Computer Communication & Collaboration (Vol. 6, Issue 3, 2018)
ISSN 2292-1028(Print) 2292-1036(Online)
DOIC:2292-1036-2018-08-002-94

Interpolation-Based Metamodels

E. Jack Chen

BASF Corporation

100 Park Avenue, Florham Park, New Jersey 07932, USA

E-mail: e.jack.chen@basf.com

Abstract: A metamodel is a simplified mathematical description of a simulation model that represents the system's input-output relationship with a function. In many situations, we may not need a single formula to describe the systems being simulated. This paper discusses interpolation-based metamodels, which are useful for providing simple estimates at non-design points to communicate the input-output relationship. The algorithm dynamically increases the sample size and the number of design points so that the estimates obtained via the metamodel satisfy the pre-specified precision. An experimental performance evaluation demonstrates the validity of interpolation-based metamodels.

Keywords: Kriging Metamodels, Experimental Design, Metamodels, Interpolation Methods

1. Introduction

Many simulation studies have been used to investigate system characteristics, such as the mean and the variance of certain system performances of the system under study. In many cases, users are interested in the system responses under different design points (input combinations, scenarios). Thus, a series of design points need to be evaluated, i.e., we need to construct a metamodel (or a response surface, a surrogate model), see Law (2014) for more information on metamodels. For example, we are interested in the mean waiting time (or system time) of queuing systems with different traffic intensities. The metamodel can be "used as a proxy for the full-blown simulation itself in order to get at least a rough idea of what would happen for a large number of design points," Law (2014). Metamodels have diverse applications and have many types, e.g., a mathematical function or an algorithm represents input and output relations. As such, metamodeling typically involves studying the output and input relationships and then fitting right metamodels to represent that behavior.

The purpose of constructing metamodels is to estimate or approximate the response surface (see Section 3.1). We could then use the metamodel to learn how the response curve would behave over various input-parameter combinations, e.g., output sensitivity. Sensitivity analysis is the study of how the uncertainty in the output of a system or mathematical model can be attributed to different sources of uncertainty in its inputs and may be one of the goals of metamodeling. This approach is "helpful when the simulation is very large and costly, precluding exploration of all but a few input-parameter combinations" (Law 2014). Note that the metamodel is designed to provide the overall tendency of performance measures rather than accurate estimates at all input-parameter combinations. These metamodels can also be used for visualization. Graphical representations of metamodels are useful for providing a simple and easy form to communicate the input-output relationship. In general, the goal is to evaluate the effect of different values of input variables on a system. However, the interest is sometimes in finding the optimal value for input variables in terms of the system outcomes. In these cases, the goal is to find optimal values for the input variables. As such metamodeling can be used to find the best input variables that produce desired outcomes in terms of response variables, see, e.g., Rahimi Mazrae Shahi, et al. (2015). Zakerifar, et al. (2011) use kriging metamodeling in multiple-objective simulation optimization. Couckuyt, et al. (2010) apply metamodel optimization to electromagnetic problems.

Kriging (also known as Gaussian process) metamodels (Krige 1951) from geostatistics is a metamodeling methodology that estimates the value at non-design points by interpolation for which the interpolated values are modeled by a Gaussian process governed by prior covariances. It was originally developed to estimate values at non-design point of deterministic models, nevertheless, it has been extended for non-deterministic (stochastic) models to take into account the randomness of the response at design points. Kriging is a classical frequentist approach estimating the response by obtaining the best linear unbiased estimator of the response. Traditional approximation methods only predict a single function value; however, Gaussian process methods predict a complete normal distribution for each point. The predicted distribution imparts the probability that a function value occurs. Kriging requires choice of a correlation function (i.e., spatial dependence) to compute the weights, which depend on the distances between the input combination to be estimated and the existing input combinations (design points) already observed. The Inverse Distance Weighting (IDW) interpolator assumes that each input point has a local influence that diminishes with distance. It weights the points closer to the processing area greater than those further away. A specified number of points, or all points within a specified radius can be used to determine the output value of each location. Use of this method assumes the variable being estimated decreases in influence with distance from its sampled design points.

Kriging assumes that the distance or direction between sample points reflects a spatial correlation that can be used to explain variation in the surface. Kriging is a multistep process; it includes exploratory statistical analysis of the data, variogram modeling, creating the surface, and (optionally) exploring a variance surface. The variogram modeling is an important step of the kriging process because it determines the kriging weights. Kriging is most appropriate when there is a spatially correlated distance or directional bias in the data. It is often used in soil science and geology. Kriging assumes that the closer the input combinations are, the stronger positively correlated the responses are. The choice of correlation function should be motivated by the underlying system being modelled. However, the (true) correlation function is unknown and both its type and parameter values must be estimated. Note that it is computationally expensive to construct and evaluate the weights and some of the weights may be negative. Kriging is a complex process and requires extensive knowledge of spatial statistics. Oliver and Webster (2014) point out that the ease of using Kriging without understanding the theoretical background may lead to producing unreliable and even misleading results. A crucial requirement for sound Kriging is a plausible function for the spatial covariances, which must be estimated reliably and modelled with valid mathematical functions. Before using kriging, the researchers should have a thorough understanding of its fundamentals and assess the appropriateness of the data for kriging. See van Beers and Kleijnen (2008) for an overview of Kriging.

