



Non-Stationary Kriging For Design Optimization

David John James Toal, Andy J Keane

► To cite this version:

David John James Toal, Andy J Keane. Non-Stationary Kriging For Design Optimization. Engineering Optimization, 2011, pp.1. 10.1080/0305215X.2011.607816 . hal-00734526

HAL Id: hal-00734526

<https://hal.science/hal-00734526>

Submitted on 23 Sep 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Non-Stationary Kriging For Design Optimization

Journal:	<i>Engineering Optimization</i>
Manuscript ID:	GENO-2010-0296.R3
Manuscript Type:	Original Article
Date Submitted by the Author:	15-Jul-2011
Complete List of Authors:	Toal, David; University of Southampton, School of Engineering Sciences Keane, Andy; University of Southampton, Engineering Sciences
Keywords:	Non-stationary, Kriging, Optimization
Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.	
non-stationary surrogate modelling for design optimization4.tex	

SCHOLARONE™
Manuscripts

RESEARCH ARTICLE

Non-Stationary Kriging For Design Optimization

D.J.J. Toal,^{a*} A.J. Keane^a

^aUniversity of Southampton, Southampton, SO17 1BJ, United Kingdom;
(Received 00 Month 200x; final version received 00 Month 200x)

Traditional surrogate modeling techniques, such as kriging, have been employed quite effectively within design optimizations. However, such models can fail to accurately reproduce non-stationary responses. The following paper explores the application of non-stationary kriging to design optimization and attempts to determine its applicability with regard to the optimization of both stationary and non-stationary objective functions. A series of analytical test problems and an engineering design problem are used to compare the performance of non-stationary and adaptive partial non-stationary kriging to traditional stationary kriging.

Keywords: non-stationary kriging; surrogate modeling; optimization;

1. Introduction

The application of surrogate modeling, or metamodeling, techniques within design optimization has grown in popularity in recent years. A surrogate model constructed of an objective function using a relatively small number of initial sample points can be exhaustively searched in place of the true objective function. Promising designs can then be evaluated using the true objective function, the surrogate updated and then searched again. The construction and application of surrogate models is particularly advantageous when the objective function of interest is expensive to evaluate thereby making an exhaustive search, which directly employs the true objective function, unfeasible.

Kriging, first developed by Krige (1951) and popularized by Sacks *et al.* (1989) for the construction of surrogate models of deterministic computational experiments, is a popular surrogate modeling technique due to its ability to represent objective functions accurately whilst providing an error estimate of the predictor. While kriging has proved effective in the optimization of a number of engineering design problems (Jones *et al.* (1998), Sakata *et al.* (2003), D'Angelo and Minisci (2005), Hoyle *et al.* (2006), Forrester *et al.* (2006)), traditional kriging by formulation is stationary in nature and can therefore have difficulty

*Corresponding author. Email: djjt@soton.ac.uk

accurately representing non-stationary responses. This inadequacy may therefore have a detrimental effect on the performance of an optimization when the objective function is non-stationary in nature. Non-stationarity in this case refers to significant changes in the “smoothness” (Xiong *et al.* (2007)) of the function throughout the design space which can only really be captured accurately if the covariance function is permitted to vary over the design space. A simple example of such a function is that used by Xiong *et al.* and is illustrated in Figures 1(b) and 2. Here the function’s “smoothness” can be observed to vary significantly between $x \in [0, 0.3]$ and $x \in [0.3, 1.0]$.

Within the literature there are a number of different strategies for the representation of non-stationary data: the direct formulation of non-stationary covariance functions, the “moving window approach”, the “divide and conquer” approach and non-linear mapping.

The direct formulation of specific non-stationary covariance functions has been implemented by a number of authors in the creation of non-stationary surrogate models (Gibbs (1997), Paciorek and Schervish (2004), Pintore and Holmes (2004)). Paciorek and Schervish, for example, introduced a non-stationary version of the Matérn stationary covariance. Pintore and Holmes decomposed a stationary covariance function using Fourier and Karhunen-Loève expansions to generate a spectrum which evolved over the design space thereby resulting in a non-stationary model. However, as is usually the case with such direct formulations they are quite complex and are normally only applied to sample data with relatively few dimensions.

Moving window approaches approximate non-stationary functions by adopting a stationary model which moves over the input space. Rather than model the whole non-stationary function it is assumed that the region of interest within the window is stationary and can be modelled as such. Haas (1990) employed both this approach and various alternative formulations of kriging within a moving window. However, determining the size of the window is not a trivial operation and additional costs may be introduced when determining the hyperparameters as the window moves through the space. If employed within a design optimization framework this may substantially increase the cost of any exhaustive search of the surrogate. Typically design optimization involves the creation of a surrogate using a relatively sparse data set, the wisdom of creating a local surrogate using an even smaller subset of the data could therefore be questioned.

The divide and conquer approach, otherwise know as a mixture of experts, is similar to the moving window approach in that the non-stationary function is recreated by a local stationary model. Whereas a moving window uses points on either side of the region of interest, the divide and conquer approach splits the design space into a series of fixed subsets within which the response is assumed stationary. The non-stationary function is therefore represented by a patchwork of stationary models. Fuentes (2001) for example represents a non-stationary response as the weighted average of a series of local stationary models. Rasmussen and Ghahramani (2002) and Kim *et al.* (2005) divide the design space into a series of regions each with its own stationary model. Such tiling approaches are adept at reducing the model tuning cost for large datasets as the data within each tile is independent of others. However, as with a moving window, the partitioning of the relatively small sample typical of a design optimization in such a manner may result in poor performance. Depending on the methodology used, such tiled models may also suffer from discontinuities at their boundaries and the optimal position of these boundaries may be difficult to determine.

The non-linear mapping approach, introduced by Sampson and Guttorp (1992), attempts to deform the input space in such a way that the non-stationary response can actually be represented by a stationary model. Sampson and Guttorp proposed the map-

ping of sample points from one space to another via the minimization of a stress function. The resultant mapping was then extended to a smooth function via thin plate splines thereby allowing the entire input space to be mapped. Smith (1996), on the other hand, approached the fitting of a radial basis function representation of the non-linear mapping using a maximization of the likelihood. While perhaps costly when applied to large datasets the non-linear mapping approach could be considered more appropriate when constructing models using a relatively sparse sample set, as is the case in design optimization. The non-linear mapping approach also results in a single continuous model of the non-stationary function which can be readily searched for update points.

Non-stationary modeling techniques, such as those described above, are typically applied within geostatistics (Atkinson and Lloyd (2007)) and in the modeling of environmental processes (Fuentes (2001), Haas (1990, 2002), Smith (1996)). These applications usually consider models of typically two or three variables and utilize relatively large amounts of data. Rarely have non-stationary modeling techniques been applied to problems of the order of those typically encountered during engineering design optimization where problems with tens, or perhaps hundreds of variables, are commonplace. The techniques described above are typically quite difficult to apply successfully to large scale problems, extending the non-linear mapping approach (Smith (1996), Sampson and Guttorp (1992), Gibbs (1997)) to higher dimensions, for example, can result in an over parameterization of the mapping leading to poor results. Recently Xiong *et al.* (2007) proposed a modification of the non-linear mapping approach with the express purpose of applying it to engineering problems and it is this strategy in particular which forms the basis of the non-stationary kriging strategies presented within this paper.

While Xiong *et al.* (2007) considered the suitability of non-stationary kriging with regard to the representation of both stationary and non-stationary analytical test functions, the application of non-stationary kriging within the overall design optimization process was not considered. The following paper aims to address this by comparing the performance of both stationary and non-stationary kriging to the optimization of a series of 11 analytical test functions and an engineering design problem.

The paper commences by introducing stationary kriging which is then expanded to non-stationary kriging employing a variation of the scheme presented by Xiong *et al.* (2007). An in depth study into the accuracy of the surrogate models produced using both stationary and non-stationary kriging is then carried out using 11 analytical test functions with a range of sampling densities. The performance of both modeling strategies within an optimization framework is then considered using the same analytical functions. The conclusions drawn from these tests lead to the consideration of an adaptive partial non-stationary (APNS) kriging model which aims to capture the best features of both wholly stationary and non-stationary kriging. The performance of this strategy is then evaluated once again using the analytical test functions. The paper concludes with the application of the three techniques to a truss design problem.

2. Surrogate Modelling

The formulation of a non-stationary kriging model that employs a non-linear mapping scheme is directly related to that of traditional stationary kriging. The following section therefore presents the formulation of stationary kriging and non-stationary kriging, focusing in particular, on the non-linear mapping scheme of Xiong *et al.* (2007).

2.1. Stationary Kriging

The standard kriging formulation is well documented within the literature (Sacks *et al.* (1989), Jones (2001), Simpson *et al.* (2001), Queipo *et al.* (2005), Forrester and Keane (2009)) and the performance of kriging has been extensively compared to other stationary surrogate modeling techniques (Sakata *et al.* (2003), Jin *et al.* (2001), Peter and Marcelet (2008), Clarke *et al.* (2005)). The complete derivation of the traditional stationary kriging model will therefore not be considered within this paper. Instead, focus is given to the important formulae and those results which are also applicable to non-stationary kriging. For a more complete derivation of kriging the interested reader may wish to consult the literature.

The construction of a Kriging model assumes that when the difference between two design points, \mathbf{x}_i and \mathbf{x}_j is small then the difference between the resulting objective function values, $y(\mathbf{x}_i)$ and $y(\mathbf{x}_j)$, will be small. This can be modelled statistically by assuming that the correlation between two sets of random variables, $Y(\mathbf{x}_i)$ and $Y(\mathbf{x}_j)$ is given by,

$$\text{Corr}[Y(\mathbf{x}_i), Y(\mathbf{x}_j)] = \exp\left(-\sum_{l=1}^d 10^{\theta^{(l)}} \|\mathbf{x}_i^{(l)} - \mathbf{x}_j^{(l)}\|^{\mathbf{p}^{(l)}}\right). \tag{1}$$

where $\theta^{(l)}$ and $\mathbf{p}^{(l)}$ represent the hyperparameters of the l^{th} variable. Values of θ and \mathbf{p} are chosen to maximize the likelihood on the observed dataset, \mathbf{y} , which after simplification (Jones (2001)) becomes,

$$\phi = -\frac{n}{2} \ln(\hat{\sigma}^2) - \frac{1}{2} \ln(|\mathbf{R}|), \tag{2}$$

with the maximum likelihood estimates of variance, $\hat{\sigma}^2$, and mean, $\hat{\mu}$, given by

$$\hat{\sigma}^2 = \frac{1}{n}(\mathbf{y} - \mathbf{1}\hat{\mu})^T \mathbf{R}^{-1}(\mathbf{y} - \mathbf{1}\hat{\mu}) \tag{3}$$

and

$$\hat{\mu} = \frac{\mathbf{1}^T \mathbf{R}^{-1} \mathbf{y}}{\mathbf{1}^T \mathbf{R}^{-1} \mathbf{1}}, \tag{4}$$

where \mathbf{y} is the vector of objective function values. As the concentrated likelihood function, Eq. 2, is dependent only upon the correlation matrix \mathbf{R} , formed from Eq. 1, the kriging hyperparameters can be optimized in order to maximize Eq. 2.

Depending upon the strategy used, the optimization to maximize the likelihood function can prove to be rather costly as each likelihood evaluation requires a $O(n^3)$ factorization of the correlation matrix. A global optimization technique, such as a genetic algorithm, will effectively find the global optimum but will require many more evaluations of the likelihood than, for example, a local gradient descent search. Such local searches may, however, become trapped and not reliably locate the global optimum. Recently, efforts have been made to accelerate the likelihood optimization through the application of a hybridized search scheme. Toal *et al.* (2011), for example, recently formulated an adjoint of the likelihood function and employed it within a hybridized particle swarm.

The adjoint of the correlation matrix,

$$\bar{\mathbf{R}} = \frac{1}{2\hat{\sigma}^2} \mathbf{R}^{-T} (\mathbf{y} - \mathbf{1}\hat{\mu})^T (\mathbf{y} - \mathbf{1}\hat{\mu})^T \mathbf{R}^{-T} - \frac{1}{2} \mathbf{R}^{-T}, \quad (5)$$

can be used to calculate the partial derivatives of the hyperparameters,

$$\frac{\partial \phi}{\partial \theta^{(l)}} = \ln 10 \sum_{ij} -10^{\theta^{(l)}} \|\mathbf{x}_i^{(l)} - \mathbf{x}_j^{(l)}\|^{p^{(l)}} \mathbf{R}_{ij} \bar{\mathbf{R}}_{ij} \quad (6)$$

and

$$\frac{\partial \phi}{\partial p^{(l)}} = \sum_{ij} -10^{\theta^{(l)}} \|\mathbf{x}_i^{(l)} - \mathbf{x}_j^{(l)}\|^{p^{(l)}} \ln \|\mathbf{x}_i^{(l)} - \mathbf{x}_j^{(l)}\| \mathbf{R}_{ij} \bar{\mathbf{R}}_{ij}. \quad (7)$$

Once an optimized set of hyperparameters have been determined they can be employed to predict the response of the kriging model at an unsampled point. The prediction at an unknown design point, \mathbf{x}^* , first requires the calculation of a vector of correlations, \mathbf{r} , between the unknown point and the sample points using the optimized hyperparameters. With this correlation vector the prediction is given by,

$$y(\mathbf{x}^*) = \hat{\mu} + \mathbf{r}^T \mathbf{R}^{-1} (\mathbf{y} - \mathbf{1}\hat{\mu}). \quad (8)$$

Using this predictor the kriging model can be exhaustively searched to find promising designs. An estimate of the mean square error of the kriging model can also be calculated at an unknown point,

$$s^2(\mathbf{x}^*) = \sigma^2 [1 - \mathbf{r}^T \mathbf{R}^{-1} \mathbf{r}], \quad (9)$$

where, in this case, the small term due to the uncertainty in the estimate of the mean has been omitted. This error metric can be used to update the kriging model in regions of highest error thereby resulting in a strategy which attempts to increase the model's global accuracy rather than locate an optimal design. Both the mean square error and the predictor are also used in the calculation of another popular kriging infill criterion, expected improvement (Jones *et al.* (1998)) which is calculated via,

$$E[I(\mathbf{x}^*)] = \frac{(y_{\min} - y(\mathbf{x}^*))}{2} \left[1 + \operatorname{erf} \left(\frac{y_{\min} - y(\mathbf{x}^*)}{s\sqrt{2}} \right) \right] + \frac{s}{\sqrt{2\pi}} \exp \left[\frac{-(y_{\min} - y(\mathbf{x}^*))^2}{2s^2} \right], \quad (10)$$

where $\operatorname{erf}()$ denotes the error function and y_{\min} denotes the minimum true objective function value found so far. An exhaustive search of the expected improvement over a design space attempts to locate an update point which is most likely to result in an improvement over the current best design, y_{\min} . Expected improvement has been employed in the optimization of a number of engineering problems (Hoyle *et al.* (2006)) and has been extended for use in multi-objective optimizations (Keane (2006)).

2.2. Non-Stationary Kriging

The non-linear mapping scheme of Xiong *et al.* (2007) is a considerable simplification upon the methodologies employed by others within the literature (Smith (1996), Samp-

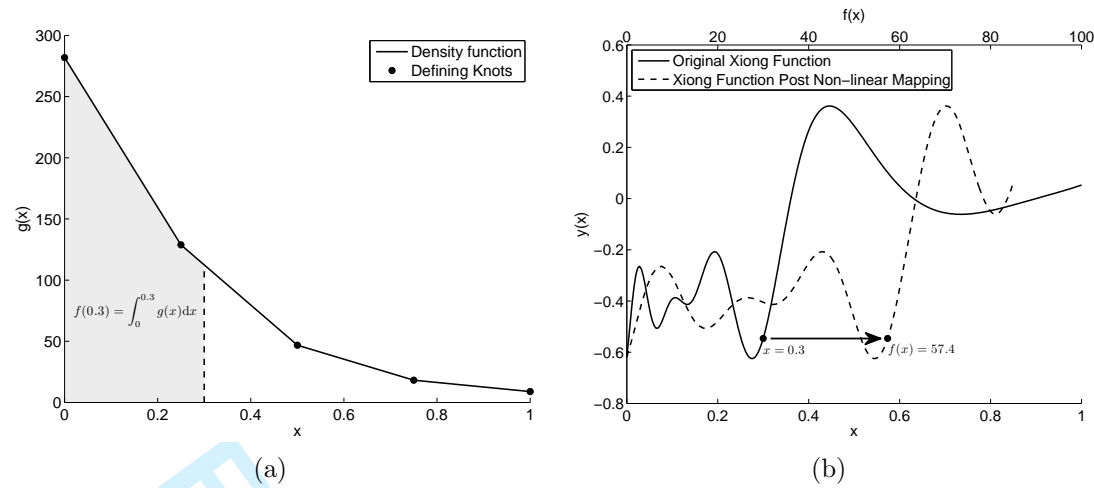


Figure 1. Illustration of the piecewise linear representation of the density function (a) and the resulting non-linear mapping of the Xiong function (b)

son and Guttorp (1992), Gibbs (1997)). Instead of the non-linear mapping being defined as the integration of a multivariate density function, itself represented by a radial basis function, a univariate piecewise linear representation of the density function is considered. This formulation drastically reduces the number of unknown modeling parameters while simultaneously reducing the cost of the integration required within the non-linear mapping.

