Engineering with Computers (2012) 28:161–178
DOI 10.1007/s00366-011-0229-7

ORIGINAL ARTICLE

Extended Gaussian Kriging for computer experiments in engineering design

Wenze Shao · Haisong Deng · Yizhong Ma · Zhuihui Wei

Received: 26 January 2011 / Accepted: 10 June 2011 / Published online: 8 July 2011

Abstract Metamodeling or surrogate modeling is becoming increasingly popular for product design optimization in manufacture industries. In this paper, an extended Gaussian Kriging method is proposed to improve the prediction performance of widely used ordinary Kriging in engineering design. Unlike the forgoing approaches, the proposed method places a variance-varying Gaussian prior on the unknown regression coefficients in the mean model of Kriging and makes prediction at untried design points based on the principle of Bayesian maximum a posterior. The achieved regression mean model is adaptive, therefore capable of capturing more effectively the overall trend of computer responses and leading to a more accurate metamodel. Particularly, the regression coefficients in the mean model are estimated by a fast numerical algorithm, making extended Gaussian Kriging implemented roughly as efficient as ordinary Kriging. Experiment results on several examples are presented, showing remarkable improvement in prediction using extended Gaussian Kriging over ordinary Kriging and several other metamodeling methods.

Keywords Computer experiments · Metamodeling · Kriging · Gaussian stochastic process · Design of experiments

W. Shao (✉)
Department of Management Science and Engineering,
Nanjing University of Science and Technology,
200 Xiaolingwei, Nanjing 210094, China
e-mail: shaowenze1010@yahoo.com.cn

H. Deng
School of Mathematics and Statistics, Nanjing Audit University,
Nanjing 211815, China

Y. Ma
Department of Management Science and Engineering,
Nanjing University of Science and Technology,
Nanjing 210094, China

Z. Wei
School of Computer Science and Technology,
Nanjing University of Science and Technology,
Nanjing 210094, China

1 Introduction

Currently, computer modeling and experiments are becoming increasingly popular for product design optimization in manufacture industries, in order to produce better and cheaper products more quickly. However, the cost of running computer simulation experiments is nontrivial in spite of the continued growing computing power. For some computer experiments, 12 h or even considerably longer time are required to produce one single simulation response. For example, it is reported that Ford Motor Company takes about 36–160 h to perform one crash simulation [1]. Another known example in literature is the piston slap noise experiment, taking 24 h for every run of computer experiments [2]. In the past 20 years, to improve the computational efficiency and reduce the cost of experiments, meta-modeling techniques have been developed to produce a metamodel or surrogate model, i.e., "model of the model", as an approximation of the complex and expensive computer code.

In literature, existing metamodels include: polynomial regressions [3, 30], radial basis functions [4, 31, 33, 35], multivariate adaptive regressive splines [54], Gaussian stochastic process models [4, 6, 13, 14, 18], artificial neural networks [5, 55, 56], and support vector regression [57]. Among these methods, the polynomial regressions are probably the most mature and the easiest metamodeling

techniques and have been utilized extensively in combination with the Taguchi approach in engineering design [5]. In more recent years, Gaussian stochastic process models, also known as Gaussian Kriging [6, 7], have also been studied extensively in the statistical communities. The popularity of Kriging consists in that, computer experiments are often deterministic and Kriging is able to provide an interpolating approximation model. However, the Kriging method has found limited use in engineering design optimization compared against polynomial regressions. Lack of readily available software to fit Kriging models in early days is one inherent reason [5]. While, the key reason lies in that fitting a Kriging model is more complex than fitting a regression polynomial. The low level of robustness of the approach to sampled points is also an important reason. Nevertheless, when the computer simulators in engineering are deterministic, and in particular, highly nonlinear in a great number of factors (more than 50 or even 100), polynomial regression modeling becomes insufficient and Kriging modeling may be a better choice in spite of the added complexity [5, 8].

A basic Kriging model is universal Kriging, consisting of a Gaussian stochastic process model and a regression mean model. The basis functions or important variables in the regression mean are assumed to be known, which is not feasible in practical applications. As a result, ordinary Kriging becomes the most commonly used Kriging method [8–14]. Ordinary Kriging utilizes a constant mean to capture the overall trend of computer simulation responses, however, resulting in a poor prediction at new design points if there are some strong trends [14]. Till now, Kriging has been used to approximate and analyze different computer experiments in engineering design. As examples, [15] uses Kriging for the design of a High Speed Civil Transport airport, [16] uses Kriging for the helicopter rotor structural design, and [17] uses Kriging for the design of a variable thickness piezoelectric bimorph actuator.

To improve the prediction accuracy of Kriging, several advanced Kriging metamodeling approaches are proposed more recently. For example, Joseph et al. [18] proposed a metamodel called blind Kriging. Blind Kriging achieves higher prediction accuracy compared against ordinary Kriging and universal Kriging. However, a scheme of computationally intensive Bayesian forward selection is involved in the process of estimation. In addition, the correlation parameters in blind Kriging have to be estimated two times, making the computational complexity of blind Kriging at least twice as that of ordinary Kriging. Linkletter et al. [19] also proposed a meta-modeling approach through combining Bayesian variable selection with Kriging. Although the variable selection procedure therein is easy to implement, it is computationally demanding as involving Kriging, too. Besides, Linkletter et al. [19] focused primarily on screening of important design factors, rather than the objective of improvement in prediction. Joseph [20] deduced a metamodel called limit Kriging. Limit Kriging may be viewed as the limit of ordinary Kriging to a certain degree, however, achieving limited improvement in prediction and appearing to be somewhat numerically unstable as claimed. Besides, Li and Sudjianto [2] proposed a penalized likelihood approach, i.e., smoothly clipped absolute deviation (SCAD) penalized Kriging, to improve ordinary Kriging. Their research motivation is that, the correlation parameters in ordinary Kriging are often obtained by the maximum likelihood approach where the likelihood function near the optimum may be flat in some situations, often leading to estimates of the parameters having quite large variance. Notice that, however, the penalized likelihood approach is essentially rooted in ordinary Kriging, where the mean is just a constant.

In this paper, an extended Gaussian Kriging method is proposed to improve the prediction accuracy of Kriging metamodeling. Unlike the forgoing approaches, the new method imposes a variance-varying Gaussian prior on the unknown regression coefficients in the mean function of universal Kriging and makes prediction at untried design points based on the principle of Bayesian maximum a posterior. The achieved regression mean model is adaptive, thereby able to capture more effectively the overall trend of computer responses and lead to a more accurate metamodel. The underlying parameters in Kriging are estimated iteratively, and particularly, the regression coefficients are estimated by a fast numerical algorithm; the benefit of exploiting the fast algorithm is that the computational complexity in fitting extended Gaussian Kriging is roughly the same as that of the most commonly used ordinary Kriging. Experimental results on several examples are presented, showing remarkable improvement in prediction using extended Gaussian Kriging over ordinary Kriging and several other metamodeling methods.

One more point is to claim that our proposed method is particularly suitable to those much complex computer experiments. In such situations, collecting a large sample may be very much difficult, and a more accurate approximation model from a small sample size becomes crucial. The extended Gaussian Kriging is just one of such potential choices. However, as the sample size of computer runs increases, the efficiency of our proposed method as well as ordinary Kriging is to decrease due to the inversion of the large-scale correlation matrix involved in fitting Kriging.

The remainder of the paper is organized as follows: Sect. 2 gives a brief review of metamodeling methodologies for computer experiments. In Sect. 3, a Bayesian Kriging metamodeling approach is proposed using Bayesian modeling and posterior inference. Section 4 describes a

fast algorithm solving the regression coefficients and other related computational aspects. Section 5 gives experimental results on several empirical applications, validating the performance of our proposed method. Conclusions are finally given in Sect. 6.

2 Metamodeling methods for computer experiments in engineering design

This study concentrates on the deterministic computer experiments in many engineering design fields; that is, the same input settings yield the same computer responses. Suppose a vector of design variable is denoted as $x = (x_1, x_2, \ldots, x_p)' \in D \subseteq R^p$, and the true nature of a computer analysis code is denoted as $y = f(x)$. Then the general metamodeling problem can be stated in the following: Given a set of computer runs $(x^{(1)}, x^{(2)}, \ldots, x^{(n)})$ produced by the corresponding expensive computer code, the aim is to seek an accurate and computationally inexpensive surrogate model $y = g(x)$ over a given design domain D, satisfying

$$y_k = g(x^{(k)}) = f(x^{(k)}), \quad k = 1, 2, \ldots, n. \tag{1}$$

That is, the model $y = g(x)$ is an interpolating global metamodel of the true deterministic computer code. In general, metamodeling is determined by two key issues, namely, the experimental designs and the metamodeling procedures.

An experimental design represents a sequence of experiments to be performed, expressed in terms of factors (design variables) set at specified levels (predefined values) and represented by a matrix where the rows denote experiment runs and the columns denote factor settings [5]. The experimental design techniques were originally developed for physical experiments and now are being applied to the design and analysis of computer experiments (DACE). Classical experimental designs for analysis of physical experiments include [1, 5] full factorial design, fractional factorial design, Plackett-Burman designs, central composite designs (CCD), Box-Behnken designs, and orthogonal arrays. As [1] wrote, "these classic methods tend to spread the sample points around boundaries of the design space and leave a few at the center of the design space". While, as deterministic computer experiments involve systematic error rather than random noise as in physical experiments, researchers advocate that experimental designs for computer experiments should be space filling [21]. Unlike those classical methods, the space-filling methods tend to fill the design space equally rather than to concentrate on the boundary. Koehler and Owen described several Bayesian and Frequentist space-filling designs [22], including maximum entropy designs, mean squared-error designs, minimax and maximin designs, Latin hypercube designs, randomized orthogonal arrays, and scrambled nets. Four types of space-filling methods are relatively more often used in the literature. They are, respectively, Latin hypercube designs [23, 24], uniform designs [25], Hammersley sequences [26], and orthogonal arrays [27, 28]. A comparison of these methods can be referred to ref. [29]. Besides, sequential and adaptive sampling methodologies have recently gained popularity because of the difficulty of knowing the "appropriate" sampling size a priori. Jin et al. [21] compared several different sampling methods, finding that sequential sampling allows engineers to control the sampling process and it is generally more efficient than one-stage sampling.

Though many metamodeling approaches have been proposed for computer experiments in engineering design, we review here the three that are most prevalent in the literature: polynomial regressions, radial basis functions, and Kriging.