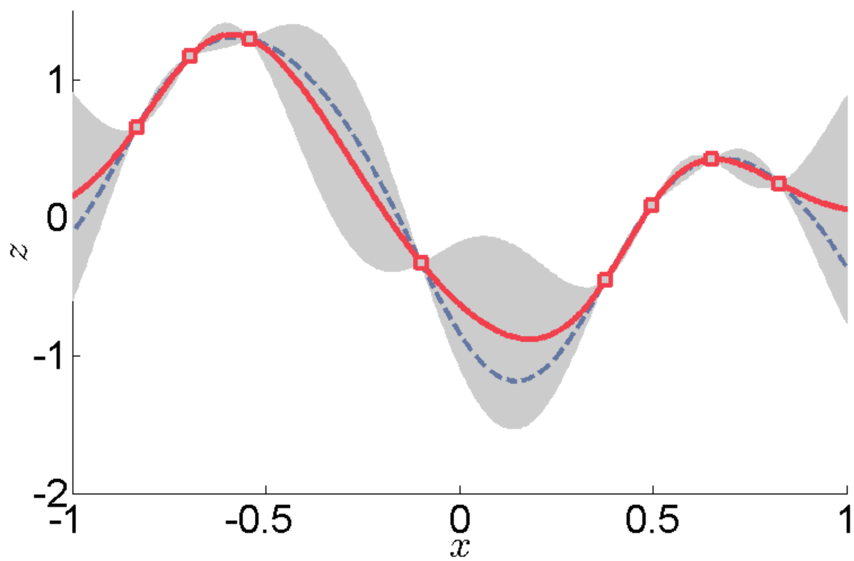

Figure 1: Example of one-dimensional data interpolation by Kriging

Figure 1 is an example of one-dimensional data interpolation by Kriging, with confidence intervals (Antro5 2018). Squares indicate the location of the design points. The Kriging interpolation, shown as a solid curve, runs along the means of the normally distributed confidence intervals shown in gray. Note that this is a deterministic example and the responses at design points are true values. Furthermore, the confidence interval half widths generally increase as the locations depart further away from design points. The dashed curve shows a spline that is smooth but departs significantly from the expected intermediate values given by those means (which are point estimators by Kriging and not necessarily the true values). Note that the chosen design points are not evenly spaced and the estimated responses at non-design points by Kriging are closed to estimated responses by linear interpolation in many segments, where the confidence interval half widths are generally smaller.

Chen (2009) develops a procedure to construct a metamodel with non-functional form (i.e., interpolation-based metamodels) of quantiles. The technique can be used to construct metamodels of various system characteristics. Chen and Li (2014) extend and refine the technique to investigate the performance of interpolation-based metamodels. This general-purpose procedure is sequential and selects design points as more information becomes available. The simulation run length at each design point is determined sequentially and independently. The stopping rules are based on the required precision measure.

Metamodeling requires determination of the number and placements of design points and estimation of their responses to which the metamodel is fitted. Design of experiments should be used during the metamodel construction to improve the construction process and the metamodel's quality. In this paper we investigate the issue of finding both the number and placements of design points to achieve the required precision. To account for the stochastic nature of the output, two additional metamodels are constructed: lower and upper bounds. For non-design points, the interpolated value is a linear interpolation of two bounding design points, not a weighted average of all design points. Hence, the response at non-design points can be estimated computationally inexpensively. This approach is a special case of the correlation function LIN in the Matlab Kriging toolbox developed by Lophaven, et al. (2002).

We believe that without the knowledge of the underlying systems, a simple estimate obtained via linear interpolation is as good as that obtained via weighted average of all observations. Other interpolation techniques such as Lagrange interpolation, spline interpolation can also be used.

The paper is organized as follows. Section 2 reviews some theoretical basis of metamodels. Section 3 illustrates our strategy of constructing interpolation-based metamodels. Section 4 lists the empirical results from experiments using interpolation-based metamodels to estimate output performance measures. Concluding remarks follow in Section 5.

2. Metamodel

A simulation model can be thought of as a function that turns input parameters into output performance measures (Law 2014). For example, if we use simulation to estimate μ_ρ (the mean system time) of the M/M/1 queuing process with certain traffic intensity $0 < \rho < 1$, we could in principle write $\mu_\rho = G(\rho)$ for some function G that is stochastic and unknown. Simulation can be used to evaluate G for numerical input values of ρ. We are interested in μ_ρ for any point in the one-dimensional region defined by the traffic intensity ρ. Since ρ is continuous, there are an infinite number of values and it is impossible to evaluate all possible ρ. On the other hand, when the parameter is not continuous, it may be too costly to evaluate all of them. Instead, we evaluate the function at certain input combinations and fit the results to some curves. The response of other input combinations is then approximated via the fitted curves (surfaces).

When a metamodel is specified to be a standard regression model, the independent variables for regression are the simulation input parameters and the dependent variable is the response of interest (Law 2014). For example, we would try to approximate the response-surface function $G(\rho)$ with a simple explicit formula involving ρ. It is nice to have a single formula to represent the responses of a system, but the formula is generally unknown and is likely complex. Fitting the grid points (design points) to a formula also introduces error (noise) into the estimates. Response at design points obtain via the resulting metamodel may be different than those are observed. Furthermore, we are investigating stochastic systems and the responses at design points are estimates and are not the true values.

In the interpolation-based-metamodel approach, instead of obtaining a formula or a regression model, it treats the collection of design points themselves, responses at design points, and the set of fitted curves as metamodels. Consequently, the metamodel is not described by a single (simple) formula. The responses at non-design points are estimated via interpolation. Note that the standard normal distribution function does not have a closed form and tables of cumulative distribution values of selected points are listed. Cumulative distribution value of points not listed in the tables are then obtained through linear interpolation. Furthermore, for stochastic systems, the randomness of the responses at design points is described by confidence intervals. Consequently, three response surfaces will be constructed: the lower bounds, the expected values, and the upper bounds.

It is likely that the form and complexity of the metamodel vary substantially for different systems. Thus, the metamodeling procedure needs to include a scheme for model selection given the simulation data, obtain a metamodel that is of the least complexity but adequate to characterize the underlying response surface. Furthermore, once a general tendency of the system is reflected in the metamodel, the researchers can add any point of interest as new design points to obtain more precise estimates.

