Non-Deterministic Kriging for Probabilistic Systems with Mixed Continuous and Discrete Input Variables

Jayasekara Jayasekara[1] and Sabarethinam Kameshwar Ph.D.[2]

[1]Ph.D. student, Department of Civil and Environmental Engineering, Louisiana State University, Baton Rouge, LA 70803, Email: jjayas2@lsu.edu

[2]Assistant Professor, Department of Civil and Environmental Engineering, Louisiana State University, Baton Rouge, LA 70803, Email: skameshwar1@lsu.edu (corresponding author)

Abstract

This paper presents a Non-Deterministic Kriging method to approximate the response of probabilistic systems with mixed continuous and discrete input variables. The proposed method approximates both epistemic (extrinsic) and aleatory (intrinsic) uncertainties in addition to the mean response of a system. Kriging is a popular metamodeling method for approximating the responses of computationally demanding systems along with prediction variances. However, conventional Kriging fails to perform with non-deterministic datasets with replications. The recently developed Non-Deterministic Kriging (NDK) method addresses those challenges in the continuous input space. Currently, Kriging is often used for approximations in probabilistic systems with mixed continuous and discrete input variables as well. Therefore, this study aims to fill the gap in the NDK method for probabilistic systems with mixed continuous and discrete input variables. Herein, the aleatory uncertainty is assessed using Locally Weighted Regression (LWR). The proposed method employs a combination of continuous and discrete kernels to capture the effects of mixed inputs. The effectiveness of the newly proposed NDK method was demonstrated using a set of probabilistic analytical cases and engineering applications. The proposed method

provides separable information about aleatory and epistemic uncertainties, which are beneficial in design optimizations and sequential explorations of probabilistic systems, especially with large-scale experiments and computer simulations with randomness.

Keywords: Non-Deterministic Kriging, Mixed variables, Continuous and Discrete variables, Probabilistic Systems, Machine Learning

1. Introduction

Metamodels, also called surrogate models, have been widely used to replace computationally expensive simulations (Asher et al., 2015; Mukhopadhyay et al., 2015) and physical experiments (Kabir et al., 2017; Stuckner et al., 2021). These techniques are mostly used when the underlying model has difficulties such as informatic complexity, computing efficiency, and code coupling (Delage et al., 2022). Kriging (Krige, 1951; Matheron, 1962), Neural Networks (Kohonen, 1982; Widrow and Hoff, 1960), Support Vector Machines (SVM) (Boser et al., 1992; Vapnik, 1999), Radial Basis Function (RBF) (Broomhead and Lowe, 1988), Multivariate Adaptive Regression Splines (MARS) (Friedman, 1991), and Polynomial Response Surface models (Wang et al., 2001) are some of the commonly used metamodeling techniques. In addition to these types of metamodels, sometimes different metamodels are combined, and hybrid modeling platforms are formed to achieve effective solutions based on the requirements of the application (Yin et al., 2018; Zhang et al., 2012).

Kriging is considered one of the most investigated surrogate models among these surrogate modeling techniques due to its attractive interpolative and stochastic characteristics (Jiang et al., 2018; Zhou et al., 2018). For instance, Kriging can provide the prediction uncertainty at each unsampled point. This feature is especially useful in refining the surrogate model in sequential

design strategies. Furthermore, Kriging requires a smaller training sample size due to strong interpolation among sampled points (Khan, 2011; Welch et al., 1992). Kriging was initially introduced for geostatistics (Krige, 1951; Matheron, 1962). However, it has been expanded to a wider spectrum of engineering problems due to its interpolative and probabilistic characteristics (Di Maio et al., 2022; Koziel and Pietrenko-Dabrowska, 2022; Su et al., 2019; Trochu et al., 2022; X. Zhang et al., 2020). Although the performance of surrogate models varies due to different conditions, Qian et al. (2020) show that compared to available surrogate models, Kriging models have been able to perform with higher accuracy and robustness, especially when black box-type functions show high non-linearity. For example, a comparison done by Kianifar and Campean (2020) depicts that Kriging outperforms polynomials and radial basis functions with highly nonlinear underlying functions regardless of the problem scale and size of training data samples. Furthermore, Abbas et al. (2018) highlight that Kriging shows more robust performance in spatial predictions compared to neural networks. In another work by Kahrizi et al., (2022) for investigating the characteristics of porous concrete, Kriging showed the least errors in estimations compared to the polynomial response surface method and radial basis function. Although Kriging can outperform other metamodels, it is vital to be aware that Kriging does have several disadvantages as well. For example, Kriging fails to perform accurately when the underlying function shows non-stationary trends (Bae et al., 2019). Furthermore, Kriging models can be computationally demanding with a larger number of input variables (Saves et al., 2021). Also, compared to other available surrogate models, such as polynomial regression models, Kriging models are not easy for users to interpret (Kianifar and Campean, 2020).

