International Journal for Uncertainty Quantification, 4 (3): 205–223 (2014)

IMPROVEMENTS TO GRADIENT-ENHANCED KRIGING USING A BAYESIAN INTERPRETATION

Jouke H. S. de Baar, Richard P. Dwight, & Hester Bijl*

TU Delft, Kluyverweg 1 (10.18), 2629 HS Delft, The Netherlands

Original Manuscript Submitted: 12/17/2012; Final Draft Received: 06/24/2013

Cokriging is a flexible tool for constructing surrogate models on the outputs of computer models. It can readily incorporate gradient information, in which form it is named gradient-enhanced Kriging (GEK), and promises accurate surrogate models in >10 dimensions with a moderate number of sample locations for sufficiently smooth responses. However, GEK suffers from several problems: poor robustness and ill-conditionedness of the surface. Furthermore it is unclear how to account for errors in gradients, which are typically larger than errors in values. In this work we derive GEK using Bayes' Theorem, which gives an useful interpretation of the method, allowing construction of a gradient-error contribution. The Bayesian interpretation suggests the "observation error" as a proxy for errors in the output of the computer model. From this point we derive analytic estimates of robustness of the method, which can easily be used to compute upper bounds on the correlation range and lower bounds on the observation error. We thus see that by including the observation error, treatment of errors and robustness go hand in hand. The resulting GEK method is applied to uncertainty quantification for two test problems.

KEY WORDS: *Gaussian random fields, maximum likelihood, fluid mechanics*

1. INTRODUCTION

The goal of surrogate modeling is to construct an accurate and cheap approximation of the output of an expensive computer model, over a large parameter space, using as few observations of the full model as possible. Surrogate models are useful for a variety of applications, e.g., global optimization, uncertainty quantification, data assimilation, and model updating. The availability of derivatives of the output with respect to the parameters has the potential to reduce the cost of building surrogates considerably [1]. Adjoint versions of computer models, which return derivatives of a single scalar output with respect to any number of input parameters at constant cost [2], make the use of gradient information especially attractive for high-dimensional parameter spaces [3–6].

A promising framework for surrogate modeling using gradients is gradient-enhanced Kriging (GEK); a special case of Cokriging in which the covariables are the derivatives of the primary variable, and the gradient relationship is established using a specific form of the variogram.[1] While functional, GEK has several issues that can make it unreliable in practice: (1) It is not clear how to include gradient errors in GEK, that is, how to include the intrinsic errors of the computer simulations; (2) the numerical positive-definiteness of the gain influences the robustness of the analysis; and (3) the computational cost of a GEK analysis can be significant.

Using a Bayesian formulation of GEK, we propose solutions to the first two problems. The third problem is mainly due to the computational cost of estimating the hyperparameters [8], however this is beyond the scope of the present work—it has been treated in related work [9, 10] and is the subject of future research. Let us have a closer look at the first two problems:

First, gradients from computer models have potentially high levels of error [3, 4]. If this error is not accounted for when building the surrogate, an oscillatory response results. Whereas Kriging can regress observed values, until now

[1] Both Kriging and radial basis functions can be considered as special cases of support vector regression [7].

*Correspond to Jouke H. S. de Baar, E-mail: j.h.s.debaar@tudelft.nl, URL: http://www.lr.tudelft.nl/

2152–5080/14/$35.00

it has been unclear how the errors in the observed gradients should be modeled in GEK. Our contribution here is to derive GEK in a Bayesian framework, given which the gradient error is seen to be an "observation error" occurring in the observation-likelihood term, and therefore is appropriately modeled using the "likelihood." Given an estimate of gradient error, this allows us to build surrogate models with variable-quality gradients (see Section 2).

The second serious issue with GEK is robustness. A critical step in the algorithm is the inversion of the gain matrix generated from the variogram (denoted $\boldsymbol{A} \overset{\text{def}}{=} \boldsymbol{R} + \boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T$ in the following), for which a Cholesky factorization is commonly used. This operation is succeptable to floating-point round-off errors, and regularly fails when $\boldsymbol{A}$ has small eigenvalues. We develop a new analytic estimate for the smallest eigenvalue of $\boldsymbol{A}$ in the case of GEK, and by requiring that this estimate is greater than the machine epsilon $\varepsilon_{\text{machine}}$, we obtain constraints on the variogram correlation range, and the observation error. For example, we conclude that if the observation error is greater than roughly $\sqrt{\varepsilon_{\text{machine}}}$, then $\boldsymbol{A}$ is numerically invertable for all correlation ranges (see Section 3).

With insight into these two problems, GEK can be implemented in a robust manner and becomes a practical tool. It is demonstrated as such for two uncertainty quantification applications.

A Matlab implementation of the GEK method described in this paper is available online at our web site: aerodynamics.lr.tudelft.nl/∼bayesiancomputing.

1.1 Background of Kriging

Kriging was proposed independently in the field of geology by [11] and in the field of meteorology by [12], as a means of spatially interpolating sparsely sampled observations. It is equally well-suited to interpolation in general parameter spaces, in which context it is useful for optimization, and as a general-purpose surrogate model. A detailed account on the origins of Kriging is given by [13], while a complete overview is provided by [14] and [15]. A popular implementation of Kriging is provided by the Matlab Toolbox DACE [16]. A recent development for extremely large data sets is fixed rank Kriging [17].

In the present work we advocate regressing the output of computer models, which are often noisy, especcialy in the case of gradient information. In general we claim that primary interest typically lies, not in a perfect surrogate of an imperfect computer model, but in prediction (with error estimates) of a real process. Although usually the output of a computer code might well be considered free of noise [18], this approach becomes problematic when including gradient information: The gradient information might be considerably noisy due to off-design conditions, "difficult" physics like shocks regions [4], or approximation techniques [3]. Therefore the question arises how to treat noisy data in Kriging. In Kriging's original application as a predictor of mineral ore grades, ores such as gold would be found in small nuggets: a "nugget effect" appeared as a granularity-scale discontinuity within a homogeneous field [11].[2] This nugget effect is the result of the superposition of several structures of different scales known as a "gigogne structure," such that: *"In a general way, all nugget effects are reflections of a transition structure, the dimensions of which are considerably exceeded by the working scale: the details and the characteristics of this prior [microscale] structure have long since ceased to be perceptible, and the larger scale has barely preserved a single parameter—the nugget constant—which gives a kind of overall undifferentiated measure of the 'intensity' of this hidden structure."* [19]. This concept of a gigogne structure in the variogram is present in the work of Khintchin, Wold, and Kolgomorov [20–22]. Accordingly, Cressie considers a nugget constant $c_0 > 0$ due to microscale variations, however, he also points out that for an experimental variogram where *"... continuity is expected at the microscale, the only reason for $c_0 > 0$ is measurement error"* [13]. Cressie's distinction between nonsmooth fields with microscale variations and smooth fields with observation errors is crucial. The presence of observation errors has been discussed in the work of Wiener, Gandin, Thompson, Henderson, and recently by Wikle and Berliner [12, 23–26]. The modifications of Kriging suitable for modeling experimental noise are also suitable for modeling discretization error and modeling discrepancy. Although these errors are not random like experimental noise (an example of aleatory uncertainty), they are unknown (epistemic uncertainty), and our lack of knowledge can be represented as a probability distribution.

[2]Confusingly, when matrix regularization is required for numerical stability, the small diagonal added to the gain matrix is also known as a "nugget."

Gradient information can be incorporated into the method in two ways: (1) assuming gradient observations and value observations are colocated, one can construct new artificial value observations by linear extrapolation over a distance δ_ξ. This is usually done in each coordinate direction at each observation. Standard Kriging is then applied to this extended dataset. This approach—known as indirect GEK—is simple, but is sensitive to the choice of δ_ξ, results in very ill-conditioned gain matrices for small δ_ξ, and cannot consider observation errors [4, 27, 28]. Therefore we consider the second approach: (2) including the gradient as a covariable in Cokriging (see Section 2.2).

Combining error estimates and gradient information in Cokriging is challenging, as the interpretation of several elements of Cokriging is unclear; see the discussion in [29] and [30]. A critical question is, what is the most appropriate way to represent observation errors? This question can be clarified by formulating the problem in a Bayesian framework. Several Bayesian derivations of Kriging have been made [31–36], and a particularly clear presentation is that of [26]. In this last reference we find a clear distinction between the variogram and the observation error: The variogram is the covariance of the Bayesian prior, while the observation errors are contained in the likelihood. As such, the observation errors are clearly not part of the variogram. With this simple distinction between prior and likelihood the Bayesian framework enables us to treat observation errors in gradient-enhanced Kriging (Section 2.2).