2.1 The Natural Estimators

We assume that the process under study is stationary, i.e., the covariance depends only on the distance between observations but not on time. Let $F(\cdot)$ and $F_N(\cdot)$ denote respectively the true and the (sampled) empirical steady-state cumulative distribution function of the simulation-output process under study, where N is the simulation run length (we assume a discrete-time output process X_1, X_2, ...). For purpose of analysis, it is convenient to express $F_N(\cdot)$ as $F_N(x) = (1/N)\sum_{i=1}^{N} I_{(-\infty,x]}(X_i)$, where $I_{(-\infty,x]}(X_i) = 1$ if $X_i \le x$ and $I_{(-\infty,x]}(X_i) = 0$ if $X_i > x$.

In addition to the autocorrelations of the stochastic process output sequence approaching zero as the lag between observations increases (e.g., ϕ-mixing, see next section), we require two things for our method to work. First, $F_N(\cdot)$ must converge to $F(\cdot)$ as $N \to \infty$. In random (i.e., independent) sampling, $X_1, X_2, ..., X_N$ is a random sample of size N. If $N \to \infty$, then $F_N(\cdot)$ will tend with probability one to $F(\cdot)$. Second, if a statistic T is estimating some characteristic property Ψ of the distribution, then this characteristic must satisfy certain smoothness properties (e.g., continuity and differentiability). Most characteristics such as moments and percentiles are smooth and so their estimators have distributions that can be estimated. Our procedure may fail when T is a statistic estimating a function that is not smooth.

The property Ψ could be, for example, the mean, the variance, or a quantile. The natural point estimator for Ψ, denoted by $\hat{\Psi}$, is typically the sample mean, the sample variance, a sample quantile, or a simple function of the relevant order statistics, chosen to imitate the performance measure Ψ. Furthermore, the natural estimators are appropriate for estimating any Ψ, even in the presence of autocorrelation. This follows since $F_N(\cdot)$ converges to $F(\cdot)$. See Chen (2016) for detail.

2.2 Determining the Required Sample Size at Design Points

Recall that we are investigating stochastic systems, hence, the responses (performance measure) at design points need to be estimated by many observations. Although asymptotic results are often applicable when the amount of data is "large enough," the point at which the asymptotic results become valid generally depends on unknown factors. Furthermore, the quality of metamodels is heavily dependent upon the quality of the estimated responses at the design points.

We use the procedure of Chen (2012) to estimate responses at any given design points. The procedure is valid regardless the output process is independent and identically distributed normal. However, the procedure requires that the output sequences satisfy the ϕ-mixing conditions (Billingsley 1999). Intuitively, a stochastic process is ϕ-mixing if its distant future event is essentially independent of its present and past events. We use the Quasi-Independent (QI) procedure (Chen 2012) to determine simulation run length at each design point. When data are correlated, the QI procedure will progressively increase the simulation run length until a sequence of n samples (taken from the original output sequence) appear to be independent, as determined by the runs-up/runs-down tests (Knuth 1998). If the n systematic samples (from lag l observations of a stochastic output process) appear to be independent, then the simulation run length N is set to nl. The QI procedure assumes that simulation run length nl is a sufficient condition for asymptotic results to become valid.

The simulation run length becomes longer as the autocorrelation becomes stronger and the simulation run lengths at different design points are likely to be different. The achieved default precision determined by this strategy may not meet the specified requirement. In the next section, we will discuss our strategy to increase the simulation run length to meet the accuracy/precision requirement at design points.

3. Interpolation-Based Metamodels

This section introduces our approach of constructing interpolation-based metamodels. The strategy is to determine the number and placements of design points as well as the simulation run lengths at design points.

3.1 Design of Experiments and Construction of Metamodels

A metamodel will have greater accuracy and precision when constructed with a larger set of design points. On the other hand, a larger set of design points requires more simulation efforts. Furthermore, the required number of design points for constructing metamodels to achieve the pre-specified accuracy and precision depends on the underlying systems. We can sequentially determine the required number of design points by observing the fitted curves. It is known that sequential procedures are more efficient and may require less samples than fixed-sample-size procedures. Unlike fixed-sample-size procedures, sequential procedures determine the required sample size dynamically and can adapt to different environments. Placements of design points also affect the quality of the motamodel.

Figure 2 displays the flowchart of the procedure. We select the initial design points in a space-filling way. Assuming the range of interest of the (controllable) parameter ρ is $[\rho_L, \rho_U]$, we initially simulate the system with five evenly spaced design points, e.g., $\rho_1 = \rho_L < \rho_2 < \cdots < \rho_5 = \rho_U$. Let G_i be the simulated (estimated) response of input ρ_i. Let $\hat{G}_i$ (for $i = L+1, L+3, \ldots, U-3, U-1$) denote the estimate obtained via linear interpolation between design points ρ_{i-1} and ρ_{i+1}. Additional design points are added by adaptive sampling (or infill criterion). An infill criterion is a function that measures how interesting a data point is in the design space. If the relative precision $\left|\hat{G}_i - G_i\right| \big/ \left|G_i\right| < 4\gamma$ is satisfied for all i, then the procedure returns the collection of all the histograms as the interpolation-based metamodel; otherwise design point $\rho_{j_1} = (\rho_{i-1} + \rho_i)/2$ or $\rho_{j_2} = (\rho_i + \rho_{i+1})/2$ will be added for each i that the precision requirement is not satisfied. This can be considered cross validation, which leaves out one observation and then predicts it based on the remaining observations. Design point ρ_{j_1} will be added when $\left|(G_i - G_{i-1})/(\rho_i - \rho_{i-1})\right| > \left|(G_{i+1} - G_i)/(\rho_{i+1} - \rho_i)\right|$. Otherwise, design point ρ_{j_2} will be added. This is a greedy algorithm favoring the new design point where there is the largest improvement in precision, see Jones, et al. (1998). Note that the value γ is upper bound of the relative precision of estimates obtained at design points, i.e., $\left|\hat{X}_\rho - \mu_\rho\right| \big/ \mu_\rho \le \gamma$, where $\hat{X}_\rho$ is the final point estimator of μ_ρ. Using the value 4γ in the relative precision $\left|\hat{G}_i - G_i\right| \big/ \left|G_i\right| < 4\gamma$ is somewhat arbitrary. The rationale is