2.1 Polynomial regressions

In this article, PR refers to the low-order polynomial regressions: first-order (linear) and second-order (quadratic) polynomials, which have been used successfully for prediction in many engineering design applications [30]. By now, polynomial regressions are probably the fastest and simplest metamodeling technique. For low curvature, a linear polynomial (LP) can be used as in (2); for significant curvature, a quadratic polynomial (QP) including all two-factor interactions can be used as in (3):

$$\text{PR} - \text{LP}: \quad y = g(x) = \beta_0 + \sum_{k=1}^{p} \beta_k x_k \tag{2}$$

$$\text{PR} - \text{QP}: \quad y = g(x) = \beta_0 + \sum_{k=1}^{p} \beta_k x_k + \sum_{k=1}^{p} \beta_{kk} x_k^2 + \sum_{k=1}^{p} \sum_{j=k+1}^{p} \beta_{kj} x_k x_j \tag{3}$$

The parameters in (2) and (3) are unknown coefficients usually determined by the least squares approach. However, a PR merely produces a regression fit rather than an interpolative fit satisfying (1); it is not suitable for creating global metamodels for those highly non-linear, multimodal, and multi-dimensional response surfaces which are frequently encountered in engineering design, either [1]. To improve the prediction accuracy of PR, various local or successive methods have been developed to overcome the deficiency of the global PR in the past several decades, at the expense of additional sampling points and/or representing only a small portion of design space [31]. These methods, however, are not appropriate to the

computationally expensive computer simulators, such as the case of vehicle crash simulations [32]. For these reasons, global metamodels with high accuracy and limited sampling points are highly desirable. A more complete discussion of response surfaces and least squares fitting can be referred to ref. [30]. We use the PR in our comparative study in Sect. 5 purely as a benchmark approach, considering its popularity and widespread use in the industry.

2.2 Radial basis functions

The RBF are methodologies for exact interpolation of data in multi-dimensional space [33], originally developed for fitting irregular topographic contours of geographical data [34]. The RBF is a family of models constructed by the same methodology but distinguished by the either linear or nonlinear basis functions. Specifically, a general form of RBF approximation to the true computer code $y = f(x)$ can be formulated as

$$y = g(x) = \sum_{k=1}^{n} \beta_k \gamma\left(\left\|x - x^{(k)}\right\|\right) \tag{4}$$

where the term $||x - x^{(k)}||$ is the Euclidean norm, γ is a basis function, and $\boldsymbol{\beta}_k$ is the coefficient for the kth basis function. From (4), it is observed that a RBF model is an interpolative approximation, thus satisfying (1). The overall accuracy of RBF essentially depends on the selected basis function for a given set of data samples. The most widely used basis functions include linear, cubic, thin-plate spline (TPS), Gaussian, multi-quadric (MQ), and inverse MQ functions [31]. In recent years, multiple comparative studies have demonstrated that the RBF are particularly appropriate to highly nonlinear problems, in both case of large or limited number of samples, and more accurate than polynomial regressions in general [31, 35].

The RBF are good for highly nonlinear responses; however, they are shown to be inappropriate for linear responses [36]. Thus, an augmented RBF model is proposed through adding a linear polynomial, given as [31, 36]

$$y = g(x) = \sum_{k=1}^{n} \beta_k \gamma\left(\left\|x - x^{(k)}\right\|\right) + \sum_{j=1}^{p} \lambda_j b_j(x) \tag{5}$$

where $bj(x)$ is a linear polynomial function, and λ_j is the corresponding unknown coefficient. Since (5) is underdetermined, the orthogonality condition is further imposed on $\lambda_j (j = 1, 2, \ldots, p)$ so that

$$\sum_{k=1}^{n} \beta_k b_j\left(x^{(k)}\right) = 0, \text{ for } j = 1, 2, \ldots, p \tag{6}$$

Combining (5) and (6), coefficients $\beta = [\beta_1, \beta_2, \ldots, \beta_n]'$ and $\lambda = [\lambda_1, \lambda_2, \ldots, \lambda_p]'$ can be estimated by the following system of equations:

$$\begin{pmatrix} \Upsilon & B \\ B' & 0 \end{pmatrix} \begin{pmatrix} \beta \\ \lambda \end{pmatrix} = \begin{pmatrix} Y \\ 0 \end{pmatrix} \tag{7}$$

where $Y = [y_1, y_2, \ldots, y_n]'$, $\Upsilon_{j,k} = \gamma\left(||x^{(j)} - x^{(k)}||\right)$, and $\mathrm{B}_{j,k} = b_j\left(x^{(k)}\right)$. In addition, Mullur and Messac [37] recently developed an extended RBF model by adding extra terms to a regular RBF model. It has been shown that the extended RBF is more flexible, in that it provides better smoothing properties and generally it yields more accurate metamodels than the typical RBF [37, 38].

2.3 Kriging

Recently, Kriging is particularly recommended if the computer code to be studied is deterministic and highly nonlinear [5], in that Kriging is capable of dealing with a great number of design factors (more than 50 or even 100) and providing an interpolating approximation model. More importantly, Kriging is by now the least assuming method in this paper in terms of the range of function forms it can approximate. The general form of Kriging metamodel is to represent the true computer analysis code $y = f(x)$ as a combination of a polynomial model plus departures of the form

$$y = f(x) = \mu(x) + Z(x) \tag{8}$$

In (8), $\mu(x)$ is a regression function modeling the drift of the process mean over the design space. Particularly, when the mean term $\mu(x)$ is a linear combination of some known function of x as in Eq. (9), the resulting Kriging model is called universal Kriging.

$$\mu(x) = \sum_{k=0}^{J} \beta_k \psi_k(x) \tag{9}$$

In (9), the terms $\{\psi_k(x)\}_{k=0}^{J}$ are some known functions, e.g., low-order polynomial functions, and $\{\beta_k\}_{k=0}^{J}$ are unknown parameters to be estimated.

The second term in (8), i.e., $Z(\boldsymbol{x})$, is a stationary Gaussian stochastic process modeling residuals from the linear regression, with mean 0, and covariance

$$\mathrm{cov}(Z(x^{(l)}), Z(x^{(k)})) = \sigma^2 \varphi(x^{(l)} - x^{(k)}) \tag{10}$$

In most cases, φ in (10) is a prespecified correlation function of the "distance" between $\boldsymbol{x}^{(l)}$ and $\boldsymbol{x}^{(k)}$, controlling the smoothness of the resulting Kriging model, the influence of other nearby points, and the differentiability of the surface. Koehler and Owen [22] provided an insightful review of four functions commonly utilized in Kriging when approximating a deterministic computer model and the impact of parameter selection on the properties of these functions. In this paper, the correlation function is chosen as the Gaussian function

$$\psi(x^{(l)} - x^{(k)}) = \prod_{j=1}^{p} \exp\left(-\theta_j |x_j^{(l)} - x_j^{(k)}|^2\right) \tag{11}$$

where $\{\theta_j\}_{j=1}^{p}$ are the correlation parameters.

The parameters $\{\beta_k\}_{k=0}^{J}$, $\{\theta_j\}_{j=1}^{p}$ and σ^2 in (8) are frequently estimated by the maximum likelihood approach [4, 6, 13, 14, 18]. The corresponding predictor of universal Kriging can be formulated as

$$y = g(x) = \psi'(x)\beta + \varphi'(x)\Phi^{-1}(\theta)(y - \Psi\beta) \tag{12}$$

where $\boldsymbol{\psi}(\boldsymbol{x}) = [\psi_0(\boldsymbol{x}), \psi_1(\boldsymbol{x}),\ldots,\psi_J(\boldsymbol{x})]'$, $\boldsymbol{\varphi}(\boldsymbol{x}) = [\varphi(x - x^{(1)}), \varphi(x - x^{(2)}),\ldots,\varphi(x - x^{(n)})]'$,$\boldsymbol{\Phi}(\boldsymbol{\theta})$ is the correlation matrix with elements $\Phi_{l,k} = \varphi(x^{(l)} - x^{(k)})$, and

$$\beta = \left(\Psi'\Phi^{-1}\Psi\right)^{-1}\Psi'\Phi^{-1}Y \tag{13}$$

Furthermore, given $\{\theta_j\}_{j=1}^{p}$ and σ^2, the variance of the predictor (12) can be obtained as

$$s(x) = \sigma^2\left(1 + u'\left(\Psi'\Phi^{-1}\Psi\right)^{-1}u - \varphi'\Phi^{-1}\varphi\right) \tag{14}$$

where $u = \boldsymbol{\Psi}'\boldsymbol{\Phi}^{-1}\boldsymbol{\varphi} - \boldsymbol{\psi}$.

It is observed that, universal Kriging uses known functions to capture the trend of various computer simulators. However, we seldom have a priori knowledge of the trends in the data and specifying them using known functions may introduce inaccuracies. Therefore, ordinary Kriging as shown in (15) becomes the most popular Kriging model, which uses only a constant to capture the overall trend

$$y = f(x) = \mu_0(x) + Z(x) \tag{15}$$

However, ordinary Kriging will lead to a poor prediction at new design points if there are some strong trends [14]. Thereby, neither universal Kriging nor ordinary Kriging is satisfactory. And it is for this reason that several improved Kriging metamodeling approaches are recently reported in the literature [2, 18–20], as commented in the Sect. 1

3 Extended Gaussian Kriging: Bayesian modeling and inference for Kriging

From the above, we know that neither universal Kriging nor ordinary Kriging is satisfactory to capture the overall trend of computer responses. In fact, the basis functions in universal Kriging can be viewed as the variables greatly affecting the computer responses. Thus, the drawbacks of universal Kriging and ordinary Kriging may be addressed as follows: in practical applications, universal Kriging can seldom avoid including the unimportant variables in the mean function of universal Kriging, while ordinary Kriging has completely ignored the important variables. To improve the prediction accuracy of Kriging metamodeling, an extended Gaussian Kriging method is proposed to balance between ordinary Kriging and universal Kriging. In this section, low-order basis functions are still assumed in the mean model of the extended Gaussian Kriging; however, they are considered to be no more uniformly important as in universal Kriging. Specifically, unlike previous improved Kriging methods, the new method imposes a straightforward variance-varying Gaussian prior density on the unknown regression coefficients in the mean model, making us more adaptively capture the overall trend of computer responses, and therefore leading to a more accurate Kriging metamodel.

3.1 Bayesian modeling for the adaptive mean function

In this subsection, the specification of the prior $p(\boldsymbol{\beta})$ is considered, relying on our prior knowledge on the regression coefficients. Ideally, $p(\boldsymbol{\beta})$ should reflect our expectation of adaptively choosing the important variables from $\{\Psi_k(\boldsymbol{x})\}_{k=0}^{J}$ for a more accurate mean model. In other words, the estimated regression coefficients are expected to be sparse, i.e., $p(\boldsymbol{\beta})$ is a sparseness-promoting Bayesian prior. Commonly exploited sparse priors in the literature include the student's t distribution and the double exponential distribution [39–41]. However, they are not efficient for the final Bayesian inference in this paper. On the one hand, intensive computation is desired to estimate parameters involved in the above distributions. As we know, three parameters are to be estimated in the student's t distribution, i.e., the degrees of freedom, the location, and the scale parameters; and a scale parameter is to be estimated in the double exponential distribution which has no analytical solution. On the other hand, neither the student's t distribution nor the double exponential distribution allows for a tractable Bayesian analysis, since neither one is conjugate to the Gaussian likelihood to be specified in the next subsection.

