Date: 12/12/2014

Locally-Optimized Covariance Kriging Window Size Optimization

Daniel L. Clark, Jr.

The propose of this paper is to formulate and implement an optimization routine for selecting Locally Optimized Covariance Kriging (LOC-Kriging) window’s. LOC-Kriging is a proposed method for estimating a global non-stationary covariance structure by aggregating individual local stationary covariance Kriging structure by aggregating individual local stationary covariance Kriging structures at different locations. Once the individual structures or windows cover the entire design space they are aggregated with a predetermined individual membership weighting function. The proposed non-stationary covariance structure improves the estimations produced by Kriging for data collected unevenly and adaptively, and reduces the expected mean squared error. To analyze this optimized method the predictions produced will be compared to the predictions of the stationary Ordinary Kriging method. The LOC-Kriging will be shown to produce smaller global maximum standard deviations, as well as, smaller root mean squared error for numerical example functions with data collected adaptively and unevenly.

# Motivation and Description

I

n an early stage of engineering design exploration, engineers perform extensive simulation based on design explorations that include multidisciplinary design optimizations and uncertainty quantifications. Because of the increasing scale and complexity of simulation models, the computational cost of a design exploration often becomes one of the major challenges. To alleviate the computational burden, various surrogate models are used instead of expensive actual simulation models [1-6]. Among many different surrogate model methods, Kriging has gained its popularity due to its high accuracy and flexibility of representing non-linear system responses [7,8].

Kriging is basically an optimal interpolation method combining both the trend regression and the realization of a local random process that is typically represented by a covariance function with zero mean. As one of the benefits of using Kriging, an expected Mean Square Error (MSE) is calculated along with a response prediction at any location of interest. In the Efficient Global Optimization (EGO) method [8], the expected improvement defined as a function of MSE and a current minimum response plays a key role to deploy adaptive data points sequentially in the process of Kriging based design optimization. The Efficient Global Reliability Analysis (EGRA) [9] also uses EGO to construct an adaptive Kriging model of a limit state boundary for efficient probability integration. The Multiple Surrogate Efficient Global Optimization (MSEGO) method [10] uses MSE from Kriging as a universal uncertainty estimate for other types of surrogate models.

However, the estimated MSE from a traditional Kriging is calculated based on a stationary covariance structure that is constructed by a correlation function among the given samples. In most engineering applications, the correlation is defined by a Gaussian function, in which an optimum correlation parameter is determined by using the maximum likelihood approach. Typically, the covariance function is assumed to be stationary over the sample domain of interest, which means there is only one optimal correlation parameter. Generally, the stationary Kriging is adequate when the samples are uniformly distributed across the space. However, when data is scattered unevenly, as in most of efficient adaptive sampling techniques, Kriging may fail to find a globally well-fitted stationary covariance structure. Even for a system with moderate nonlinearity, as in most of engineering problems, the estimated error bounds (±3σ) with adaptive data samples are severely amplified compared to that of uniform samples as shown in Fig. 1. By using the stationary covariance, it is assumed that the stochastic random process has a uniform fluctuation between samples that can be captured by a single correlation function over the entire domain.