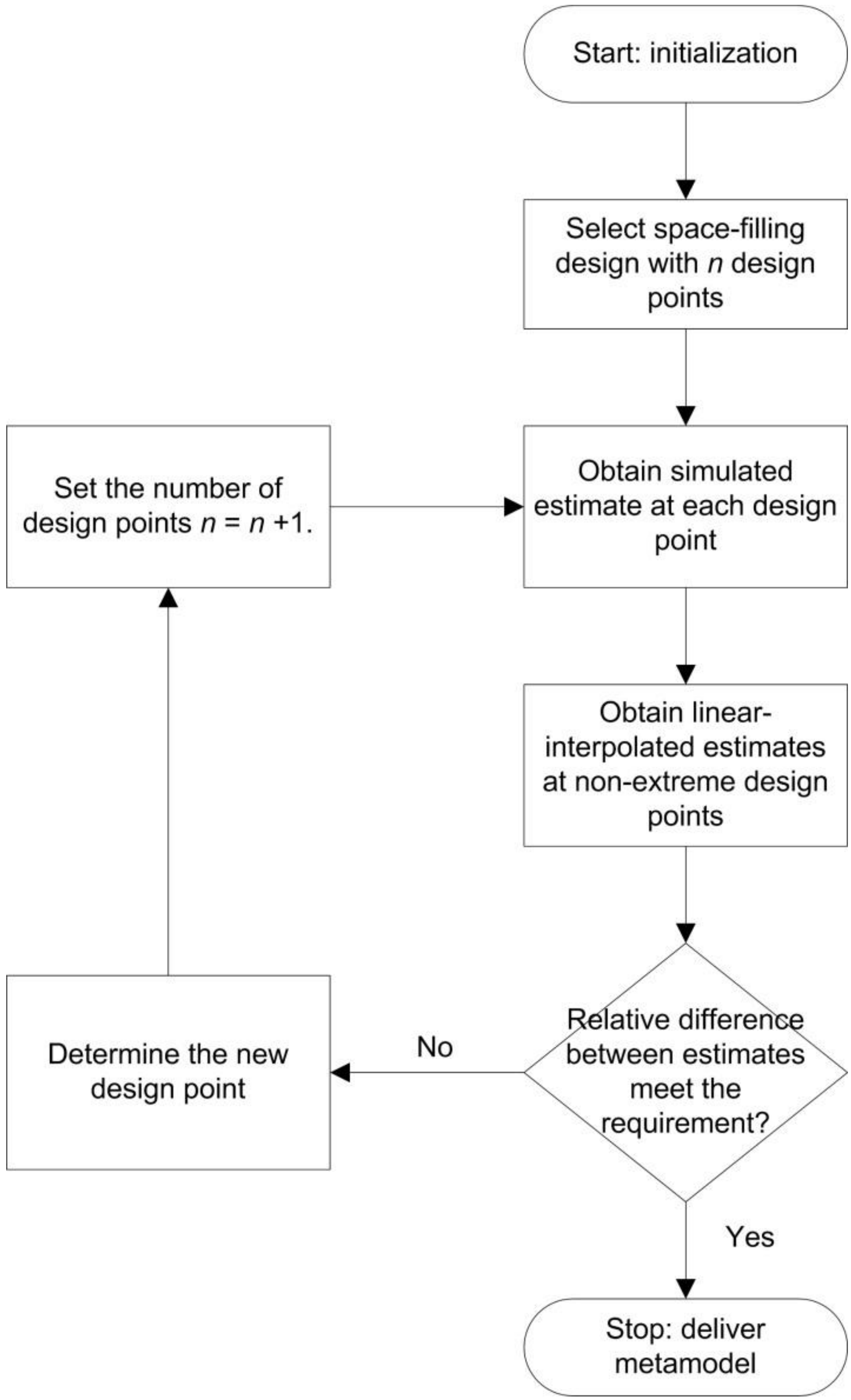

Figure 2: Flowchart of constructing a linear-interpolation metamodel

that when ρ_i's are evenly spaced, the distance between the design points ρ_{i-1} and ρ_i is half of the distance (between ρ_{i-1} and ρ_{i+1}) used to estimate $\hat{G}_i$ and the angle of the line between design points is also reduced by no less than half.

The step of adding more design points will be performed repeatedly until the specified precision on the selected estimators is achieved. Thus, the final number of design points is not known until the simulation is complete. Basically, we will allocate more input combinations around design points where the performance measure changes at different rate. Our preliminary experimental results indicate that the required number of design points to achieve the specified precision is not excessive (see Section 4).

3.2 Constructing Confidence Intervals

Because we are constructing metamodels of stochastic systems, with a given design point (e.g., traffic intensity ρ), we use the Batch-Means (BM) procedure of Chen (2012) to estimate the confidence interval of the mean of a stochastic process. Batch means are sample means of subsets of consecutive subsamples from a simulation output sequence.

In the non-overlapping batch-means method, the simulation output sequence $\{X_i : i = 1, 2, ..., N\}$ is divided into b adjacent non-overlapping batches, each of size m. For simplicity, we assume that N is a multiple of m so that $N = bm$. The method of BM is a well-known technique for estimating the variance of point estimators computed from simulation experiments. The BM method tries to reduce autocorrelation by batching observations. The BM variance estimator (for estimating the variance of sample means, i.e., Var[$\bar{X}_j$]) is simply the sample variance of the mean estimator $\bar{X}_j$ computed from the sample (batch) means of subsets of consecutive subsamples.