In this paper, we choose $p(\boldsymbol{\beta})$ to be the relevance vector machine or sparse Bayesian learning [42], denoted by

$$p(\beta|\gamma) = \prod_{k=0}^{J} p(\beta_k|\gamma_k) = \prod_{k=0}^{J} \mathrm{N}(\beta_k|0, \gamma_k^{-1}) \tag{16}$$

In (16), $\gamma = (\gamma_0, \gamma_1, \ldots, \gamma_J)'$, $\mathrm{N}(\beta_k|0, \gamma_k^{-1})$ is a zero-mean Gaussian prior placed on β_k with varying inverse variance γ_k. Notice that the similar prior density has also been applied to signal compressed sensing more recently [43, 44]. The rationality of (16) is interpreted as follows. In fact, each γ_k in (16) may be considered to be a random variable, and a Gamma prior can be imposed independently on each unknown inverse variance, obtaining

$$p(\gamma|a,b)=\prod_{k=0}^{J}\Gamma(\gamma_k|a,b) \tag{17}$$

In the theory of Bayesian statistics, $p(\gamma|a,\ b)$ is called hyperprior, and a, b are hyperparameters to be estimated. By marginalizing over the inverse variances γ, the overall prior on β is then evaluated as

$$p(\beta|\gamma)=\prod_{k=0}^{J}\int_0^{\infty}\mathrm{N}(\beta_k|0,\gamma_k^{-1})\Gamma(\gamma_k|a,b)d\gamma_k \tag{18}$$

When β_k plays the role of observed data and $N\left(\beta_k|0,\gamma_k^{-1}\right)$ is a likelihood function, the hyperprior $\Gamma(\gamma_k|a,b)$ is the conjugate prior for γ_k, and consequently, the integral in (18) can be evaluated analytically, corresponding to a student's t distribution [42]. With appropriate choices of a and b, the student's t distribution is strongly peaked about $\beta_k=0$, and therefore the prior in (18) favors most β_k being zero, i.e., it is a sparseness-promoting prior. Therefore, (16) is a simplified version of the above two-level hierarchical modeling, i.e., a uniform density is imposed for each inverse variance. Yet, the simplified hierarchical modeling much releases us estimating the hyperparameters.

3.2 Prediction via Bayesian inference and analysis

Given computer outputs $\boldsymbol{y}=(y_1,\ldots,y_n)'$ collected at n design points$\{\boldsymbol{x}^{(1)},\ldots,\boldsymbol{x}^{(n)}\}$, the aim is to make prediction at an untried design point $\boldsymbol{x}^0$. Specifically, the prediction can be formulated as

$$y^0=g(x^0)=\arg\max_{y^0}p(y^0|\boldsymbol{y},\boldsymbol{\beta},\boldsymbol{\theta},\sigma^2) \tag{19}$$

Because $\{y_1,\ldots,y_n\}$ are realizations of a stationary Gaussian stochastic process, $p(\boldsymbol{y}|\boldsymbol{\beta},\ \boldsymbol{\theta},\ \sigma^2)$ is the multivariate Gaussian likelihood function of the computer outputs $y=(y_1,\ldots,y_n)'$. That is,

$$p(y|\beta,\theta,\sigma^2)=\frac{1}{(2\pi\sigma^2)^{n/2}\sqrt{|\Phi(\theta)|}}\exp\left\{-\frac{(y-\Psi\beta)'\Phi^{-1}(\theta)(y-\Psi\beta)}{2\sigma^2}\right\} \tag{20}$$

Therefore, it is not hard to obtain the joint distribution of $y^0=g(x^0)$ and $y=(y_1,\ldots,y_n)'$ as follows:

$$\begin{pmatrix}y^0\\ y\end{pmatrix}\Bigg|\beta,\theta,\sigma^2\sim N\left(\begin{pmatrix}\psi'(x^0)\\ \Psi\end{pmatrix}\beta,\ \sigma^2\begin{pmatrix}1 & \varphi'(x^0)\\ \varphi(x^0) & \Phi\end{pmatrix}\right) \tag{21}$$

where $\psi(x^0)=[\psi_0(x^0),\psi_1(x^0),\ldots,\psi_J(x^0)]'$, and $\varphi(x^0)=[\varphi(x^0-x^{(1)}),\varphi(x^0-x^{(2)}),\ldots,\varphi(x^0-x^{(n)})]'$. Then, based on the conditional distribution of the multivariate normal [4], $p(y^0|y,\beta,\theta,\sigma^2)$ can be formulated as follows:

$$y^0|y,\beta,\theta,\sigma^2\sim N\left(\psi'(x^0)\beta+\varphi'(x^0)\Phi^{-1}(y-\Psi\beta),1-\varphi'(x^0)\Phi^{-1}\varphi(x^0)\right) \tag{22}$$

Hence, the maximizer in (19) is just the expectation of the multivariate normal distribution in (22), that is,

$$y^0=g(x^0)=\psi'(x^0)\beta+\varphi'(x^0)\Phi^{-1}(y-\Psi\beta) \tag{23}$$

Similar to (14) in universal Kriging, the prediction variance of the proposed predictor (23) can be also clarified. It only needs firstly formulating the density $p(y^0|y,\theta,\sigma^2)$, and then calculating the variance. Specifically, $p(y^0|y,\theta,\sigma^2)$ can be written as

$$\begin{aligned}p(y^0|y,\theta,\sigma^2)&=\int p(y^0,\beta|y,\theta,\sigma^2)d\beta\\&=\int p(y^0|y,\beta,\theta,\sigma^2)p(\beta|y,\theta,\sigma^2)d\beta\end{aligned} \tag{24}$$

It is observed that (23) relies on the correlation parameters $\boldsymbol{\theta}$ and the regression coefficients $\boldsymbol{\beta}$. Hence, parameters $\boldsymbol{\beta}$, $\boldsymbol{\theta}$, σ^2 need estimating to achieve the final prediction in (23). Based on the principle of Bayesian maximum a posterior (MAP), the parameters $\boldsymbol{\beta}$, $\boldsymbol{\theta}$, σ^2 can be estimated by

$$\begin{aligned}\{\beta,\theta,\sigma^2\}&=\arg\max_{\beta,\theta,\sigma^2}p(\beta,\theta,\sigma^2|y)\\&=\arg\max_{\beta,\theta,\sigma^2}\{p(y|\beta,\theta,\sigma^2)p(\beta)p(\theta)p(\sigma^2)\}\end{aligned} \tag{25}$$

From the above, it is known that $p(y|\beta,\theta,\sigma^2)$ corresponds to a multivariate Gaussian likelihood function of the computer outputs $\boldsymbol{y}$, and $p\beta$ is chosen as a multivariate normal distribution with varying variance in Subsect. 3.1. Thereby, only $p(\boldsymbol{\theta})$ and $p(\sigma^2)$ need specifying in the following. For computational efficiency, the two priors are expected to be parameter-free. Any reasonable non-informative prior densities may be imposed on the two parameters, such as Jeffrey's noninformative prior [40, 45], constant prior [39, 40], and so on. In this paper, a constant prior is, respectively, placed on the two parameters; that is,

$$p(\theta)=\prod_{j=1}^{p}p(\theta_j)\sim 1 \tag{26}$$

$$p(\sigma^2)\sim 1 \tag{27}$$

Substituting (16), (20), (26), and (27) into (25), parameters $\boldsymbol{\beta}$, $\boldsymbol{\theta}$ and σ^2 can be then estimated by

$$\beta|\gamma=\arg\max_{\beta}\{p(y|\beta,\theta,\sigma^2)p(\beta|\gamma)\} \tag{28}$$

$$\sigma^2 = \arg\max_{\sigma^2} p(y|\beta,\theta,\sigma^2) \tag{29}$$

$$\theta = \arg\max_{\theta} p(y|\beta,\theta,\sigma^2) \tag{30}$$

Besides, the inverse variances γ can be estimated by the following evidence maximization procedure, formulated as

$$\gamma = \arg\max_{\gamma} \log p(y,\gamma|\theta,\sigma^2) \tag{31}$$

where

$$L(\gamma) \triangleq \log p(y,\gamma|\theta,\sigma^2) = \log \int p(y|\beta,\theta,\sigma^2)p(\beta|\gamma)d\beta \tag{32}$$

4 Computational issues and fast algorithm

This section mainly focuses on the implementation of the optimization problems formulated in Sect. 3. In particular, a fast suboptimal iterative algorithm is presented for the extended Gaussian Kriging. Other detailed computational issues are also to be discussed and clarified in this section.

4.1 Computation and optimization

Above all, reformulate (20) to ease the computation in the following. Since the correlation function φ in $\boldsymbol{\Phi}(\boldsymbol{\theta})$ is of Gaussian type, $\boldsymbol{\Phi}(\boldsymbol{\theta})$ is symmetric and positive definite, implying that $\boldsymbol{\Phi}(\boldsymbol{\theta})$ can be rewritten in a factorized form using the Cholesky decomposition

$$\Phi(\theta) = C(\theta)C'(\theta)$$

where $\boldsymbol{C}(\boldsymbol{\theta})$ is an upper triangular matrix, called the Cholesky factor. Multiply the factor $\boldsymbol{C}^{-1}$ on both sides of the equation $\boldsymbol{y} = \boldsymbol{\Psi\beta} + z$, where $z = (z_1,\ldots, z_n)'$ are the realizations of the Gaussian stochastic process $z(\boldsymbol{x})$ at design points $\{\boldsymbol{x}^{(1)},\ldots,\boldsymbol{x}^{(n)}\}$, and obtain

$$C^{-1}y = C^{-1}\Psi\beta + C^{-1}z$$

Denote $\tilde{y} = C^{-1}y, \tilde{\Psi} = C^{-1}\Psi, \tilde{z} = C^{-1}z$, and obtain

$$\tilde{y} = \tilde{\Psi}\beta + \tilde{z} \tag{33}$$

Then, $\tilde{y}$ is distributed to a multivariate normal PDF with an identity covariance matrix; (20) can be reformulated as

$$p(\tilde{y}|\beta,\theta,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left\{-\frac{(\tilde{y}-\tilde{\Psi}\beta)'(\tilde{y}-\tilde{\Psi}\beta)}{2\sigma^2}\right\} \tag{34}$$

Now, consider (28) for the Bayesian MAP estimate of the regression coefficients $\boldsymbol{\beta}$. Since the likelihoods (20) and (34) are equivalent to each other as given $\boldsymbol{\beta}$, $\boldsymbol{\theta}$, σ^2, (28) can be reformulated as

$$\beta|\gamma = \arg\max_{\beta}\{p(\tilde{y}|\beta,\theta,\sigma^2)p(\beta|\gamma)\}$$

Because the prior density $p(\beta|\gamma)$ is conjugate to the likelihood $p(\tilde{y}|\beta,\theta,\sigma^2)$, it is easy to find that the Bayesian posterior for the regression coefficients can be expressed analytically as a multivariate normal distribution, i.e.,

$$p(\beta|\tilde{y},\theta,\sigma^2,\gamma) \propto p(\tilde{y}|\beta,\theta,\sigma^2)p(\beta|\gamma) \propto N\left(\mu,\sum\right)$$

where

$$\mu = \sigma^{-2}\sum\tilde{\Psi}'\tilde{y},$$
$$\sum = \left(\Lambda + \sigma^{-2}\tilde{\Psi}'\tilde{\Psi}\right)^{-1},$$

and $\Lambda = diag(\gamma_0,\gamma_1,\ldots,\gamma_J)$. Hence, provided the transformed outputs $\tilde{y}$ and the parameters θ,γ,σ^2, the MAP estimate of the regression coefficients is

$$\beta|\gamma = \arg\max_{\beta}\{p(\tilde{y}|\beta,\theta,\sigma^2)p(\beta|\gamma)\} = \sigma^{-2}\sum\tilde{\Psi}'\tilde{y} \tag{35}$$

For the standard deviation σ^2 and correlation parameters $\boldsymbol{\theta}$, due to their being placed on the constant prior, the optimization in (29) and (30) corresponds to the maximum likelihood estimation as in universal Kriging. The MAP estimates of parameters σ^2, $\boldsymbol{\theta}$ are therefore given by

$$\sigma^2 = \frac{1}{n}(\tilde{y}-\tilde{\Psi}\beta)'(\tilde{y}-\tilde{\Psi}\beta) \tag{36}$$

$$\theta = \arg\min_{\theta}\left\{\log\left(\frac{1}{n}(\tilde{y}-\tilde{\Psi}\beta)'(\tilde{y}-\tilde{\Psi}\beta)\right) + \frac{1}{n}\log(\det(\Phi))\right\} \tag{37}$$

Obviously, there is no analytical solution to (36). In the literature, several numerical algorithms have been exploited to solve (36), e.g., the Newton–Raphson method [2], the quasi-Newton method [46], the Fisher scoring algorithm [2], adaptive simulated annealing [46], and the pattern search method [46, 47]. The former three methods use gradient information and yield local optimizers. Though the pattern search is not directly gradient-based, it is still a local optimization method. Notice that the pattern search method tends to be better appropriate to highly nonlinear or discontinuous functions. Comparatively, adaptive simulated annealing is more robust in finding solutions to functions where multimodal and long near-optimal ridge features exist, e.g., (37). While, adaptive simulated annealing is computationally intensive as it is a Monte Carlo method [46]. Considering both issues of computational efficiency and accuracy, the Fisher scoring algorithm and the pattern search are chosen as candidate optimization algorithms for solving (37).