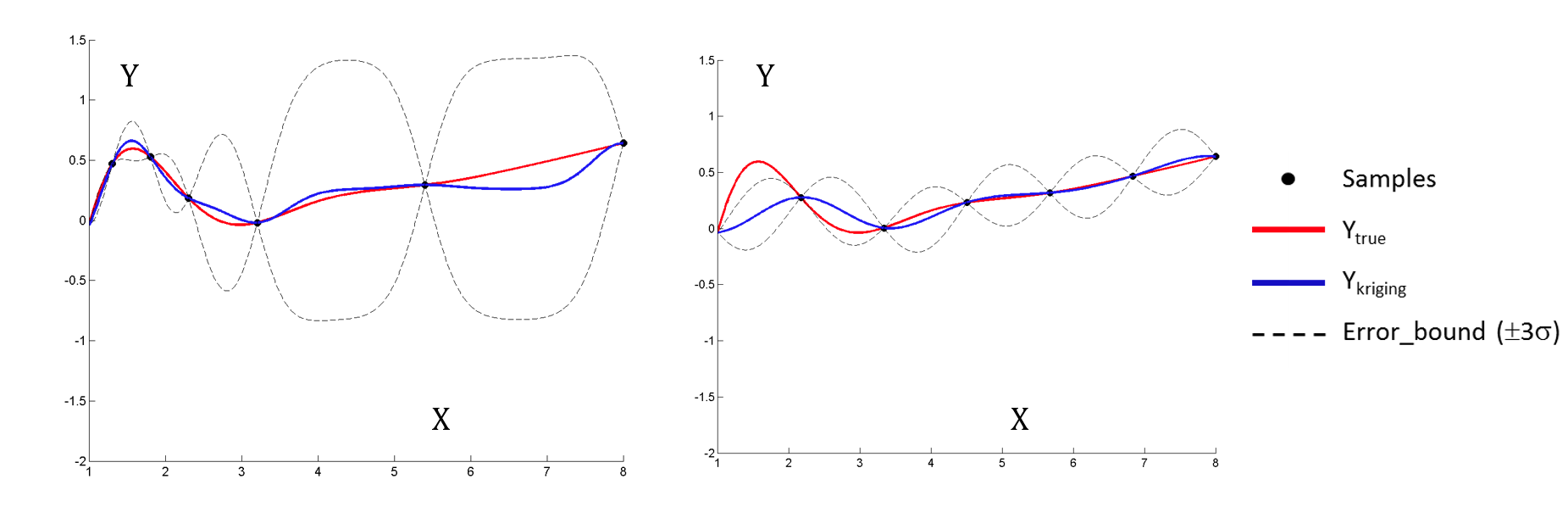


Figure . Stationary Kriging responses with ±3σ error bounds. *Adaptive points (left) and evenly distributed points (right).*

The assumption is invalid for adaptively and unevenly collected samples used to trace nonlinear responses of an engineering system. Regarding the issue with unevenly distributed data, many different non-stationary covariance based Kriging (NS-Kriging) methods were proposed by researchers, especially in the fields of geostatistics and environmental processes [11-16]. However, most of the proposed methods were developed for situations that have a significant number of samples with low-dimensional geostatistical problems.

In this paper, to address the challenges with unevenly distributed samples without a significant computational cost increase, a new methodology, locally-optimized covariance based Kriging (LOC-Kriging) method is proposed. The main idea of the proposed method is to approximate a non-stationary covariance structure by using multiple stationary structures built in locally optimized covariance windows. Individual local Kriging models are constructed within the local coverage by the identified local stationary structures. The prediction of LOC-Kriging is obtained by aggregating the local estimations considering a membership weighting function at a point of interest.

The paper is organized as follows. In Section 2, the proposed methodology of Locally-Optimized Covariance Kriging is presented, including: the optimization formulation, membership weighting function and aggregation formulation. The Necessary data and description of analysis method are presented in Section 3. In the final paper additional sections, Section 4 and Section 5 will be added. Section 4 will include the solution to the optimization problem for a simple 1 and 2 dimensional case, and Section 5 will include the findings and further discussions.

# Proposed Methodology of Locally-Optimized Covariance Kriging (LOC-Kriging)

To alleviate the computational difficulties while addressing the NS covariance structure with adaptively collected data in a practical engineering application, locally-optimized covariance based Kriging (LOC-Kriging) is proposed, in which the non-stationary covariance structure is approximated by using multiple local stationary structures.

LOC-Kriging constructs a finite number of local stationary structures where the optimal combination of sizes is simultaneously optimized. This occurs under the assumption that the window centers are given from density based cluster analysis. The prediction of LOC-Kriging is obtained by combining multiple local stationary models based on an aggregation membership function.