In Fig. 1 we have classified works on Kriging according to three criteria: the conventional statistical or Bayesian framework, the treatment of gradients, and the treatment of observation errors. To our knowledge there is no existing work that combines a treatment of errors and gradients. Although we presently consider the robustness of Cokriging, we note that the theoretical results developed for the strict positive-definiteness and condition number of $\boldsymbol{A}_c$ also apply to standard Kriging. Furthermore, an operation closely related to simple Kriging is the analysis step in a Kalman filter [26, 37], where "Kalman gain regularization" or "covariance regularization" can be necessary for matrix inversion [38]. Another related technique is kernel dimension reduction, where a "regularization term" is required to enable matrix inversion. Importantly [39] "are not aware of theoretically justified methods of choosing these parameters; this is an important open problem." We observe that in both cases the present approach of estimating the robustness analytically could readily be applied in order to improve understanding of these regularization methods.

1.2 Application: Efficient Uncertainty Quantification

Our motivation for the above-described developments of GEK is to be able to perform efficient uncertainty quantification. Parameters of physical problems almost always have some epistemic uncertainty associated with them—usually

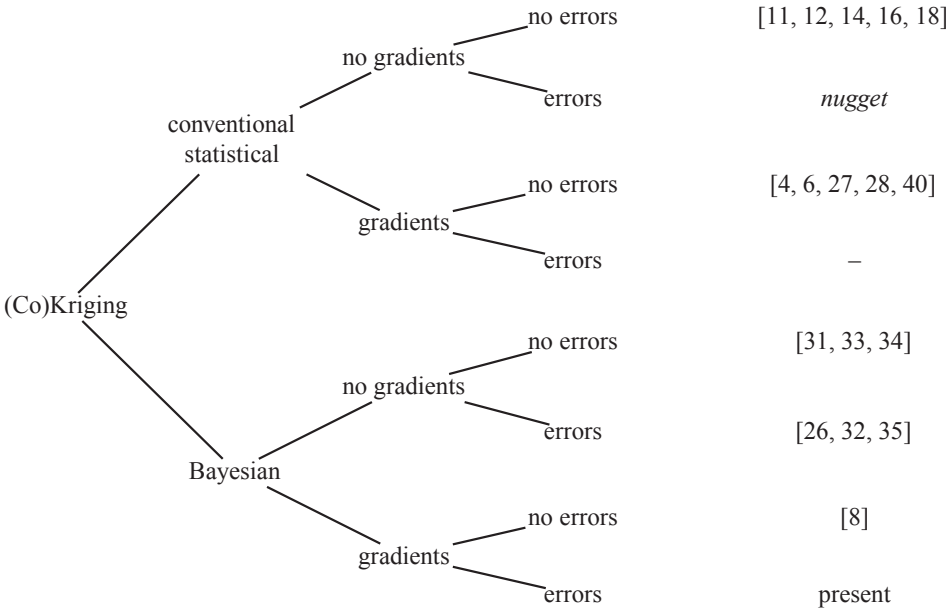


FIG. 1: References to literature on different Kriging and GEK techniques.

represented as a probability density function (pdf). It is desirable to know what the consequent uncertainty is in the model output, and to determine this with as few runs of the model as possible. Efficient existing approaches—such as polynomial chaos and probabilistic collocation—use high-order polynomial interpolations of the model response [40–44]. They are susceptible to oscillations for noisy model observations, and scale poorly to high dimensions. A GEK surrogate offers an alternative approach, using gradients from adjoint solves [1–3, 5, 45–47] to be efficient in high dimensions. Here, we show how GEK can be used in a way that incorporates the gradient error information.

2. DERIVATION OF KRIGING WITH GRADIENTS AND ERRORS

In *simple Kriging* we assume that the hyperparameters—the mean, variance, and correlation range contained in the prior—are known [14, 48]. As can be seen in Fig. 1, various Bayesian derivations of Kriging have been made, however, a particularly clear Bayesian derivation of Kriging has been made by [26]. We extend this derivation to *simple GEK*, where we express the relation between the values and gradients in the prior covariance matrix. We then extend the method to *ordinary GEK* by making maximum likelihood estimates of the hyperparameters [31].

Presently we consider a normally distributed likelihood. This is adequate if our information on an observation is limited to a value and standard deviation [49]. We assume that the prior is also normally distributed, and we find the parameters of this distribution from maximum likelihood estimates. In the case that one has more detailed information, this can of course result in a non-normally distributed likelihood and prior [50].

2.1 Conventional Derivation of GEK

According to [4, 6, 7, 27, 28] one can consider a Gaussian process:

$$y(\xi) = \mu + Z(\xi), \tag{1}$$

with mean μ and parameter ξ. The GEK predictor for this Gaussian process is given by

$$\hat{y}_c(\xi) = \mu_c + \boldsymbol{\psi}_c^T \boldsymbol{\Psi}_c^{-1}(\boldsymbol{y} - \mu_c), \tag{2}$$

where $\boldsymbol{\psi}_c$ and $\boldsymbol{\Psi}_c$ can be generated from a correlation function. More details of this approach and of the "indirect" GEK approach can be found in [4, 7, 27, 28].

In the conventional approach the observations are considered to be exact. However, the results of an expensive computer simulation often have significant errors. Typically, the errors in the gradients are larger than the errors in the values [3, 4]. This is not accounted for in the conventional GEK predictor.

2.2 Bayesian Derivation of One-Dimensional GEK

Along the lines of [26] we are interested in the n discrete values $\boldsymbol{x}$, for which we assume the normal prior distribution

$$\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{P}). \tag{3}$$

For simple Kriging, mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{P}$ are considered known. We observe $N = n - 1$ values $\boldsymbol{y}$, a subset of $\boldsymbol{x}$. The observations have the normal likelihood

$$\boldsymbol{y}|\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{H}\boldsymbol{x}, \boldsymbol{R}), \tag{4}$$

given a known observation matrix $\boldsymbol{H}$ (with $\boldsymbol{H}\boldsymbol{x}$ a subset of $\boldsymbol{x}$) and observation error covariance matrix $\boldsymbol{R}$. Applying Bayes' Theorem,

$$p(\boldsymbol{x}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})}{p(\boldsymbol{y})}, \tag{5}$$

the posterior mean is given by [26, 51–53]

$$E(\boldsymbol{x}|\boldsymbol{y}) = \boldsymbol{\mu} + \boldsymbol{K}(\boldsymbol{y} - \boldsymbol{H}\boldsymbol{\mu}), \tag{6}$$

while the posterior covariance is given by

$$\mathrm{var}(\boldsymbol{x}|\boldsymbol{y}) = (\boldsymbol{I} - \boldsymbol{K}\boldsymbol{H})\boldsymbol{P}, \tag{7}$$

where we have the gain matrix

$$\boldsymbol{K} \stackrel{\text{def}}{=} \boldsymbol{P}\boldsymbol{H}^T(\boldsymbol{R} + \boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T)^{-1}. \tag{8}$$

Then $E(x_n|\boldsymbol{y})$ and $\mathrm{var}(x_n|\boldsymbol{y})$ are the simple Kriging estimates for a value x_n given the observations $\boldsymbol{y}$ [14]. Note that to this point, the derivation is independent of the number of dimensions.

We now depart from [26] by including gradients and considering the one-dimensional case, i.e., each value $\boldsymbol{x}$ is supplemented with one gradient $\boldsymbol{x}_{;\xi}$, where the subscript $;\xi$ indicates a partial derivative with respect to a single parameter ξ (the number of parameters determines the dimensionality). In the case of 1-dimensional GEK we are now interested in the $2n$ compiled (notation: c) values $\boldsymbol{x}_c = [\boldsymbol{x}, \boldsymbol{x}_{;\xi}]$. Along the lines of Eqs. (3)–(8) we find the posterior mean:

$$E(\boldsymbol{x}_c|\boldsymbol{y}_c) = \boldsymbol{\mu}_c + \boldsymbol{K}_c(\boldsymbol{y}_c - \boldsymbol{H}_c\boldsymbol{\mu}_c), \tag{9}$$

while the posterior covariance is given by

$$\mathrm{var}(\boldsymbol{x}_c|\boldsymbol{y}_c) = (\boldsymbol{I} - \boldsymbol{K}_c\boldsymbol{H}_c)\boldsymbol{P}_c, \tag{10}$$

with gain

$$\boldsymbol{K}_c \stackrel{\text{def}}{=} \boldsymbol{P}_c\boldsymbol{H}_c^T(\boldsymbol{R}_c + \boldsymbol{H}_c\boldsymbol{P}_c\boldsymbol{H}_c^T)^{-1}. \tag{11}$$

Then $E(x_{c,n}|\boldsymbol{y}_c)$, $\mathrm{var}(x_{c,n}|\boldsymbol{y}_c)$, $E(x_{c,2n}|\boldsymbol{y}_c)$, and $\mathrm{var}(x_{c,2n}|\boldsymbol{y}_c)$ are the one-dimensional GEK predictors for the values and gradients, respectively.

