 times with the same sampling size ($N_s$) but different designs of optimal Latin hypercube sampling.

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

### *4.6. Improvement on quantifying prediction uncertainty*

Kriging is considered as a very attractive metamodelling technique not only because of its ability to fit a wide variety of functional behaviour, 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, 34]). Apley *et al.* [35] used the prediction error variance (and covariance) to quantify metamodel interpolation uncertainty in robust design. In Jin *et al.* [36], the prediction error variance is used as one optimality criterion in sequential sampling. We are interested in examining whether the use of non-stationary covariance could improve the quantification of prediction uncertainty by Kriging model. In this work, the MSE of the best linear unbiased predictor (BLUP) from Sacks *et al.* [4] is used to assess the prediction error variance

$$\mathrm{MSE}[\hat{y}(x)] = \sigma^2 \left[ 1 - (\mathbf{h}^{\mathrm{T}}(x)\mathbf{r}^{\mathrm{T}}(x)) \begin{pmatrix} \mathbf{0} & \mathbf{H}^{\mathrm{T}} \\ \mathbf{H} & \mathbf{R} \end{pmatrix} \begin{pmatrix} \mathbf{h}(x) \\ \mathbf{r}(x) \end{pmatrix} \right] \tag{14}$$

where $x$ is the site to be predicted; $\mathbf{r}(x)$ is the correlation vector between $x$ and $N$ sampling points $\mathbf{x}_1, \ldots, \mathbf{x}_N$; $\mathbf{R}$ is the correlation matrix of $N$ sampling points. In this work, $\mathbf{r}(x)$ and $\mathbf{R}$ are



DOI: 10.1002/nme

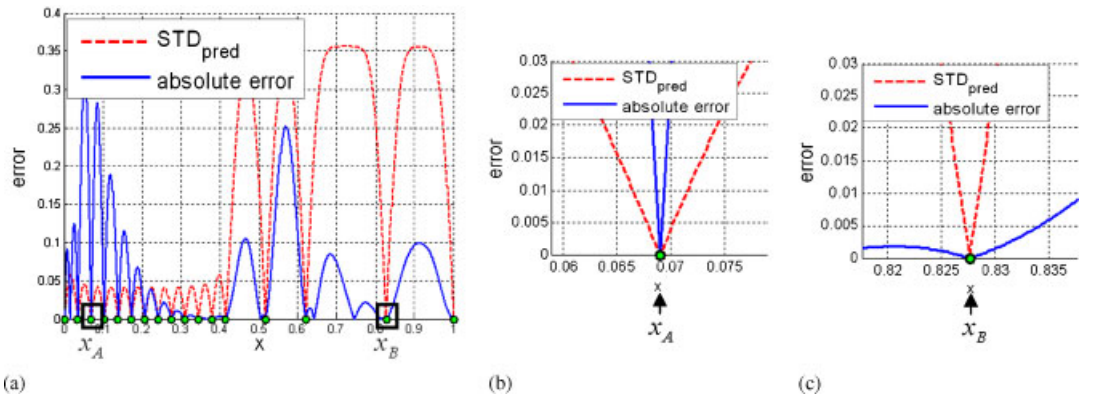


Figure 13. The theoretical standard deviation of prediction error *vs* the actual absolute prediction error yielded by stationary Kriging.