The formulation of Xiong *et al.* (2007) is similar to that of the stationary model presented in Eq. 1, however, instead of a direct correlation between \mathbf{x}_i and \mathbf{x}_j the correlation between their respective non-linear mappings is considered,

$$\text{Corr}[Y(\mathbf{x}_i), Y(\mathbf{x}_j)] = \exp\left(-\sum_{l=1}^d 10^{\theta} \|f(\mathbf{x}_i^{(l)}) - f(\mathbf{x}_j^{(l)})\|^{\mathbf{p}^{(l)}}\right), \quad (11)$$

where

$$f(\mathbf{x}^{(l)}) = \int_0^{\mathbf{x}^{(l)}} g(x') dx', \quad (12)$$

defines a non-linear mapping where the density function, $g(x)$, is represented by a piecewise linear function of K pieces defined by $K + 1$ knots of density function value 10^{η_k} and position ζ_k . The integral of Eq. 12 therefore reduces to a simple analytical expression for each point.

Figure 1 illustrates both the piecewise linear representation of the density function and the effect of this density function on the non-linear mapping of the Xiong function. In Figure 1(a) the density function is represented by four piecewise lines defined by five equally spaced knots. Integrating across this density function from $x = 0$ to $x = 1$ maps x to $f(x)$ as shown in Figure 1(b). This non-linear mapping deforms the original Xiong function in such a manner that it can be more easily modelled using a stationary correlation function.

Figure 1 also illustrates the simplicity of the density function integration using the method of Xiong *et al.* (2007). Here, for example, the non-linear mapping of $x = 0.3$ is simply the integral of the density function between $x = 0$ and $x = 0.3$ which is a simple

analytical expression. With the area under the i^{th} piecewise section given by,

$$A_i = \frac{1}{2}(\zeta_{i+1} - \zeta_i)(10^{\eta_{i+1}} + 10^{\eta_i}), \quad (13)$$

and the gradient and intercept of the i^{th} section given by,

$$b_i = \frac{10^{\eta_{i+1}} - 10^{\eta_i}}{\zeta_{i+1} - \zeta_i} \quad \text{and} \quad a_i = 10^{\eta_{i+1}} - b_i \zeta_{i+1}, \quad (14)$$

respectively, the integral of the density function becomes,

$$f(x) = \frac{1}{2}(x - \zeta_j)(10^{\eta_x} + 10^{\eta_j}) + \sum_{l=1}^{j-1} A_l, \quad (15)$$

where j denotes the piecewise section that the x falls within and 10^{η_x} is,

$$10^{\eta_x} = b_i x + a_i. \quad (16)$$

As per the traditional hyperparameters of a stationary kriging model both the knot locations and density function values are unknowns and therefore require optimization. Likelihood maximization can be employed to determine these unknowns using equations 2, 3 and 4 with the stationary correlation function replaced by that of Eq. 11.

The optimization of these non-stationary hyperparameters can be simplified to some degree by fixing the knot locations, Xiong *et al.* (2007) for example, equally space the knots in each direction. Alternatively the knots can be placed according to experience or prior knowledge of the non-stationary structure of the true function as in, for example, the prediction of time variant responses (Toal and Keane (2011)).

An adjoint of the hyperparameters of a non-stationary kriging model can also be derived thereby accelerating the optimization process. The calculation of $\bar{\mathbf{R}}$ remains identical to Eq. 5 while the partial derivatives of the hyperparameters become,

$$\frac{\partial \phi}{\partial \theta} = -10^\theta \ln 10 \sum_{l=1}^d \left(\sum_{ij} \|f(\mathbf{x}_i^{(l)}) - f(\mathbf{x}_j^{(l)})\|^{p^{(l)}} \mathbf{R}_{ij} \bar{\mathbf{R}}_{ij} \right), \quad (17)$$

with

$$\frac{\partial \phi}{\partial p^{(l)}} = \sum_{ij} -10^\theta \|f(\mathbf{x}_i^{(l)}) - f(\mathbf{x}_j^{(l)})\|^{p^{(l)}} \ln \|f(\mathbf{x}_i^{(l)}) - f(\mathbf{x}_j^{(l)})\| \mathbf{R}_{ij} \bar{\mathbf{R}}_{ij}, \quad (18)$$

and

$$\frac{\partial \phi}{\partial \eta_k^{(l)}} = \sum_{ij} -10^\theta p^{(l)} \|f(\mathbf{x}_i^{(l)}) - f(\mathbf{x}_j^{(l)})\|^{(p^{(l)}-2)} \left[f(\mathbf{x}_i^{(l)}) - f(\mathbf{x}_j^{(l)}) \right] \left[\frac{\partial f(\mathbf{x}_i^{(l)})}{\partial \eta_k^{(l)}} - \frac{\partial f(\mathbf{x}_j^{(l)})}{\partial \eta_k^{(l)}} \right] \mathbf{R}_{ij} \bar{\mathbf{R}}_{ij}, \quad (19)$$

where

$$\frac{\partial f(\mathbf{x}^{(l)})}{\partial \eta_k} = \frac{\partial A_1}{\partial \eta_k} \quad \text{if } k = 1 \quad (20)$$

$$\frac{\partial f(\mathbf{x}^{(l)})}{\partial \eta_k} = \frac{\partial A_k}{\partial \eta_k} + \frac{\partial A_{k-1}}{\partial \eta_k} \quad \text{if } k \leq L - 1 \quad (21)$$

$$\frac{\partial f(\mathbf{x}^{(l)})}{\partial \eta_k} = \frac{1}{2}(\mathbf{x}^{(l)} - \zeta_L)(10^{\eta_L} \ln 10 + \mathbf{x}^{(l)} \frac{\partial b_L}{\partial \eta_k^{(l)}} + \frac{\partial a_L}{\partial \eta_k^{(l)}}) + \frac{\partial A_{L-1}}{\partial \eta_k^{(l)}} \quad \text{if } k = L \quad (22)$$

$$\frac{\partial f(\mathbf{x}^{(l)})}{\partial \eta_k^{(l)}} = \frac{1}{2}(\mathbf{x}^{(l)} - \zeta_L)(\mathbf{x}^{(l)} \frac{\partial b_L}{\partial \eta_k^{(l)}} + \frac{\partial a_L}{\partial \eta_k^{(l)}}) \quad \text{if } k = L + 1 \quad (23)$$

$$\frac{\partial f(\mathbf{x}^{(l)})}{\partial \eta_k^{(l)}} = 0 \quad \text{if } k \geq L + 2, \quad (24)$$

and L refers to the L^{th} piecewise section that $\mathbf{x}^{(l)}$ falls within. Given that the density function is represented by a series of K straight lines of intercept a , gradient b and integral A then,

$$\frac{\partial a_L}{\partial \eta_L} = -\zeta_{L+1} \frac{\partial b_L}{\partial \eta_L} \quad \text{and} \quad \frac{\partial a_L}{\partial \eta_{L+1}} = 10^{\eta_{L+1}} \ln 10 - \zeta_{L+1} \frac{\partial b_L}{\partial \eta_{L+1}}, \quad (25)$$

$$\frac{\partial b_L}{\partial \eta_L} = \frac{10^{\eta_L} \ln 10}{\zeta_{L+1} - \zeta_L} \quad \frac{\partial b_L}{\partial \eta_{L+1}} = \frac{10^{\eta_{L+1}} \ln 10}{\zeta_{L+1} - \zeta_L}, \quad (26)$$

and

$$\frac{\partial A_{k-1}}{\partial \eta_k} = \frac{1}{2}(\zeta_k - \zeta_{k-1})10^{\eta_L} \ln 10 \quad \frac{\partial A_{k-1}}{\partial \eta_{k-1}} = \frac{1}{2}(\zeta_k - \zeta_{k-1})10^{\eta_{L+1}} \ln 10. \quad (27)$$

The formulation of the predictor of a non-stationary kriging model is identical to that presented in Eq. 8 with the exception that the new point must undergo the same non-linear mapping as the sample points before calculating any correlations. Similarly the formulations of both the mean square error and the expected improvement infill criterion are identical to that of stationary kriging, Eqs. 9 and 10.

The formulation presented in Eq. 11 is slightly different to that originally presented by Xiong *et al.* (2007). Unlike Xiong *et al.* the hyperparameter governing the degree of smoothness in the model, p , is permitted to vary and an additional global θ hyperparameter is included as this was found to help prevent ill-conditioning of the correlation matrix at higher dimensions.

Figure 2 helps to illustrate the effect of including p within the hyperparameter optimization of a simple kriging model using the one dimensional non-stationary function of Xiong *et al.* (2007). As per the results of Xiong *et al.*, a stationary kriging model with p fixed at two results in an inaccurate surrogate. However, when p is permitted to vary the additional flexibility that this introduces results in a more accurate model. The results presented in Table 1 reflect this with the r^2 correlation increasing and both the root mean square error (RMSE) and maximum absolute error (MAE) reducing when p is permitted to vary.

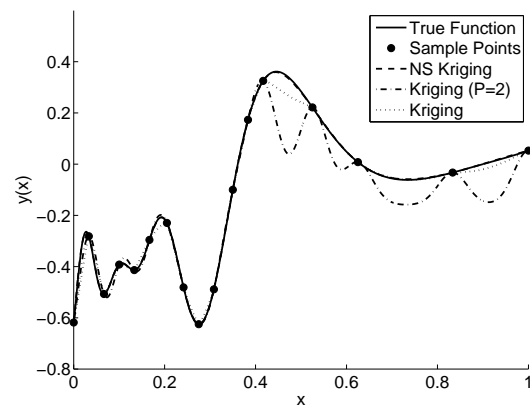


Figure 2. A one dimensional demonstration of the accuracy of non-stationary kriging N.B. the non-stationary kriging model almost exactly represents the true objective function (sampling plan taken from Xiong *et al.* (2007))

Table 1. Accuracy of three surrogate models representing the Xiong function.

Strategy	No. of Tuning Parameters	r^2	RMSE	MAE
Stationary Kriging ($p = 2$)	1	0.9353	0.0858	0.3012
Stationary Kriging ($p \neq 2$)	2	0.9930	0.0248	0.1117
Non-Stationary Kriging (5 knots)	7	1.0000	0.0019	0.0097

Based upon this simple demonstration of the importance of the additional flexibility introduced and evidence to suggest that fixing $p = 2$ introduces numerical instabilities into the likelihood calculation (Quttineh and Holmström (2009)), p is permitted to vary in both the stationary and non-stationary models considered within this paper. Figure 2 also helps to demonstrate the predictive capabilities of a non-stationary kriging model. The non-stationary kriging model is indiscernible from the true function in Figure 2 while the results of Table 1 help to demonstrate the increase in accuracy of the stationary models. The r^2 correlation of the non-stationary model is approximately equal to one, while both the RMSE and MAE are reduced by more than an order of magnitude.

Assuming an identical distribution of knots in each ordinate direction results in an efficient implementation of the non-linear mapping. The cost of evaluating the resulting likelihood function is only slightly greater than that for the equivalent stationary kriging model, typically less than 5%.

3. Stationary Versus Non-stationary Kriging

Given the surrogate modeling strategies presented in Section 2, it is necessary to assess their performance with regard to both their global accuracy and performance within an optimization.

3.1. Surrogate Model Accuracy

Although Xiong *et al.* (2007) presented some comparisons between the accuracy of stationary and non-stationary kriging models, only test problems with at most two variables

were considered and, as described previously, p was fixed at two. We therefore compare the stationary and non-stationary models presented previously using 11 analytical test problems ranging from one to 10 variables with a variety of sampling densities.

The analytical formula for each test function is presented in Appendix A along with 1D, 2D or 2D section plots of each function. As can be observed from these figures the test functions exhibit a variety of levels of non-stationarity. The Forrester, Branin, six hump camelback, Hartmann $H_{3,4}$, Hartmann $H_{6,4}$ and Trid functions are all relatively stationary in nature and exhibit a uniform wavelength over the design space. The Xiong, Mystery, Paciorek, Shekel and Michalewicz functions however, exhibit a variation in wavelength and frequency throughout the design space.

Fourier analysis can be used to help indicate the non-stationarity of the responses. While the Forrester function results in a single important frequency, the Xiong function results in three important frequencies appearing between $0 \leq x \leq 0.5$ and only one between $0.5 \leq x \leq 1.0$. Figure 1(b) illustrates that the wavelength of the function is much shorter in the first half of the design space than the other. It is, of course, this variation which causes the problems shown in Figure 2 when fitting the response using a stationary model.

The Mystery function is a similar case with the most important frequency of the Fourier analysis varying across the design space. Moving from $x_1 = 0$ to $x_1 = 1.0$ in Figure 2(b) sees the frequency of the response vary from 1Hz to 3Hz. Likewise, the response of the Shekel function, Figure 4(b), varies in frequency from the variable bounds towards the center of the design space. In the case of the Michalewicz function, Figure 5(b), the upper and lower bounds have been altered to result in a function akin to a step response. This, as demonstrated by Toal and Keane (2011), is something which a stationary model can have great difficulty representing.

In comparing both strategies a total of 50 different design of experiments (DoE) are generated via a random Latin Hypercube for each test function. This helps to negate the impact of the sampling plan on the performance of each method and provides a meaningful average of each performance metric. Identical sampling plans are used for each of the models thereby comparing like with like. Models are constructed from sampling plans of three different sampling densities, five, seven and 10. The number of points within a sampling plan is equal to the number of dimensions multiplied by the sample density, a sample density of five when applied to a two variable problem will therefore result in a 10 point random Latin Hypercube.

The hyperparameters of both the stationary and non-stationary models are tuned in an identical manner. A genetic algorithm (GA) with a population size of 50 points run for 40 generations, is followed by a local terminal search via sequential quadratic programming (SQP) employing the adjoints of the likelihood.

The accuracy of each surrogate model is evaluated through a comparison of the model's prediction of the objective function to the true objective function at a series of test points. Test points are generated from a random Latin Hypercube with a sampling density of 200, predictions of the Branin function, for example, are therefore compared to the true function at 400 points. Given the true and predicted objective function values, three performance metrics are calculated, the r^2 correlation, the RMSE and the MAE. Results for each of these tests are presented in Tables B2, B3 and B4 of Appendix B, along with results of the adaptive partial non-stationary (APNS) strategy, see Section 4. Table B1 presents a brief overview of the surrogate modeling strategies tested along with formulae to calculate the total number of hyperparameters optimized in their construction.

The results of Appendix B indicate that generally for problems of relatively low dimen-

sionality there is little difference in the accuracy of the models produced via stationary and non-stationary kriging with two knots. The mean r^2 correlations, RMSE and MAE for the Xiong, Branin, Mystery, six hump camelback and Paciorek functions are all relatively similar with perhaps only a few instances where one model outperforms the other. The r^2 correlation of the non-stationary model for the Paciorek function, for example, is consistently higher for each of the three sample sizes but the difference in RMSE and MAE is minimal.

However, things are different when the analytical problems of higher dimensionality are considered. From the four variable Shekel function to the 10 variable Trid function, the two knot non-stationary model clearly out performs the stationary kriging model. Although the RMSE and MAE are similar for the Shekel function the r^2 correlation is improved for sample densities of both five and 10. The r^2 correlation increases from 0.231 with the stationary model to 0.346 with the non-stationary model when a sample density of 10 is used. Given the similar standard deviations this is a significant increase. The r^2 correlation of the prediction of the Hartmann $H_{6,4}$ function is also improved significantly by the non-stationary model, almost doubling for sampling densities of both five and seven and increasing by 60% for a sampling density of 10. The RMSE and MAE are also improved slightly when the non-stationary model is employed.

Each of the three performance metrics indicate that the two knot non-stationary predictor of the 10 variable Michalewicz function is superior to the stationary predictor. All of the r^2 correlations are better than those obtained by the stationary model while both the RMSE and MAE show noticeable reductions. Both the RMSE and MAE for the models constructed using sample densities of seven and ten are approximately half that obtained by the stationary model given identical sample plans.

The two-knot non-stationary model performs significantly better on the 10 variable Trid function, the r^2 correlation is approximately twice that of the stationary model. Employing a sampling density of 10 points, for example, causes the r^2 increase from 0.527 to 0.999. This improvement in performance is reflected by both the RMSE and MAE. The mean RMSE decreases from 8316.1 to 252.9, likewise the mean MAE reduces by two orders of magnitude from 35386.2 to 375.8.

The non-stationary kriging formulation of Xiong et al. is capable of representing more elaborate definitions of the non-linear mapping function through the addition of more knots in each piecewise linear representation. Tables B2, B3 and B4 of Appendix B also contain results for a non-stationary kriging model constructed with a five knot non-linear mapping in each ordinate direction. The hyperparameters for these models are tuned in an identical manner to those of the two knot models.

The r^2 correlation results for each of the test functions, with the exception of the 10 variable Michalewicz function, are all significantly lower than the corresponding two knot non-stationary model. Likewise the RMSE and MAE results for each of these tests are worse than for the equivalent two knot model. These results therefore demonstrate the dangers of an over parameterization of the non-linear mapping. The additional flexibility of the model has made the hyperparameter optimization much more difficult and the “optimum” hyperparameters found result in a model less accurate than that produced via a simpler non-linear mapping.

Applying the five knot non-stationary model to the Michalewicz function, however, results in a superior model to that produced by the two knot model with a higher mean r^2 correlation for each of the three sample densities and lower mean RMSE and MAE. These results highlight one of the major difficulties of non-stationary kriging, namely how to determine an appropriate fidelity of the non-linear mapping function. Too complex

a model risks an over parameterization while a simpler model risks the creation of an under performing surrogate.