Several types of Kriging models, such as simple Kriging (Li et al., 2009), ordinary Kriging (Kumar et al., 2023), stochastic Kriging (Ankenman et al., 2008), universal Kriging, and regression

Kriging (Picheny et al., 2013) are currently used based on metamodeling requirements and characteristics of Kriging models. Since geographical distance was the primary input variable in the early stages, Kriging models were initially used only with continuous variables in geostatistics. However, many systems in engineering problems consist of continuous and discrete design variables. Therefore, due to the need to use Kriging over systems that consist of mixed input variables, several studies extended the use of Kriging models to include discrete input variables as well (Pelamatti et al., 2020; Saves et al., 2022). In these methods, different correlation kernels were used to adapt the existing Kriging models to accommodate the presence of discrete input variables. Hypersphere decomposition kernel (Pelamatti et al., 2019), Latent variable kernel (Tao et al., 2021), Compound symmetry kernel (Roustant et al., 2020), and coregionalization matrix kernel (Pelamatti et al., 2021) are a few of such discrete kernels.

As mentioned earlier, conventional Kriging provides the uncertainty in the prediction at unsampled points, by fitting residuals from a global regression model (Loquin and Dubois, 2010). This can be identified as the extrinsic or epistemic uncertainty that is imposed on the problem when developing the Kriging model. It accounts for the modeling error of the Kriging model. Epistemic uncertainty, which is likely due to the lack of knowledge or information about the underlying physics of the problem, can be reduced by increasing the size of training samples and updating the estimations (Choi et al., 2006). There are several adaptive sampling-based design exploration methods that use the prediction variance or epistemic uncertainty given by Kriging. The efficient global optimization method (Jones et al., 1998; Yi and Taflanidis, 2023), quantile-based design and optimization (Nazeeh et al., 2023), sequential Kriging optimization (Hao et al., 2010), value-based global optimization (Moore et al., 2014), and reliability analysis methods (Hong et al., 2022; Kitahara et al., 2021) are a few of such methods that employ the prediction

variance. However, prediction variance given by conventional Kriging with probabilistic and replicated data can be misleading. Mostly, physical experiments have inherent randomness that results in aleatory or intrinsic uncertainty within data. Furthermore, under simulation-based design exploration methods, computational models are also developed as variable-fidelity models that consist of computational inaccuracy or uncertainty (Bae et al., 2019). Aleatory uncertainty arises mostly due to natural variability in the parameters of a physical system and cannot be reduced in a similar way to epistemic uncertainty (Palar et al., 2019; Zhuang and Pan, 2012). Estimations obtained through insufficient quantification of aleatory uncertainty can incline towards overconfident predictions (Mortazavi et al., 2012). Therefore, the use of conventional Kriging over non-deterministic (noisy) data samples from physical experiments, variable-fidelity models, natural phenomena, or non-stationary underlying functions could mislead the predictions, especially in design exploration methods.

Given the need for estimation of aleatory uncertainty within the prediction variance in Kriging models, a Non-Deterministic Kriging (NDK) method was derived recently as a flexible method that approximates both epistemic and aleatory uncertainties associated with the Kriging model and the underlying probabilistic function, respectively (Clark, 2019). The proposed NDK methodology is considered numerically more stable than other Kriging models as it captures uncertainty bounds more efficiently. However, the NDK proposed by Clark only considered continuous input variables. There is a lack of studies that cover the approximation of both epistemic and aleatory uncertainties using the NDK method for systems with mixed input variables, which is necessary as engineering systems can consist of both continuous and discrete variables. Furthermore, in design optimization models, the use of discrete variables can replace a number of continuous variables, thus reducing the computational time (Rosness, 1993). The

114 introduction of such a discrete variable can increase the randomness of a system and its data.
115 Furthermore, especially with experimental designs, replications are obtained at the same design
116 point due to the inherent randomness in natural phenomena (Dunn, 2019). As revealed by the
117 review of existing work related to surrogate model-assisted optimization under uncertainty,
118 conventional Kriging models that are developed to work with mixed continuous and discrete input
119 variables fail to address the challenges arising due to non-deterministic and replicated data.
120 Although Stochastic Kriging for mixed input spaces (Lopez et al., 2022) aims at estimating the
121 randomness present in data, it requires a larger number of replications at a design point. This
122 process is not always possible with engineering system design explorations, especially with large-
123 scale physical experiments or computer simulations.