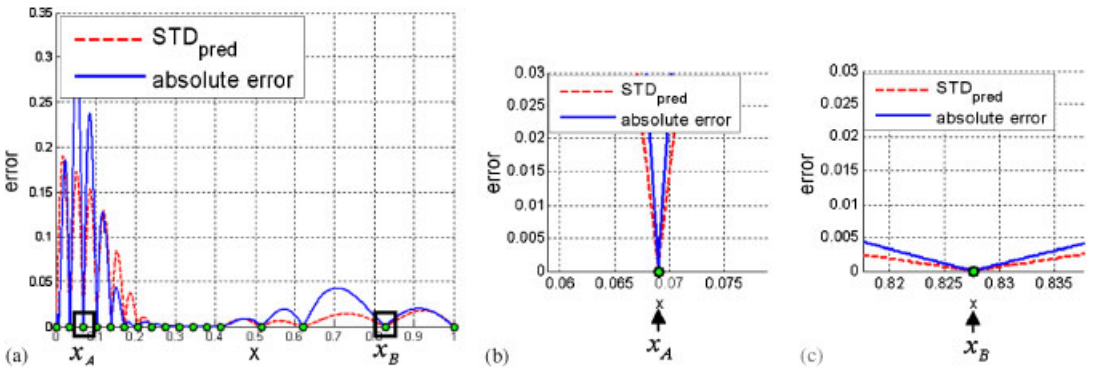


Figure 14. The theoretical standard deviation of prediction error *vs* the actual absolute prediction error yielded by non-stationary Kriging.

calculated by the non-stationary correlation function in Equation (7) instead of the stationary one in Equation (3).

The one-dimensional example in Section 4.1 (Figure 6) is used here for demonstrative purpose. To verify whether the theoretical prediction error variance provided by the Kriging modelling 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. $[\mathrm{MSE}(\hat{y}(x))]^{1/2}$, denoted as $\mathrm{STD}_{\mathrm{pred}}(x)$) are provided in Figures 13 and 14, respectively, for the stationary Kriging and the non-stationary Kriging. In both Figures 13 and 14, panels (b) and (c) are two zoom-in plots of panel (a) around two selected sampling points $x_{\mathrm{A}}$ ($=0.0690$) and $x_{\mathrm{B}}$ ($=0.8276$). It is observed from panel (a) that, overall, the actual and theoretical prediction error quantifications are in much better agreement for the non-stationary Kriging model than for the stationary Kriging model.



DOI: 10.1002/nme

Moreover, from the zoom-in plots in panels (b) and (c) of Figure 14, we see that the theoretical prediction error standard deviation from the non-stationary Kriging model has different behaviour 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, Figures 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 $\mathrm{STD}_{\mathrm{pred}}(x)$ increases at nearly identical rates (the slightly slower increase in $\mathrm{STD}_{\mathrm{pred}}(x)$ 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 non-stationary Kriging model is able to capture this phenomenon. Figures 14(b) and (c) show that the theoretical prediction uncertainty for the non-stationary Kriging model increases at much higher rates in regions in which the response is rougher than in regions in which the response is smoother.

## 5. CLOSURE

In this paper, we investigate the development of Kriging models with non-stationary covariance structure for metamodelling in the design of engineering systems based on data from computer experiments. Non-stationary 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 Gibbs' non-linear map approach is proposed, with a sparser, yet flexible, parameterization. Through both mathematical and engineering examples we demonstrate that Kriging modelling based on the proposed non-stationary covariance representation is flexible enough to capture the changing smoothness behaviour of the response.

By testing over multiple functions under different sampling settings, we demonstrate that, in terms of prediction accuracy, the non-stationary Kriging method shows good robustness against various function behaviours and sampling sizes. When the performance exhibits obvious varying smoothness levels, the non-stationary Kriging model is able to effectively capture local features and significantly enhance the prediction accuracy. We also demonstrate that for general functions that do not exhibit strong changing smoothness behaviour, the non-stationary Kriging model performs no worse than those with non-stationary covariance structures.

The non-stationary 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 irregularities. We also investigate the benefit of the non-stationary covariance to prediction uncertainty quantification. It is demonstrated that, when the response surface has non-stationary behaviour, the non-stationary Kriging yields more effective quantification of prediction error variance than the stationary Kriging.

Despite the argument that our proposed non-stationary 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.