3.2. Optimization Performance

To determine the optimization performance of both stationary and non-stationary kriging the previous analytical test functions are used once again. The optimization of each function commences with a surrogate model created from the $5d$ sampling plan used in the previous investigation. This model is searched in order to maximize the expected improvement, Eq. 10. Depending on the dimensionality of the problem one or more points are evaluated using the true objective function value and added to the sampling plan. Analytical test problems with four or fewer variables have a single update point evaluated at a time, while those with six or 10 variables have five or 10 update points evaluated based on a K-MEANS cluster analysis.

With the sampling plan updated, the hyperparameters of the surrogate model are re-trained and new update points searched for using the expected improvement criterion. This process repeats until a predefined budget of $10d$ update points have been evaluated. As with the previous comparison of surrogate model accuracy each optimization is repeated 50 times, each with a different design of experiments (DoE). The results of these optimizations are presented in Table C1 of Appendix C along with the true global minima. Due to the poor results of the five knot non-stationary model in the previous tests only the optimization performance of the two knot non-stationary model is considered here.

As per the results concerning the accuracy of the surrogate models, the optimizations using both the stationary and non-stationary models are quite similar for those problems with one, two or three variables. There is no significant difference in the final optimum obtained when optimizing the Forrester, Xiong, Branin, Mystery, six hump camelback, Paciorek or Hartmann $H_{3,4}$ functions. With the exception of the Mystery function the mean optima obtained by the non-stationary optimization are within less than 5×10^{-3} of those obtained by the stationary optimization while the magnitude of the standard deviations are also similar. Both optimizations also tend to result in objective functions within 3.8×10^{-3} of the value at the true global minimum in of each of these tests.

From the four variable Shekel function onwards there is no longer such a parity between the strategies. However, unlike the results of Section 3.1, the non-stationary model does not consistently outperform the stationary model, instead the performance of each strategy appears to be more problem dependant. The non-stationary model outperforms the stationary model when optimizing both the Shekel and Trid functions while the stationary model performs better on the Hartmann $H_{6,4}$ function.

When optimizing the Shekel function the mean objective function obtained by the non-stationary kriging model is much lower than that obtained by the stationary model. The stationary optimization obtained a mean objective function of -4.68 while the non-stationary optimization obtained -5.27. It should be noted however that there is much more variance in the final objective function obtained by the non-stationary optimization. Therefore, although it performs better on average, there is less consistency between the results.

When optimizing the Hartmann $H_{6,4}$ function the stationary optimization performs better than the non-stationary optimization. The mean final objective function is -3.15 compared to the mean objective function of -2.82 obtained using the non-stationary optimization. The stationary optimization even results in a more consistent optimization

as indicated by the standard deviation in the final objective function value.

When applied to the Michalewicz function both optimizations obtained a similar mean final objective function, -4.92 when using the stationary model and -4.93 when using the non-stationary model. However, the variance in the objective function is slightly higher when employing a non-stationary model at 0.166 compared to 0.125 when using a stationary model.

The non-stationary optimization significantly outperforms the stationary optimization on the Trid function. When employing a stationary model a mean objective function of -20.7 is obtained, whereas when a non-stationary model is employed the mean objective function is substantially reduced to approximately -57.7. Employing a non-stationary model also results in a reduction in the variance of the final optima, the standard deviation reduces from approximately 75.6 to 59.9.

Given the previous differences in the accuracy of the initial surrogate models it is perhaps initially surprising that the stationary model performs so well when the corresponding functions are optimized. The r^2 correlation, RMSE and MAE for the Hartmann $H_{6,4}$ function are better when employing the non-stationary model and yet when a stationary model is utilized within an optimization consistently better results are obtained. Likewise the results for the Michalewicz function indicate that a non-stationary model initially represents the function better however both strategies perform similarly within an optimization.

This apparent disparity can be explained by considering the surrogates presented in Figure 2. As noted previously, the non-stationary model is clearly more globally accurate than either stationary model. However, Figure 2 also demonstrates that even though the stationary models are less accurate, they do approximate the region of the global optimum very well. Searching any of the models for an update point will therefore produce similar results.

Based on the results presented here it could therefore be inferred that a stationary model performs just as well as a non-stationary model within an optimization as long as either the region around the global optimum, or the general trend towards it, are captured well by the stationary model. This implies that the performance gain offered by non-stationary kriging when employed within an optimization process is very much problem dependent.

4. Adaptive Partial Non-Stationary Kriging

The previous results raised a number of issues with the application of non-stationary kriging in the representation and optimization of black box functions. At lower dimensions non-stationary kriging offers little performance advantage in either the representation or optimization of the analytical test functions considered, however, as dimensionality increases some disparity in the performance of each strategy is observed. The performance of each strategy within an optimization appears to be determined by the nature of the objective function as some test functions are better optimized by the stationary model and others by the non-stationary model. The question is therefore how to determine when one model should be used over another and how to determine an appropriate level of complexity for the density function parameterization.

In the following section we consider a simple adaptive partial non-stationary kriging strategy to investigate the potential of automatically determining the best model for a black box objective function. It should be noted that this strategy is not presented as the

Table 2. Description of the Adaptive Partial Non-Stationary (APNS) Kriging process.

Step	
1	Construct a stationary kriging model (optimize the hyperparameters)
2	Split the sample data into four random groups
3	Remove each group from the sample data and calculate a prediction of the missing data
4	Using the predictions for each sample point, calculate the r^2 correlation
5	Introduce non-stationarity into the variable with the largest θ
6	Construct the partial non-stationary kriging model
7	Calculate the r^2 correlation
8	If the r^2 correlation improves add an additional knot otherwise move to the next variable
9	If all variables have been tried and no further improvement in r^2 is obtained use the best model

definitive solution to the issue of appropriate model selection, rather it serves to demonstrate the potential impact of such an approach on design optimization performance. The proposed adaptive strategy is assessed with regard to both the accuracy of the surrogate models produced and the performance within an optimization.

4.1. Surrogate Modeling Strategy

The adaptive partial non-stationary kriging strategy considered here commences with the construction of a stationary kriging model from the sample data. The hyperparameters of this model are tuned using the method employed in Section 3.1. This stationary model provides a baseline with which to compare each of the subsequent partial non-stationary models. In a similar process to Viana *et al.* (2009), the sample data used in the construction of the model is randomly partitioned into four approximately equal groups. Each group of points is removed from the sample set in turn and a kriging predictor, based upon the remaining data and the previously determined hyperparameters, is used to predict the missing data. Repeating the process for the four groups leads to a prediction for all of the sample points and permits the calculation of the r^2 correlation. This correlation value is then used to determine if the subsequent non-stationary model is more accurate.

Once constructed, each dimension of the stationary model is considered in turn for the introduction of non-stationarity. In a similar manner to Welch *et al.* (1992) the magnitude of the θ parameter is used to indicate the relative importance of a variable. Starting with the most important variable, the correlation function is modified and a two knot representation of the density function is introduced. The correlation function of this partial non-stationary model is therefore a combination of Eq. 1 for those dimensions remaining stationary and Eq. 11 for the dimension considered as non-stationary. The hyperparameters are optimized, the partial non-stationary model constructed and the r^2 correlation calculated as before. This r^2 correlation is then compared to that obtained by the stationary model, if an improvement is observed the non-stationarity is accepted and the number of knots in the density function representation increased. Again the hyperparameters are tuned and the model performance assessed, if the model is an improvement the number of knots is increased again, if not then the next most important variable is considered for the introduction of non-stationarity.

This process can be repeated for either a predefined number of cycles or until the θ value falls below a predefined limit therefore ensuring that only the most active variables are considered for the introduction of non-stationarity. Table 2 summarizes the APNS kriging process.

4.2. Surrogate Model Accuracy

As per the comparisons in Section 3.1 the accuracy of the adaptive partial non-stationary (APNS) kriging model is evaluated over 11 analytical test functions using a range of different sampling densities. To maintain consistency the same sampling plans used in the previous investigations are employed here again. The results for the APNS strategy are also presented in Tables B2, B3 and B4 of Appendix B.

The results presented in Appendix B indicate that generally the performance of the APNS strategy falls somewhere in between that of the non-stationary and stationary strategies with the r^2 correlation, RMSE and MAE generally better than the worst performing of the other two strategies, whichever that may be. The results for the Forrester function, for example, are better than the stationary model, but not as good as the purely non-stationary model. Likewise, when sample densities of seven and ten are considered, for the six hump camelback function the APNS results are better than the non-stationary model but not quite as good as the stationary model.

These results indicate that by employing an adaptive strategy when constructing the surrogate model the deficiencies of selecting one modeling technique over another prior to the construction of a surrogate model of an unknown function can be mitigated to some extent. The best example of this is the strategies performance with respect to the Trid function. By employing the APNS kriging strategy, this reduction in general performance is reduced quite significantly. Employing the strategy on a $10d$ sample plan, for example, increases the r^2 correlation from 0.527 to 0.966 which is much closer to the purely non-stationary model's 0.999.

The brute force strategy employed within the APNS strategy to select an appropriate level of non-stationarity can result in a significant increase in the cost over the simpler stationary or non-stationary models presented previously. This is due to the numerous global hyperparameter optimizations and cross-validations and the increase in complexity of the non-linear mapping when differing density function representations are employed in each ordinate direction. The cost can also be problem dependent and increase with dimensionality. An APNS model of the Forrester function constructed using five sample points will on average cost 135% more to construct than the equivalent stationary model. However, an APNS model of the 10 variable Michalewicz function constructed from 50 sample points will, on average, cost 620% more. A more efficient approach to the optimization of the level of non-stationarity within the model would, of course, drastically reduce these costs.

4.3. Optimization Performance

As with the stationary and non-stationary strategies, the APNS kriging strategy is intended to be employed within a design optimization framework. It is therefore necessary to assess the optimization performance of such a strategy.

As per the previous performance comparisons of Section 3.2 each of the 11 analytical test functions is optimized a total of 50 times, each commencing from a different $5d$ random latin hypercube sampling plan. Again a total of $10d$ updates are evaluated via a maximization of the expected improvement. The results of these optimizations are also presented in Table C1 of Appendix C.

As per the results of Section 3.2, stationary kriging, non-stationary kriging and the APNS kriging strategy all perform similarly for problems with one, two or three variables. From the Forrester function through to the Hartmann $H_{3,4}$ function the mean objective

functions obtained by each strategy are very similar with differences only in the third decimal place. As with the other two strategies the APNS kriging approach obtains a final optimum which is typically very close to the true global optimum.

The performance of the APNS kriging strategy on problems with more than three variables reflects the results pertaining to the accuracy of the surrogate models produced. That is to say the performance falls somewhere between the stationary model and non-stationary model. In Section 3.2 the Shekel function was demonstrated to be optimized more effectively via a non-stationary kriging model. Here the APNS kriging strategy outperforms the stationary approach and attains a mean objective function similar to that of the non-stationary approach but with a reduction in the variance.

In Section 3.2, the Hartmann $H_{6,4}$ function was better optimized by the stationary kriging approach. Here the performance of the APNS strategy approaches that of the stationary strategy achieving a mean objective function of -3.051 to -3.148 obtained using the stationary strategy. The optimization of the Michalewicz function demonstrated similar results for both the stationary and non-stationary models, likewise the APNS strategy results in similar performance. The non-stationary optimization of the Trid function was far superior to that of the stationary optimization, with both the mean and variance in the objective function reducing considerably. The APNS kriging strategy performs similarly well, with the objective function reducing from -20.7 when using a stationary kriging model to -49.1.

Given a black box function with an unknown prior level of stationarity an adaptive partial non-stationary kriging approach could therefore be said to offer the best of both worlds. Through adapting the level of stationarity present within the model to the current objective function and sampling plan the adaptive approach appears to offer a more consistent level of optimization performance.

5. 40D Truss Attenuation Optimization

To further investigate the performance of stationary, non-stationary and adaptive partial non-stationary kriging based design optimizations consider now a 40 variable truss optimization problem. A two-dimensional truss structure similar to those employed on satellites (Nair and Keane (1999)), shown in Figure 3(a), is represented by 42 Euler-Bernoulli beams each with a total of two finite elements. A unit force excitation is applied to node number 1 over a 100-200Hz frequency range. Each of the nodes are free to move with the exception of the leftmost nodes which are fixed. The objective is to minimize the band-averaged vibration attenuation at the tip compared to the baseline design shown in Figure 3(a). Each of the 20 nodes are permitted to move both horizontally and vertically within a 0.9×0.9 box, also shown in Figure 3(a). This results in a 40 variable optimization of a complex multi-modal non-stationary function, as indicated by surface plot in Figure 3(b) of the response of the structure to changes in the position of the ninth node.

Each of the three optimizations commence from an identical 200 point design of experiments defined by a random Latin hypercube. A further 400 true objective function evaluations are permitted as updates to the surrogate model as the optimization progresses. These updates are evaluated in batches of 10 and are defined via a maximization of the expected improvement criterion using a GA followed by a local terminal search from the best points of 10 clusters resulting from a K-Means cluster analysis. With such a large number of sample points and such a high dimensional problem, inevitably the cost of the optimization of the kriging hyperparameters can become an issue. To reduce this

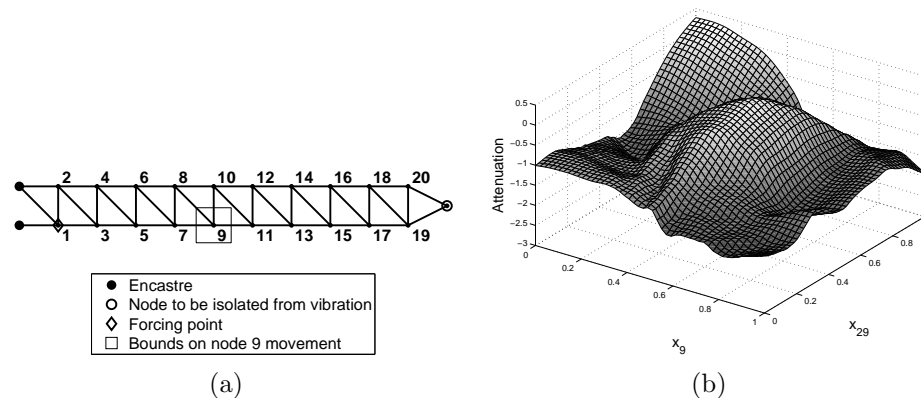


Figure 3. The baseline truss geometry (a) and a 2D example of the design space (b)

Table 3. Average final objective function obtained by the three optimization strategies.

Kriging		NS Kriging		APNS Kriging	
Mean	Std.	Mean	Std.	Mean	Std.
-29.727dB	4.646dB	-32.338dB	5.807dB	-32.239dB	2.764dB

cost the hyperparameters are reassessed after alternate batches of updates to the model (Toal *et al.* (2008)) and the hyperparameters are tuned using only the best 200 design points, with all points used to form the predictor. The performance of each optimization strategy is averaged over 10 optimizations commencing from a different initial design of experiments. The average and standard deviation in objective function obtained by each of the three strategies are presented in Table 3 with the mean optimization histories presented in Figure 4.

Upon comparison of the results of Table 3 it is clear that the non-stationary kriging strategy outperforms stationary kriging. Given the same initial starting points, the non-stationary kriging strategy produces an average attenuation of -32.3dB, compared to -29.7dB produced via the stationary strategy. The stationary strategy does however result in a slight improvement in the variance in the quality of these final designs. As per Section 4, the adaptive partial non-stationary kriging strategy results in similar mean designs to that of the non-stationary strategy but with a significant reduction in the variance in the quality of these designs. The mean objective function is -32.2dB, slightly less than the non-stationary approach, but the standard deviation in these designs is 2.76dB compared to 5.81dB. The optimization histories of Figure 4 indicate that the non-stationary approach appears to converge at a faster rate than the stationary kriging approach, however after the first 100 updates the strategy appears to have difficulty improving further upon the current best solution. While the rate of converge of the APNS strategy is slightly slower than that of the non-stationary strategy, after approximately 150 updates it appears to have converged almost completely to an objective function similar to that of the final solution of the non-stationary strategy.

Examples of “optimum” geometries resulting from each of the three optimization strategies are presented in Figure 5 along with the frequency responses of each design over an extended 1-300Hz range. It is quite obvious that each of the three strategies converge to different solutions. This result highlights the multi-modal nature of the design space and reflects the results observed by Nair and Keane (1999) where it was noted that

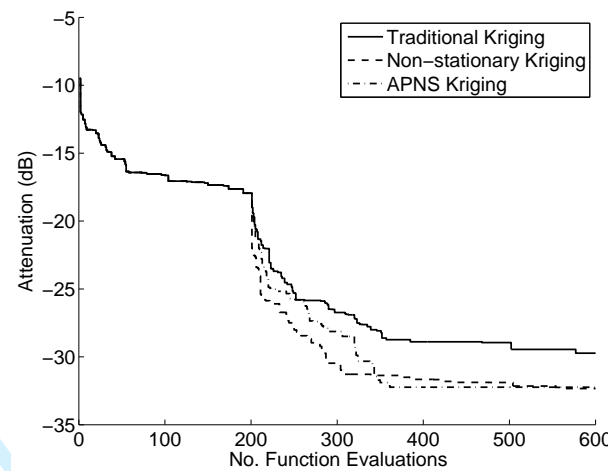


Figure 4. Optimization history for the 40D attenuation problem

quite different designs can result in similar objective function values. As demonstrated by the results of Figures 5(b), 5(c) and 5(d) each of the optimizations has resulted in a noticeable decrease in the displacement of the tip node across the optimized frequency domain. As expected from the results of Nair and Keane (1999) each of the final designs offers little improvement over the baseline design outside the 100-200Hz range as these frequencies were not considered for optimization.