124 The present study was motivated by the application of Kriging on probabilistic systems in
125 mixed continuous and discrete input spaces, especially with systems that are expensive to evaluate.
126 For instance, there are time-consuming physical experiments that demand a large amount of
127 resources. In such instances, it is not effective to have a larger number of replications to account
128 for noise present within the system. Although there exists an NDK method for such probabilistic
129 systems, it cannot be employed in mixed input space. Therefore, this study aimed to fill the gap in
130 the need for an NDK method for probabilistic systems with mixed continuous and discrete input
131 variables. The proposed NDK can accommodate replicated data, especially with physical
132 experiments. Also, the proposed method does not require a large number of replications as in
133 existing Stochastic Kriging models. Since this NDK method quantifies both aleatory and epistemic
134 uncertainties, the prediction variance of the NDK model can be used for design optimization under
135 uncertainty for systems with non-deterministic data and non-stationary trends. Currently, the use
136 of conventional Kriging can yield erroneous results with such systems. Furthermore, the use of the

proposed NDK model with mixed continuous and discrete inputs is beneficial in Kriging-associated design explorations with time-consuming large-scale physical experiments and computationally expensive computer simulations.

The next section of the paper provides a review of existing work of surrogate model-assisted uncertainty quantification. Section 3 presents the proposed framework, including steps in the NDK methodology for mixed input variables. Section 4 presents the application of the proposed method to a set of numerical examples with continuous and discrete variables. Additionally, this section discusses the goodness-of-fit of the estimated mean and uncertainties of numerical cases using a set of goodness-of-fit measures. The next section presents the application of the proposed method to two engineering systems. The last sections of the paper provide the discussion and conclusion of the study.

2. Related work on Surrogate assisted Uncertainty Quantification (UQ)

Surrogate model-assisted optimization has been popular due to the computational burden of optimization under uncertainty with traditional approaches. Accounting for uncertainties stands paramount in any type of design optimization strategy that employs surrogate models, especially with the modeling error induced by surrogate models. A significant number of past studies have covered and developed uncertainty quantification in optimization strategies with surrogate models in the continuous input space. However, comparatively, there are still only a few studies that focus on the mixed continuous-discrete input variables to the best knowledge of the authors. Elaborating on early work on surrogate-assisted mixed variable optimization under uncertainty, Sriver and Chrissis (2004) used surrogate function approximations in a novel framework, which is a combination of generalized pattern search and ranking and selection method, for systems with inherent variation. Furthermore, a novel hierarchical hybrid fuzzy neural network was presented

by Wang et al. (2007) to represent systems with mixed input variables. This method uses a set of fuzzy sub-systems that aggregate discrete input variables into intermediate variables and plugs them into neural networks for approximations along with continuous input variables. However, surrogate models utilized in these methods do not provide insights into the uncertainty in predictions similar to modeling uncertainty presented by Kriging models.

Recent literature shows a growing interest in mixed variables constrained optimization approaches coupled with modified versions of Kriging. Huang et al., (2023) used Latent-variable Gaussian process (aka Kriging) (LVGP) modeling with Bayesian optimization for optimal search. This method matches discrete inputs into a Latent space and uses the Latent distances for Kriging. An et al., (2022) also used Kriging similarly for multi-objective optimization, under noise. However, their Kriging model was built only considering model uncertainty while measurement noise was added to the objective later as an assumption. In another work, An et al., (2021) used Kriging for reliability-based design optimization, where continuous and discrete variables were initially decoupled using the total probability theorem. After that Kriging was used only with continuous space, ultimately resulting in a category-wise approach based on discrete levels. In contrast, Pelamatti et al., (2021) also worked on using Kriging for optimization in the mixed continuous-discrete design space combinedly. This work employs several discrete kernels such as hypersphere decomposition kernel, compound symmetry kernel, and coregionalization matrix kernel within the Kriging model to combine with continuous kernels. Furthermore, their work is extended to consider the heteroscedasticity of Gaussian process variance in the discrete input variables. However, their method does not consider the randomness present within the data set. There can be erroneous estimations in both the mean response and prediction of the model when this method is used with probabilistic systems.