DOI: 10.1002/nme

## APPENDIX A

*Function 1*

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

*Function 2* ('*Mystery Function*', [37])

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

*Function 3* [12]

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

*Function 4*

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

*Function 5* [36]

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

*Function 6*

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

*Function 7* [12]

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

*Function 8* ('*Six-hump Function*', [37])

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

*Function 9* ('*Branin Function*', [36])

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



DOI: 10.1002/nme

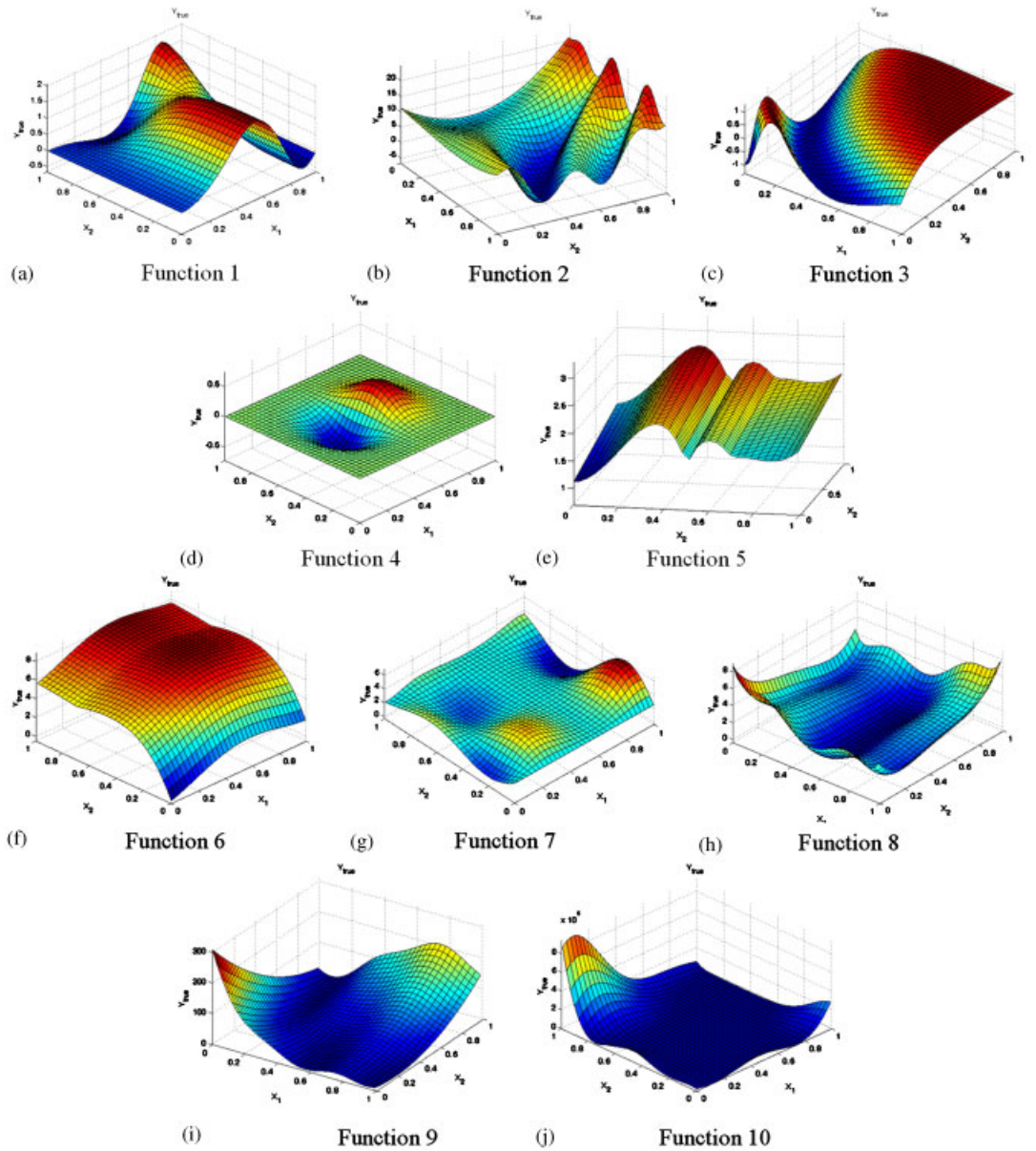


Figure A1. The 3-D plots of Functions 1–10 (a–e: there are obvious non-stationary behaviour; f–j: there are no obvious non-stationary behaviour).