6. Conclusions

A comparison of the performance of stationary and non-stationary kriging with respect to the construction of surrogate models for the purpose of design optimization has been presented and a novel adaptive partial non-stationary kriging strategy has been introduced.

The presented results indicate that on problems with one, two or three variables both stationary and non-stationary kriging perform similarly, effectively supporting optimizing of each of the analytical test functions considered. However, differences in the performance of the models do become apparent as the dimensionality and complexity of the optimization problem increases. These differences appear to be problem dependant, with optimizations carried out using stationary kriging performing better than those using non-stationary kriging on some problems and vice versa.

This difficulty in determining an appropriate model for a given black box objective function lead to the consideration of an adaptive partial non-stationary (APNS) kriging model which alters the degree of non-stationarity of the model in each ordinate direction in order to improve overall model accuracy based on a cross-validation metric. This surrogate modeling strategy, while not presented as an outright solution to the problem of model selection due to its computational inefficiencies, provides an indication of the potential benefits of such a strategy within design optimization. The adaptive partial non-stationary kriging strategy was capable of spanning the difference in the performance of the pure stationary and non-stationary kriging strategies. When utilized within design optimizations the APNS strategy approached the performance of the better of the other two strategies, demonstrating that, for black box functions, an adaptive non-stationary approach can result in more consistent performance.

Figure 5. Example truss designs from each of the three optimizations and the frequency response of each design

Each of the three kriging strategies were employed in the optimization of a 2D truss structure. The objective function, the minimization of the band-averaged vibration attenuation at the tip of the structure, is non-stationary and multi-modal in nature. Both the non-stationary and APNS kriging approaches produced better designs than the stationary kriging approach while the APNS approach resulted in an improvement in the consistency between final designs.

Although further research is necessary to reduce the additional cost of the presented adaptive strategy, the current paper goes some way to demonstrate that such an approach may have tangible benefits over a traditional kriging based design optimization process.

Acknowledgements

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no.

234344 (www.crescendo-fp7.eu).

References

Atkinson, P. and Lloyd, C., 2007. Non-Stationary Variogram Models for Geostatistical Sampling Optimisation: An Empirical Investigation Using Elevation Data. *Computers & Geosciences*, 33 (10), 1285–1300.

Clarke, S., Griedsch, J., and Simpson, T., 2005. Analysis of Support Vector regression for Approximation of Complex Engineering Analyses. *Journal of Mechanical Design*, 127 (6), 1077–1087.

D’Angelo, S. and Minisci, E., 2005. Mult-Objective Evolutionary Optimization of Subsonic Airfoils by Kriging Approximation and Evolution Control. In: *2005 IEEE Congress on Evolutionary Computation*, Vol. 2, 1262–1267.

Forrester, A., Bressloff, N., and Keane, A., 2006. Optimization Using Surrogate Models and Partially Converged Computational Fluid Dynamics Simulations. *Proceedings of the Royal Society A*, 462 (2071), 2177–2204.

Forrester, A. and Keane, A., 2009. Recent advances in surrogate-based optimization. *Progress in Aerospace Sciences*, 45 (1-3), 50–79.

Fuentes, M., 2001. A High Frequency Kriging Approach For Non-Stationary Environmental Processes. *Environmetrics*, 12, 469–483.

Gibbs, M., 1997. *Bayesian Gaussian Processes for Regression and Classification*. Ph.D. Dissertation, University of Cambridge.

Haas, T., 1990. Lognormal and moving Window Methods of Estimating Acid Deposition. *Journal of the American Statistical Association*, 85 (412), 950–963.

Haas, T., 2002. New Systems for Modeling, Estimating, and Predicting A Multivariate Spatio-temporal Process. *Environmetrics*, 13 (4), 311–332.

Hoyle, N., Bressloff, N., and Keane, A., 2006. Design Optimization of a Two-Dimensional Subsonic Engine Air Intake. *AIAA Journal*, 44 (11), 2672–2681.

Jin, R., Chen, W., and Simpson, T., 2001. Comparative Studies of Metamodelling Techniques Under Multiple Modelling Criteria. *Structural and Multidisciplinary Optimization*, 23 (1), 1–13.

Jones, D., 2001. A Taxonomy of Global Optimization Methods Based on Response Surfaces. *Journal of Global Optimization*, 21 (4), 345–383.

Jones, D., Schonlau, M., and Welch, W., 1998. Efficient Global Optimization of Expensive Black-Box Functions. *Journal of Global Optimization*, 13 (4), 455–492.

Keane, A., 2006. Statistical Improvement Criteria for Use in Multiojective Design Optimization. *AIAA Journal*, 44 (4), 879–891.

Kim, H., Mallick, B., and Holmes, C., 2005. Analyzing Nonstationary Spatial Data Using Piecewise Gaussian Processes. *Journal of the American Statistical Association*, 100 (470), 653–668.

Krige, D., 1951. A Statistical Approach to Some Basic Mine Valuation Problems on the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Engineering Society of South Africa*, 52 (6), 119–139.

Nair, P. and Keane, A., 1999. Design Optimization of Space Structures With Nonperiodic Geometries for Vibration Suppresion. In: *Proceedings of the 40th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference, St. Louis, MO*.

Paciorek, C. and Schervish, M., 2004. Nonstationary Covariance Functions for Gaussian

- Process Regression. In *Advances in Neural information Processing Systems 16*, 273–280.
- Peter, J. and Marcelet, M., 2008. Comparison of Surrogate Models for Turbomachinery Design. *WSEAS Transactions on Fluid Mechanics*, 3 (1), 10–17.
- Pintore, A. and Holmes, C., 2004. Spatially Adaptive Non-Stationary Covariance Functions via Spatially Adaptive Spectra. *Technical Report, University of Oxford, U.K.*
- Queipo, N., *et al.*, 2005. Surrogate-based analysis and optimization. *Progress in Aerospace Sciences*, 41, 1–28.
- Quttineh, N. and Holmström, K., Implementation of a One-Stage Efficient Global Optimization (EGO) Algorithm. , 2009. , Technical report, Linköping University.
- Rasmussen, C. and Ghahramani, Z., 2002. Infinite Mixtures of Gaussian Process Experts. In *Advances in Neural information Processing Systems 14*, 881–888.
- Sacks, J., *et al.*, 1989. Design and Analysis of Computer Experiments. *Statistical Science*, 4 (4), 409–435.
- Sakata, S., Ashida, F., and Zako, M., 2003. Structural Optimization Using Kriging Approximation. *Computer Methods in Applied Mechanics and Engineering*, 192 (7-9), 923–939.
- Sampson, P. and Guttorp, P., 1992. Nonparametric Estimation of Nonstationary Spatial Covariance Structure. *Journal of the American Statistical Association*, 87 (417), 108–119.
- Simpson, T., *et al.*, 2001. Metamodels for Computer-based Engineering Design: Survey and Recommendations. *Engineering with Computers*, 17 (2), 129–150.
- Smith, R., 1996. Estimating Nonstationary Spatial Correlations. *Technical Report, Cambridge University.*
- Toal, D., Bressloff, N., and Keane, A., 2008. Kriging Hyperparameter Tuning Strategies. *AIAA Journal*, 46 (5), 1240–1252.
- Toal, D., *et al.*, 2011. The Development of a Hybridized Particle Swarm for Kriging Hyperparameter Tuning. *Engineering Optimization (Accepted for Publication)*.
- Toal, D. and Keane, A., 2011. Non-Stationary Kriging Prediction of the Performance of Turbomachinery Components. In: *European Turbomachinery Conference, Istanbul, Turkey, March 21st – 25th*.
- Viana, F., Haftka, R., and Steffen, V., 2009. Multiple Surrogates: How Cross-Validation Errors Can Helps Us To Obtain The Best Predictor. *Structural and Multidisciplinary Optimization*, 39 (4), 439–457.
- Welch, W., *et al.*, 1992. Screening, Predicting, and Computer Experiments. *Technometrics*, 34 (1), 15–25.
- Xiong, Y., *et al.*, 2007. A Non-Stationary Covariance-Based Kriging Method for Meta-modelling in Engineering Design. *International Journal for Numerical Methods in Engineering*, 71 (6), 733–756.

Appendix A. Analytical Test Functions

Forrester Function:

$$y(x) = (6x - 2)^2 \sin(12x - 4), \quad x \in [0, 1] \quad (\text{A1})$$

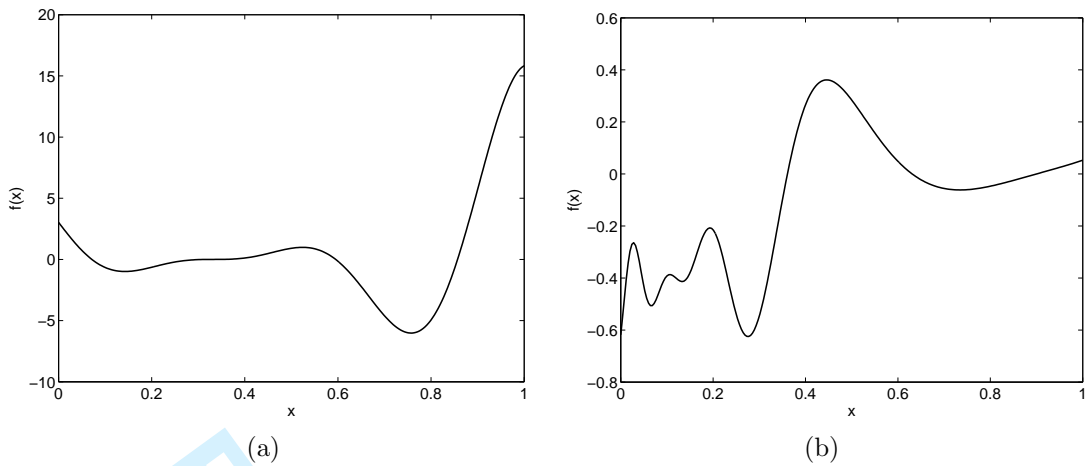


Figure A1. Analytical (a) Forrester and (b) Xiong test functions

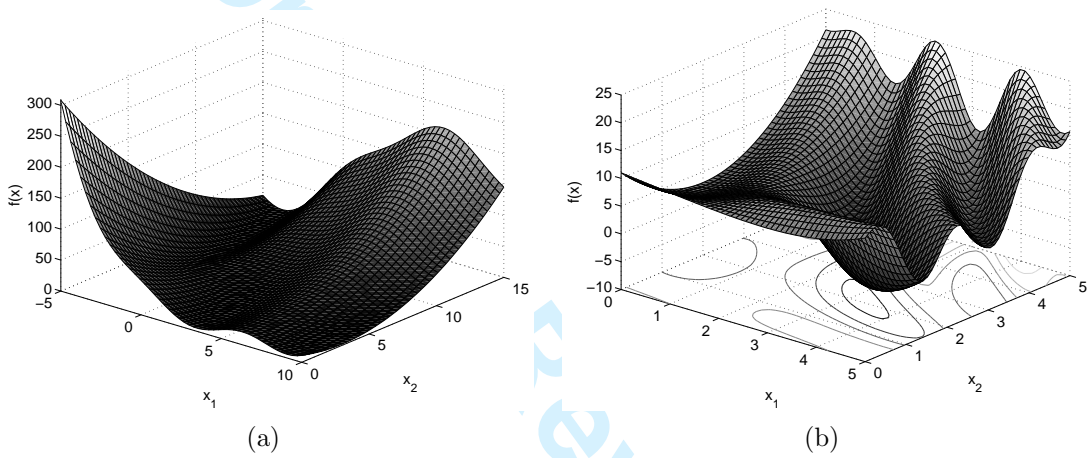


Figure A2. Analytical (a) Branin and (b) Mystery test functions

Xiong Function:

$$y(x) = \sin[30(x - 0.9)^4] \cos[2(x - 0.9)] + (x - 0.9)/2, \quad x \in [0, 1] \tag{A2}$$

Branin Function:

$$y(x) = (x_2 - \frac{5}{4\pi^2}x_1^2 + \frac{5}{\pi}x_1 - 6)^2 + 10(1 - \frac{1}{8\pi})\cos(x_1) + 10, \\ x_1 \in [15, 10] \quad x_2 \in [0, 15] \tag{A3}$$

Mystery Function:

$$y(x) = 3 + 0.01(x_2 - x_1^2)^2 - x_1 + 2(2 - x_2)^2 + 7\sin(0.5x_1)\sin(0.7x_1x_2), \\ x_1 \in [0, 5] \quad x_2 \in [0, 5] \tag{A4}$$

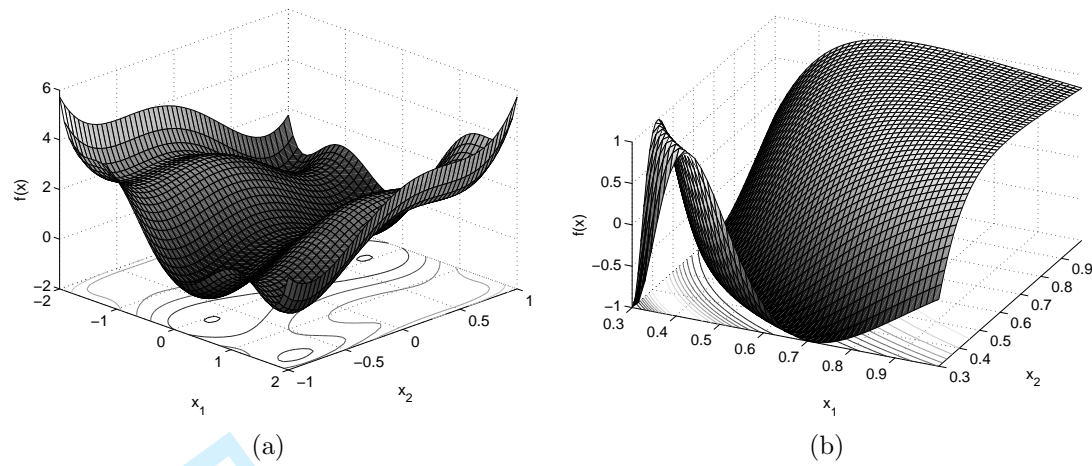


Figure A3. Analytical (a) six hump camelback and (b) Paciorek test functions

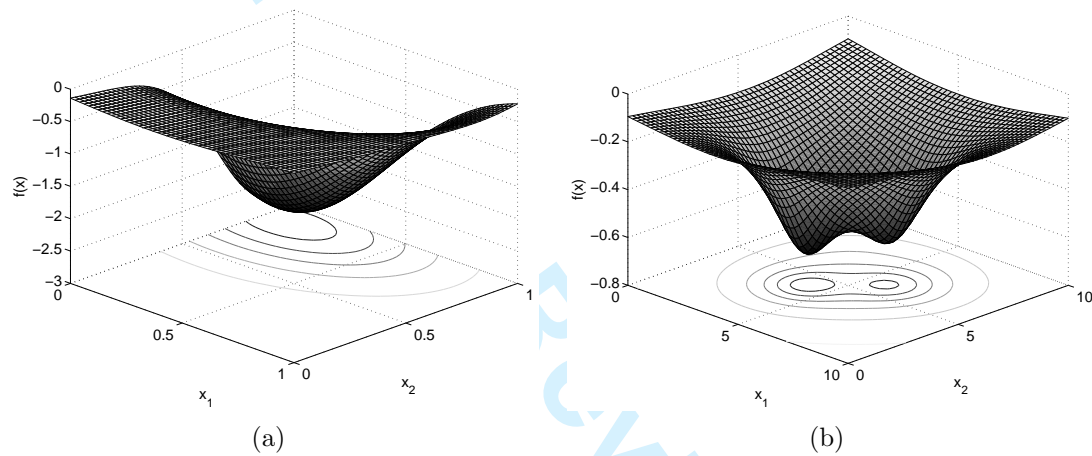


Figure A4. Two dimensional slices through the (a) Hartmann $H_{3,4}$ (all other variables equal 0.5) and (b) Shekel test functions (all other variables equal 5)

Six Hump Camelback Function:

$$y(x) = 4x_1^2 - 2.1x_1^4 + x_1^6/3 + x_1x_2 - 4x_2^2 + 4x_2^4, \quad x_1 \in [-2, 2] \quad x_2 \in [-1, 1] \quad (A5)$$

Paciorek Function:

$$y(x) = \sin \frac{1}{x_1x_2}, \quad x_1 \in [0.3, 1] \quad x_2 \in [0.3, 1] \quad (A6)$$