To estimate the inverse variances γ, one has to compute the function $L(\gamma)$ as in (32). Since we have

$p(\tilde{y}|\beta,\theta,\sigma^2)p(\beta|\gamma) \propto N(\mu,\sum)$, $L(\gamma)$ can be written as follows:

$$
\begin{aligned}
L(\gamma) &\triangleq \log \int p(\tilde{y}|\beta,\theta,\sigma^2)p(\beta|\gamma)d\beta \\
&= \log\left\{\left(\frac{1}{2\pi}\right)^{n/2}|\sigma^2 I_{n\times n}+\tilde{\Psi}\Lambda^{-1}\tilde{\Psi}'|^{-1/2}\right. \\
&\quad \left.\cdot\exp\left\{-\frac{1}{2}\tilde{y}'(\sigma^2 I_{n\times n}+\tilde{\Psi}\Lambda^{-1}\tilde{\Psi}')^{-1}\tilde{y}\right\}\right\} \\
&= \log\left\{\left(\frac{1}{2\pi}\right)^{n/2}|\Omega|^{-1/2}\cdot\exp\left\{-\frac{1}{2}\tilde{y}'\Omega^{-1}\tilde{y}\right\}\right\}
\end{aligned}
$$

where $I_{n\times n}$ is an identity matrix, $\Omega = \sigma^2 I_{n\times n}+\tilde{\Psi}\Lambda^{-1}\tilde{\Psi}'$, and

$$
\begin{aligned}
|\Omega| &= |\Lambda|^{-1}\cdot|\sigma^2 I_{n\times n}|\cdot|\Lambda+\sigma^{-2}\tilde{\Psi}'\tilde{\Psi}| = |\Lambda|^{-1} \\
&\quad \cdot|\sigma^2 I_{n\times n}|\cdot\left|\sum^{-1}\right|, \\
\Omega^{-1} &= (\sigma^2 I_{n\times n}+\tilde{\Psi}\Lambda^{-1}\tilde{\Psi}')^{-1} \\
&= \sigma^{-2}I_{n\times n}-\sigma^{-2}\tilde{\Psi}(\Lambda+\sigma^{-2}\tilde{\Psi}'\tilde{\Psi})^{-1}\tilde{\Psi}'\sigma^{-2} \\
&= \sigma^{-2}I_{n\times n}-\sigma^{-2}\tilde{\Psi}\sum\tilde{\Psi}'\sigma^{-2}.
\end{aligned}
$$

Thus,

$$
\gamma_k = \arg\max_{\gamma_k} L(\gamma) = \frac{1-\gamma_k\sum_{kk}}{\mu_k} \tag{38}
$$

where μ_k is the kth element of μ, and $\sum_{kk}$ is the kth diagonal element of $\sum$.

Finally, using (35), (36), (37) and (38), an iterative algorithm for proposed extended Gaussian Kriging can be obtained by updating each of the parameters $\boldsymbol{\gamma}$, $\boldsymbol{\beta}$, $\boldsymbol{\sigma}^2$, $\boldsymbol{\theta}$ per iteration.

4.2 Fast algorithm

In Subsect. 4.1, estimation of parameters $\boldsymbol{\gamma}$, $\boldsymbol{\beta}$, $\boldsymbol{\sigma}^2$, $\boldsymbol{\theta}$ requires iterative computation of (35), (36), (37), and (38). Notice that σ^2 is easy to calculate utilizing (36), and the correlation parameters $\boldsymbol{\theta}$ can be calculated by the efficient Fisher scoring algorithm or the pattern search method. However, the estimation of β and γ is not cheap because matrix inversion in (35) and (38) requires $O((J+1)^3)$ computations. Therefore, the algorithm presented in Sect. 4.1 cannot be directly applied to practical metamodeling tasks, but is able to serve us as the starting point in developing a fast algorithm in the following. The kernel idea of the fast algorithm is the reduction of the dimension of $\boldsymbol{\Sigma}$ in (35) and (38). According to the expectation that the regression coefficients $\boldsymbol{\beta}$ are sparse and most inverse variances of $\{\gamma_k\}_{k=0}^{J}$ should be set equal to zero, thus leading to a lower dimension matrix $\boldsymbol{\Sigma}$. Therefore, the fast algorithm can be implemented by starting an empty matrix $\boldsymbol{\Psi}$ (that is, all the inverse variances are equal to zero), and iteratively and adaptively adding a single variable every time to the matrix $\boldsymbol{\Psi}$.

The fast algorithm is presented as follows: consider the function $L(\boldsymbol{\gamma})$ (omitting a constant)

$$
L(\gamma) = -\frac{1}{2}\log|\Omega| - \frac{1}{2}\tilde{y}'\Omega^{-1}\tilde{y} \tag{39}
$$

To estimate each inverse variance γ_k while holding other inverse variances fixed, (39) is rewritten as

$$
\begin{aligned}
L(\gamma) &= -\frac{1}{2}\left[\log|\Omega_{-k}| + \tilde{y}'\Omega_{-k}^{-1}\tilde{y}\right] \\
&\quad + \frac{1}{2}\left[\log\gamma_k - \log(\gamma_k+s_k) + \frac{q_k^2}{\gamma_k+s_k}\right] \\
&\triangleq L(\gamma_{-k}) + l(\gamma_k)
\end{aligned} \tag{40}
$$

where $\boldsymbol{\Omega}_{-k}$ denotes that the contribution of kth basis in $\tilde{\Psi} = [\tilde{\psi}_0, \tilde{\psi}_1, \ldots, \tilde{\psi}_J]$, i.e., $\tilde{\psi}_k$, is not included, and

$$
\begin{aligned}
\Omega &= \sigma^2 I_{n\times n}+\tilde{\Psi}\Lambda^{-1}\tilde{\Psi}' = \sigma^2 I_{n\times n}+\sum_{k=0}^{J}\gamma_k^{-1}\tilde{\psi}_k\tilde{\psi}_k' \\
&= \Omega_{-k}+\gamma_k^{-1}\tilde{\psi}_k\tilde{\psi}_k', \\
\Omega^{-1} &= \Omega_{-k}^{-1} - \frac{\Omega_{-k}^{-1}\tilde{\psi}_k\tilde{\psi}_k'\Omega_{-k}^{-1}}{\gamma_k+\tilde{\psi}_k'\Omega_{-k}^{-1}\tilde{\psi}_k}, \\
|\Omega| &= |\Omega_{-k}|\cdot|1+\gamma_k^{-1}\tilde{\psi}_k'\Omega_{-k}^{-1}\tilde{\psi}_k|,
\end{aligned}
$$

Besides, s_k, q_k in (40) are defined as $s_k = \tilde{\psi}_k'\Omega_{-k}^{-1}\tilde{\psi}_k$, $q_k = \tilde{\psi}_k'\Omega_{-k}^{-1}\tilde{y}$. Thus, γ_k can be estimated by just maximizing $l(\gamma_k)$ since $L(\gamma_{-k})$ does not depend on γ_k. After careful calculation, we obtain

$$
\begin{aligned}
&\gamma_k = \frac{s_k^2}{q_k^2-s_k}, \text{ if } q_k^2-s_k>0 \\
&\gamma_k = \infty, \text{ if } q_k^2-s_k\le 0
\end{aligned} \tag{41}
$$

As $\gamma_k=\infty, \psi_k$ is pruned out from the matrix Ψ, and $\boldsymbol{\beta}_k$ is set equal to zero. Notice that, similar to [39, 48], s_k, q_k can be efficiently calculated by

$$
\begin{aligned}
S_k &= \sigma^{-2}\tilde{\psi}_k'\tilde{\psi}_k - \sigma^{-4}\tilde{\psi}_k'\tilde{\Psi}\sum\tilde{\Psi}'\tilde{\psi}_k \\
Q_k &= \sigma^{-2}\tilde{\psi}_k'\tilde{y} - \sigma^{-4}\tilde{\psi}_k'\tilde{\Psi}\sum\tilde{\Psi}'\tilde{y} \\
s_k &= \gamma_k S_k/(\gamma_k - S_k) \\
q_k &= \gamma_k Q_k/(\gamma_k - S_k)
\end{aligned}
$$

where $\boldsymbol{\Sigma}$ and $\tilde{\Psi}$ include only the columns that have been included in the regression matrix$\tilde{\Psi}$. Since the inverse variances γ can be easily estimated by (41), $\boldsymbol{\beta}$ can be calculated rather efficiently by (35) with low dimensional matrices $\boldsymbol{\Sigma}$ and $\tilde{\Psi}$. Additionally, the initial values of parameters $\gamma, \beta, \sigma^2, \theta$ to the fast algorithm are given as follows: $\gamma^{(0)}=0$, $\beta^{(0)}=0$, $\sigma^{2(0)}=0$, $\theta^{(0)}=0$.

4.3 Discussion on computational complexity

Another important issue is the computational complexity of the overall extended Gaussian Kriging. It is noticed that the related parameters in extended Gaussian Kriging are estimated iteratively; and particularly, the regression coefficients are estimated adaptively by a fast algorithm in Subsect. 4.2. However, since the correlation parameters in extended Gaussian Kriging are estimated by the Fisher scoring algorithm or pattern search method as in ordinary Kriging, the computational efficiency of extended Gaussian Kriging and ordinary Kriging is to decrease as the sample size of computer runs increases, due to the inversion of the correlation matrix involved in the process of fitting Kriging. As a matter of fact, the similar problem also exists in other recent Kriging metamodeling methods, e.g., limit Kriging [20], SCAD penalized Kriging [2], blind Kriging [18], and so on. However, it is believed that the prediction performance of extended Gaussian Kriging can be still superior to that of ordinary Kriging and universal Kriging. That is because extended Gaussian Kriging can provide a more accurate mean model adaptively capturing the overall trend of computer responses.