Let $\bar{X}_{\rho,j}$ denote the estimator of μ_ρ in the jth batch, i.e., $\bar{X}_{\rho,j} = (1/m)\sum_{i=(j-1)m}^{jm} X_i$, for $j = 1, 2, ..., b$. Note that the batch means are independent when the batch size m is large enough. We use $\hat{X}_\rho = (1/b)\sum_{j=1}^{b} \bar{X}_{\rho,j}$ (the average of batch means) as a point estimator of μ_ρ. Note that the average of batch means $\hat{X}_\rho$ equals the grand mean, i.e., the average of all samples $\sum_{i=1}^{N} X_i / N$. By the central limit theorem, a confidence interval (CI) for μ_ρ using the independently and identically distributed normal $\bar{X}_{\rho,j}$'s can be approximated using standard statistical procedures. That is, the ratio $T = (\hat{X}_\rho - \mu_\rho)/(S/\sqrt{b})$ has an approximate t distribution with $(b-1)$ d.f. (degrees of freedom), where $S^2 = (1/(b-1))\sum_{j=1}^{b}(\bar{X}_{\rho,j} - \hat{X}_\rho)^2$ is the usual unbiased estimator of the variance of $\bar{X}_{\rho,j}$'s. Let $h = t_{b-1,1-\alpha_2/2} S/\sqrt{b}$ be the half width of $100(1-\alpha_2)\%$ CI, where $t_{b-1,1-\alpha_2/2}$ is the $(1-\alpha_2/2)$ quantile of the t distribution with $(b-1)$ d.f. $(b \geq 2)$. Then the $100(1-\alpha_2)\%$ CI for μ_ρ is defined as $(\hat{X}_\rho \pm h)$. Let α and γ respectively, be the user specified absolute and relative precision. If the half width of CI does not meet the user specified absolute or relative precision (i.e., $w < a$ or $\gamma' |\bar{\hat{x}}_\rho|$), then we can progressively increase the number of b until the CI meets the specified requirement. We assume that the run length that achieves $h < \gamma' \left| \bar{\bar{x}}_\rho \right|$ also achieves $\left| \hat{X}_\rho - \mu_\rho \right| / \mu_\rho \le \gamma$ with $\gamma = \gamma'$. This assumption is somewhat arbitrary but our primary results support this. Furthermore, $\hat{X}_\rho - h \le \mu_\rho \le \hat{X}_\rho + h$. Hence, $\left| \hat{X}_\rho - \mu_\rho \right| \le h$ and $\bar{\hat{X}}_\rho \approx \mu_\rho$.

We construct the curve of means as well as the curves of the lower and upper confidence limits of means at the design points via the Quasi-Independent (QI) procedure (Chen 2012). The mean estimator and CI at non-design points can then be obtained via linear interpolation. Let $\hat{X}_{\rho_a}$ and $\hat{X}_{\rho_b}$ respectively denote the simulated estimate of the time in system of the M/M/1 queuing process with traffic intensity ρ_a and ρ_b. Then the mean estimate with traffic intensity $\rho_a < \rho < \rho_b$ is obtained by $\hat{X}_\rho = \hat{X}_{\rho_a} + \left[(\rho - \rho_a)/(\rho_b - \rho_a) \right] \left(\hat{X}_{\rho_b} - \hat{X}_{\rho_a} \right)$. The confidence limits of μ_ρ are computed in the same manner.

Let ρ_i for $i = 1, 2, ..., n$ be the final design points. Let w_i be the (Kriging metamodel) weights for estimating the response at ρ. Then the Kriging estimate is $\hat{X}_\rho = \sum_{i=1}^{n} w_i \hat{X}_{\rho_i}$. Consequently, the linear-interpolated estimate is a special case with $w_{\rho_b} = (\rho - \rho_a)/(\rho_b - \rho_a)$, $w_{\rho_a} = 1 - w_{\rho_b}$, and $w_i = 0$ for $i \ne a \ne b$.

3.3 Using Common Random Numbers

The goal of the simulation run length in the previous section aims to ensure that the relative deviation of the estimates is within γ. However, this accuracy is not guaranteed at non-design points. Recall that the goal of our metamodel construction strategy is mainly to provide the general trend of the responses.

The technique of common random numbers (CRNs) (see, e.g., Law (2014)) can be used to induce (positive) correlation of responses among design points. Consequently, the deviation of estimates to the corresponding unknown true values at design points is positively correlated. That is, the relative position of each point on the surface is the same when the entire response surface is lower or elevated. Using CRNs in simulation meant to compare the outputs of different input combinations while all other “circumstances” are the same in real world.

3.4 A Roadmap to Higher Dimensions

Higher dimensional models are of more interest. For two dimensions (i.e., two variables), bilinear interpolation (an extension of linear interpolation) can be used. The key idea is to perform linear interpolation first in one direction, and then again in the other direction.

GIS Resources (2018) list several interpolation methods. The following two methods can be used in interpolations at higher dimensions: the Natural Neighbour IDW (NNIDW) and the PointInterp function.

Like IDW, NNIDW is a weighted-average interpolation method. However, instead of finding an interpolated point’s value using all the input points weighted by their distance, Natural Neighbors interpolation creates a Delauney Triangulation of the input points and selects the closest nodes that form a convex hull around the interpolation point, then weights their values by proportionate area. This method is most appropriate where sample data points are distributed with uneven density. It is a good general-purpose interpolation technique and has the advantage that you do not have to specify parameters such as radius, number of neighbours or weights.