2.2.1 The Error Covariance Matrix

The objective of this derivation was to include the value and gradient errors. The Bayesian perspective makes clear that these belong in the likelihood. Specifically, the matrix

$$\boldsymbol{R}_c = \begin{pmatrix} \epsilon_1^2 & & & & & & & \\ & \epsilon_2^2 & & & & & & \\ & & \ddots & & & 0 & & \\ & & & \epsilon_N^2 & & & & \\ & & & & \nu_1^2 & & & \\ & & & & & \nu_2^2 & & \\ & & 0 & & & & \ddots & \\ & & & & & & & \nu_N^2 \end{pmatrix}, \tag{12}$$

contains all individual random observation errors ϵ_i and ν_i, of the values and the gradients, respectively. The user can specify or optimize for these errors, as is illustrated in the test problems in Sections 4.1 and 4.2, respectively.

2.2.2 Approximate Covariance Function

Instead of defining the covariances through $f(\xi)$ we will use an approximate covariance function. Consider the lag $h_{ij} \stackrel{\text{def}}{=} |\xi_j - \xi_i|$ and estimate

$$\boldsymbol{P}_{ij} \stackrel{\text{def}}{=} \boldsymbol{P}(\xi_i, \xi_j) = r(h_{ij}). \tag{13}$$

In the following we apply the Gaussian covariance function:

$$r(h_{ij}) = \sigma^2 \exp\left(-\frac{h_{ij}^2}{2\theta^2}\right), \tag{14}$$

with correlation range θ and variance σ^2.

2.2.3 The Covariance between Values and Gradients

Up till now we have been treating the values and the gradients as independent quantities. We will now see what relation we would actually like them to have and express this relation in the prior covariance. In one-dimensional GEK, the prior covariance matrix

$$\boldsymbol{P}_c = \begin{pmatrix} \boldsymbol{P}^{00} & \boldsymbol{P}^{10} \\ \boldsymbol{P}^{01} & \boldsymbol{P}^{11} \end{pmatrix}, \tag{15}$$

contains the following submatrices: $\boldsymbol{P}^{00}$ is the covariance matrix of the values, $\boldsymbol{P}^{11}$ is the covariance matrix of the gradients, and $\boldsymbol{P}^{01}$ and $\boldsymbol{P}^{10}$ are the cross covariances (such that lower indices are for the different samples and upper indices are for the submatrices). Various derivations of these submatrices in a conventional framework can be found, see for example [6, 7, 27, 28]. More recently a concise derivation in the context of Gaussian random fields has been given [54], which fits directly in the Bayesian framework and gives

$$\begin{aligned} \boldsymbol{P}^{00}_{ij} &= r(h_{ij}), \\ \boldsymbol{P}^{10}_{ij} &= -\frac{\partial}{\partial h} r(h_{ij}), \\ \boldsymbol{P}^{01}_{ij} &= \frac{\partial}{\partial h} r(h_{ij}), \\ \boldsymbol{P}^{11}_{ij} &= -\frac{\partial^2}{\partial h^2} r(h_{ij}). \end{aligned} \tag{16}$$

With these approximations we can construct the prior covariance matrix, given in (15), which completes our Bayesian derivation of simple one-dimensional GEK.

2.3 Example: One-Dimensional GEK with Observation Errors

Consider the following simple test function:

$$\begin{aligned} x(\xi) &= \sin(\pi\xi), \\ x_{;\xi}(\xi) &= \pi\cos(\pi\xi). \end{aligned} \tag{17}$$

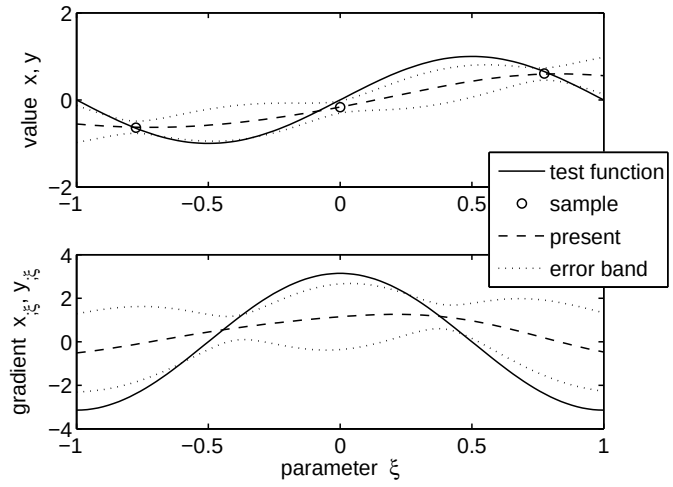
As can be seen in Fig. 1, various Bayesian derivations have been made, however, a particularly clear Bayesian derivation of Kriging has been made by [26]. We extend this derivation to *simple GEK*, where we express the relation between the values and gradients in the prior covariance matrix. We then extend the method to *ordinary GEK* by making maximum-likelihood estimates of the hyperparameters [31]. We consider the test function on $\xi \in [-1, 1]$. We assume that the prior mean $\mu = 0$, the prior variance $\sigma^2 = 0.5$, and the prior correlation range $\theta = 0.5$ are known. We sample the test function at three nodes, with an observation error of $\epsilon = 0.2$ and $\nu = 0.2$.

If we only sample the function values, we find the surrogate model shown in Fig. 2(a). Note that due to the observation error the error band—given by the square root of the diagonal entries of $\mathrm{var}(\boldsymbol{x}|\boldsymbol{y})$ (7)—is not zero at the sample locations. If we sample both the values and the gradients, we find the surrogate model in Fig. 2(b). We see that after including the gradient information the surrogate model resembles the function more closely, while the error band is reduced significantly.

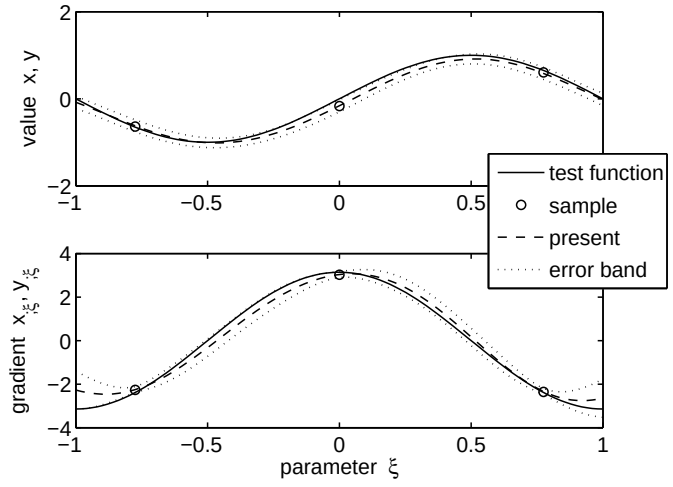
2.4 Extension to Multidimensional GEK

The extension to multidimensional GEK is straightforward. The GEK predictor equations (9)–(11) are in fact independent of the number of dimensions and remain unchanged.

However, for the m-dimensional case with m parameters $\xi_1, \xi_2, \ldots, \xi_m$, each value is supplemented with m gradients, which results in the $(m+1)N$ compiled observations $\boldsymbol{x}_c = [\boldsymbol{x}, \boldsymbol{x}_{;\xi_1}, \ldots, \boldsymbol{x}_{;\xi_m}]$.



(a)



(b)

FIG. 2: Simple Kriging and GEK surrogate models, obtained by sampling a test function. In (a) we only sample the values, while in (b) we sample both the values and the gradients.