*Function 10* (‘*Goldstein–Price Function*’, [36])

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



DOI: 10.1002/nme

*Function 11*

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

## 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; $\overline{y}$ represents the mean of $y_i$

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

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

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

## NOMENCLATURE

| | |
|---|---|
| $Z(\mathbf{x})$ | the Gaussian process indexed by $\mathbf{x}$ |
| $\boldsymbol{\beta}^{\mathrm{T}}\mathbf{h}(\mathbf{x})$ | the polynomial regression part in Kriging model |
| $C_{\text{stat}}$ | stationary covariance |
| $\rho_{\text{stat}}$ | stationary correlation |
| $C_{\text{non-stat}}$ | non-stationary covariance |
| $\rho_{\text{non-stat}}$ | non-stationary correlation |
| $\boldsymbol{\Theta}$ | hyperparameter set |
| $J$ | number of function basis centres |
| $L$ | 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 |
| $\xi_k$ | knot in the piecewise density function |
| $K$ | number of pieces in the piecewise density function |
| $k$ | index of function piece ($k = 1, 2, \ldots, K$); index of knots or hyperparameters ($k = 0, 1, \ldots, K$) |



DOI: 10.1002/nme

ACKNOWLEDGEMENTS

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

REFERENCES

1. Friedman JH. Multivariate adaptive regression splines. *The Annals of Statistics* 1991; **19**(1):1–141.
2. Box GEP, Hunter WG, Hunter JS. *Statistics for Experimenters*. Wiley: New York, 1978.
3. Hardy RL. Multiquadratic equations of topography and other irregular surfaces. *Journal of Geophysical Research* 1971; **76**:1905–1915.
4. Sacks J, Welch WJ, Mitchell TJ, Wynn HP. Design and analysis of computer experiments. *Statistical Science* 1989; **4**(4):409–435.
5. Jin R, Chen W, Simpson T. Comparative studies of metamodelling techniques under multiple modeling criteria. *Journal of Structural Optimization* 2001; **23**(1):1–13.
6. Wang G, Dong Z, Aitchison P. Adaptive response surface method—a global optimization scheme for computation-intensive design problems. *Journal of Engineering Optimization* 2001; **33**(6):707–734.
7. Booker AJ, Dennis JE, Frank PD, Serafini DB, Torczon V, Trosset MW. A rigorous framework for optimization of expensive functions by surrogates. *Structural Optimization* 1999; **17**:1–13.
8. Barthelemy J-FM, Haftka RT. Approximation concepts for optimum structural design—a review. *Structural Optimization* 1993; **5**:129–144.
9. Wang N, Ge P. Study of metamodelling techniques and their applications in engineering design. *ASME-MED Manufacturing Science and Engineering* 1999; **10**:89–95.
10. Simpson TW, Peplinski J, Koch PN, Allen JK. Metamodels for computer-based engineering design: survey and recommendations. *Engineering with Computers* 2001; **17**(2):129–150.
11. 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* 1991; **86**:953–963.
12. Paciorek CJ. Nonstationary Gaussian processes for regression and spatial modelling. *Ph.D. Dissertation*, Carnegie Mellon University, Pittsburgh, PA, U.S.A., 2003.
13. Schmidt AM, O'Hagan A. Bayesian inference for nonstationary spatial covariance structure via spatial deformations. *Journal of the Royal Statistical Society*, *Series B* 2003; **65**:745–758.
14. Kim HM, Mallick BK, Holmes CC. Analyzing nonstationary spatial data using piecewise Gaussian processes. *Journal of the American Statistical Association* 2005; **100**(470):653–668.
15. Lin Y, Mistree F, Allen JK, Tsui K-L, Chen V. Sequential metamodelling in engineering design. *AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, Paper Number AIAA-2004-4304, Albany, NY, U.S.A., 2004.
16. 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* 2005; **62**(15):2104–2126.
17. Santner TJ, Williams BJ, Notz WI. *The Design and Analysis of Computer Experiments*. Springer: New York, 2003.
18. Sampson P, Guttorp PD. Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association* 1992; **87**:108–119.
19. Gibbs MN. Bayesian Gaussian processes for regression and classification. *Ph.D. Dissertation*, University of Cambridge, U.K., 1997.
20. MacKay DJC. *Introduction to Gaussian Processes*. *Neural Networks and Machine Learning*, vol. 168. Springer: Berlin, 1998; 133–165.
21. Higdon D. A process-convolution approach to modelling temperatures in the North Atlantic ocean. *Journal of Environmental and Ecological Statistics* 1998; **5**:173–190.
22. Pintore A, Holmes CC. Non-stationary covariance functions via spatially adaptive spectra. *Technical Report*, University of Oxford, U.K., 2004.
23. Stein M. Nonstationary spatial covariance functions. *Technical Report No. 21*, Center for Integrating Statistical and Environmental Science, University of Chicago, IL, U.S.A., 2005.
24. Gramacy RB, Lee HKH, Macready WG. Parameter space exploration with Gaussian process trees. *Proceedings of the 21st International Conference on Machine Learning*, Banff, Canada, 2004; 353–360.