The PointInterp function allows more control over the sampling neighborhood. The influence of a sample on the interpolated grid cell value depends on whether the sample point is in the cell's neighborhood and how far from the cell being interpolated is located. Points outside the neighborhood have no influence. The weighted value of points inside the neighborhood is calculated using an inverse distance weighted interpolation or inverse exponential distance interpolation. This method interpolates a raster using point features but allows for different types of neighborhoods. Neighborhoods can have shapes such as circles, rectangles, irregular polygons, annuluses, or wedges.

4. Empirical Experiments

In this section, we use some experiment results from Chen and Li (2014) to demonstrate how exploratory metamodels can be built.

4.1 Choosing the Design Points

We want to construct a metamodel of the system time of the M/M/1 queuing system for traffic intensity $0.1 \le \rho \le 0.9$. We begin with five evenly spaced traffic intensities (i.e., pilot runs) and sequentially increase the number of design points until the number of design points reaches 10, the same number as in van Beers and Kleijnen (2008) for comparison.

Table 1: Chosen Design Points

Procedure	Traffic Intensity									
LI	0.1	0.3	0.4	0.5	0.6	0.7	0.75	0.8	0.85	0.9
V&K	0.1	0.3	0.5	0.7	0.8	0.85	0.875	0.8875	0.89375	0.9

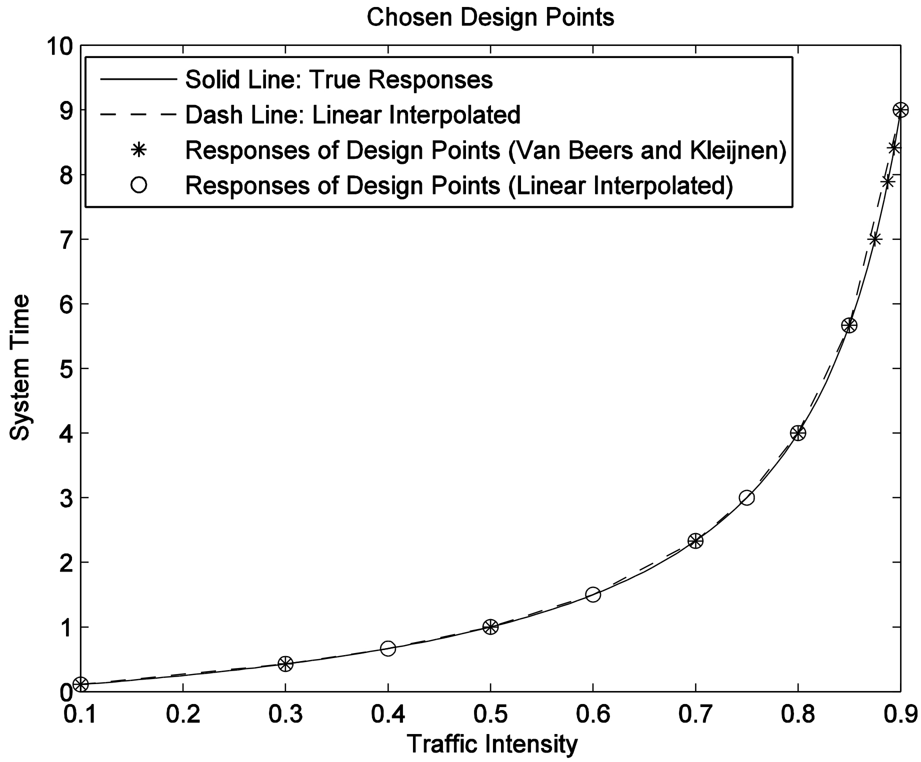

Figure 3: Plots of the Linear Interpolated Versus True MM1 System Time

Let $\gamma_i = \left|\hat{G}_i - G_i\right| \big/ \left|G_i\right|$. To adapt to this stopping rule, we add the design point $\rho_{j_1} = \left(\rho_{j-1} + \rho_j\right) \big/ 2$ or $\rho_{j_2} = \left(\rho_j + \rho_{j+1}\right) \big/ 2$ when $\gamma_j = \max \gamma_i$, where i is the index of design points. Table 1 lists the final 10 design points. The LI and V&K rows, respectively, are for the linear interpolated and van Beers and Kleijnen (2008). The LI approach places design points more evenly because the strategy basically aims to minimize the maximum of γ_i and selects relatively few design points in the area where the responses change linearly as the input changes. On the other hand, the approach of V&K aims to minimize the integrated mean square error and is computationally expensive. Consequently, they place all the additional design points between traffic intensities 0.7 and 0.9, where the responses have larger values and larger variances. Figure 3 displays the true responses and the linear-interpolated responses.

4.2 Estimating Means of Queuing Systems via Metamodels

In this experiment, we estimate the steady-state waiting time in M/M/1 queuing systems. We are interested in the response with traffic intensities between 0.7 and 0.95. The waiting times are generally shorter when the traffic intensities are lower. The proposed procedure selects the following design points (or traffic intensities): 0.7, 0.7625, 0.825, 0.8875, 0.9188, and 0.95. Traffic intensity 0.9188 ((0.8875 + 0.95)/2) is included in the design points because the relative precision of the estimates at traffic intensity 0.8875 and from linear interpolation at traffic intensities 0.825 and 0.925 is larger than the specified relative precision $4\gamma = 40\%$ (see Section 3.1).