The prior covariance matrix is now given by

$$\boldsymbol{P}_c = \begin{pmatrix} \boldsymbol{P}^{00} & \boldsymbol{P}^{10} & \dots & \boldsymbol{P}^{m0} \\ \boldsymbol{P}^{01} & \boldsymbol{P}^{11} & & \boldsymbol{P}^{m1} \\ \vdots & & \ddots & \vdots \\ \boldsymbol{P}^{0m} & \boldsymbol{P}^{1m} & \dots & \boldsymbol{P}^{mm} \end{pmatrix}, \tag{18}$$

with

$$\boldsymbol{P}_{ij}^{kl} = \frac{\partial^2}{\partial \xi_k \partial \xi_l} r(\xi_{k,j} - \xi_{l,i}). \tag{19}$$

2.5 Estimating the Hyperparameters

We will now extend our derivation to *ordinary GEK*. In this case we find estimate the values of the hyperparameters μ, σ, and θ conditional on the observations $\boldsymbol{y}_c$. Recall that the $N_c \times N_c$ matrix

$$\boldsymbol{A}_c \stackrel{\text{def}}{=} (\boldsymbol{R}_c + \boldsymbol{H}_c \boldsymbol{P}_c \boldsymbol{H}_c^T). \tag{20}$$

The standard approach to estimating the hyperparameters is to make a maximum likelihood estimate (MLE) [4, 31, 55–57]. With a prior belief $\theta \sim \mathcal{U}(\theta_{\min}, \theta_{\max})$, we maximize the log likelihood:

$$\ln p(\theta|\mu, \sigma, \boldsymbol{y}_c) = \frac{-N_c \ln(2\pi) - \ln|\boldsymbol{A}_c| - (\boldsymbol{y}_c - \boldsymbol{\mu}_c)^T \boldsymbol{A}_c^{-1} (\boldsymbol{y}_c - \boldsymbol{\mu}_c)}{2} + \text{constant}, \tag{21}$$

which is equivalent to minimizing

$$L(\theta) = \ln|\boldsymbol{A}_c| + (\boldsymbol{y}_c - \boldsymbol{\mu}_c)^T \boldsymbol{A}_c^{-1} (\boldsymbol{y}_c - \boldsymbol{\mu}_c), \tag{22}$$

with respect to $\theta \in [\theta_{\min}, \theta_{\max}]$. In the case of GEK the drift vector $\boldsymbol{f}_c$ has N entries equal to 1 followed by zeros for the remaining entries, furthermore the matrix $\boldsymbol{A}_0 = \sigma^{-2} \boldsymbol{A}_c$. Optimizing (22) for all hyperparameters can be unattractive for a small number of observations, therefore one can instead optimize the concentrated likelihood [7] or consider statistically normalized (i.e., zero drift) data and optimize the restricted likelihood [31, 58]. In Sections 4.1 and 4.2 we optimize the restricted likelihood; also, we consider a signal-to-noise ratio SNR $= \sigma^2/\epsilon^2$ instead of considering σ^2 and ϵ^2 separately, since in (8) and (11) it is only important to provide the correct relative error.

Due to the estimated hyperparameters we find some important changes in the posterior, as given in Eqs. (9) and (10). The expression for the posterior mean (9) remains unchanged, while the posterior variance (10) should be multiplied with the factor [57]

$$\frac{N_c - q}{N_c - q - 2} = \frac{N_c - 1}{N_c - 3}, \tag{23}$$

where for ordinary GEK the number of drift coefficients $q = 1$, as μ is a scalar. We observe that these alterations are quite significant for a small number of observations N_c. Note that these alterations are not accounted for in conventional statistical Kriging, which for small N leads to an underestimation of the variance of the prediction [36, 57].

3. ROBUSTNESS OF KRIGING AND GEK

It is well known that for large N the Kriging gain matrix becomes increasingly ill-conditioned [59, 60]. Here we quantify this effect and relate it to nonpositive definiteness of the gain matrix $\boldsymbol{A}$, which leads to inaccurate solutions of the linear system $\boldsymbol{A}^{-1}(\mathbf{y} - \boldsymbol{\mu})$.

Consider the $N \times N$ gain matrix:

$$\boldsymbol{A} \stackrel{\text{def}}{=} (\boldsymbol{R} + \boldsymbol{H}\boldsymbol{P}\boldsymbol{H}^T). \tag{24}$$

In each iteration of the optimization of (22) with respect to θ, we have to evaluate $\ln|\boldsymbol{A}_c|$ and $\boldsymbol{A}^{-1}(\boldsymbol{y} - \boldsymbol{\mu})$. We can reduce the cost of these evaluations by making the Cholesky factorization:

$$\boldsymbol{A} = \boldsymbol{U}^T \boldsymbol{U}, \tag{25}$$

with standard numerical routines (e.g. LAPACK). From this factorization we find

$$\ln|\boldsymbol{A}| = 2 \ln|\boldsymbol{U}|, \tag{26}$$

and solve the triangular linear systems

$$\begin{aligned} \boldsymbol{U}^T \boldsymbol{v} &= (\boldsymbol{y} - \boldsymbol{\mu}), \\ \boldsymbol{U}\boldsymbol{w} &= \boldsymbol{v}, \end{aligned} \tag{27}$$

to find

$$\boldsymbol{A}^{-1}(\boldsymbol{y} - \boldsymbol{\mu}) = \boldsymbol{w}. \tag{28}$$

This approach has a limitation: to make a Cholesky factorization, $\boldsymbol{A}_c$ has to be strictly positive-definite.

Although in the application of Kriging the condition number $\kappa(\boldsymbol{A})$ is studied frequently [59, 60], we consider it to provide a very pessimistic upper bound for the accuracy of the solution of (27), and we find that in most cases the effect of the condition number is subordinate to the requirement of $\boldsymbol{A}$ being strictly positive-definite. Therefore, we focus on developing an analytical estimate of the conditions under which $\boldsymbol{A}$ is strictly positive-definite. We will restrict ourselves to uniform sampling on a one-dimensional grid. Also, we will restrict ourselves to a white noise observation error in both the values and gradients.

3.1 Positive-Definiteness for Kriging

Since $\boldsymbol{A}$ is real and symmetric, we will find $\boldsymbol{A}$ to be strictly positive-definite if all of its eigenvalues λ_i are strictly positive. Note that although a matrix may be analytically strictly positive-definite, it is not necessarily possible to make a Cholesky factorization in finite precision arithmic. We are therefore interested in $\boldsymbol{A}$ being *numerically* strictly positive-definite, and require that all of its eigenvalues are larger than machine precision $\varepsilon_{\text{machine}}$.

We can approximate the discrete eigenvalue spectrum of $\boldsymbol{A}$ by the continuous spectral density function [59]. The *smallest* eigenvalue we can represent in this spectrum corresponds to the *largest* wave number we can represent. On a uniform grid with domain size L and spacing $\Delta\xi$ this is:

$$k_{\max} = \frac{\pi}{\Delta\xi} = \frac{\pi(N-1)}{L}, \tag{29}$$

which corresponds to the Nyquist frequency. Substituting $k_{\max}$ from [59] we find

$$\min\left[\Lambda(k)\right] = \frac{\epsilon^2 L}{2\pi(N-1)} + \frac{\theta\sigma^2}{\sqrt{2\pi}}\exp\left(-\frac{\pi^2\theta^2(N-1)^2}{2L^2}\right). \tag{30}$$

We see that the lowest eigenvalue is always strictly positive, hence $\boldsymbol{A}$ is always strictly positive-definite. However, recall that for a Cholesky factorization we require that the smallest eigenvalue is larger than machine precision:

$$\frac{\epsilon^2 L}{2\pi(N-1)} + \frac{\theta\sigma^2}{\sqrt{2\pi}}\exp\left(-\frac{\pi^2\theta^2(N-1)^2}{2L^2}\right) > \varepsilon_{\text{machine}}. \tag{31}$$

In Fig. 3(a) we plot the effect of these conditions, the numerical strict positive-definiteness of $\boldsymbol{A}$. This is an example for $N = 32$ and $\sigma = 1$. We proceed to derive two sufficient conditions for numerical positive definiteness from the two terms in (31).

First, we expect $\boldsymbol{A}$ to be numerically strictly positive-definite if

$$\epsilon^2 > \frac{2\pi(N-1)}{L}\ \varepsilon_{\text{machine}}, \tag{32}$$

which is in fact the limit we encounter when moving from right to left in the upper part of Fig. 3(a), for a relatively large correlation range of $\theta/L = 1$. In Fig. 3(b) we compare this estimate with numerical results. Note the distinct increase in both the estimate and results when we move from double to single precision. This is another clear indication that we should indeed consider the limit (31) as a numerical issue. Interestingly, this estimate resembles the correlation matrix regularization term $\mu_{\text{DACE}}\boldsymbol{I}$ which is applied in DACE [16]:

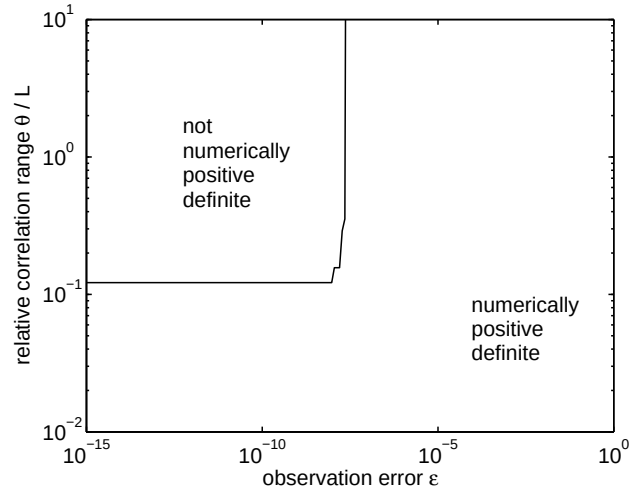
$$\mu_{\text{DACE}} = (10+N)\ \varepsilon_{\text{machine}}, \tag{33}$$

where in DACE $L \approx 1$ due to normalization. However, note that where observation errors and matrix regularization can both enhance robustness, the motivation for these two approaches is quite different.