Hartmann H_{3,4} Function:

$$y(x) = - \sum_{i=1}^4 \alpha_i \exp \left[- \sum_{j=1}^3 A_{ij} (x_j - P_{ij})^2 \right],$$

$$\text{where } \alpha = \begin{bmatrix} 1 \\ 1.2 \\ 3 \\ 3.2 \end{bmatrix}, A = \begin{bmatrix} 3.0 & 10 & 30 \\ 0.1 & 10 & 35 \\ 3.0 & 10 & 30 \\ 0.1 & 10 & 35 \end{bmatrix} \text{ and } P = \begin{bmatrix} 6890 & 1170 & 2673 \\ 4699 & 4387 & 7470 \\ 1091 & 8732 & 5547 \\ 381 & 5743 & 8828 \end{bmatrix} \times 10^{-4}$$

$$x_i \in [0, 1] \quad (A7)$$

Shekel Function:

$$y(x) = - \sum_{j=1}^5 \left[\sum_{i=1}^4 (x_i - C_{ij})^2 + \beta_j \right]^{-1}$$

$$\text{where } \beta = \frac{1}{10} [1 \ 2 \ 4 \ 4 \ 6 \ 3 \ 7 \ 5 \ 5]^T \text{ and } C = \begin{bmatrix} 4.0 & 1.0 & 8.0 & 6.0 & 3.0 & 2.0 & 5.0 & 8.0 & 6.0 & 7.0 \\ 4.0 & 1.0 & 8.0 & 6.0 & 7.0 & 9.0 & 5.0 & 1.0 & 2.0 & 3.6 \\ 4.0 & 1.0 & 8.0 & 6.0 & 3.0 & 2.0 & 3.0 & 8.0 & 6.0 & 7.0 \\ 4.0 & 1.0 & 8.0 & 6.0 & 7.0 & 9.0 & 3.0 & 1.0 & 2.0 & 3.6 \end{bmatrix}$$

$$x_i \in [0,10] \quad (A8)$$

Hartmann H_{6,4} Function:

$$y(x) = - \sum_{i=1}^4 \alpha_i \exp \left[- \sum_{j=1}^6 B_{ij} (x_j - Q_{ij})^2 \right]$$

$$\text{where } \alpha = \begin{bmatrix} 1 \\ 1.2 \\ 3 \\ 3.2 \end{bmatrix}, B = \begin{bmatrix} 10 & 3 & 17 & 3.05 & 1.7 & 8 \\ 0.05 & 10 & 17 & 0.1 & 8 & 14 \\ 3 & 3.5 & 1.7 & 10 & 17 & 8 \\ 17 & 8 & 0.05 & 10 & 0.1 & 14 \end{bmatrix} \text{ and } Q = \begin{bmatrix} 1312 & 2329 & 2348 & 4047 \\ 1696 & 4135 & 1451 & 8828 \\ 5569 & 8307 & 3522 & 8732 \\ 124 & 3736 & 2883 & 5743 \\ 8283 & 1004 & 3047 & 1091 \\ 5886 & 9991 & 6650 & 381 \end{bmatrix}^T \times 10^{-4}$$

$$x_i \in [0, 1] \quad (A9)$$

Michalewicz Function:

$$y(x) = - \sum_{i=1}^{10} \sin(x_i) \left[\sin\left(\frac{ix_i^2}{\pi}\right) \right]^{10}, \quad x_i \in [0, 1] \quad (A10)$$

Trid Function:

$$y(x) = \sum_{i=1}^{10} (x_i - 1)^2 - \sum_{i=2}^{10} x_i x_{i-1}, \quad x_i \in [-100, 100] \quad (A11)$$

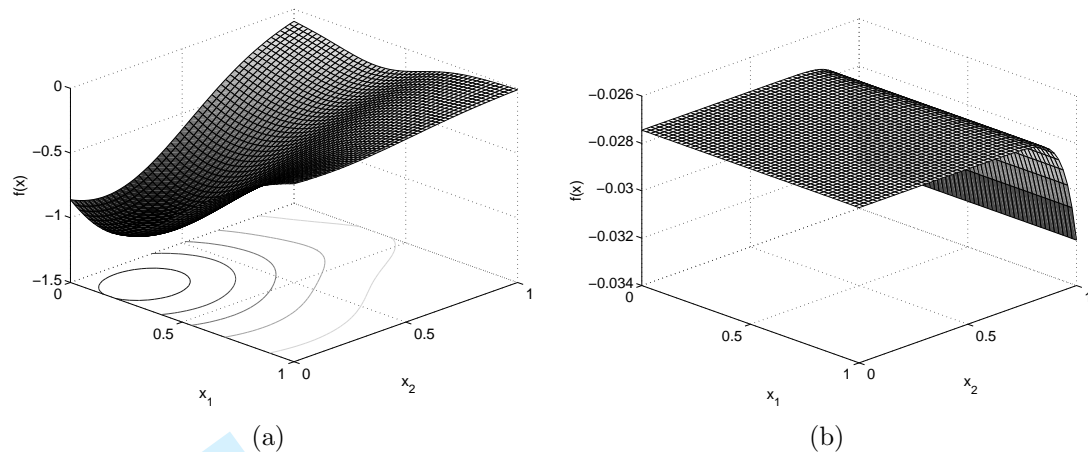


Figure A5. Two dimensional slices through the (a) Hartmann $H_{6,4}$ and (b) Michalewicz test functions (in both cases all other variables equal 0.5)

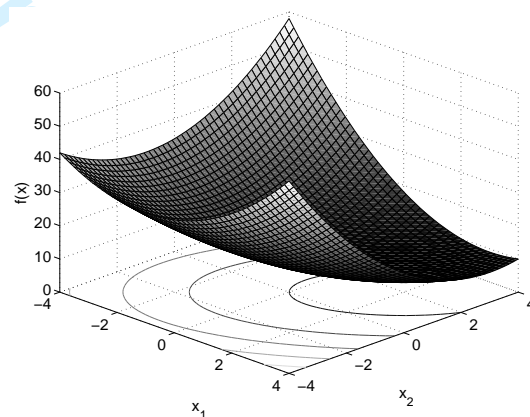


Figure A6. Two dimensional slice through the Trid test function (all other variables equal 0)

Appendix B. Surrogate Model Accuracy

Table B1. Overview of the surrogate modeling and optimization strategies including formulae for the total number of hyperparameters optimized during model construction.

Name	Description	Total no. of Kriging hyperparameters
Kriging	Traditional stationary kriging with p permitted to vary.	$2d + 1$
NS Kriging (2 knots)	Non-stationary kriging model employing 2 knots in the piecewise linear representation of the density function.	$3d + 2$
NS Kriging (5 knots)	Non-stationary kriging model employing 5 knots in the piecewise linear representation of the density function.	$6d + 2$
APNS Kriging	Adaptive partial non-stationary kriging model constructed using a combination of stationary and non-stationary correlations.	between $2d + 1$ and $5d + 2$ for each optimisation

Table B2. r^2 of the three Kriging strategies.

Objective Function	Sample Density ^a	Kriging		NS Kriging (2 knots)		NS Kriging (5 knots)		APNS Kriging	
		Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
Forrester Function	5	0.3389	0.3115	0.4485	0.3435	0.1481	0.2841	0.4081	0.3451
	7	0.6490	0.3104	0.8494	0.2514	0.5752	0.3615	0.7123	0.3016
	10	0.9983	0.0009	0.9949	0.0068	0.7164	0.3352	0.9973	0.0029
Xiong Function	5	0.5723	0.2162	0.6623	0.1686	0.2017	0.2923	0.5774	0.2255
	7	0.7259	0.1144	0.7517	0.1081	0.1573	0.2883	0.7225	0.1122
	10	0.8900	0.0555	0.8978	0.0733	0.8550	0.2885	0.9095	0.0541
Branin Function	5	0.5790	0.2654	0.5634	0.2488	0.0980	0.2305	0.5023	0.2645
	7	0.8447	0.1663	0.8262	0.1612	0.3620	0.3753	0.8174	0.1798
	10	0.9859	0.0276	0.9764	0.0335	0.8994	0.1975	0.9791	0.0378
Mystery Function	5	0.3731	0.1760	0.3193	0.1705	0.0836	0.1655	0.3491	0.1723
	7	0.4131	0.1854	0.3998	0.1442	0.0960	0.1774	0.3955	0.1626
	10	0.5243	0.1340	0.4892	0.1017	0.1673	0.2291	0.4862	0.1195
6 Hump Camelback	5	0.4361	0.2198	0.4146	0.2105	0.0865	0.1707	0.3694	0.2200
	7	0.6560	0.1874	0.6096	0.1738	0.2783	0.3016	0.6182	0.1870
	10	0.8161	0.0743	0.7916	0.0874	0.5790	0.3341	0.7968	0.0955
Paciorek Function	5	0.5673	0.1856	0.6213	0.1675	0.1475	0.2261	0.5371	0.1784
	7	0.7171	0.1086	0.7834	0.1111	0.3607	0.3412	0.6941	0.1415
	10	0.7955	0.1419	0.8353	0.0899	0.6418	0.3050	0.7873	0.1571
Hartmann H _{3,4}	5	0.5551	0.2558	0.5592	0.2368	0.2460	0.2710	0.5366	0.2267
	7	0.7668	0.1959	0.7210	0.1384	0.3882	0.3282	0.7804	0.0887
	10	0.9041	0.0492	0.8718	0.0648	0.6844	0.2925	0.8766	0.0647
Shekel Function	5	0.1076	0.1085	0.1325	0.1016	0.0259	0.0707	0.1184	0.1253
	7	0.2006	0.1926	0.1775	0.1507	0.0518	0.1039	0.2103	0.1841
	10	0.2309	0.2258	0.3464	0.2198	0.1216	0.1929	0.3236	0.2283
Hartmann H _{6,4}	5	0.0714	0.0631	0.1468	0.0917	0.0270	0.0602	0.0913	0.0899
	7	0.1012	0.0938	0.2137	0.1324	0.0438	0.0928	0.1673	0.1242
	10	0.2073	0.1807	0.3344	0.1597	0.1155	0.1448	0.2893	0.1852
Michalewicz Function	5	0.2815	0.2049	0.4446	0.2783	0.4562	0.4200	0.3491	0.2572
	7	0.8603	0.2137	0.9375	0.1900	0.9930	0.0029	0.8898	0.2416
	10	0.9861	0.0038	0.9971	0.0011	0.9990	0.0005	0.9982	0.0020
Trid Function	5	0.1029	0.1028	0.1863	0.1052	0.0551	0.0758	0.1326	0.1056
	7	0.2230	0.2402	0.5238	0.2245	0.2060	0.1903	0.3840	0.2988
	10	0.5274	0.3470	0.9999	0.0001	0.9751	0.0090	0.9655	0.1412

^aWhere the sample density is given by n/d

Appendix C. Optimization Results

Table B3. RMSE of the three Kriging strategies.

Objective Function	Sample Density	Kriging		NS Kriging (2 knots)		NS Kriging (5 knots)		APNS Kriging	
		Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
Forrester Function	5	3.749	1.038	3.255	1.265	4.203	0.940	3.452	1.240
	7	2.482	1.234	1.449	1.129	2.777	1.263	2.178	1.265
	10	0.233	0.092	0.318	0.199	2.063	1.379	0.267	0.124
Xiong Function	5	0.180	0.044	0.165	0.042	0.239	0.049	0.182	0.053
	7	0.147	0.031	0.142	0.036	0.244	0.045	0.150	0.032
	10	0.093	0.025	0.089	0.028	0.082	0.064	0.083	0.025
Branin Function	5	37.034	13.670	37.748	12.861	54.399	9.345	41.630	41.588
	7	21.353	10.304	22.876	10.163	43.352	15.315	23.274	10.558
	10	6.021	4.127	7.850	5.306	14.614	10.781	6.978	5.390
Mystery Function	5	5.165	0.942	5.482	0.854	5.940	0.549	5.332	0.822
	7	4.953	0.970	5.038	0.854	5.824	0.615	5.1089	0.807
	10	4.339	0.707	4.554	0.528	5.498	0.649	4.594	0.652
6 Hump Camelback	5	0.947	0.245	0.968	0.216	1.132	0.093	1.034	0.233
	7	0.690	0.196	0.743	0.188	0.999	0.235	0.735	0.203
	10	0.493	0.106	0.529	0.109	0.721	0.264	0.519	0.125
Paciorek Function	5	0.515	0.133	0.530	0.237	0.679	0.109	0.600	0.248
	7	0.402	0.129	0.353	0.146	0.588	0.199	0.435	0.200
	10	0.309	0.092	0.280	0.081	0.411	0.171	0.317	0.110
Hartmann H _{3,4}	5	0.673	0.223	0.678	0.219	0.850	0.131	0.697	0.203
	7	0.461	0.187	0.519	0.142	0.750	0.203	0.463	0.114
	10	0.293	0.073	0.340	0.095	0.513	0.206	0.335	0.095
Shekel Function	5	0.120	0.049	0.118	0.048	0.104	0.0205	0.121	0.048
	7	0.108	0.033	0.110	0.030	0.104	0.0215	0.109	0.035
	10	0.107	0.039	0.098	0.040	0.100	0.0270	0.102	0.042
Hartmann H _{6,4}	5	0.467	0.075	0.435	0.075	0.428	0.049	0.458	0.065
	7	0.446	0.057	0.407	0.071	0.423	0.046	0.418	0.066
	10	0.394	0.070	0.358	0.064	0.401	0.032	0.370	0.069
Michalewicz Function	5	0.506	0.092	0.430	0.151	0.370	0.213	0.479	0.132
	7	0.184	0.122	0.097	0.114	0.045	0.009	0.131	0.146
	10	0.067	0.009	0.030	0.006	0.019	0.004	0.022	0.009
Trid Function	5	14553.4	1394.2	13704.6	1303.4	14383.6	958.1	14397.4	1341.3
	7	12669.2	2680.2	9534.7	2554.9	12814.4	1387.6	11153.9	3548.2
	10	8316.1	5251.1	252.9	76.0	2152.7	383.6	871.0	2540.0

Table B4. MAE of the three Kriging strategies.

Objective Function	Sample Density	Kriging		NS Kriging (2 knots)		NS Kriging (5 knots)		APNS Kriging	
		Mean	Std.	Mean	Std.	Mean	Std.	Mean	Std.
Forrester Function	5	12.497	4.563	11.197	5.295	14.344	3.177	11.729	5.196
	7	9.721	5.122	6.134	4.707	11.590	4.576	8.804	5.122
	10	1.640	0.722	1.885	0.943	9.706	4.921	1.778	0.802
Xiong Function	5	0.492	0.093	0.476	0.114	0.533	0.092	0.499	0.129
	7	0.465	0.110	0.458	0.126	0.544	0.053	0.474	0.116
	10	0.313	0.089	0.325	0.090	0.280	0.104	0.308	0.074
Branin Function	5	194.2	65.1	195.7	62.8	236.9	33.3	206.0	63.6
	7	149.8	66.5	154.9	64.4	214.7	57.3	154.9	65.8
	10	51.0	37.6	65.6	47.9	94.1	60.2	55.8	46.0
Mystery Function	5	14.772	3.325	15.559	3.227	16.5	2.029	15.180	3.004
	7	15.782	3.322	16.233	3.203	16.7	2.221	15.936	3.404
	10	14.257	2.037	14.727	2.144	15.9	2.010	14.603	2.495
6 Hump Camelback	5	3.658	0.850	3.684	0.782	4.126	0.549	3.672	0.909
	7	3.393	0.729	3.560	0.729	4.056	0.841	3.487	0.758
	10	2.665	0.840	2.787	0.852	3.169	0.928	2.661	0.899
Paciorek Function	5	2.231	0.687	2.559	1.215	1.796	1.070	2.713	1.427
	7	2.099	0.644	2.187	0.658	2.074	1.179	2.362	1.022
	10	1.879	0.476	2.158	0.812	2.227	1.112	2.014	0.542
Hartmann H _{3,4}	5	2.445	0.659	2.561	0.696	2.866	0.428	2.633	0.630
	7	1.988	0.633	2.161	0.578	2.670	0.524	2.061	0.574
	10	1.425	0.432	1.598	0.464	2.186	0.515	1.622	0.494
Shekel Function	5	0.986	0.145	0.965	0.138	1.021	0.073	0.984	0.143
	7	0.933	0.0896	0.936	0.103	0.990	0.075	0.919	0.099
	10	0.946	0.132	0.918	0.116	0.988	0.096	0.932	0.113
Hartmann H _{6,4}	5	2.665	0.160	2.633	0.304	2.646	0.164	2.702	0.177
	7	2.585	0.168	2.549	0.304	2.625	0.143	2.584	0.177
	10	2.504	0.263	2.400	0.291	2.597	0.251	2.410	0.332
Michalewicz Function	5	1.872	0.366	1.621	0.573	1.457	0.818	1.778	0.502
	7	0.677	0.424	0.380	0.454	0.197	0.055	0.483	0.509
	10	0.2691	0.039	0.128	0.025	0.083	0.021	0.102	0.040
Trid Function	5	60770.6	6828.0	57128.9	7792.6	61160.3	4380.6	59429.2	6943.2
	7	54361.9	11766.0	42030.7	11498.4	54551.9	7223.8	48093.0	14191.7
	10	35486.2	22905.4	375.8	212.2	9635.0	2283.5	3253.2	11571.0

Table C1. Analytical test function optimization results.