## Local Window Size Optimization Formulation

To avoid an over-parameterization in the global regression and to minimize the finite number of the local models, the local window sizes are optimized as follows:

Find (1a)

Minimize (1b)

Subject to: (1c)

(1d)

(1e)

Where *ω* is the local window size measured by the ratio between the current window and the entire design sapce, *ωmin* is the required minimum size, *Ψ(θ)* is the Kriging likelihood function, *Nωmin* and *Nω* are the numbers of the minimum samples and current samples in the local window respectively, and *λ* is the global coverage parameter with associated upper and lower bounds*, λmin*and *λmax.* The local window size, *ω* can be viewed as a hypersphere radius in a multidimensional problem. The minimization of the sum of likelihoods, *Ψ(θ)i* is selected as the objective function because likelihood has a strong foundation in Kriging and it is believed to be a satisfactory fitting parameter. The global coverage parameter, *λ* is the percent of data points shared by two or more windows, however, this term does not become active unless all data sample points are covered. Based on the user’s intention, different weightings can also be applied to the window size.

Any prior information or knowledge regarding the system local behavior can be implemented into the minimum window size requirement. In the constraints, the minimum number of sample points also needs to be considered to obtain a meaningful optimum correlation parameter with respect to the order of the basis function in the global trend term and the total number of available samples. The coverage parameters upper and lower bounds should be selected so sensible overlap occurs.

The optimization formulated in Eq. 1 is a multi-dimensional optimization problem with *L* design parameters, *ωi.* The multi-dimensional objective function is expected to be nonlinear, therefore the solution is highly influenced by the starting location. In this study, the Design of Experiments, Latin Hypersphere Sampling method is used to determine an approximate global optimuim starting point, so that local gradient-based optimization via fmincon’s interior-point algorithm can be used to find the true optimum.

## Membership Weighting

In the construction of the local Kriging model, the bell-shaped membership function shown in Fig. 2 is used to apply the degree of membership to samples in order to maintain the continuity of the predicted response across finite local window boundaries. The samples within the range of full membership have unit weightings. Over the full membership, there is the transitional range defined by α×ω, where α is the scale factor of the transition range. In this study, the Gaussian function is used to vary the membership between unity and zero within the transition range. Essentially, any transition function can be used, such as a linear, spline, or an exponential function.

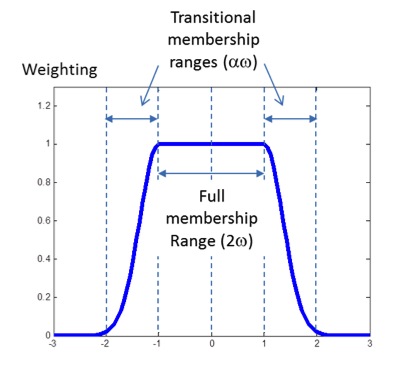


Figure . Membership function for a LOC window.

With the membership weightings on the samples, the local Kriging model is constructed by the generalized least square regression as:

(2)

where is the diagonal matrix of the membership weights. The weighted regression coefficient vector, is calculated by Eq. 3.

(3)

Where . The LOC-Kriging prediction,, at an untried location, *x*, is obtained as:

(4)

where . With multiple local windows over the entire sample domain, the following section shows how to construct and aggregate local Kriging models for response estimation.

## Aggregation Strategy

With unevenly distributed samples within in a design domain of interest, multiple LOC-Krigings will be constructed to cover the entire domain. In this study, the following steps are used to build LOC-Krigings simultaneously:

1. Deploy virtual points using space-filling sampling method such as Optimal Latin Hypercube Sampling.
2. Select center points for the individual local winodws.
3. Optimize the local window sizes and construct the LOC-Krigings.
4. Aggregate the LOC-Krigings.