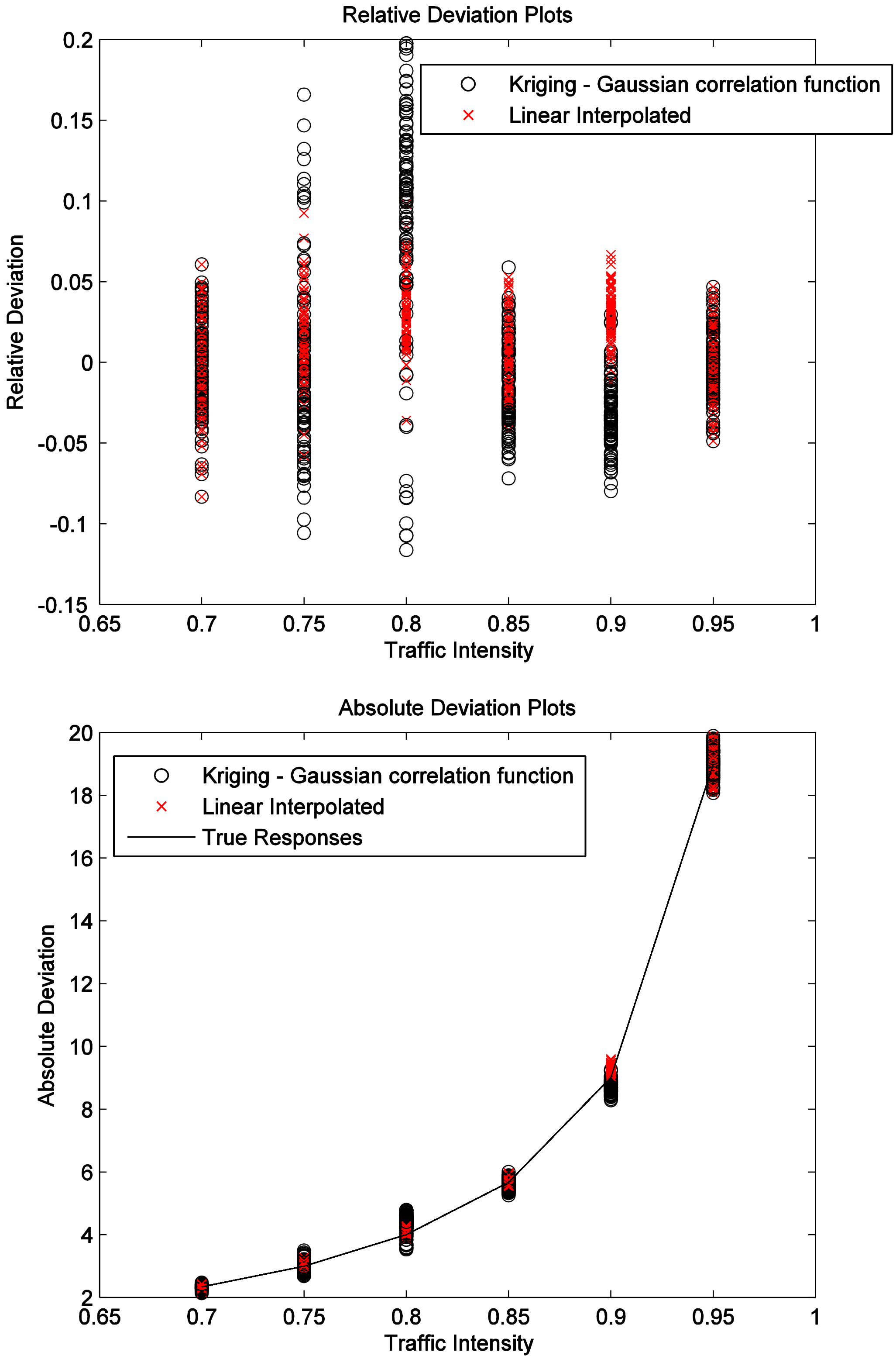

Figure 4: Plots of the Estimates for MM1 via Kriging with Gaussian correlation function and linear interpolation

Two types of plots are constructed to display graphically the 100 realizations of each estimator in a simulation study: (1) relative deviation plots and (2) absolute deviation plots, in which estimates are plotted around their true values. Figure 4 shows the results. The 100 response estimates (estimated waiting time) at each design point are obtained using the QI procedure of Chen (2012). The design points and their corresponding responses are used to build the metamodels. We then estimate (predict) the responses at traffic intensities 0.7, 0.75, 0.8, 0.85, 0.9, and 0.95 and compare the predictions to the expected responses of the M/M/1 process. In the absolute deviation plots, the solid curve represents a piecewise linear version of the true expected-response curve across various traffic intensities and the mean estimates are plotted as points. Note that a piecewise linear version of mean estimates can provide information regarding the general tendency of means and a rough idea of mean sensitivity as the traffic intensity changes.

For queuing processes, the mean estimates obtained through linear interpolation are slightly biased high, which can be explained by the curves in Figure 4. The mean vs. ρ curve is concave upward, i.e., $F(\rho)$ is a convex function. As the value of ρ deviates further away from the design points, e.g., $\rho = 0.8$ and $\rho = 0.9$, the accuracy of the estimates gets worse. Note that the (absolute or relative) precision can be improved by increasing the number of design points.

Table 2: Estimated 90% CI Half Width (HW) and Coverage

Model	Traffic Intensity	0.7	0.75	0.8	0.85	0.9	0.95
MM1	Average HW	0.095	0.118	0.151	0.203	0.294	0.578
	Coverage	0.9	0.9	0.64	0.09	0.69	0.86

Table 2 shows the estimated confidence intervals for M/M/1. The column labeled “Avg HW” is the average of the 90% CI half width calculated from the 100 realizations of the estimator. The “Coverage” column lists the proportion of these 100 CIs that cover the true mean. The CI coverage at the design points (i.e., $\rho = 0.7$ and 0.95) are around the nominal value of 0.90, which indicates the estimated CI half widths are also accurate. Note that the estimates at design points are obtained by the QI procedure (Chen 2012) and the achieved coverages are consistent with the reported performance of the QI procedure. However, the CI coverages of ρ that are further away from the design points (e.g., $\rho = 0.8$, 0.85, and 0.9) are less than the nominal value, especially at $\rho = 0.85$ where the coverages are zero or just above zero. This is expected because the point estimators are greater than the true values; while the half widths are about the right length. We do not think this is a major drawback of the procedure. The main purpose of the metamodel is to gain the overall tendency of expected responses of the system under study and is not to obtain accurate CIs throughout the entire traffic-intensity range of interest. Note that accurate CIs can be obtained by designating the traffic intensity of interest ρ as a design point. Furthermore, with the help of the plots, we can predict approximately the accuracy of the interpolated estimates.