Objective Function	Kriging		NS Kriging		APNS Kriging		Global Minimum
	Mean	Std.	Mean	Std.	Mean	Std.	
Forrester Function	-6.021	9.04×10^{-4}	-6.021	2.78×10^{-4}	-6.021	1.49×10^{-4}	-6.021
Xiong Function	-0.621	1.70×10^{-2}	-0.625	7.48×10^{-4}	-0.624	2.10×10^{-3}	-0.625
Branin Function	0.402	6.40×10^{-3}	0.405	7.54×10^{-3}	0.403	6.53×10^{-3}	0.398
Mystery Function	-6.495	2.51×10^{-2}	-6.431	1.66×10^{-1}	-6.459	7.74×10^{-2}	-6.514
6-Hump Camelback Function	-1.029	4.93×10^{-3}	-1.029	6.61×10^{-3}	-1.030	2.24×10^{-3}	-1.032
Paciorek Function	-1.000	7.08×10^{-4}	-1.000	4.78×10^{-5}	-1.000	3.44×10^{-4}	-1.000
Hartmann H _{3,4}	-3.863	4.37×10^{-4}	-3.861	3.15×10^{-3}	-3.860	4.18×10^{-3}	-3.863
Shekel Function	-4.681	2.750	-5.270	3.250	-5.273	2.947	-10.153
Hartmann H _{6,4}	-3.148	0.275	-2.818	0.570	-3.051	0.415	-3.322
Michalewicz Function	-4.924	0.125	-4.930	0.166	-4.936	0.141	-5.062
Trid Function	-20.730	75.654	-57.727	59.920	-49.112	62.746	-200

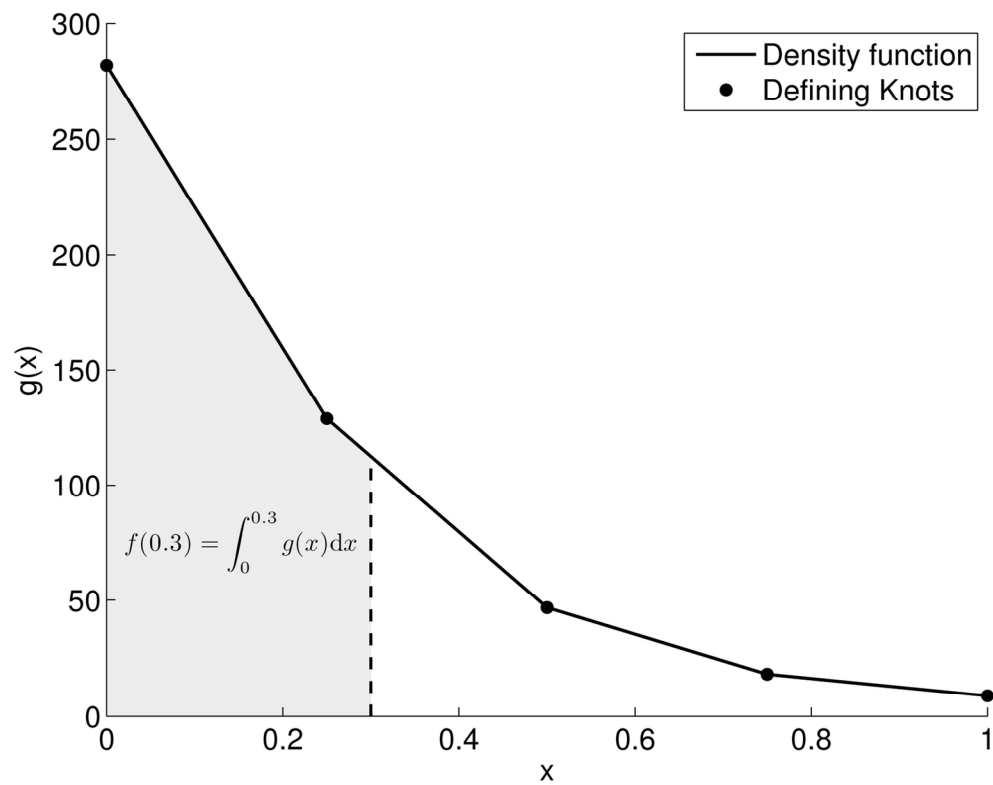


Illustration of the piecewise linear representation of the density function (a) and the resulting non-linear mapping of the Xiong function (b)
142x112mm (300 x 300 DPI)

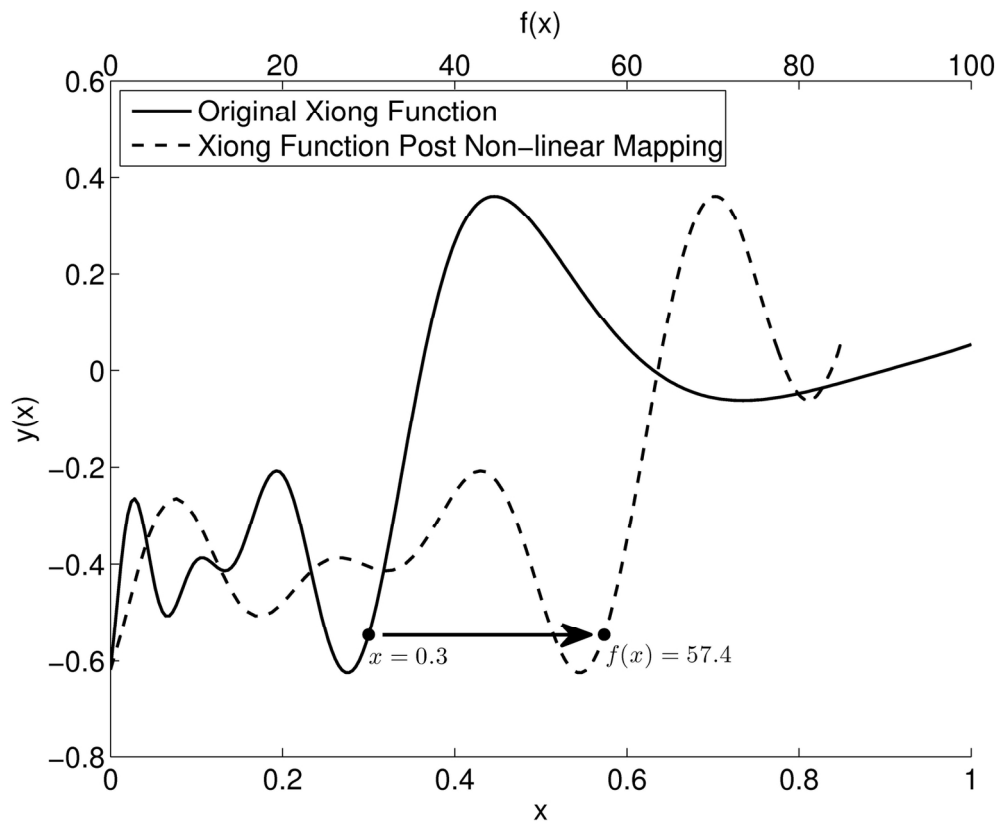
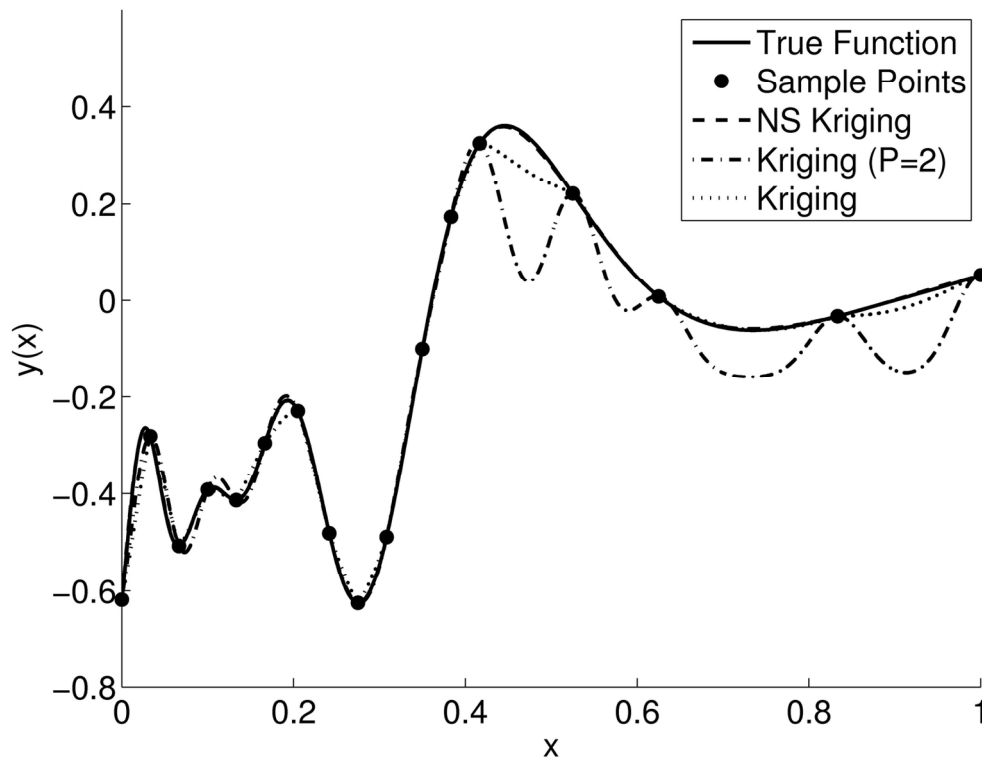
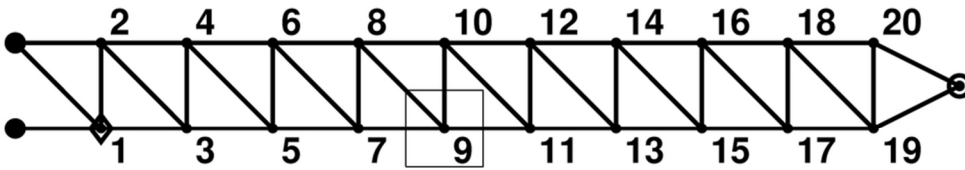


Illustration of the piecewise linear representation of the density function (a) and the resulting non-linear mapping of the Xiong function (b)
150x123mm (300 x 300 DPI)

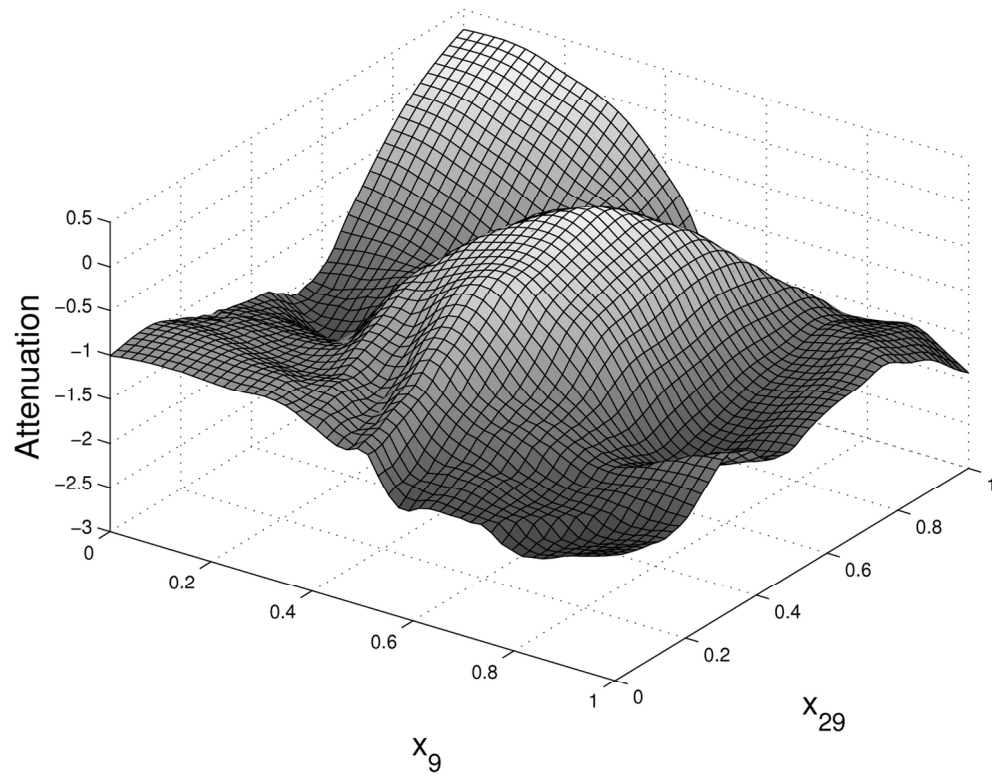


A one dimensional demonstration of the accuracy of non-stationary kriging N.B. the non-stationary kriging model almost exactly represents the true objective function (sampling plan taken from Xiong et al.)
143x112mm (300 x 300 DPI)

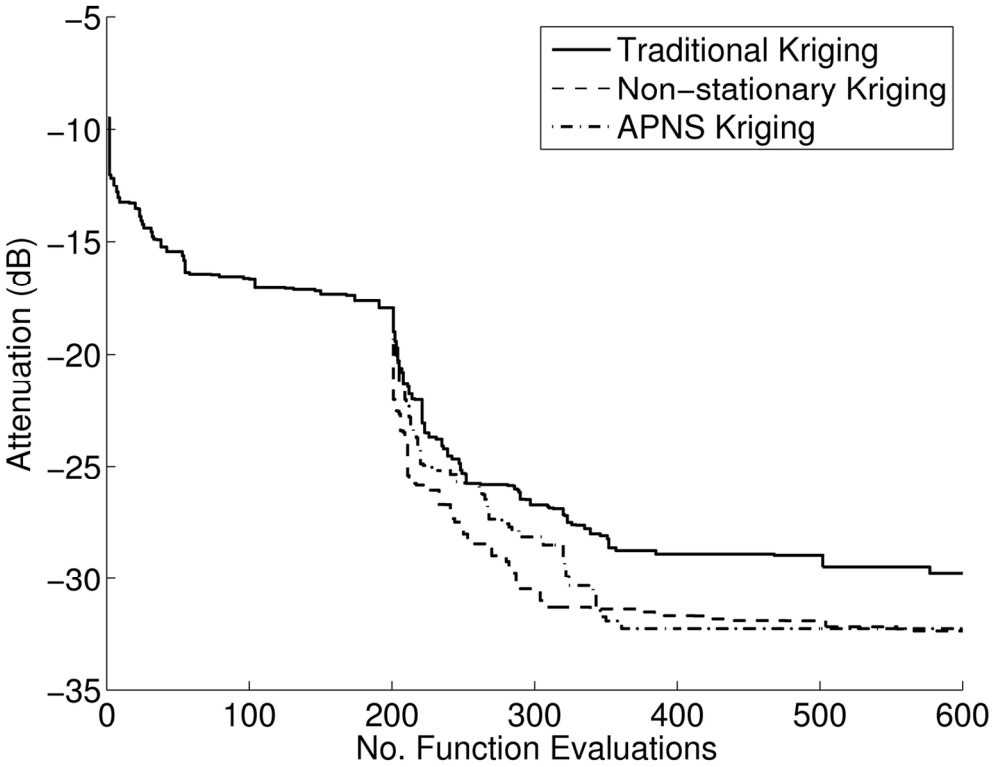


- Encastrate
- Node to be isolated from vibration
- ◆ Forcing point
- Bounds on node 9 movement

The baseline truss geometry (a) and a 2D example of the design space (b)
81x39mm (300 x 300 DPI)

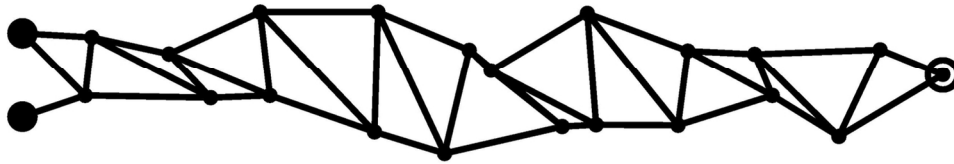


The baseline truss geometry (a) and a 2D example of the design space (b)
142x110mm (300 x 300 DPI)

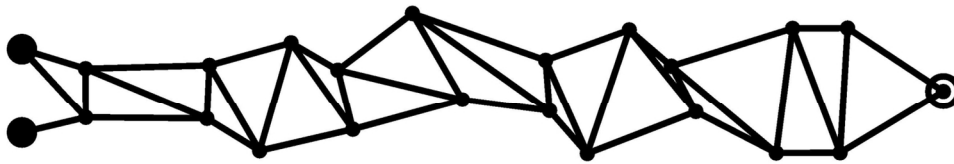


Optimization history for the 40D attenuation problem
143x112mm (300 x 300 DPI)