The coverage of individual LOC-Krigings can be mutually overlapped with each other. Depending on the system response behavior and sampling point arrangement, the different number of LOC windows will be required to cover the entire design domain. For the response prediction at a point, , the multiple predictions from the identified LOC-Krigings are aggregated by using a performance weighting function as:

(5)

where is the number of LOC-Krigings; is the *i*th LOC-Kriging response prediction, and is the *i*th performance weighting factor that will be determined by considering available model assessment measurements.

The individual LOC-Kriging model can be quantitatively assessed by using measurements, such as: a distance from a local model center to an estimation point, estimated variance and confidence intervals, cross-validation errors, and any other local model quality index. In this study, we determined the performance weightings based on the distance function between an estimation location and the centers of active LOC-Krigings. The distance weighting function is defined as:

(6)

where in which is the center of the jth LOC-Kriging window. The same aggregation is also applied for the estimated variance.

# Analysis Method and Solution

In this section, two numerical functions are presented and the performance measures are defined so LOC-Kriging can be compared quantitatively against stationary Kriging. The first example, a simple 1D mathematical problem will be investigated more in-depth. The second example, a 2D mathematical problem was removed from the paper due to length requirments, please email the author for more information.

To measure the performance of the different methods, the root mean squared error (RMSE) between the true and stationary Kriging surfaces is calculated by Eq. 7. Where *N* is the total number of test data points, is the true response, and is the Kriging prediction.

(7)

## One-Dimensional Example

The simple one-dimensional mathematical example as given by Eq. 8 is considered first. The true response, the stationary Ordinary Kriging response, and estimated error bounds are shown in Fig. 3. As compared in Fig. 1, the error bounds are unnecessarily amplified because of the unevenly distributed samples.

(8)

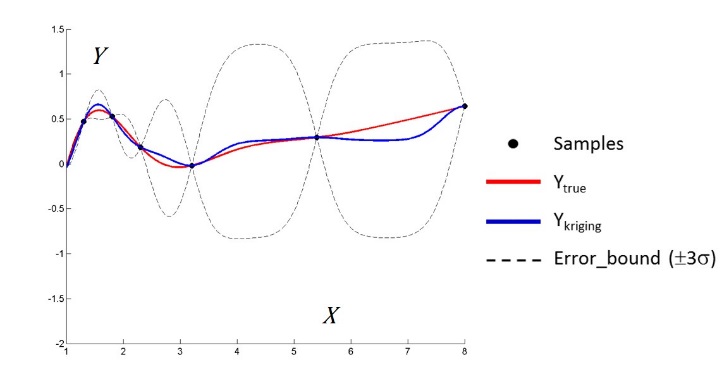


Figure . Mathematical example 1. *True response and stationary Kriging estimates.*

Based on the process described in the previous section with control parameters (ωmin = 0.2, *N*ω\_min = 3, λmin = 0.1 and λmax= 0.2) two local covariance structures are identified with their own range of coverage as shown in Fig. 4. The triangle marks in the figure indicate the center locations of local windows, and the ranges are named with the number in the order of their identification. As shown in the figure, the ranges can be overlapped to each other.