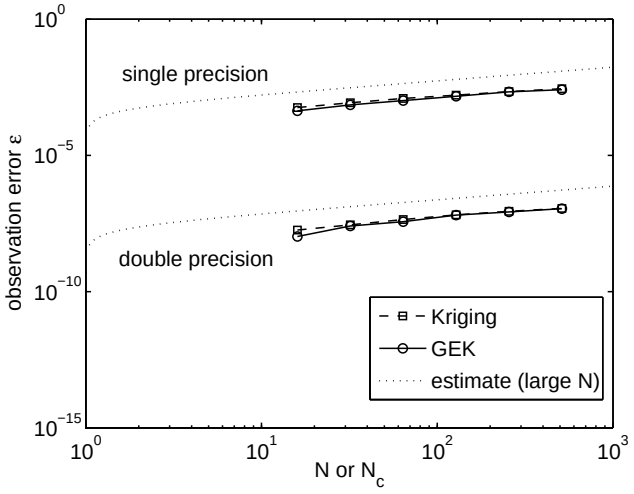
Second, we expect $\boldsymbol{A}$ to be numerically strictly positive-definite if

$$\frac{\theta\sigma^2}{\sqrt{2\pi}}\exp\left(-\frac{\pi^2\theta^2(N-1)^2}{2L^2}\right) > \varepsilon_{\text{machine}}. \tag{34}$$

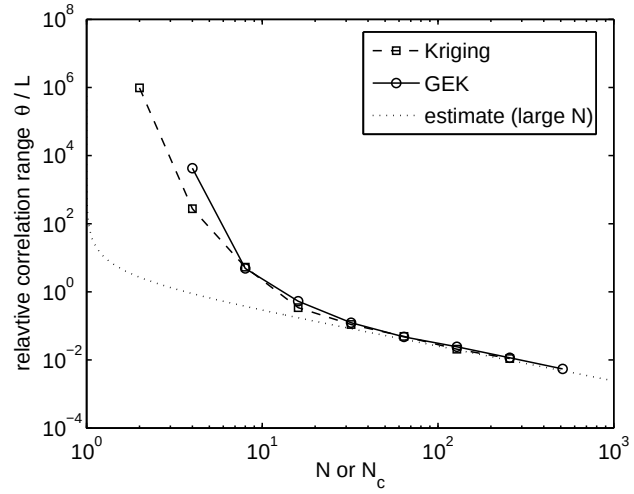
This is the limit we encounter when moving upward in the left part of Fig. 3(a), for a observation error of $\epsilon = 0$. In Fig. 3(c) we compare this estimate with numerical results.



(a)



(b)



(c)

FIG. 3: (a) Numerical strict positive definiteness of the matrix $\boldsymbol{A}$, (b) the minimum observation error for a relatively large correlation range $\theta/L = 1$, and (c) the maximum correlation range for zero observation error.

3.2 Positive-Definiteness for GEK

One can apply the estimates (32) and (34) to GEK, as long as one takes into account that the derivative information increases the Nyquist frequency [23], as is also illustrated in Fig. 2. Therefore, one N should be replaced by N_c, such that

$$\epsilon^2 > \frac{\sqrt{2\pi}(N_c - 1)}{L}\ \varepsilon_{\text{machine}}, \tag{35}$$

and

$$\frac{\theta\sigma^2}{\sqrt{2\pi}}\exp\left(-\frac{\pi^2\theta^2(N_c-1)^2}{2L^2}\right) > \varepsilon_{\text{machine}}. \tag{36}$$

The accuracy of these estimates is illustrated in Figs. 3(a) and 3(b).

It is now straightforward to arrive at a proposal for a robust implementation. In most cases, if an observation error is included, the actual observation error (or "noise") will be sufficiently large to satisfy (35). In special cases with high-quality data, where neither (35) nor (36) is satisfied, one can ensure robustness by adding a synthetic error given by (35) during the analysis.

4. UNCERTAINTY QUANTIFICATION FOR TWO TEST PROBLEMS

In Section 4.1 we consider the heat advection diffusion equation to illustrate how GEK mitigates the curse of dimensionality as well as the effect of noisy gradients. In Section 4.2 we consider turbulent flow over an airfoil to illustrate the MLE estimation of the amount of gradient noise.

The cost of the dual solve (relative to the primal solve) in terms of CPU time is found to depend on the type of problem. Therefore, we consider it more informative to provide the cost in terms of number of solves. In both test problems, we report the cost as the total number of solves, so for example a case with 10 samples would lead to a cost of 10 solves in the case of Kriging and to a cost of 20 solves in the case of GEK.

4.1 The Heat Advection Diffusion Equation

Consider the one-dimensional heat advection diffusion equation:

$$T_{;t} + \text{Pe}\ T_{;z} - T_{;z;z} = 0, \tag{37}$$

for the temperature T at the spatial position $z \in [0, 1]$ and time t. The boundary conditions and initial condition are

$$\begin{aligned} T(z = 0, t) &= 1, \\ T(z = 1, t) &= 0, \\ T(z, t = 0) &= 1 - z, \end{aligned} \tag{38}$$

while the Péclet number is given by

$$\text{Pe} = \frac{Z\ w}{\alpha}, \tag{39}$$

with spatial domain size Z, convecting velocity w, and thermal diffusivity α. The Péclet number is treated as the piecewise linear interpolation of a finite number of uncorrelated random parameters, each with uniform distribution $\text{Pe}_i \sim \mathcal{U}(-4, 18)$. The quantity of interest is the steady-state midpoint temperature $T_{z=0.5}$. The random input distribution of Pe_i results in a probability density function (pdf) for the output $T_{z=0.5}$: presently our aim is to compute the skewness of this pdf with an accuracy of 1%, with a 10^5 sample Monte Carlo simulation as a reference.

4.1.1 Discretization for Primal and Dual Solve

We discretize Eq. (37) with a second-order central difference scheme with first-order implicit time stepping, where we note that time accuracy is not an issue here. As an example, the solutions for three different Péclet numbers on a spatial grid size of $N_z = 1024$ are shown in Fig. 4(a).

Apart from the temperature $T_{0.5}$ we compute the gradients $T_{0.5;\mathrm{Pe}_i}$ with respect to the Péclet numbers Pe_i. We compute the gradients from an adjoint method with complex step derivatives [2, 4, 45–47]. This requires a dual solve at approximately the same cost as the primal solve. The spatial convergence of both the value and gradient is illustrated in Fig. 4(b). Already for small grids we see that we are in the exponential range, with second-order convergence.

4.1.2 Curse of Dimensionality

The number of solves required to compute the skewness of the output pdf with the target accuracy of 1% increases rapidly when we increase the number of random variables, an effect known as the curse of dimensionality. We expect that this curse of dimensionality will be mitigated by the use of adjoint-based gradient information. However, first we note that we observe substantial variability in the surrogates for both Kriging and GEK with respect to the choice of Latin-hypercube sample points (LHS). To account for this variability, and extract meaningful trends, we construct response surfaces based on 100 different LHS designs.

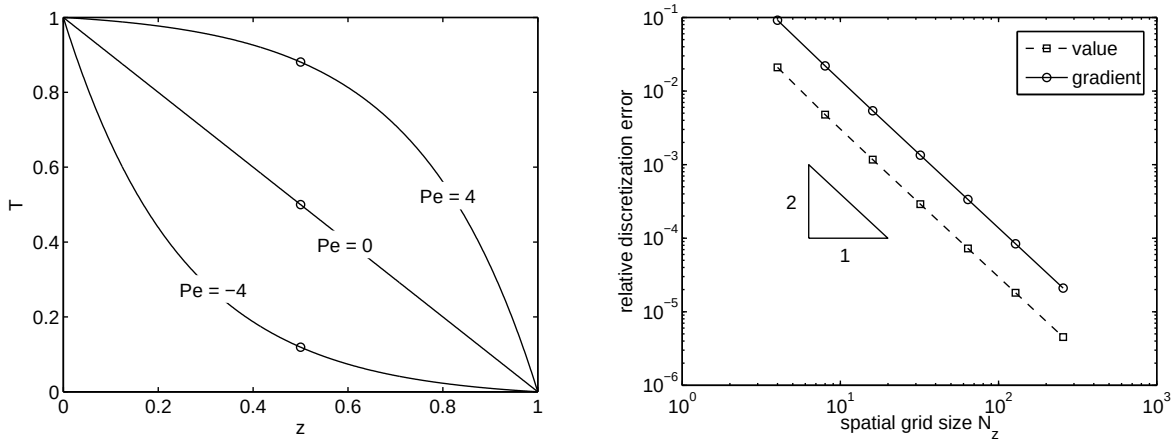
For an increasing number of random variables, the cost for different LHS designs is summarized by the 25%, 50%, and 75% quantiles plotted in Fig. 5. For this tests case, generally, GEK is found to be cheaper (i.e., require less solves) than Kriging. This effect becomes stronger when the number of random variables increases: for four random variables, GEK is roughly twice as cheap as Kriging, using only some 30 instead of 60 solves.