Actually, our extended Gaussian Kriging is particularly appropriate to those much more complex computer experiments. In such situations, collecting a large or moderate sample may be very difficult, and a more efficient and accurate approximation model from a small sample becomes much crucial. To be specific, our extended Gaussian Kriging has competitive computational efficiency compared against ordinary Kriging, in that a fast algorithm is exploited to estimate the regression coefficients in the mean model and the correlation parameters are estimated only once as ordinary Kriging. More importantly, experimental results in Sect. 5 demonstrate that extended Gaussian Kriging is capable of achieving higher prediction accuracy. However, the correlation parameters in blind Kriging have to be estimated two times [18], making the computational complexity of blind Kriging at least twice as that of ordinary Kriging. In the meanwhile, a scheme of computationally intensive Bayesian forward variable selection is also involved in the process of estimation. Though limit Kriging [20] achieves the similar computational efficiency as ordinary Kriging, the improvement in prediction accuracy is rather limited. Furthermore, SCAD penalized Kriging [2] improves ordinary Kriging at the cost of estimating the tuning parameters involved in the penalized likelihood. As for non-Kriging metamodeling methods in Sect. 2, model construction can be more efficient than ordinary Kriging because there only needs estimating the regression coefficients in PR and (augmented) RBF. Nevertheless, less efficiency both in computation and prediction is to be expected as applied to those more complex problems with high dimension and highly nonlinear performance [5].

5 Examples

5.1 Example 1: piston slap noise experiment

Automobile customer satisfaction depends highly on the level of satisfaction that a customer has with the vehicle's engine. The noise, vibration, and harshness characteristics of an automobile and its engine are the critical elements of customer dissatisfaction. Specifically, piston slap noise is focused in this case study. Generally, the piston slap noise results from the piston secondary motion which is caused by a combination of transient forces and moments acting on the piston during engine operation and presence of clearances between the piston and the cylinder liner. The combination results in both a rotation and a lateral movement of the piston, causing the piston to impact the cylinder wall at regular intervals. It is just these impacts resulting in the objectionable engine noise known as piston slap. A thorough and detailed description on the experiment can be found in Hoffman et al. [49].

For this study, a computer experiment needs performing by varying variables to minimize the piston slap noise. In the experiment, the piston slap noise (y) is taken as the output variable, and the clearance between the piston and the cylinder liner (x_1), the location of peak pressure (x_2), the skirt length (x_3), the skirt profile (x_4), the skirt ovality (x_5), and the pin offset (x_6) are taken as the input variables. It is to be noted that the clearance between the piston and the cylinder liner (x_1) and the location of peak pressure (x_2) are noise factors. Besides, the uniform design [25] is used to construct an experimental design for the computer experiment with 12 runs. The designed inputs and outputs of the computer experiment are given in Table 1. More details on the experimental setup can be found in Li et al. [2].

In the following, extended Gaussian Kriging is used to construct a metamodel as an approximation to the computationally expensive computer experiment. The candidate basis functions in the mean model are assumed to be the linear and quadratic polynomial functions, as well as the two-factor interactions, that is,

$$\begin{aligned}&\psi_0(x) = 1,\ \psi_1(x) = x_1, \ldots,\ \psi_p(x) = x_p,\\ &\psi_{p+1}(x) = x_1^2, \ldots,\ \psi_{2p}(x) = x_1 x_p,\\ &\psi_{2p+1}(x) = x_2^2, \ldots,\ \psi_{3p}(x) = x_2 x_p, \ldots,\ \psi_{p(p+3)/2}(x) = x_p^2\end{aligned} \tag{42}$$

Since there are six factors in the piston slap noise experiment, there are hence totally 28 variables in the regression

Table 1 Design points and outputs of Piston slap noise experiment for metamodeling

Run	x_1	x_2	x_3	x_4	x_5	x_6	y
1	71	16.80	21.00	2.00	1.00	0.98	56.75
2	15	15.60	21.80	1.00	2.00	1.30	57.65
3	29	14.40	25.00	2.00	1.00	1.14	53.97
4	85	14.40	21.80	2.00	3.00	0.66	58.77
5	29	12.00	21.00	3.00	2.00	0.82	56.34
6	57	12.00	23.40	1.00	3.00	0.98	56.85
7	85	13.20	24.20	3.00	2.00	1.30	56.68
8	71	18.00	25.00	1.00	2.00	0.82	58.45
9	43	18.00	22.60	3.00	3.00	1.14	55.50
10	15	16.80	24.20	2.00	3.00	0.50	52.77
11	43	13.20	22.60	1.00	1.00	0.50	57.36
12	57	15.60	23.40	3.00	1.00	0.66	59.64

mean model. Besides, ordinary Kriging, universal Kriging, SCAD penalized Kriging [2], and blind Kriging [18] are also implemented for comparison. In the following, universal Kriging is denoted as universal Kriging-LP as linear polynomial functions are chosen as candidate variables, and denoted as universal Kriging-QP as (42) is chosen.

For all the Kriging methods, the correlation parameters are estimated by the Fisher scoring algorithm [2]. Augmented radial basis functions (A-RBF) using several commonly used basis functions are also tried for this example, including linear, cubic, Gaussian, multi-quadric, and inverse multi-quadric functions [31]. The related parameters in these basis functions are tuned to be optimal. Besides, PR–LP and PR–QP are implemented for comparison, considering their popularity and widespread use in the industry.

To assess the performance of different metamodeling approaches, additional 100 runs are provided by another designed computer experiments. Two criteria are used to measure how well the prediction performs. They are, respectively, the root mean square error (RMSE) and the median of absolute error (MAR), defined as

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{y}(x^{(i)}) - y(x^{(i)})\right)^2}$$

$$\text{MAR} = median\{|\hat{y}(x^{(i)}) - y(x^{(i)})|, i = 1, 2, \ldots, N\}$$

where N is the number of total testing design points $\{x^{(1)}, \ldots, x^{(N)}\}$, $\hat{y}(x)$ is the predicted response at a design point $\boldsymbol{x}$, and $y(x)$ is the true response. The RMSE and MAR of different metamodeling methods are listed in Tables 2 and 3. Besides, estimated regression coefficients of different approaches and correlation parameters in different Kriging methods are outlined in Tables 4 and 5.

The RMSE for extended Gaussian Kriging is 1.1174, and equals 1.2807, 1.4813, 1.7307, 1.4864, and 4.5407 for blind Kriging, SCAD penalized Kriging, and ordinary Kriging, universal Kriging-LP, and universal Kriging-QP. The MAR for extended Gaussian Kriging is 0.7852, and MAR equals 0.8560, 1.0588, 1.3375, 1.1730, and 3.2739 for blind Kriging, SCAD penalized Kriging, and ordinary Kriging, universal Kriging-LP, and universal Kriging-QP. Both the RMSE and MAR demonstrate that extended Gaussian Kriging achieves a fairly better prediction performance than other Kriging methods. Though SCAD penalized Kriging focuses on the more careful calculation of correlation parameters, it is still limited in improving the prediction accuracy due to its mean function just a constant. It is also observed that extended Gaussian Kriging performs slightly better than blind Kriging though a

Table 2 RMSE and MAR of Kriging and polynomial regression methods in Piston slap noise experiment

Methodology	RMSE	MAR
Extended Gaussian Kriging	1.1174	0.7852
Blind Kriging	1.2807	0.8560
SCAD penalized Kriging	1.4813	1.0588
Ordinary Kriging	1.7307	1.3375
Universal Kriging-LP	1.4864	1.1730
PR-LP	2.6435	2.1827
Universal Kriging-QP	4.5407	3.2739
PR-QP	6.4818	3.6618

Table 3 RMSE and MAR of augmented radial basis functions (A-RBF) using several basis function in piston slap noise experiment

Basis function name	RMSE	MAR
Linear	1.8027	1.2465
Cubic	1.8656	1.3308
Gaussian	1.4856	1.1730
Multi-quadric	1.8090	1.2480
Inverse multi-quadric	1.4856	1.1730

Table 4 Estimates of regression coefficients via extended Gaussian Kriging and other approaches

Regression coefficients	β_0	β_1	β_2	β_3	β_4	β_5	β_6
Extended Gaussian Kriging	56.4017	0.0000	0.0000	0.0000	0.0000	0.0000	0.7091
Universal Kriging-QP	0.1022	1.4626	0.7100	1.1644	0.1247	0.0987	0.0687
PR-QP	0.0532	1.0783	0.3920	0.6146	0.0319	0.0603	0.0426
Universal Kriging-LP	66.4217	0.0473	0.0142	−0.4285	−0.4997	−0.4809	−0.5044
PR-LP	22.7278	0.0520	0.2463	0.9180	−0.0746	0.6657	5.8775
Regression Coefficients	β_7	β_8	β_9	β_{10}	β_{11}	β_{12}	β_{13}
Extended Gaussian Kriging	−0.0003	0.0000	0.0085	0.0000	0.0000	−0.1143	0.0000
Universal Kriging-QP	−0.0021	−0.0096	−0.0090	0.0221	−0.1874	−0.5087	−0.1107
PR-QP	−0.0057	0.0447	−0.0521	0.3155	−0.0676	−0.4605	-0.0001
Regression coefficients	β_{14}	β_{15}	β_{16}	β_{17}	β_{18}	β_{19}	$yyy\beta_{20}$
Extended Gaussian Kriging	0.0000	0.0000	0.0000	0.0000	0.0000	0.1112	0.0000
Universal Kriging-QP	0.0819	−0.6582	0.4131	0.6499	−0.0549	0.2541	0.1713
PR-QP	−0.1070	−0.3061	−0.1152	0.3947	0.1138	−0.4295	0.3616
Regression coefficients	β_{21}	β_{22}	β_{23}	β_{24}	β_{25}	β_{26}	β_{27}
Extended Gaussian Kriging	−0.3900	1.3089	−0.1931	−2.4253	0.0000	0.0000	9.1991
Universal Kriging-QP	0.5288	0.4252	0.5232	0.1471	−0.2915	0.1562	0.0666
PR-QP	0.3967	0.0295	-0.0157	-0.0142	0.0136	0.0870	0.0397

Bayesian forward variable selection strategy has been specifically incorporated into Kriging. And universal Kriging-QP has achieved the worst performance among several Kriging methods, in that there are many unimportant variables included in the mean model as shown in Table 4, which actually deteriorate the prediction performance. That is also why universal Kriging-LP has performed better than universal Kriging-QP in this example.

The RMSE for PR-LP is 2.6435, and the MAR equals 2.1827. The RMSE for PR-QP is 6.4818, and the MAR equals 3.6618. Compared against the Kriging methods, it is clear that PR-LP (QP) achieves poorer prediction performance than universal Kriging-LP (QP). The reason lies in that there is a Gaussian stochastic process modeling residuals from the linear regression in universal Kriging-LP (QP). Besides, Kriging is interpolative, whereas PR is not. Compared against A-RBF, PR is of also lower prediction accuracy. The RMSE for the linear basis function is 1.8027 and equals 1.8656, 1.4856, 1.8090, and 1.4856 for the cubic, Gaussian, multi-quadric, and inverse multi-quadric basis functions. Their MAR equals 1.2465, 1.3308, 1.1730, 1.2480, and 1.1730, correspondingly. The superiority of A-RBF to PR also consists in that a RBF model passes through all the sampling points exactly, whereas PR does not. In the meanwhile, it is observed that Gaussian and inverse multi-quadric basis functions achieve the same RMSE and MAR and perform a little better than other three basis functions.