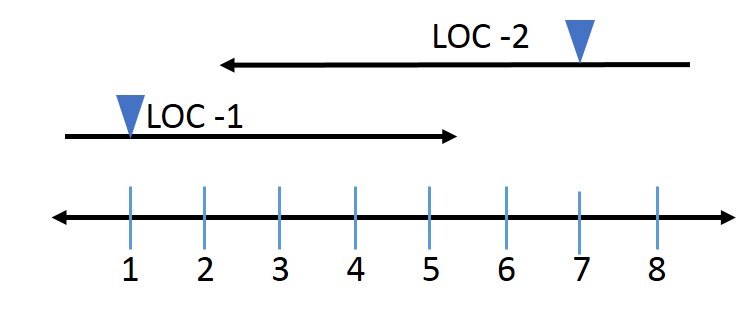
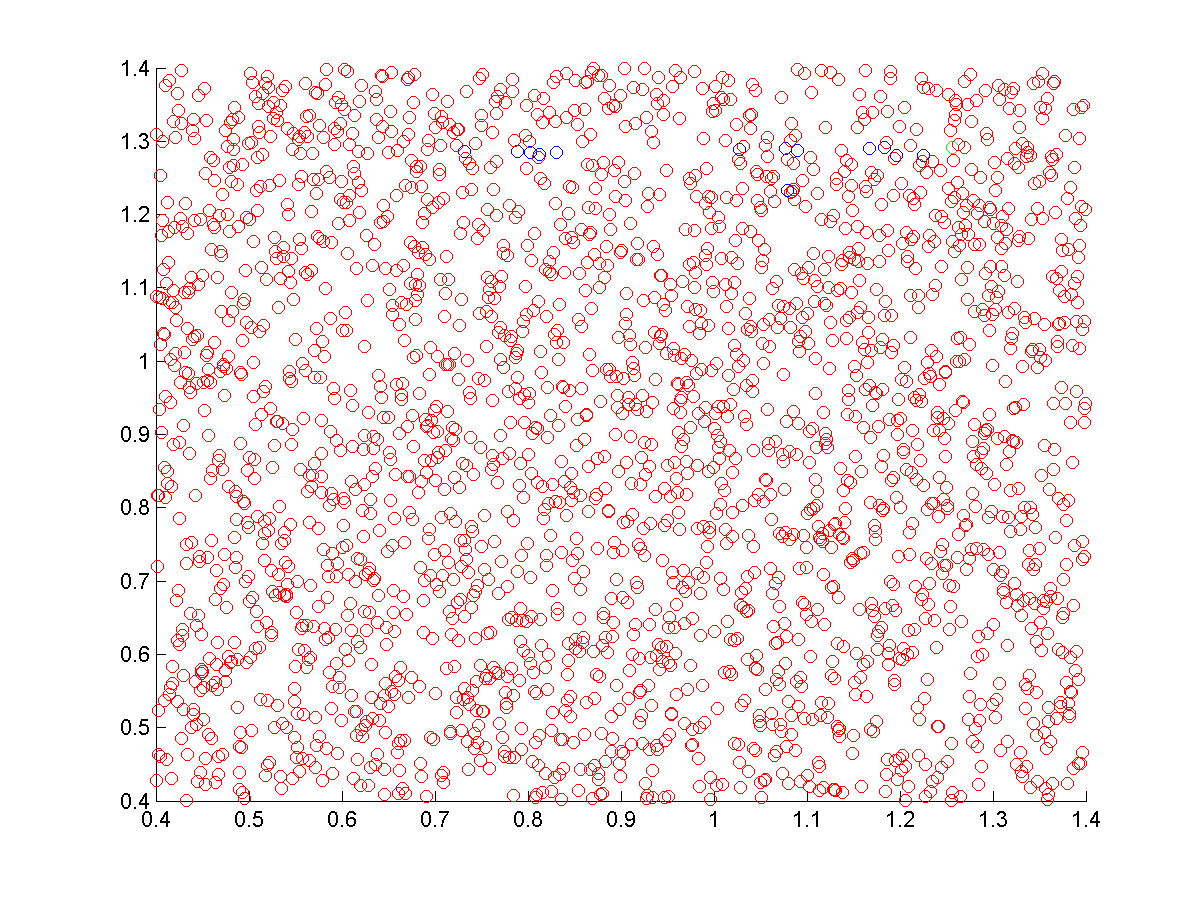


Figure . LOC Ranges. *LOC-1 and LOC-2 respective ranges with center location signified by a triangle.*

Originally a low sample LHS design was going to be used to determine the starting location for a gradient based optimizer. However, from the feasiable region plot Fig. 5, most if not all of the initial points are thrown out. This is because the modified version of DACE used to create the models occasionally generates a negative MSE. Unforutnitly there is not a accepted method to fix this therefore, the models are consider infeasible. So instead of using a gradient based optimizer, the best LHS design point of the 40000 points is selected. This is more computationaly expensive, however, it ensures a true model is selected as the optimum. The optimal location is indicated by a black star with ɷ1 = 0.6285 and ɷ2 = 0.646.