If the CI coverage at non-design points is a concern, a conservative adjustment can be used to increase the coverage of the CI estimated via interpolation, see Chen and Li (2014).

4.3 Comparison of Interpolation-Based Metamodels

The Kriging metamodel in its original form is for deterministic models, it is a non-sequential procedure and uses fixed design points. We discuss extended Kriging metamodel for stochastic metamodels, it uses sequential procedure to determine placements and the number of design points. The stopping rule based on the specified accuracy works well for the LI metamodel. We believe that the LI metamodel is a viable choice when there is no prior knowledge of the characteristic of the underlying system.

van Beers and Kleijnen (2008) point out that the optimal weights are stochastic and ignoring the randomness of the estimated optimal weights tends to underestimate the true variance of the Kriging predictor.

5. Conclusions

In order to obtain valid asymptotic, some estimates require more observations than others. The quasi-independent algorithm works well in determining the required simulation run length for a valid asymptotic approximation. The empirical results of Chen (2012) indicate that the QI procedure works well in achieving the pre-specified precision. QI procedures can be executed without user intervention. Furthermore, the procedure is a data-based algorithm, i.e., the procedure can be embodied in a software package whose input is the data $\left(X_1,...,X_n\right)$ and whose output is the estimate. The procedure is easy to understand, simple to implement, and is a powerful tool for estimating system characteristics.

Interpolation-based metamodels providing overall tendency of responses and can be constructed with a set of carefully selected design points. The metamodel can be used as a proxy for the full-blown simulation itself to get at least a rough idea of what would happen for a large number of input-parameter combinations. The goal of metamodeling can also be to find the output sensitivity in terms of input combinations or the optimal input combinations. Furthermore, our approach is a general-purpose procedure and does not assume any structure of the underlying system being simulated.

When the underlying systems are unknown, it is not clear which correlation or covariance functions should be used to obtain the greatest accuracy or precision. Hence, instead of using correlation or covariance functions which require intensive computation, it is beneficial to use linear-interpolation in these situations, which requires less computation and may have better accuracy and precision. Other approaches are using Lagrange interpolation, NNIDW and the PointInterp function.

References

[1]. Antro5 at English Wikipedia, CC BY-SA 3.0, [Online] Available at: https://commons.wikimedia.org/w/index.php?curid=52740780 (August 31th, 2018).

[2]. Billingsley, P. (1999), *Convergence of probability measures*, 2nd ed. John Wiley and Sons, Inc., New York.

[3]. Chen, E. J. (2009), “Metamodels for estimating quantiles of systems with one controllable parameter”, *Simulation: Transactions of the Society for Modeling and Simulation*, 85:307-317.

[4]. Chen, E. J. (2012), “A stopping rule using the quasi-independent sequence”, *Journal of Simulation*, 6(2):71-80.

[5]. Chen, E. J. (2016), *Crafts of Simulation Programming*, World Scientific Publishing Co Pte Ltd.

[6]. Chen, E. J., Li, M. (2014), “Design of experiments for interpolation-based metamodels”, *Simulation Modelling Practice and Theory*, 44:14-25.

[7]. Couckuyt, I., Declercq, F., Dhaene, T., Rogier, H. (2010), “Surrogate-based infill optimization applied to electromagnetic problems”, *International Journal of RF and microwave computer-aided engineering*, 20(5):492-501.

[8]. GIS Resources. [Online] Available at: http://www.gisresources.com/types-interpolation-methods_3/ (August 31th, 2018).

[9]. Jones, D. R., Schonlau, M., Welch, W. J. (1998), “Efficient global optimization of expensive black-box functions”, *Journal of Global Optimization*, 13:455-492.

[10]. Knuth, D. E. (1998), *The Art of Computer Programming*, Vol. 2, 3rd ed. Addison-Wesley, Reading, Mass.

[11]. Krige, D. G. (1951), “A statistical approach to some basic mine valuation problems on the Witwatersand”, *Journal of the Chemical Metallurgical and Minting Society of South Africa*, 52:119-139.

[12]. Law, A. M. (2014), *Simulation Modeling and Analysis*, 5th ed. McGraw-Hill, New York.

[13]. Lophaven, S.N., Nielsen, H. B., Sondergaard, J. (2002), *A Matlab Kriging Toolbox,* Version 2.5, IMM Technical University of Denmark, Lyngby.

[14]. Oliver, M. A., Webster, R. (2014), “A tutorial guide to geostatistics: computing and modelling variograms and kriging, *Catena*, 113:56-69.

[15]. Rahimi Mazrae Shahi, M., Fallah Mehdipour E., Amiri, M. (2015), “Optimization using simulation and response surface methodology with an application on subway train scheduling”, *Intl. Trans. In Op. Res.*, 23:797-811.

[16]. van Beers W. C. M., Kleijnen, J. P. C. (2008), “Customized sequential designs for random simulation experiments: Kriging metamodeling and bootstrapping”, *European Journal of Operational Research*, 186:1099-1113.

[17]. Zakerifar, M., Biles, W. E., Evans, G. W. (2011), “Kriging metamodeling in multiple-objective simulation optimization”, *Simulation*, 87(10): 843-856.