DOI: 10.1002/nme

25. Morris M, Mitchell T, Ylvisaker D. Bayesian design and analysis of computer experiments: use of derivatives in surface prediction. *Technometrics* 1993; **35**:243–255.
26. Fuentes M. A new high frequency Kriging approach for nonstationary environmental processes. *Environmetrics* 2001; **12**:469–483.
27. Handcock MS, Stein ML. A Bayesian analysis of Kriging. *Technometrics* 1993; **35**(4):403–410.
28. Li R, Sudjianto A. Analysis of computer experiments using penalized likelihood Gaussian Kriging models. *Technometrics* 2003; **47**:111–120.
29. Martin JD, Simpson TW. Use of Kriging models to approximate deterministic computer models. *AIAA Journal* 2005; **43**(4):853–863.
30. Martin JD, Simpson TW. A Monte Carlo simulation of the Kriging model. *AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, Paper No. AIAA 2004-4483, Albany, NY, U.S.A., 2004.
31. Goffe WL, Ferrier GD, Rodger J. Global optimization of statistical functions with simulated annealing. *Journal of Econometrics* 1994; **60**(1/2):65–99.
32. Jin R, Chen W, Sudjianto A. An efficient algorithm for constructing optimal design of computer experiments. *Journal of Statistical Planning and Inference* 2005; **134**(1):268–287.
33. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning*, *Data Mining*, *Inference and Prediction*. Springer: New York, 2001.
34. O'Hagan A. Some Bayesian numerical analysis. In *Bayesian Statistics 4*, Bernardo JM, Berger JO, Dawid AP, Smith AFM (eds). Oxford University Press: Oxford, 1992; 345–363.
35. Apley DW, Liu J, Chen W. Understanding the effects of model uncertainty in robust design with computer experiments. *Journal of Mechanical Design* 2005; **128**(4):945–958.
36. Jin R, Chen W, Sudjianto A. On sequential sampling for global metamodelling in engineering design. *Proceedings of the ASME Design Automation Conference*, Paper Number DETC2002/DAC-34092, Montreal, Canada, September–October 2002.
37. Sasena M. Efficiency and flexibility enhancements for constrained global design optimization with Kriging approximations. *Ph.D. Dissertation*, University of Michigan, Ann Arbor, MI, U.S.A., 2002.



DOI: 10.1002/nme