2ɷ1

2ɷ2

LHS Design Points

Figure 5. Optimal Solution. *LHS design points for window sizes.*

The two LOC-Krigings, LOC-1 and LOC-2 are shown in the figure below. The same legends are used in Fig. 3. It is noted that the Kriging prediction outside of the LOC bounds will be weighted according to the membership function when they are aggregated together.

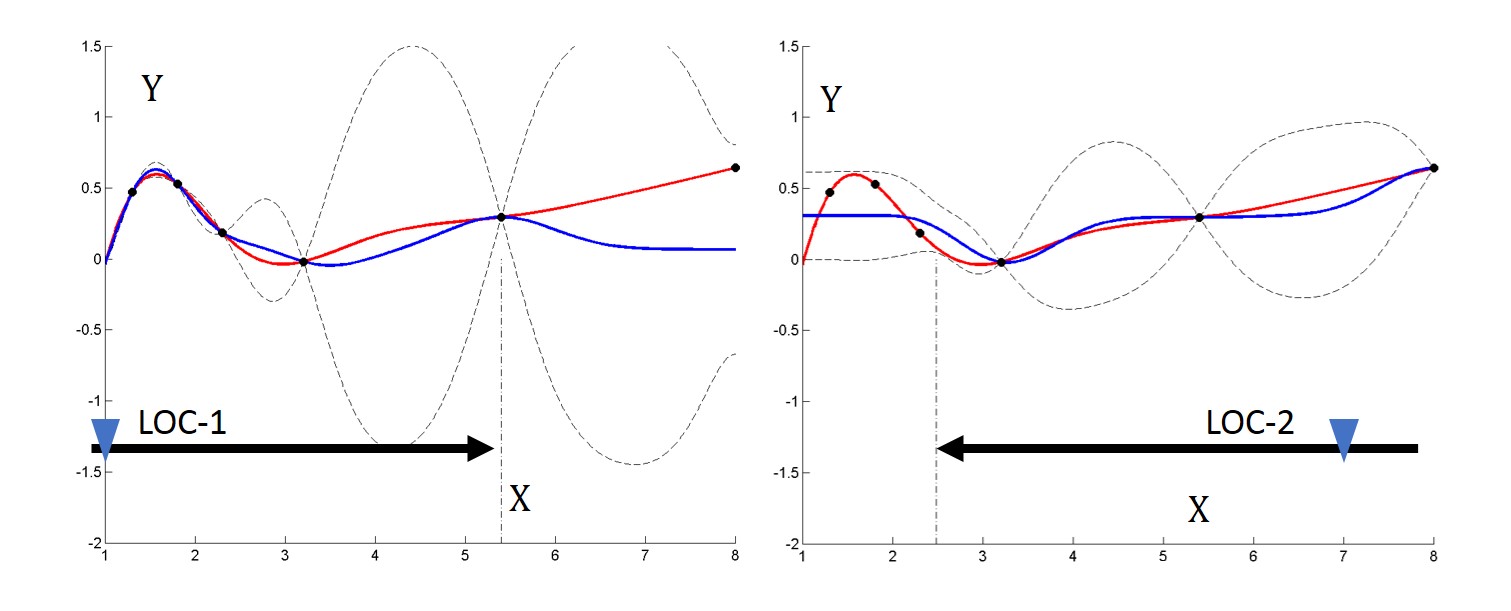


Figure . LOC Predictions. *LOC-Krigings for LOC-1 and LOC-2 with corresponding ranges signified via black arrows.*

By aggregating the two LOC-Krigings, the approximated non-stationary Kriging prediction is obtained and compared to the stationary Kriging response side by side in Fig. 7. It visibly shows LOC-Kriging produces more reasonable predictions and error bounds based on the collected samples. Credibility information could also be measured by cross validation or sample validation in sequential samplings can help to discuss the validity of the obtained error bounds.