Observed from the RMSE and MAR of all the methods discussed in this example, it is obvious that extended Gaussian Kriging achieves the best performance, not only lower in the global RMSE but also lower in the local MAR. Observed from the estimated regression coefficients shown in Table 4, there are many more zeros in the regression mean model of extended Gaussian Kriging than other Kriging and non-Kriging methods. Hence, extended Gaussian Kriging has obtained a more adaptive regression mean, thereby leading to a more accurate metamodel for expensive computer simulators.

Furthermore, Fig. 1 plots the sorted absolute residuals (AR) from universal Kriging-LP, SCAD penalized Kriging, blind Kriging, and extended Gaussian Kriging versus the AR from ordinary Kriging. From these plots, we see that SCAD penalized Kriging, blind Kriging, and extended Gaussian Kriging have uniformly improved ordinary Kriging, and extended Gaussian Kriging performs more effectively than SCAD penalized Kriging and blind Kriging. It is also observed that universal Kriging-LP achieves similar performance as SCAD penalized Kriging, whereas universal Kriging-QP completely fails in this example due to the inclusion of unimportant effects in the mean model. Besides, Fig. 2 plots the true responses versus the predicted responses obtained by extended Gaussian Kriging and A-RBF via different basis functions, showing that the predicted responses obtained by extended Gaussian Kriging (Fig. 2f) are more accurate than those predicted by A-RBF methods.

5.2 Example 2: exhaust manifold sealing experiment

In this case study, the design of an exhaust manifold is considered. The exhaust manifold is one of the engine

Table 5 Estimated correlation parameters of different Kriging methods

Methodology	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
Extended Gaussian Kriging	0.0007	0.0076	0.0114	0.0191	0.0191	0.0571
Blind Kriging	0.0100	0.0100	0.0900	1.3200	0.0100	0.4900
SCAD Penalized Kriging	0.0008	0.0000	0.0398	0.0000	0.0000	4.4979
Ordinary Kriging	0.1397	1.6300	2.4451	4.0914	4.0914	12.2253
Universal Kriging-LP	0.1397	1.6300	2.4451	4.0914	4.0914	12.2253
Universal Kriging-QP	0.0007	0.0076	0.0114	0.0191	0.0191	0.0571

components, expected to endure harsh and rapid thermal cycling conditions ranging from sub-zero to, in some cases, more than 1,000°C. To dramatically reduce the cycle of product development, a transient non-linear finite element method is proposed by Hazime et al. [50] to simulate the inelastic deformation used to predict the thermo-mechanical fatigue life of cast exhaust manifolds. The exhaust manifold assembly includes the exhaust manifold component, fastener, gasket, and a portion of the cylinder head. In addition to predicting fatigue life, the model also predicts dynamic sealing pressure on gasket to identify potential exhaust gas leaks. To optimize the gasket design to prevent leaks, a sym- metric Latin hypercube (SLH) design 17 runs and 5 factors is applied to the simulation model as shown in Table 6. More details on exhaust manifold sealing experiment can be referred to ref. [4, 50]. Besides, a set of testing data set is listed in Table 7 to assess the performance of different Kriging and non-Kriging metamodeling methods.

Various Kriging methods including extended Gaussian Kriging, SCAD penalized Kriging, universal Kriging-LP, universal Kriging-QP, and ordinary Kriging, and non-Kriging methods, i.e., augmented RBF via different basis functions are tried for the exhaust manifold sealing experiment. In this example, the correlation parameters in extended Gaussian Kriging are estimated by the pattern search method. The comparison of results in terms of RMSE and MAR is presented in Table 8.

The RMSE for extended Gaussian Kriging is 1.2269, and equals 1.2323, 1.8772, 2.3697, 2.8792, and 4.0745 for blind Kriging, SCAD penalized Kriging, and universal Kriging-QP, universal Kriging-LP, and ordinary Kriging. The MAR for extended Gaussian Kriging is 0.5762, and correspondingly, the MAR equals 0.5937, 0.6672, 1.5059, 2.0805, and 3.4323. From the RMSE and MAR, it is observed that extended Gaussian Kriging achieves higher accuracy than SCAD penalized Kriging, universal Kriging, and ordinary Kriging. It is also observed that ordinary Kriging achieves the poorest prediction performance among these Kriging methods in this example. The reason lies in that a mere constant mean can not capture the overall trend of system responses. This is also manifested by SCAD penalized Kriging which performs superior to ordinary and universal Kriging while inferior to extended Gaussian Kriging. In this example, note that extended Gaussian Kriging achieves similar prediction performance as blind Kriging. However, the computational complexity of blind Kriging is high, whereas extended Gaussian Kriging has competitive computational efficiency compared against ordinary Kriging.

Besides Kriging methods, augmented RBF methods are also carried out for comparison. Specifically, the RMSE for the linear basis function is 2.5895 and equals 2.2189, 1.4949, 2.2189, and 2.3391 for the cubic, Gaussian, multi-quadric, and inverse multi-quadric basis functions; and the corresponding MAR equals 1.5540, 1.3687, 0.7432, 1.3687, and 1.5768. Among the five basis functions, A-RBF-Gaussian achieves the best prediction in this example. Besides, compared against the SCAD penalized Kriging method, A-RBF-Gaussian performs slightly better in the term of RMSE and slightly worse in the term of MAR. Yet, extended Gaussian Kriging overwhelms SCAD penalized Kriging and A-RBF-Gaussian in both terms of RMSE and MAR. Hence, it can be concluded once again that the adaptive mean model helps to improving the prediction performance of Kriging metamodeling.

5.3 Example 3: Borehole model

The Borehole model is a classical example of computer experiments having studied by many authors to compare different methods in computer experiments, for instance, Fang and Lin [51], Fang et al. [52], and Li [53]. The borehole function is defined as

$$y = \frac{2\pi T_u[H_u - H_l]}{ln\left(\frac{r}{r_w}\right)\left[1 + \frac{2LT_u}{ln(r/r_w)r_w^2 k_w} + \frac{T_u}{T_l}\right]}$$

where y is the response variable, rw, r, Tu, Tl, Hu, Hl, L and Kw are the eight input variables, having the experimental domain as follows:

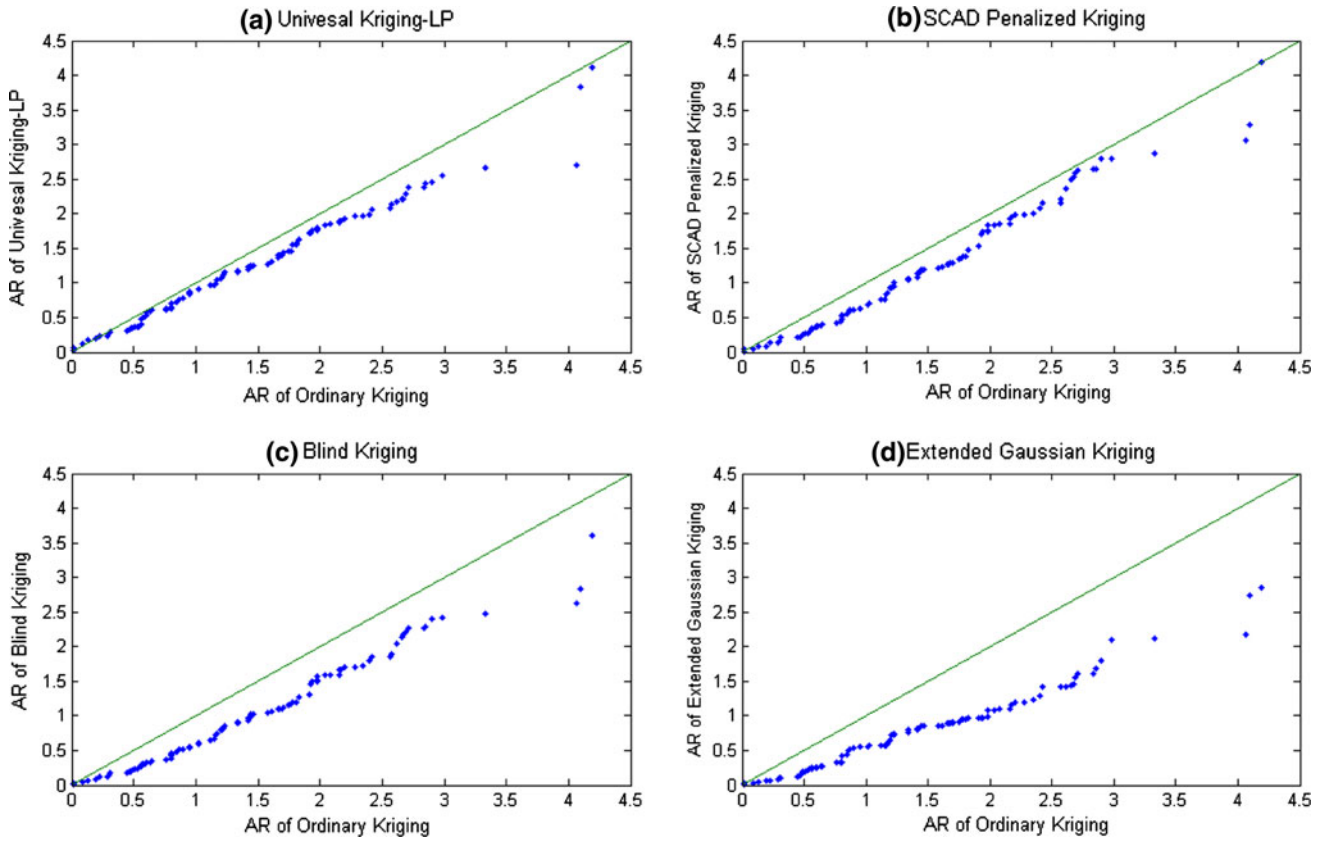

Fig. 1 Plots of absolute residuals for **a** Universal Kriging-LP, **b** SCAD Penalized Kriging, **c** Blind Kriging, and **d** Extended Gaussian Kriging

$r_w \in [0.05, 0.15], r \in [100, 50{,}000]$,
$T_u \in [63{,}070, 115{,}600], T_l \in [63.1, 116]$,
$H_u \in [990, 1{,}110], H_l \in [700, 820]$,
$L \in [1{,}120, 1{,}680], K_w \in [9{,}855, 12{,}045]$.

The same uniform design as in Li [53] is used to generate an experimental design for the Borehole function. The generated 30 design points and their corresponding outputs are depicted in Table 9. For performance assessment, 10,000 sample points are generated from the uniform distribution over the experimental domain.