4.1.3 Effect of Gradient Noise

A problem with GEK is that the gradient information can be quite noisy, causing oscillations in the surrogate model. The amount of gradient noise can be quantified by the gradient SNR:

$$\mathrm{SNR} = \frac{\mathrm{var}(\mathbf{y}_{;\xi})}{\boldsymbol{\nu}^2}. \tag{40}$$

For the case of three random variables, we add Gaussian noise to the gradients for different SNR levels. The results, again summarized by the 25%, 50%, and 75% quantiles for different LHS designs, are shown in Fig. 6. In Fig. 6,



(a) (b)

FIG. 4: (a) Solutions of the one-dimensional heat equation for different Péclet numbers, circles indicate $T_{0.5}$, and (b) spatial convergence of the value $T_{0.5}$ and a gradient $T_{0.5;\mathrm{Pe}}$.

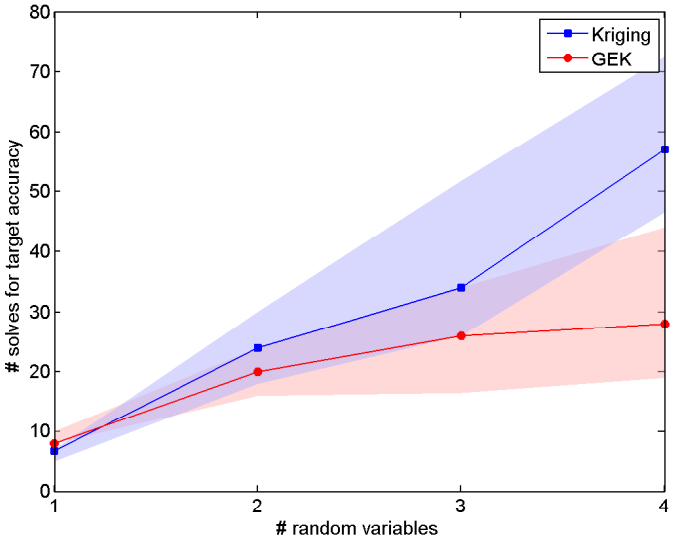


FIG. 5: The number of solves required to compute the skewness with target accuracy, for an increasing number of random variables. The results for Kriging and GEK are summarized by the 25%, 50%, and 75% quantiles for 100 different LHS-designs.

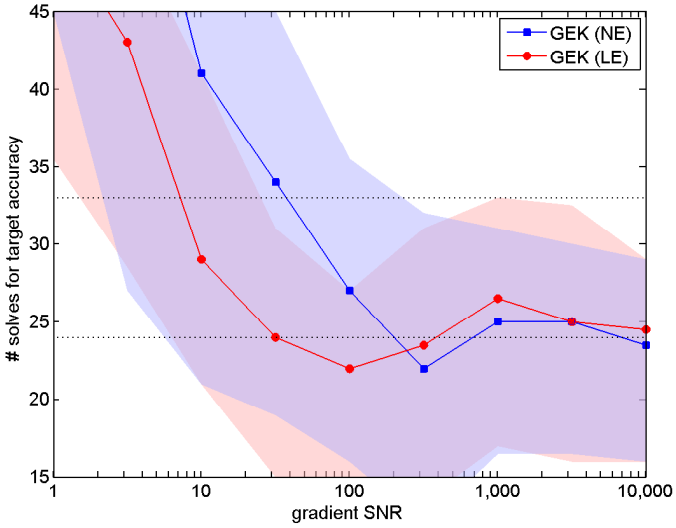


FIG. 6: The number of solves required to compute the skewness with target accuracy, for an increasing gradient SNR. The results for Kriging and GEK are summarized by the 25%, 50%, and 75% quantiles for 50 different LHS-designs. The upper dotted line is the cost of Kriging, the lower dotted line is the cost of GEK using noise-free gradients.

moving from right to left, we find different stages. At the far right, for roughly SNR > 100, we are considering very accurate gradient information. There is no need to take gradient noise into account, and therefore both GEK-NE (i.e., no estimated SNR) and GEK-LE (i.e., locally estimated SNR) are as cheap as GEK with noise-free gradients (indicated by the lower dotted line). Then, at the intermediate stage of roughly $10 <$ SNR < 100, we find that the cost of GEK-NE starts to increase as the noisy gradient information starts to deteriorate the surrogate: around a SNR of 50, the GEK-NE is as expensive as Kriging (indicated by the upper dotted line). In this same $10 <$ SNR < 100 stage, we

find that GEK-LE can perform at a significantly lower cost, as it can efficiently regress the noisy gradient information by incorporating the noise level. Note that a gradient SNR of 100 (i.e., a 10% error) is typical for complex applications, see for example [3] Finally, at the stage of SNR < 10, we find that for extremely inaccurate gradient information, the cost of both GEK-NE and GEK-LE can exceed the cost of Kriging. At such low SNR, the incorporation of the noise level fails to result in a proper regression of the gradient information, although it is not clear why GEK-LE becomes more expensive than Kriging. Even for these extremely low SNRs, GEK-LE is generally still cheaper than GEK-NE.

In the above case we knew the gradient SNR, which is not always the case. In the following section we illustrate how we can estimate a unknown gradient SNR from the data.

4.2 Turbulent Flow around a DU96–W–180 Airfoil

In the second test problem, we consider the subsonic turbulent flow over a DU96–W–180 wind turbine airfoil. Details of the airfoil are given in [61]. The quantity of interest is the lift-to-drag ratio C_l/C_d. The angle of attack is 1.45 degrees, while the Reynold's number is $\mathrm{Re} = 2 \times 10^6$. The computational two-dimensional grid contains 4×10^5 cells, while the average wall-$y^+ = 38.4$. The lift and drag are computed in OpenFOAM, using the k-ϵ turbulence model [62]. Figure 7 illustrates the turbulent kinetic energy of the flow around the airfoil.

We consider three random parameters in the k-ϵ turbulence model: C_1, C_2, and C_μ. We select these three random parameters for uncertainty quantification (UQ), because in a sensitivity analysis they showed the most pronounced effect on the simulation results. The probability density distributions for these random parameters are found from an uncertainty analysis of the k-ϵ turbulence model, and are shown in Fig. 8. We use a LHS design to sample the random parameter space, after which we create Kriging and GEK surrogates that we use to find a cumulative density function (cdf) for C_l/C_d.

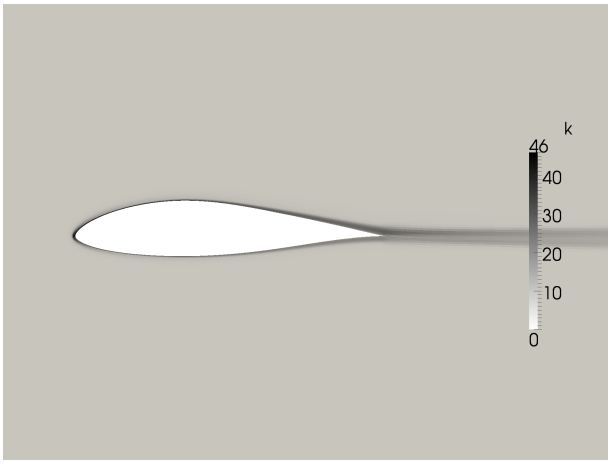


FIG. 7: Turbulent kinetic energy around a DU96–W–180 airfoil, computed in OpenFOAM.

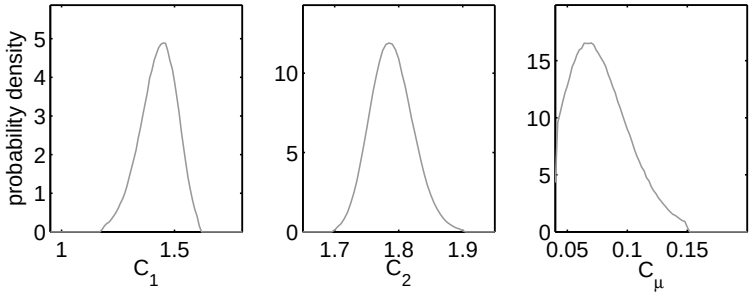


FIG. 8: Probability densities for three uncertain parameters.