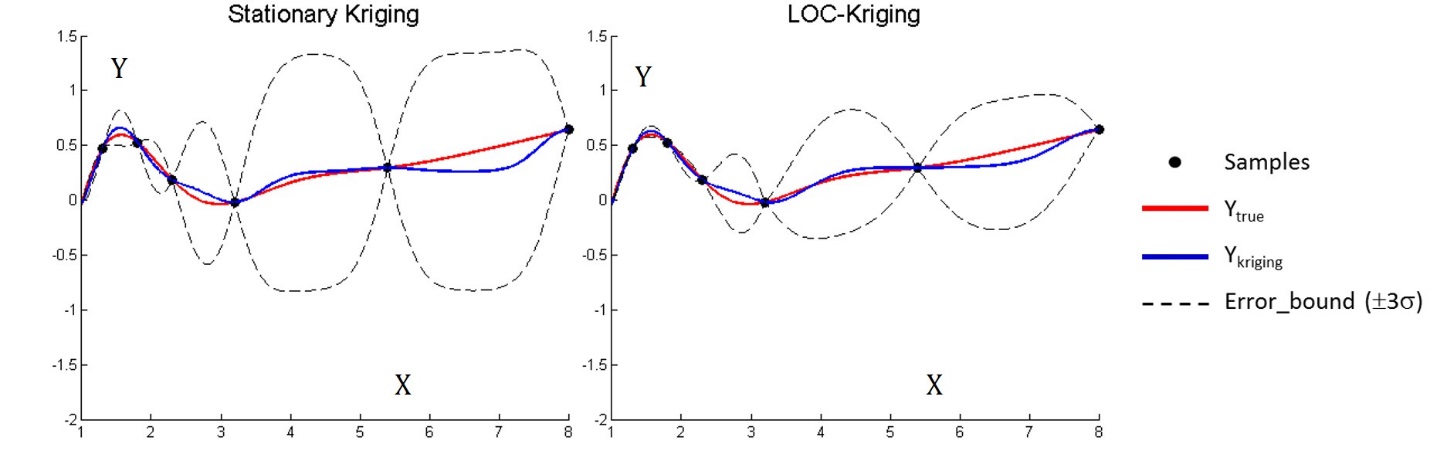
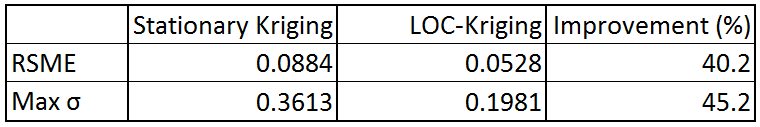


Figure . Comparison of Methods. *Stationary Kriging (left) and aggregated LOC-Kriging (right).*

The quantative results are shown in Table 1, the aggregated LOC-Kriging shows 40.2% RMSE improvement in the response prediction against the stationary Kriging. Also the maximum error reduced by 45.2%. This was achieved by relaxing the unnecessary stationary covariance requirement.

Table . Performance Comparison. *Root mean square error and maximum standard deviation performance measures between Stationary Kriging and LOC-Kriging.*



# Findings and Further Discussion

To address the non-stationary covariance structure with unevenly distributed adaptive samples, the locally optimized covariance Kriging (LOC-Kriging) method is proposed. In the proposed method, an optimization problem is formulated to find the optimal combination of window sizes that cover the entire design domain and representative local stationary covariance structures. For the identified local covariance structures, multiple stationary Krigings are constructed and aggregated by using membership weighting functions to approximate the non-stationary Kriging over the entire design domain of interest. The proposed method shows several benefits. First, since it creates separated membership sets of samples throughout the proposed approach, this is equivalent to the divide-and-conquer strategy, which can take an advantage of parallel computing and reduce the computational cost significantly. Second, it is found that by using multiple covariance structures, LOC-Kriging effectively removes the unnecessary amplified errors and provides more meaningful error bounds that can be useful in any follow-up process, such as the estimation of expected improvement for adaptive samplings. Third, it is also found that the prediction quality of LOC-Kriging is improved by relaxing the stationary covariance requirement.