In this study, extended Gaussian Kriging and several existing methods performed on the borehole model are compared for their prediction accuracy. The comparison of RMSE results is listed in Table 10. In literature, Fang and Lin [51] constructed a uniform design with 32 experiments and used the B-spline model to approximate the borehole function, achieving the RMSE 2.1095. With the same uniform design, the quadratic regression model used in Fang et al. [52] has the RMSE 0.5077. Li [53] used the SCAD penalized quadratic spline model for metamodeling. Their predictor is demonstrated to be more accurate than existing methods, and the RMSE equals 0.3335. Since universal Kriging may include unimportant variables in the mean model, the widely used ordinary Kriging, SCAD penalized Kriging, blind Kriging, and extended Gaussian Kriging are conducted on the Borehole model. And the Fisher scoring algorithm is utilized to estimate the correlation parameters. The RMSE equals 0.6587 for ordinary Kriging, equals 0.7527 for SCAD penalized Kriging, and equals 0.0044 for extended Gaussian Kriging. We see that the RMSE of extended Gaussian Kriging is much lower than that of Li's SCAD quadratic spline model, Fang and Lin's B-spline model, and Fang et al.'s quadratic regression model. It is also observed that both ordinary Kriging and SCAD penalized Kriging have achieved poor prediction performance in this example, in that the two methods completely ignore the important variables greatly affecting the system response. In fact, the adaptive mean model obtained by extended Gaussian Kriging is $\mu(x) = 10.3218 + 2.0270x_1 - 0.0015x_7$, which implies that only x_1 and x_7 are important variables. The result clearly demonstrates the importance of incorporating variable selection and hence an adaptive mean model into Kriging metamodeling.

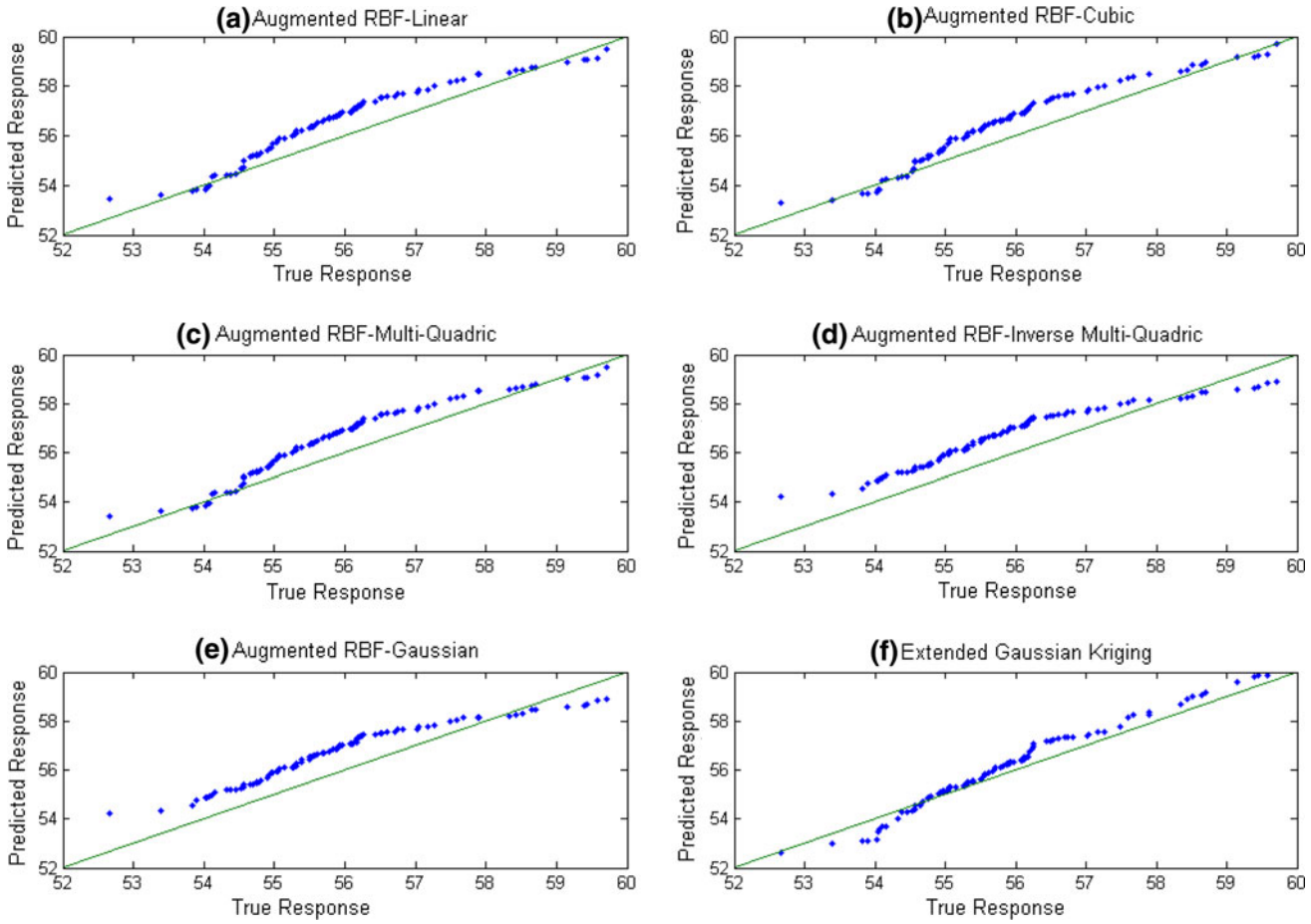

Fig. 2 Plots of the true responses versus the predicted responses obtained from the augmented RBF method with several basis functions **a** Linear, **b** cubic, **c** multi-quadric, **d** inverse multi-quadric, **e** Gaussian, and **f** extended Gaussian Kriging

Table 6 Design points and outputs of exhaust manifold sealing experiment for metamodeling

Run	x_1	x_2	x_3	x_4	x_5	y
1	0.00	0.38	0.13	0.63	0.06	14.30
2	0.75	1.00	0.63	0.06	0.44	11.08
3	0.81	0.06	0.06	0.31	0.63	14.36
4	0.13	0.19	0.31	0.25	0.69	10.01
5	0.38	0.56	0.75	0.19	1.00	11.19
6	0.94	0.25	0.56	0.00	0.25	2.70
7	0.69	0.69	0.19	0.88	0.88	19.55
8	0.44	0.88	0.00	0.44	0.19	19.68
9	0.25	0.31	0.88	1.00	0.50	9.65
10	0.56	0.13	1.00	0.56	0.81	9.63
11	0.31	0.50	0.81	0.13	0.13	7.94
12	0.06	0.75	0.44	0.50	0.75	16.98
13	0.63	0.44	0.25	0.81	0.00	19.51
14	0.88	0.81	0.69	0.75	0.31	22.88
15	0.19	0.94	0.94	0.69	0.38	18.01
16	0.50	0.00	0.38	0.94	0.56	10.20
17	1.00	0.63	0.50	0.38	0.94	17.68

Table 7 Design points and outputs of exhaust manifold sealing experiment for validation

Run	x_1	x_2	x_3	x_4	x_5	y
1	0.00	0.26	0.47	0.21	0.42	8.61
2	0.63	0.42	0.05	0.26	0.11	15.48
3	0.84	0.63	0.53	0.42	0.00	20.72
4	0.68	0.00	0.37	0.68	0.26	14.93
5	0.11	0.79	0.26	0.89	0.16	17.59
6	0.47	0.95	0.21	0.11	0.37	12.54
7	0.58	0.89	0.42	0.95	0.84	19.15
8	0.16	0.11	0.16	0.47	0.89	12.47
9	0.32	0.68	0.00	0.58	0.53	19.32
10	0.21	0.47	0.74	0.00	0.21	3.57
11	0.37	0.16	0.84	0.84	0.05	10.57
12	0.53	0.05	1.00	0.32	0.58	6.19
13	1.00	0.53	0.11	0.79	0.63	21.30
14	0.05	0.58	0.89	0.74	0.74	13.50
15	0.89	0.21	0.32	0.05	0.79	6.30
16	0.95	0.84	0.95	0.53	0.32	21.19
17	0.74	0.74	0.79	0.16	0.95	12.04
18	0.79	0.32	0.68	1.00	0.47	12.53
19	0.26	1.00	0.63	0.37	0.68	18.08
20	0.42	0.37	0.58	0.63	1.00	14.76

Table 8 RMSE and MAR of Kriging and augmented radial basis functions for exhaust manifold sealing experiment

Methodology	RMSE	MAR
Extended Gaussian Kriging	1.2269	0.5726
Blind Kriging	1.2323	0.5937
SCAD penalized Kriging	1.8772	0.6672
Universal Kriging-QP	2.3697	1.5059
Universal Kriging-LP	2.8792	2.0805
Ordinary Kriging	4.0745	3.4323
A-RBF-linear	2.5895	1.5540
A-RBF-cubic	2.2189	1.3687
A-RBF-Gaussian	1.4949	0.7432
A-RBF-multi-quadric	2.2189	1.3687
A-RBF-inverse multi-quadric	2.3391	1.5768

6 Conclusions

This paper primarily focuses on the issue of constructing surrogate models based on Kriging as cheap alternatives to computer simulators. Though several improved Kriging approaches are recently reported in the literature (e.g., SCAD penalized Kriging [2], blind Kriging [18]), an extended Gaussian Kriging method is proposed to improve the prediction accuracy of Kriging metamodeling. Unlike the forgoing approaches, the new method imposes a variance-varying Gaussian prior on the unknown regression coefficients in the mean model of universal Kriging and makes prediction at new design points based on the principle of Bayesian MAP inference. The achieved mean model is adaptive, thereby able to capture more effectively the overall trend of computer responses and lead to a more accurate meta- model. Experimental results on empirical case studies are presented, showing remarkable improvement in prediction utilizing extended Gaussian Kriging over several benchmark methods in the literature.

Future aspects on extended Gaussian Kriging include considering non-Gaussian correlation functions in Kriging

Table 9 Design points and outputs of borehole model for metamodeling

Run	$x_1 = r_w$	$x_2 = r$	$x_3 = T_u$	$x_4 = T_l$	$x_5 = H_u$	$x_6 = H_l$	$x_7 = L$	$x_8 = K_w$	y
1	0.0617	45,842	84,957	106.3017	1,108	762	1,335	9,990	30.8841
2	0.1283	35,862	93,712	81.6150	1,092	790	1,241	10,136	126.2840
3	0.0950	10,912	81,456	111.5917	1,016	794	1,223	10,063	51.6046
4	0.0883	22,555	114,725	67.5083	1,008	778	1,260	10,048	44.7063
5	0.0583	14,238	95,464	71.0350	1,068	810	1,559	9,983	17.6309
6	0.0683	4,258	88,460	88.6683	1,080	766	1,129	10,005	40.7011
7	0.0817	27,545	105,970	109.8283	1,088	802	1,428	10,034	41.9919
8	0.1350	19,228	76,203	69.2717	1,032	738	1,148	10,151	146.8108
9	0.0717	40,852	104,218	74.5617	1,000	746	1,372	10,012	29.8083
10	0.1050	39,188	72,701	101.0117	1,064	814	1,167	10,085	74.3997
11	0.0750	20,892	98,965	99.2483	1,036	754	1,671	10,019	29.8223
12	0.1383	17,565	79,704	93.9583	1,100	782	1,652	10,158	116.6914
13	0.1183	32,535	74,451	108.0650	1,076	722	1,540	10,114	101.7336
14	0.1417	44,178	83,207	72.7983	1,060	706	1,447	10,165	154.9332
15	0.1083	12,575	112,974	86.9050	1,104	730	1,484	10,092	93.2778
16	0.1117	5,922	69,199	63.9817	1,056	774	1,409	10,100	78.5678
17	0.1017	2,595	86,708	83.3783	992	714	1,633	10,078	55.4821
18	0.1150	49,168	97,215	90.4317	1,020	726	1,204	10,107	101.7270
19	0.0783	24,218	63,945	76.3250	1,096	718	1,279	10,027	56.9115
20	0.1217	7,585	107,721	78.0883	1,040	806	1,353	10,121	80.7530
21	0.0850	47,505	67,447	85.1417	1,044	798	1,615	10,041	34.6025
22	0.0983	37,525	100,717	65.7450	1,084	742	1,596	10,070	65.1636
23	0.0650	9,248	70,949	102.7750	1,012	734	1,521	9,997	24.2095
24	0.0550	30,872	109,472	95.7217	1,052	710	1,185	9,975	27.3042
25	0.0517	34,198	77,954	79.8517	1,024	786	1,465	9,968	13.5570
26	0.1450	932	102,468	104.5383	1,072	750	1,297	10,173	165.6246
27	0.0917	15,902	91,961	115.1183	1,048	702	1,391	10,056	65.8352
28	0.1250	29,208	65,697	97.4850	996	758	1,316	10,129	89.2366
29	0.1483	25,882	90,211	92.1950	1,004	818	1,503	10,180	86.2577
30	0.1317	42,515	111,222	113.3550	1,028	770	1,577	10,143	89.7999