In order to build the surrogates, we use a MLE to optimize for four hyperparameters simultaneously: the correlation ranges θ_1, θ_2, θ_μ, and the gradient SNR. The gradient noise level is assumed to scale roughly with the (locally dependent) solver residual. Note that we fix the (primary) value SNR to 10^8, reflecting a high-quality simulation result with a relative error that corresponds to the solver residual. The result of such a MLE optimization for a case with 30 solves is illustrated in Fig. 9. Clearly, the MLE does not only give a suitable result for the correlation ranges, but also for the gradient SNR. This illustrates that, if sufficient data are available, we can actually optimize for the correlation ranges and the gradient SNR simultaneously.

The final objective of this UQ analysis is to find an accurate cdf for the lift-to-drag ratio at low cost. Figure 10 shows the resulting cdf's for three different techniques—Kriging, GEK-NE, and GEK-LE—for an increasing number of solves; GEK-NE does not take the estimated SNR into account, GEK-LE takes the MLE optimized (local) error estimate into account. The black cdf is the reference cdf, obtained from 224 samples. The gray cdf's are the resulting cfd's for 10 different LHS designs. From Fig. 10 we see that for 10 samples, the cdfs obtained from Kriging, GEK-NE, and GEK-NE are inaccurate. For 14 samples, the Kriging cdf's are still inaccurate and the gradient information added in the case of GEK-NE does not improve the results due to gradient noise; however, when we do take the gradient noise into account GEK-LE improves the accuracy of the cdf's. A similar effect can be seen for 30 solves, although here all three methods can be considered reasonably accurate.

In this test problem, we found that for the turbulent flow around the DU96–W–180 airfoil we can use a MLE to optimize simulanously for the correlation ranges and the gradient-SNR. As a result, the UQ analysis becomes more efficient: already for the case of 14 samples the resulting cdf's show an improved accuracy.

5. CONCLUSION

We make a Bayesian derivation of gradient-enhanced Kriging (GEK), in which we include individual error information. In the Bayesian framework, the variogram and the relation between values and gradients is contained in the prior, while the observation error is contained in the likelihood. The observations serve to update the Kriging prediction.

When considering robustness, our analytical estimates of strict positive-definiteness with respect to the gain compare well to numerical results. The estimates of the conditions for strict positive-definiteness clearly show how the

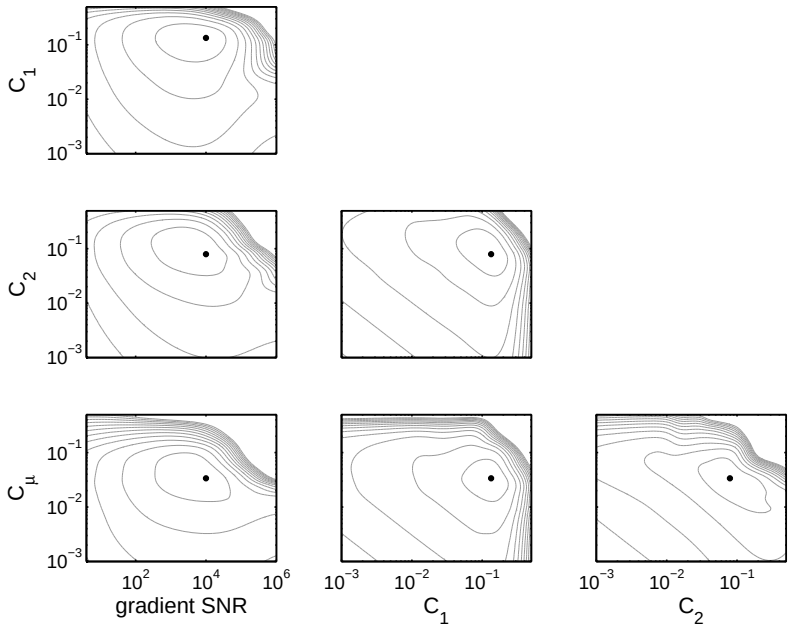


FIG. 9: Log-likelihood of the correlation ranges and the gradient-SNR.

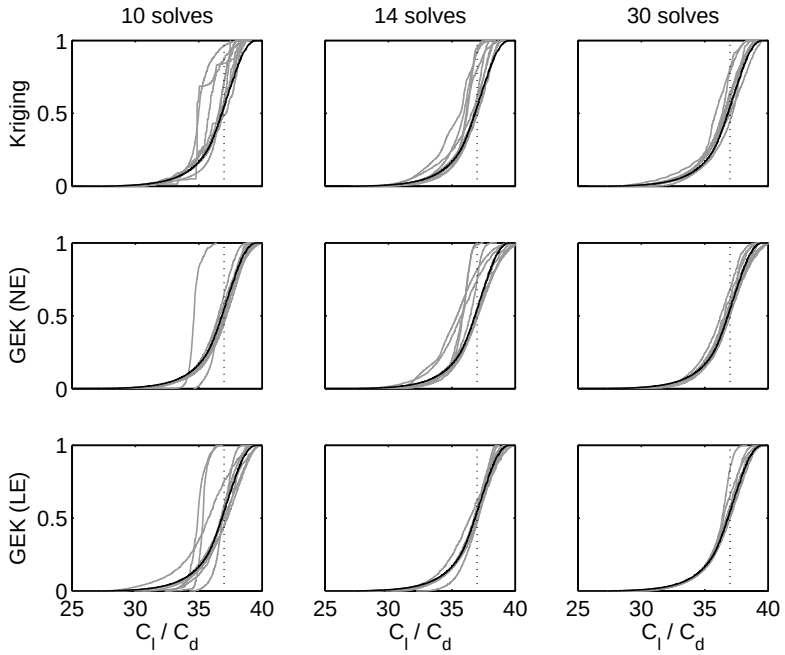


FIG. 10: Resulting cdf's (gray lines) for different methods and LHS-designs. Dotted black vertical line is deterministic result, black line is reference result.

observation error improves this property. This formalizes previous observations that adding a small value to the diagonal of the variogram can improve robustness. However, note that presently the observation errors are contained in the likelihood instead of in the variogram. The present analysis thus shows how inclusion of error information makes GEK robust.

Apart from including gradient information and treating individual observation errors, we propose to improve the Kriging process by taking into account the increase in the posterior covariance when making maximum likelihood estimates of the hyperparameters.

GEK can act as a surrogate model for uncertainty quantification of a computational fluid dynamics (CFD) simulation. Future work will focus on two areas: (a) application of GEK to a CFD problem with a $>$16-dimensional stochastic parameter space; and (b) reducing the cost of the GEK analysis.

ACKNOWLEDGMENT

This research was supported by *Technologiestichting STW* (project number 10113).

REFERENCES

1. Oden, T., Moser, R., and Ghattas, O., Computer predictions with quantified uncertainty, *SIAM News*, 43(9-10), 2010.
2. Giles, M. B. and Pierce, N. A., An introduction to the adjoint approach to design, *Flow, Turbul. Combus.*, 65:393–415, 2000.
3. Dwight, R. and Brezillon, J., Effect of approximations of the discrete adjoint on gradient-based optimization, *AIAA J.*, 44(12):3022–3071, December 2006.
4. Dwight, R. P. and Han, Z.-H., Efficient uncertainty quantification using gradient-enhanced kriging, in *11th AIAA Non-Deterministic Approaches Conf.*, Palm Springs, California, 4–7 May, 2009.