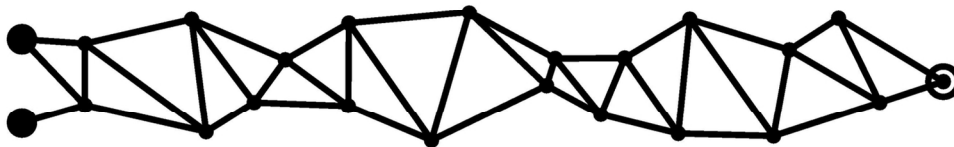
Stationary Kriging



Non-Stationary Kriging

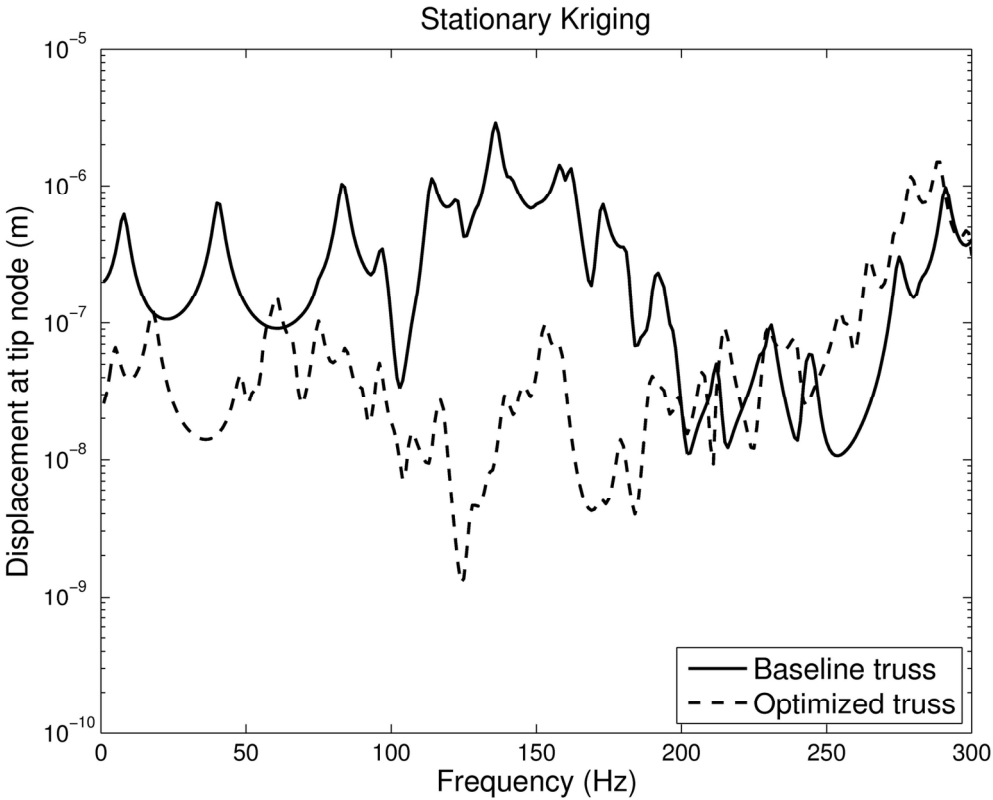


APNS Kriging

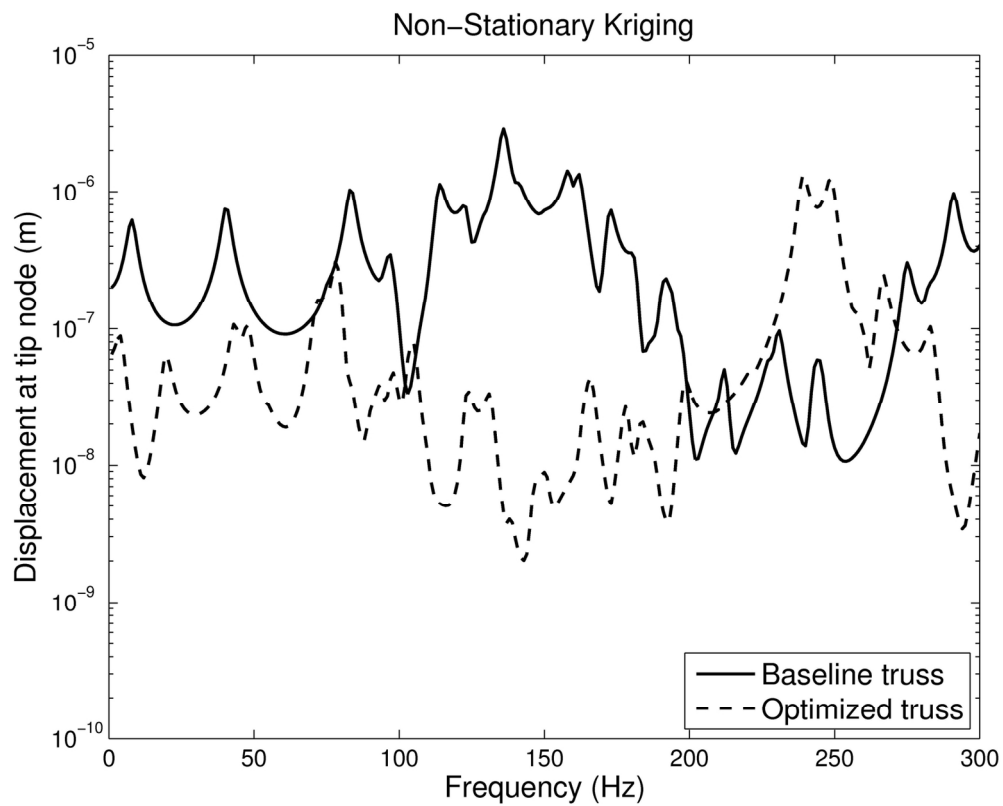


Example truss designs from each of the three optimizations and the frequency response of each design
131x147mm (300 x 300 DPI)

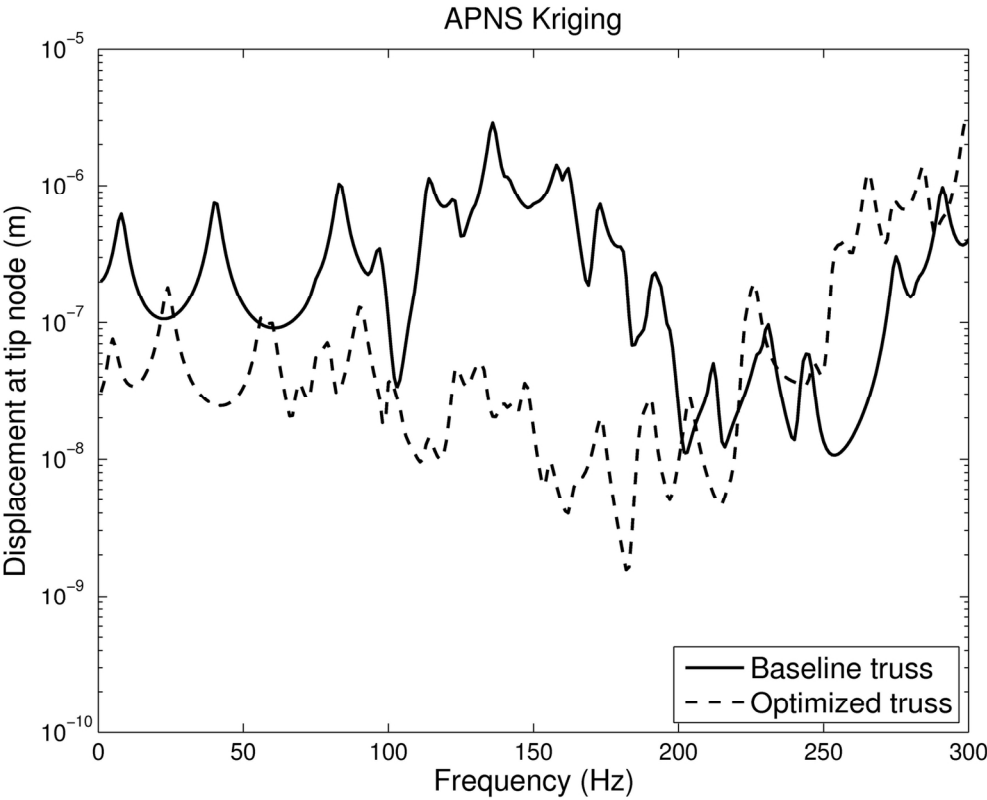




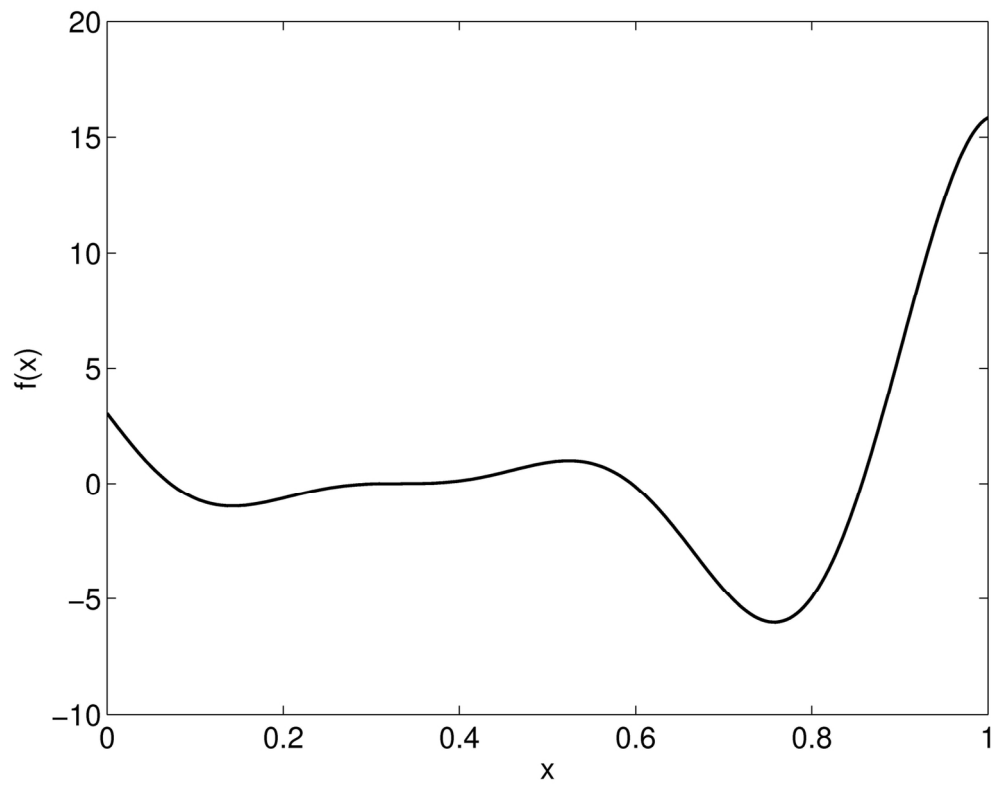
Example truss designs from each of the three optimizations and the frequency response of each design
147x119mm (300 x 300 DPI)



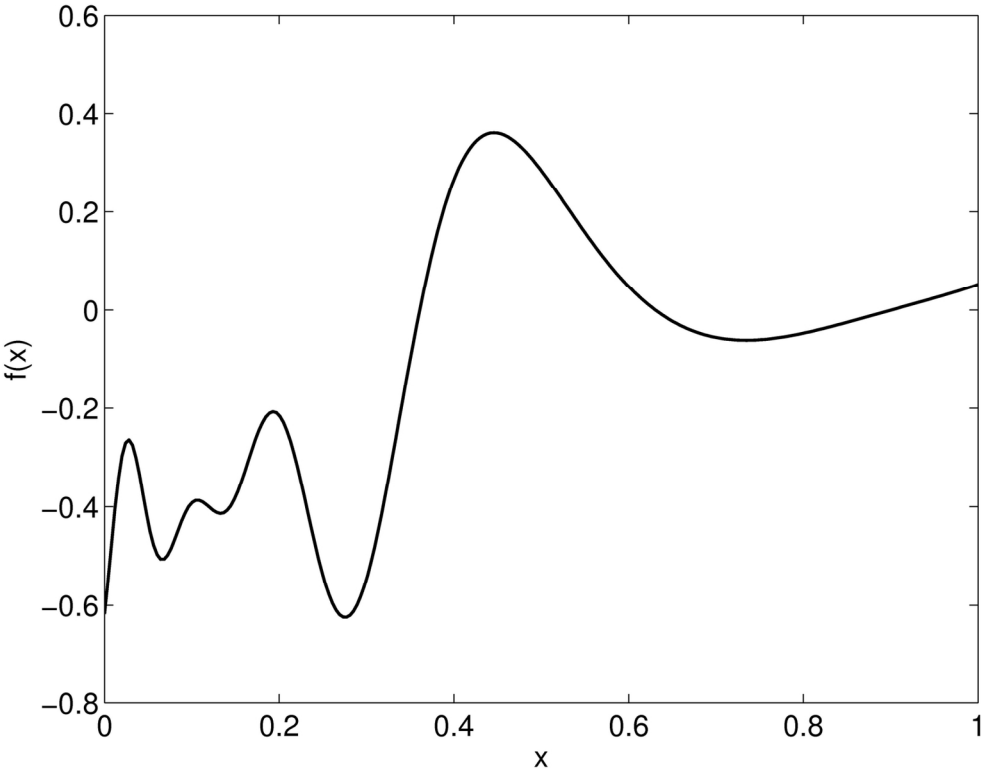
Example truss designs from each of the three optimizations and the frequency response of each design
147x119mm (300 x 300 DPI)



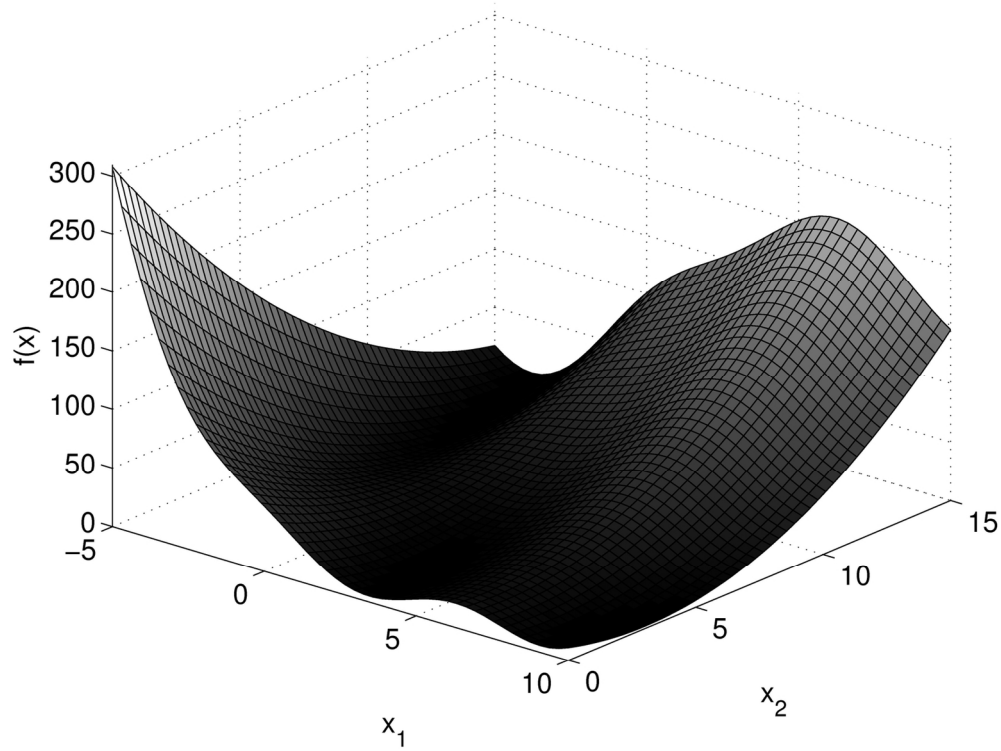
Example truss designs from each of the three optimizations and the frequency response of each design
147x119mm (300 x 300 DPI)



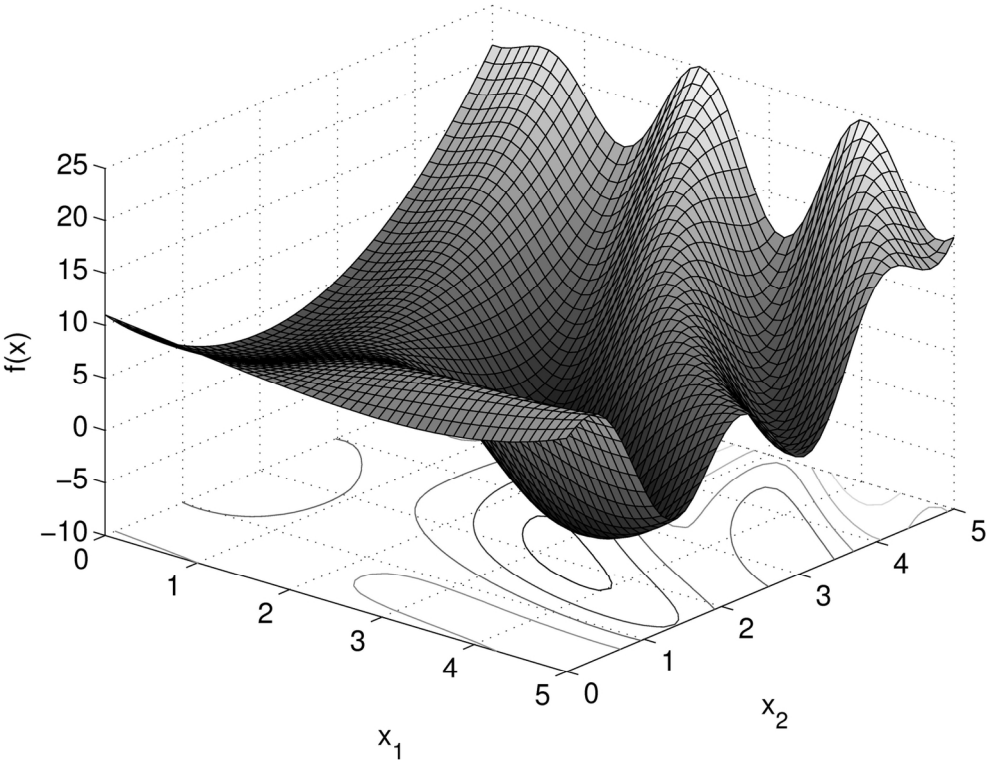
Analytical (a) Forrester and (b) Xiong test functions
142x112mm (300 x 300 DPI)