Table 10 RMSE of Different metamodeling methods for Borehole model

Method	Computer runs	RMSE
Extended Gaussian Kriging	30	0.0044
SCAD penalized Kriging	30	0.7527
Ordinary Kriging	30	0.6587
SCAD penalized spline	30	0.3335
Quadratic regression	30	0.5077
B-spline	32	2.1095

for nonsmooth metamodeling tasks; considering appropriate Bayesian priors on the correlation parameters for robust parameter estimation; considering Kriging metamodeling for stochastic computer experiments, and so on. Besides, one particular problem deserves further research, i.e., improving the computational efficiency of estimating the correlation parameters in extended Gaussian Kriging as the sample size of computer runs is large or moderate. Research on these topics is currently ongoing and will be reported elsewhere.

Acknowledgments Many thanks are given to the two anonymous reviewers for their pertinent and helpful suggestions. The research of Ma was supported by the Natural Science Foundation of China under Grant 70931002, and the work of Wei was supported by the National High Technology Research and Development Plan of China under Grant 2007AA12E100.

References

1. Wang GG, Shan S (2007) Review of metamodeling techniques in support of engineering design optimization. ASME transactions. J Mech Des 45(6):1208–1221
2. Li R, Sudjianto A (2005) Analysis of computer experiments using penalized likelihood in Gaussian Kriging models. Technometrics 47:111–120
3. Fang KT, Li R, Sudjianto A (2006) Design and modeling for computer experiments. CRC Press, New York
4. Santner TJ, Willianms BJ, Notz WI (2003) The design and analysis of computer experiment. Springer, Berlin
5. Simpson TW, Peplinski J, Koch PN, Allen JK (2001) Metamodels for computer-based engineering design: survey and recommendations. Eng Comput 17(2):129–150
6. Sacks J, Welch WJ, Mitchell TJ, Wynn HP (1989) Design and analysis of computer experiments. Stat Sci 4(4):409–435
7. Sacks J, Schiller SB, Welch WJ (1989) Designs for computer experiments. Technometrics 31(1):41–47
8. Kleijnen JPC (2005) An overview of the design and analysis of simulation experiments for sensitivity analysis. Eur J Oper Res 164(2):287–300
9. Wackernagel H (2002) Multivariate geostatistics. Springer, New York
10. Currin C, Mitchell TJ, Morris MD, Ylvisaker D (1991) Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. J Amer Statist Assoc 86:953–963
11. Welch WJ, Buck RJ, Sacks J, Wynn HP, Mitchell TJ, Morris MD (1992) Screening, predicting, and computer experiments. Technometrics 34:15–25
12. Simpson TW, Mauery TM, Korte JJ, Mistree F (2001) Kriging metamodels for global approximation in simulation-based multidisciplinary design optimization. AIAA J 39(12):2233–2241
13. Martin JD, Simpson TW (2004) On using Kriging models as probabilistic models in design. SAE Trans J Mater Manuf 5:129–139
14. Martin JD, Simpson TW (2005) On the use of Kriging models to approximate deterministic computer models. AIAA J 43:853–863
15. Giunta AA (1997) Aircraft multidisciplinary design optimization using design of experiments theory and response surface modeling. Dissertation, Virginia Polytechnic and State University, Blacksburg
16. Booker AJ, Conn AR, Dennis JE, Frank PD, Serafini D, Torczon V, Trosset M (1996) Multi-level design optimization: a Boeing/IBM/Rice collaborative project. The Boeing Company, Seattle
17. Cappelleri DJ, Frecker MI, Simpson TW, Snyder A (2002) Design of a PZT bimorph actuator using a metamodel-based approach. ASME J Mech Des 124:354–357
18. Joseph VR, Hung Y, Sudjianto A (2008) Blind Kriging: a new method for developing metamodels. ASME J Mech Des 130: 031102-1-8
19. Linkletter CD, Bingham D, Hengartner N, Higdon D, Ye KQ (2006) Variable selection for Gaussian process models in computer experiments. Technometrics 48:478–490
20. Joseph VR (2006) Limit Kriging. Technometrics 48:458–466
21. Jin R, Chen W, Simpson TW (2001) Comparative studies of metamodeling techniques under multiple modeling criteria. Struct Multidiscip Optim 23(1):1–13
22. Koehler JR, Owen AB (2003) Computer experiments. In: Ghosh S, Rao CR (eds) Handbook of Statistics. Elsevier Science, New York, pp 261–308
23. McKay MD, Bechman RJ, Conover WJ (1979) A Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 21(2):239–245
24. Park JS (1994) Optimal Latin-hypercube designs for computer experiments. J Stat Plan Inference 39:95–111
25. Fang KT, Lin DKJ, Winker P, Zhang Y (2000) Uniform design: theory and application. Technometrics 39(3):237–248
26. Kalagnanam JR, Diwekar UM (1997) An efficient sampling technique for off-line quality control. Technometrics 39(3):308–319
27. Owen A (1992) Orthogonal arrays for computer experiments, integration, and visualization. Statistica Sinica 2:439–452
28. Hedayat AS, Sloane NJA, Stufken J (1999) Orthogonal arrays: theory and applications. Springer, New York
29. Simpson TW, Lin DKJ, Chen W (2001) Sampling strategies for computer experiments: design and analysis. Int J Reliab Appl 2(3):209–240
30. Montgomery DC (2001) Design and analysis of experiments. Wiley, New York
31. Fang HB, Horstemeyer MF (2006) Global response approximation with radial basis functions. Eng Optim 38(4):407–424
32. Fang HB, Rais-Rohani M, Horstemeyer MF (2004) Multiobjective crashworthiness optimization with radial basis functions. In: Proceedings of the 10th AIAAISSMO Multidisciplinary Analysis and Optimization Conference, Paper No. AIAA-2004-4487, Albany
33. Powell MJD (1987) Radial basis functions for multivariate interpolation: a review. In: Mason JC, Cox MG (eds) Algorithm for approximation. Clarendon Press, Oxford, pp 143–167
34. Hardy RL (1971) Multiquadratic equations of topography and other irregular surfaces. J Geophys 76:1905–1915
35. Hussain MF, Barton RR, Joshi SB (2002) Metamodeling: radial basis functions, versus polynomials. Eur J Oper Res 138(1):142–154
36. Krishnamurthy T (2003) Response surface approximation with augmented and compactly supported radial basis functions. In: Proceedings of the 44th AIAAASMEASCEAHSASC Structures, Structural Dynamics, and Materials Conference, Paper No. AIAA-2003-1748, Norfolk
37. Mullur AA, Messac A (2005) Extended radial basis functions: more flexible and effective meta- modeling. AIAA J 43(6):1306–1315
38. Mullur AA, Messac A (2006) Metamodeling using extended radial basis functions: a comparative approach. Eng Comput 21:203–217
39. Babacan SD, Molina R, Katsaggelos AK (2010) Bayesian compressive sensing using Laplace priors. IEEE Trans Image Process 19(1):53–63
40. Yi NJ, Xu SZ (2008) Bayesian LASSO for quantitative trait loci mapping. Genetics 179:1045–1055
41. Park T, Casella G (2008) The Bayesian Lasso. J Am Stat Assoc 103(482):681–686
42. Tipping ME (2001) Sparse Bayesian learning and the relevance vector machine. J Mach Learn Res 1:211–244
43. Ji S, Xue Y, Carin L (2008) Bayesian compressive sensing. IEEE Trans Signal Process 56(6):2346–2356
44. Ji S, Dunson D, Carin L (2009) Multi-task compressive sensing. IEEE Trans Signal Process 57(1):92–106
45. Figueiredo M (2003) Adaptive sparseness for supervised learning. IEEE Trans Pattern Anal Mach Intell 25(9):1150–1159
46. Gano SE, Renaud JE, Martin JD, Simpson TW (2005) Update strategies for Kriging models for use in variable fidelity optimization. In: Proceedings of 1st AIAA Multidisciplinary Design Optimization Specialist Conference, AIAA, Austin
47. Lophaven SN, Nielsen HB, SØndergaard J (2002) DACE: a Matlab Kriging toolbox. Technical Report, IMM-REP-2002-12, Technical University of Denmark, Denmark
48. Tipping M, Faul A (2003) Fast marginal likelihood maximisation for sparse Bayesian models. In Bishop C, Frey B (eds), In:

Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics, Key West
49. Hoffman RM, Sudjianto A, Du X, Stout J (2003) Robust piston design and optimization using piston secondary motion analysis. SAE Transactions SAE Paper 2003-01-0148
50. Hazime RM, Dropps SH, Anderson DH, Ali MY (2003) Transient non-linear fea and tmf life estimates of cast exhaust manifolds. Society of Automotive Engineers SAE 2003-01-0918
51. Fang KT, Lin DKJ (2003) Uniform designs and their application in industry. Handbook on Statistics in industry 22:131–170
52. Fang KT, Ho WM, Xu ZQ (2000) Case studies of computer experiments with uniform design, Technical Report. Hong Kong Baptist University, Hong Kong
53. Li R (2002) Model selection for analysis of uniform design and computer experiment. Int J Reliab Qual Saf Eng 9(4):367–382
54. Friedman JH (1991) Multivariate adaptive regressive splines. The Annal Stat 19(1):1–67
55. Hajela P, Berke L (1993) Neural networks in structure analysis and design: an overview. 4th AIAA/USAF/NASA/OAI Symposium on Multidisciplinary Analysis and Optimization, September 21–23. Cleveland, OH
56. Hong J, Dagli CH, Ragsdell KM (1994) Artificial neural network and the Taguchi method application for robust Wheatstone bridge design. In: Gilmore BJ, Hoeltzel DA et al (eds) Advances in Design Automation, Minneapolis Sept 11–14, 69(2):37–41
57. Clarke SM, Griebsch JH, Simpson TW (2005) Analysis of support vector regression for approximation of complex engineering analyses. Trans ASME J Mech Des 127(6):1077–1087