5. de Baar, J. H., Scholcz, T. P., Verhoosel, C. V., Dwight, R. P., van Zuijlen, A. H., and Bijl, H., Efficient uncertainty quantification with gradient-enhanced kriging: Applications in fsi, *Proc. of the European Congress on Computational Methods in Applied Sciences and Engineering (ECCOMAS 2012)*, J. Eberhardsteiner et al. (eds.), Vienna, Austria, 10–14 Sept., 2012.
6. Han, Z.-H., Grtzb, S., and Zimmermann, R., Improving variable-fidelity surrogate modeling via gradient-enhanced krigingand a generalized hybrid bridge function, *Aerospace Sci. Technol.*, 25(1):177–189, 2013.
7. Forrester, A. I. J., Sbester, A., and Keane, A. J., *Engineering Design via Surrogate Modelling, A Practical Guide*, Wiley, 2008.
8. Morris, M. D., Mitchell, T. J., and Ylvisaker, D., Bayesian design and analysis of computer experiments: Use of derivatives in surface prediction, *Technometrics*, 35(3):243–255, 1993.
9. de Baar, J., Dwight, R., and Bijl, H., Fast maximum likelihood estimate of the kriging correlation range in the frequency domain, *IAMG Conf.*, Salzburg, 5–9 Sept., 2011.
10. de Baar, J., Dwight, R., and Bijl, H., Speeding up kriging through fast estimation of the hyperparameters in the frequency-domain, *Comput. Geosci.*, 54(0):99–106, 2013.
11. Matheron, G., Principles of Geostatistics, *Econ. Geol.*, 58:1246–1266, 1963.
12. Gandin, L., *Objective analysis of meteorological fields: Gidrometeorologicheskoe Izdatel'stvo (GIMIZ), Leningrad*, Translated by Israel Program for Scientific Translations, Jerusalem, 1965.
13. Cressie, N., The origins of kriging, *Math. Geol.*, 22(3):239–252, 1990.
14. Cressie, N., *Statistics for Spatial Data*, Wiley, New York, 1993.
15. Stein, M. L., *Interpolation of Spatial Data, Some Theory for Kriging*, Springer, New York, 1999.
16. Lophaven, S., Nielsen, H., and Søndergaard, J., DACE − A Matlab Kriging Toolbox. Report IMM-REP-2002-12, Informatics and Mathematical Modelling, Technical University of Denmark, p. 34, 2002.
17. Cressie, N. and Johannesson, G., Fixed rank kriging for very large spatial data sets, *J. R. Stat. Soc. Ser. B*, 70(1):209–226, 2008.
18. Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P., Design and analysis of computer experiments, *Stat. Sci.*, 4(4):409–423, 1989.
19. Matheron, G., *Les Cahiers du Centre de Morphologie Mathematique: The Theory of Regionalized Variables and Its Application*, Ecole Nationale Superieure des Mines de Paris, 1971.
20. Khintchine, A., Korrelationstheorie der stationren stochastischen prozesse, *Math. Ann.*, 109:604–615, 1934.
21. Wold, H., *A Study in the Analysis of Stationary Time Series*, Almqvist and Wiksell, Uppsala, 1938.
22. Kolgomorov, A., The local structure of turbulence in incompressible viscous fluid for very large reynolds numbers, translated from Dokl. Akad. Nauk SSSR (1941) 30(4), *Proc. R. Soc. Lond. Ser. A*, 434:9–13, 1991.
23. Wiener, N., *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*, Wiley, New York, 1949.
24. Thompson, P. D., Optimum smoothing of two-dimensional fields1, *Tellus*, 8(3):384–393, 1956.
25. Henderson, C., Best linear unbiased estimation and prediction under a selection model, *Biometrics*, 31:423–447, 1975.
26. Wikle, C. K. and Berliner, L. M., A Bayesian tutorial for data assimilation, *Phys. D: Nonlin. Phenom.*, 230(1-2):1–16, 2007.
27. Chung, H.-S. and Alonso, J. J., Using gradients to construct cokriging approximation models for high-dimensional design optimization problems, in *AIAA 40th Aerospace Sciences Meeting and Exhibit*, Reno Nevada, 14–17 Jan., 2002.
28. Laurenceau, J. and Sagaut, P., Building efficient response surfaces of aerodynamic functions with kriging and cokriging, *AIAA J.*, 46(2):498–507, 2008.
29. Philip, G. and Watson, D., Matheronian Geostatistics–Quo vadis?, *Mathematical Geology*, 18(1):93–117, 1986.
30. Srivastava, R., Philip, G., and Watson, D., Quo vadunt?, *Math. Geol.*, 18(1):141–146, 1986.
31. Kitanidis, P. K., Parameter uncertainty in estimation of spatial functions: Bayesian analysis, *Water Resour. Res.*, 22(4):499–507, 1986.
32. Omre, H. and Halvorsen, K., The Bayesian bridge between simple and universal Kriging, *Math. Geol.*, 21(7):767–786, 1989.
33. Handcock, M. S. and Stein, M. L., A Bayesian analysis of Kriging, *Technometrics*, 35(4):403–410, 1993.

34. Mira, J. and Sánchez, M. J., Analytical results for a Bayesian bivariate Cokriging model, *Stat. Probab. Lett.*, 58(1):97–109, 2002.
35. Kennedy, M. C. and O'Hagan, A., Bayesian calibration of computer models, *J. R. Stat. Soc. Ser. B*, 63:425–464, 2000.
36. Bayarri, M., Berger, J., Paulo, R., Sacks, J., Cafeo, J., Cavendish, J., Lin, C.-H., and Tu, J., A framework for validation of computer models, *Technometrics*, 49(2):138–154, 2007.
37. Kalman, R. E., A new approach to linear filtering and prediction problems, *Trans. ASME–J. Basic Eng.*, 82(Series D):35–45, 1960.
38. Zhang, Y. and Oliver, D., Evaluation and error analysis: Kalman gain regularization versus covariance regularization, *Comput. Geosci.*, 15(3):1–20, 2011.
39. Fukumizu, K., Bach, F. R., and Jordan, M. I., Kernel dimension reduction in regression, *Ann. Stat.*, 37(4):1871–1905, 2009.
40. Xiu, D. and Karniadakis, G., Modeling uncertainty in flow simulations via generalized polynomial chaos, *J. Comput. Phys.*, 187:137–167, 2003.
41. Babuška, I., Nobile, F., and Tempone, R., A stochastic collocation method for elliptic partial differential equations with random input data, *SIAM J. Num. Anal.*, 45(3):1005–1034, 2007.
42. Loeven, G., Witteveen, J., and Bijl, H., Probabilistic collocation: an efficient non-intrusive approach for arbitrarily distributed parametric uncertainties, in *45th AIAA Aerospace Sciences Meeting and Exhibit*, Reno, Nevada, 8–11 Jan., 2007.
43. Witteveen, J. A., Loeven, A., Sarkar, S., and Bijl, H., Probabilistic collocation for period-1 limit cycle oscillations, *J. Sound Vibr.*, 311(1-2):421–439, 2008.
44. Nobile, F. and Tempone, R., Analysis and implementation numbers for the numerical approximation of parabolic equations with random coefficients, *Int. J. Numer. Methods Eng.*, 80:979–1006, 2009.
45. Martins, J. R. R. A., Kroo, I. M., and Alonso, J. J., An automated method for sensitivity analysis using complex variables, *AIAA paper 2000-0689, 38th Aerospace Sciences Meeting*, Reno, Nevada, 10–13 Jan., 2000.
46. Nielsen, E., Lu, J., Park, M., and Darmofal, D., An implicit, exact dual adjoint solution method for turbulent flows on unstructured grids, *Comput. Fluids*, 33:1131–1155, 2004.
47. Nielsen, E. and Kleb, B., Efficient construction of discrete adjoint operators on unstructured grids by using complex variables, in *43rd AIAA Aerospace Sciences Meeting and Exhibit*, 2005.
48. Chiles, J.-P. and Delfiner, P., *Geostatistics, Modeling Spatial Uncertainty*, Wiley, New York, 1999.
49. Sivia, D., *Data Analysis, a Bayesian Tutorial*, Oxford University Press, Oxford, 1996.
50. Pilz, J. and Spöck, G., Why do we need and how should we implement Bayesian Kriging methods, *Stochast. Env. Res.Risk Assessment*, 22(5):621–632, 2008.
51. Gantmacher, F., *The Theory of Matrices*, Vol. 1, Chelsea, 1959.
52. Mardia, K., Kent, J., and Bibby, J., *Multivariate Analysis*, Academic Press, London, 1979.
53. Henderson, H. V. and Searle, S. R., On deriving the inverse of a sum of matrices, *SIAM Rev.*, 23(1):53–60, 1981.
54. Constantinescu, E. M. and Anitescu, M., Physics-based covariance models for gaussian processes with multiple outputs, *Int. J. Uncertainty Quantific.*, 3(1):47–71, 2013.
55. Mardia, K. V. and Marshall, R. J., Maximum likelihood estimation of models for residual covariance in spatial regression, *Biometrika*, 71(1):135–146, 1984.
56. Mardia, K. V., Maximum likelihood estimation for spatial models, *Proc. from the Symp. on Spatial Statistics: Past, Present, and Future*, Syracuse, pp. 203–253, 1989.
57. Kitanidis, P. K. and Lane, R. W., Maximum likelihood parameter estimation of hydrologic spatial processes by the Gauss-Newton method, *J. Hydrol.*, 79(1-2):53–71, 1985.
58. Corbeil, R. R. and Searle, S. R., Restricted maximum likelihood (reml) estimation of variance components in the mixed model, *Technometrics*, 18(1):31–38, 1976.
59. Ababou, R., Bagtzoglou, A., and Wood, E., On the condition number of covariance matrices in Kriging, estimation, and simulation of random fields, *Math. Geol.*, 26(1):99–133, 1994.

60. Davis, G. and Morris, M., Six factors which affect the condition number of matrices associated with kriging, *Math. Geol.*, 29(5):669–683, 1997.

61. Timmer, W. and Rooij, R.v., Summary of the delft university wind turbine dedicated airfoils, *Trans. ASME*, 125:488–496, 2003.

62. Jones, W. and Launder, B., The prediction of laminarization with a two-equation model of turbulence, *Int. J. Heat Mass Transfer*, 15(2):301–314, 1972.