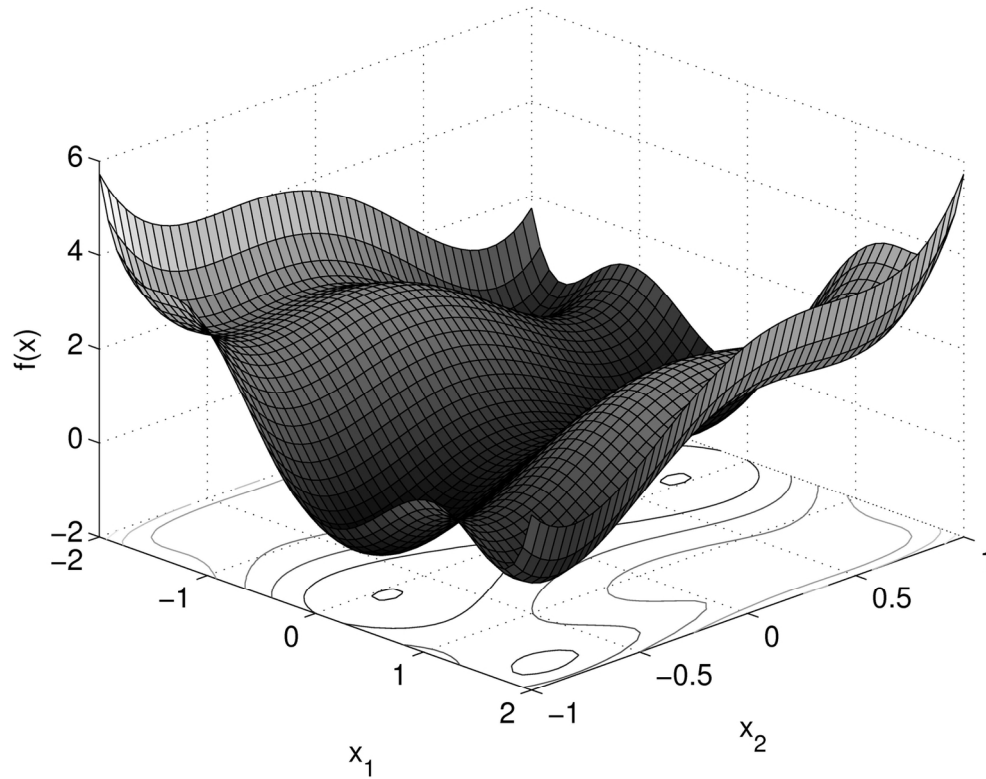
Analytical (a) Forrester and (b) Xiong test functions
142x111mm (300 x 300 DPI)



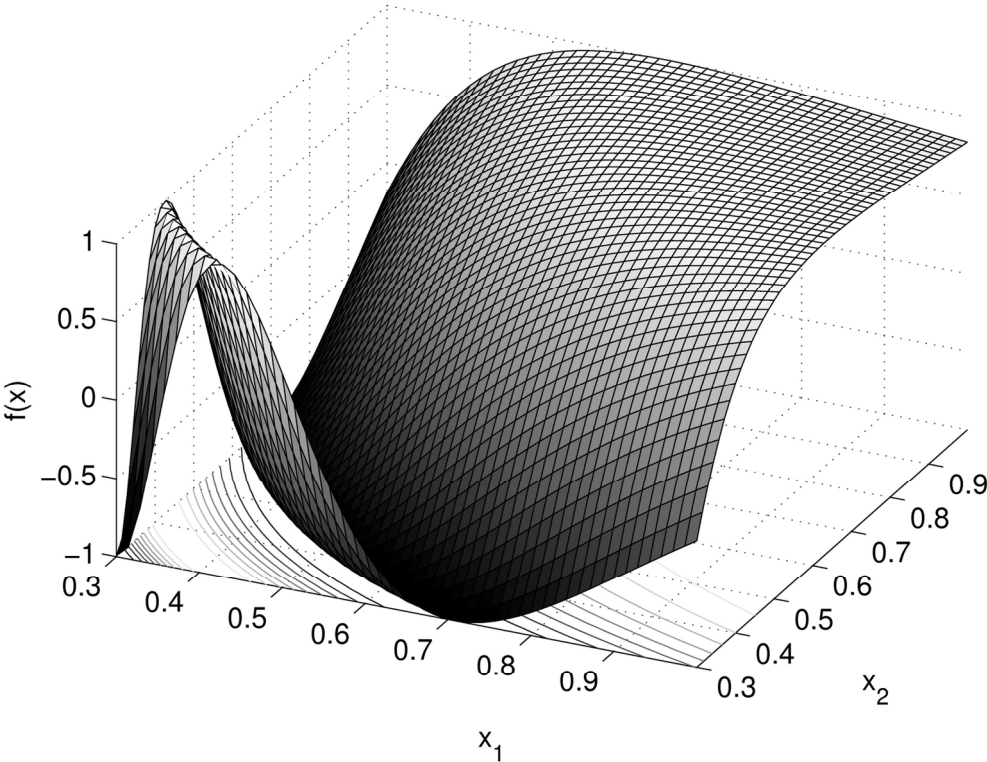
Analytical (a) Branin and (b) Mystery test functions
142x107mm (300 x 300 DPI)



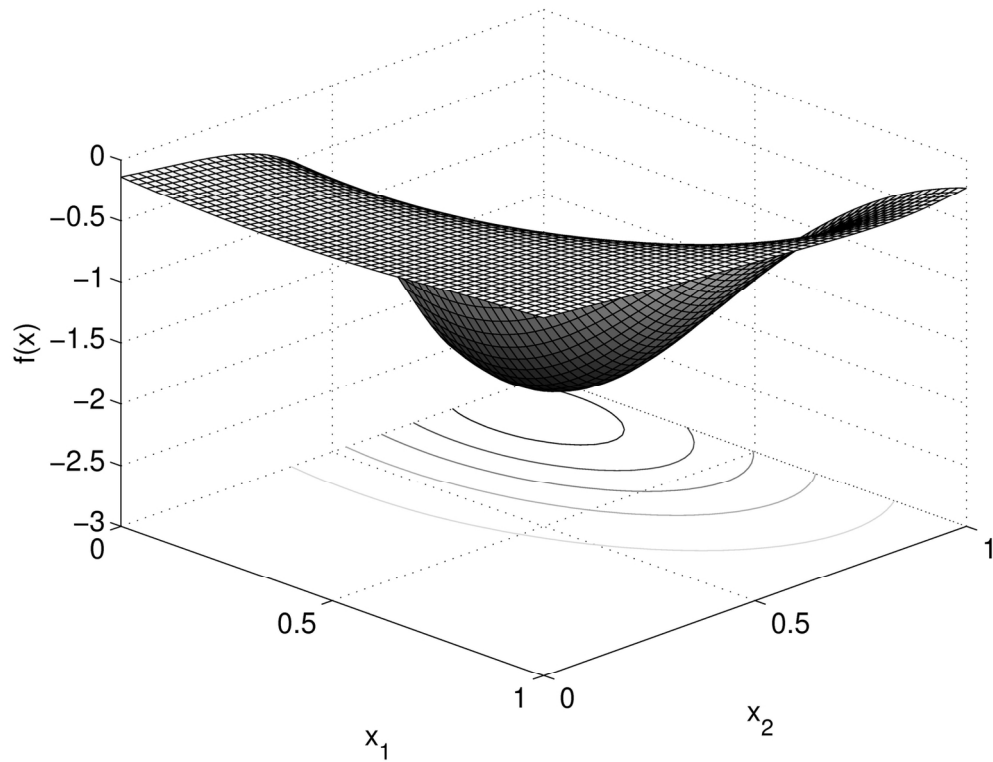
Analytical (a) Branin and (b) Mystery test functions
142x109mm (300 x 300 DPI)



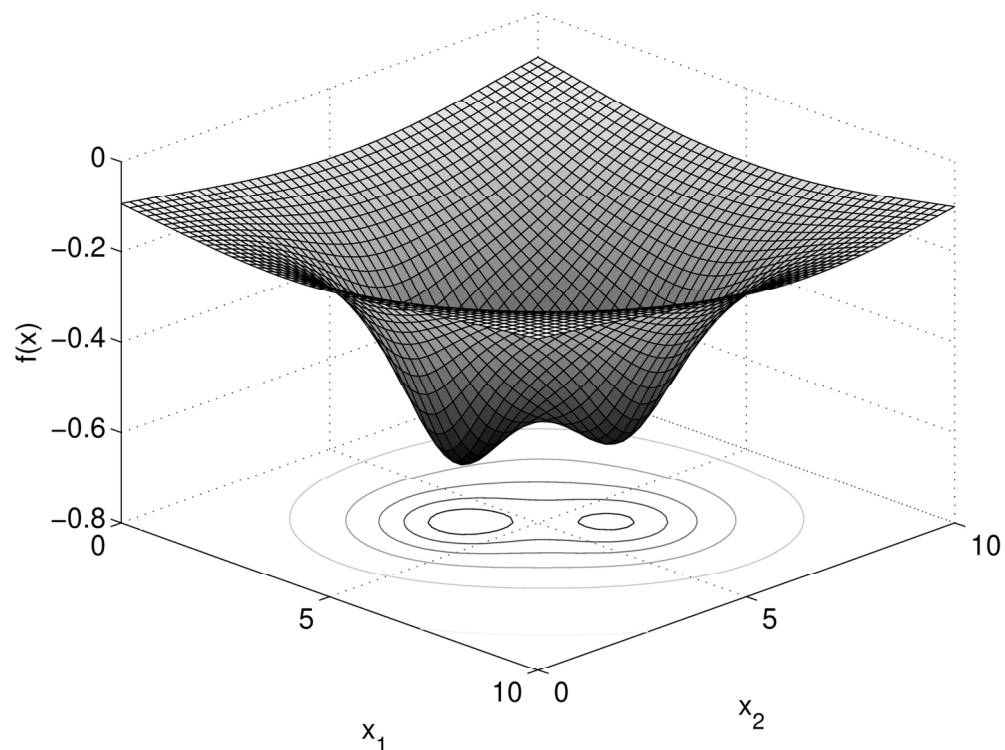
Analytical (a) six hump camelback and (b) Paciorek test functions
142x111mm (300 x 300 DPI)



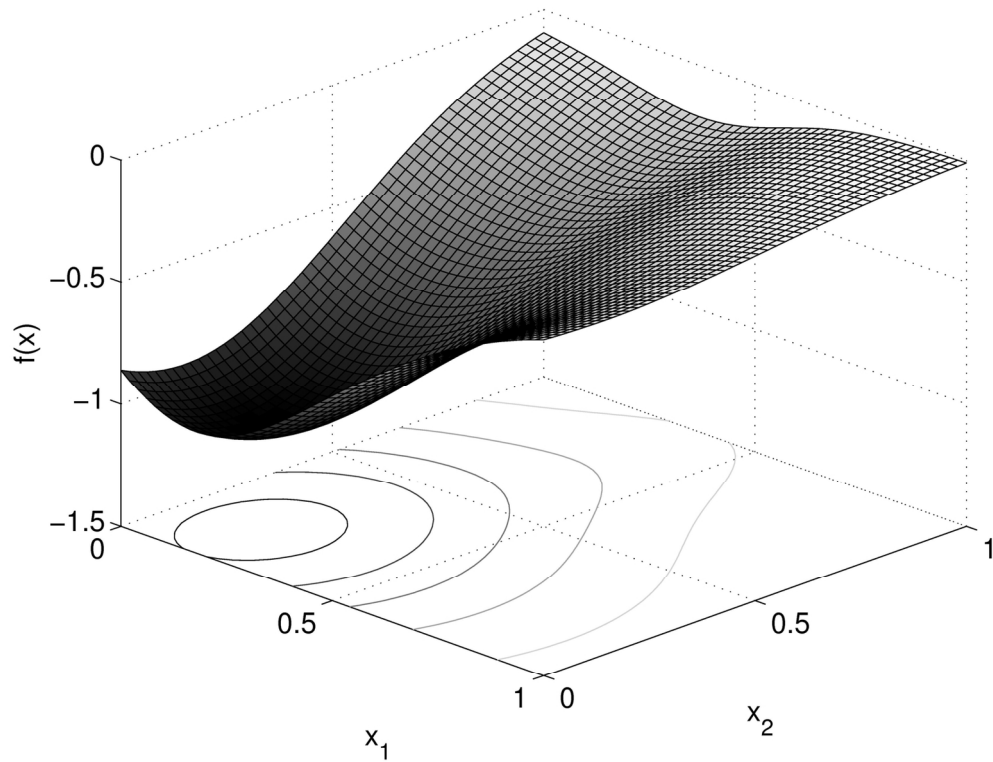
Analytical (a) six hump camelback and (b) Paciorek test functions
142x109mm (300 x 300 DPI)



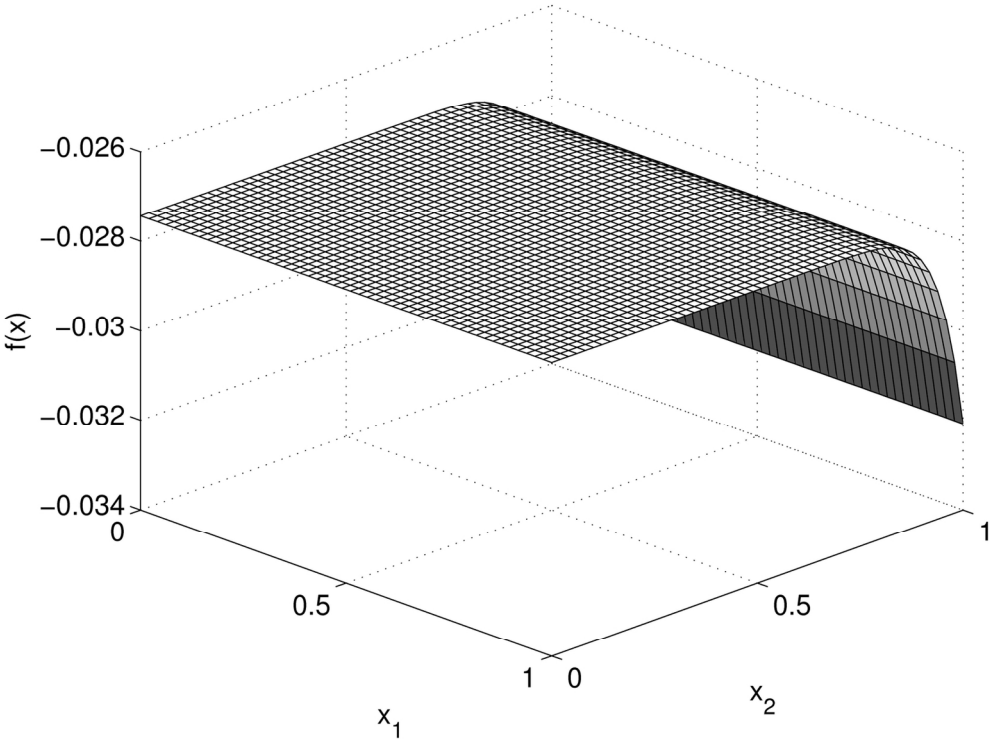
Two dimensional slices through the (a) Hartmann H34 (all other variables equal 0.5) and (b) Shekel test functions (all other variables equal 5)
142x108mm (300 x 300 DPI)



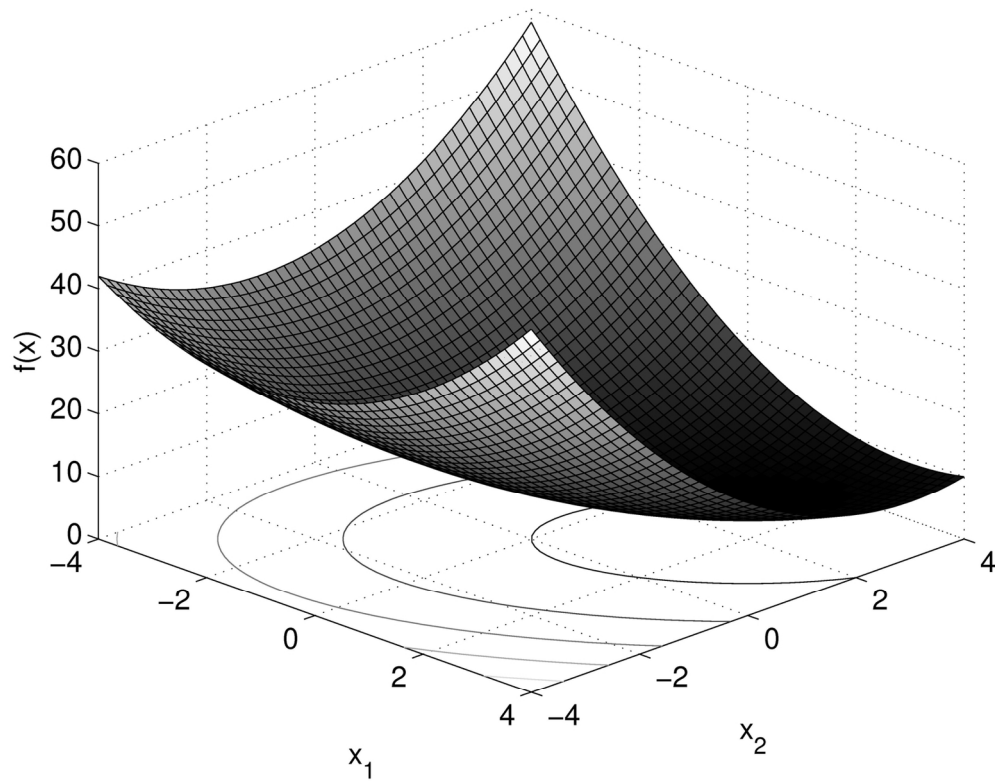
Two dimensional slices through the (a) Hartmann H34 (all other variables equal 0.5) and (b) Shekel test functions (all other variables equal 5)
142x107mm (300 x 300 DPI)



Two dimensional slices through the (a) Hartmann H64 and (b) Michalewicz test functions (in both cases all other variables equal 0.5)
142x108mm (300 x 300 DPI)



Two dimensional slices through the (a) Hartmann H64 and (b) Michalewicz test functions (in both cases all other variables equal 0.5)
142x106mm (300 x 300 DPI)



Two dimensional slice through the Trid test function (all other variables equal 0)
142x111mm (300 x 300 DPI)