# References

1Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P., “Design and analysis of computer experiments,” *Statistical Science*, Vol. 4, No 4, 1989, pp. 409–435

2Meckesheimer, M., Booker, A. J., Barton, R. R., and Simpson, T. W., “Computationally inexpensive metamodel assessment strategies,” *AIAA journal*, Vol. 40, No. 10, 2002, pp. 2053-2060.

3Simpson, T. W., Poplinski, J. D., Koch, P. N., and Allen, J. K., “Metamodels for computer-based engineering design: survey and recommendations,” *Engineering with computers*, Vol. 17, No. 2, 2001, pp. 129-150.

4Jin, R., Chen, W., and Simpson, T., “Comparative Studies of Metamodeling Techniques under Multiple Modeling Criteria,” *Structural and Multidisciplinary Optimization*, Vol. 23, No. 1, 2001, pp. 1-13.

5Forrester, A. I., and Keane, A. J., “Recent advances in surrogate-based optimization,” *Progress in Aerospace Sciences*, Vol. 45, No. 1, 2009, pp. 50-79.

6Lophaven, S. N., Nielsen, H. B., and Sondergaard, J., “DACE: A MATLAB Kriging Toolbox,” Technical Univ. of Denmark, TR IMM-TR-2002-12, Lyngby, Denmark, 2002

7Wang, G. G., and Shan, S., “Review of metamodeling techniques in support of engineering design optimization,” *Journal of Mechanical Design*, Vol. 129 No. 4, 2007, pp. 370-380.

8Jones, D. R., Schonlau, M., and Welch, W. J., “Efficient global optimization of expensive black-box functions,” *Journal of Global optimization*, Vol. 13, No. 4, 1998, pp.455-492.

9Bichon, B. J., Eldred, M. S., Swiler, L. P., Mahadevan, S., and McFarland, J. M., “Efficient global reliability analysis for nonlinear implicit performance functions,” *AIAA journal*, Vol. 46, No. 10, 2008, pp. 2459-2468.

10Viana, F. A., Haftka, R. T., and Watson, L. T., “Efficient global optimization algorithm assisted by multiple surrogate techniques,” *Journal of Global Optimization*, Vol. 56,No. 2, 2013, pp. 669-689.

11Atkinson, P. M., and Lloyd, C. D., “Non-stationary variogram models for geostatistical sampling optimisation: An empirical investigation using elevation data,” *Computers & Geosciences*, Vol. 33, No. 10, pp. 1285-1300.

12Sampson, P., and Guttorp, P., “Nonparametric Estimation of Nonstationary Spatial Covariance Structure,” *Journal of the American Statistical Association*, Vol. 87, No. 417, 1992, pp. 108–119.

13Haas, T., “Lognormal and moving Window Methods of Estimating Acid Deposition,” *Journal of the American Statistical Association*, Vol. 85, No. 412, pp. 950–963.

14Harris, P., Charlton, M., and Fotheringham, A. S., “Moving window kriging with geographically weighted variograms,” *Stochastic Environmental Research and Risk Assessment*, Vol. *24, No.* 8, 2010, pp. 1193-1209.

15Gramacy, R. B., and Lee, H. K., “Cases for the nugget in modeling computer experiments,” *Statistics and Computing*, Vol. 22, No. 3, 2012, pp. 713-722.

16Darbeheshti, N., and Featherstone, W. E., “Non-stationary covariance function modelling in 2D least-squares collocation,” *Journal of Geodesy*, Vo. 83, No. 6, 2009, pp. 495-508