In the most recent work, Moustapha et al. (2022) used Kriging in a multi-objective robust optimization approach for problems with mixed variables. This method employed quantiles of the objective functions that are determined through a Kriging model, to allow both optimality and robustness. In a similar work for a multi-objective optimization process, An et al., (2022b) used Kriging to predict the objective function values in a Monte Carlo Simulation. Furthermore, there are studies that focused on potential issues with Kriging-assisted optimization as well. For example, since a larger number of mixed continuous and discrete variables might be employed in engineering optimization processes or practical applications, the number of hyperparameters in the Kriging model also increases substantially. To address this issue in Kriging models, Saves et al., (2021) proposed a hyperparameter reduction process based on the partial least squares method. This method employed an adaptive procedure for selecting the number of hyperparameters in a Kriging model.

However, the performance of the conventional Kriging approach in these studies can be affected due to specific conditions associated with underlying response function and available data of probabilistic systems. For instance, Kriging approximations will fail to fit the underlying response accurately when available data is non-deterministic or replicated. Furthermore, the covariance structure of the Kriging model will be inaccurate when the underlying black-box function has non-stationary trends. This can result in an overly dampened or amplified prediction variance. Therefore, with a limited size of non-deterministic data or replicated data, conventional Kriging becomes misleading in design explorations. In the continuous space, there are a few methods for these challenges. As a solution for randomness in data, in Regression Kriging, hyperparameter optimization is performed with an added noise parameter, thus approximating sample data within constant noise bounds similar to regression analysis (Hengl et al., 2004;

206 Picheny et al., 2013). To the best of the authors' knowledge, this has not been extended to the
207 mixed continuous and discrete input space. Furthermore, Stochastic Kriging (SK) was proposed
208 using two uncertainty sources, epistemic and aleatory, to work with non-deterministic and
209 replicated data (Ankenman et al., 2008). In SK, aleatory uncertainty is estimated at each design
210 point using a separate ordinary Kriging model that employs replications. Lopez et al., (2022)
211 extended SK to work with mixed continuous and discrete input variables in an adaptive stochastic
212 efficient global optimization approach. However, even SK performs with the assumption that there
213 are enough replications at each design point. Therefore, with insufficient data, aleatory uncertainty
214 prediction becomes inconsistent and unreliable.

215 As a remedy for these challenges, Clark, (2019) proposed a Non-Deterministic Kriging
216 method that works in the continuous input space. Compared to conventional Kriging, this method
217 relaxes the interpolation requirement in the presence of randomness in data. It uses the aleatory
218 variance as a regularization factor in computations, thus increasing the accuracy in the prediction
219 of both mean and modeling uncertainties. Most importantly, the NDK model provides both
220 epistemic and aleatory uncertainties for use in applications such as design explorations based on
221 the requirements. However, this proposed method only works in the continuous input space and
222 cannot be used with discrete input variables. This paper improves the existing NDK method further
223 to work in the mixed continuous and discrete input space.

224 **3. Non-Deterministic Kriging (NDK) with mixed input variables**

225 This section of the paper presents the proposed NDK methodology for mixed continuous
226 and discrete input variables. In general, continuous variables like structural dimensions, fluid
227 velocity, and force are mapped to the space of real numbers within a defined interval. On the other
228 hand, discrete variables are identified as design characteristics such as material type, which have

229 a finite number of choices (Musiol, 1997). Furthermore, discrete variables are categorized as
230 ordinal and nominal variables as well. An ordinal variable consists of multiple categories that can
231 be ordered while a nominal variable consists of multiple categories which cannot be orderly
232 arranged. For example, hurricane category which is based on wind speed can be considered an
233 ordinal variable while material type is a nominal variable. However, in this study, no such
234 difference is considered between these two types.

235 The proposed NDK model aims at estimating the mean response in a probabilistic system,
236 with mixed continuous and discrete input variables, while estimating both epistemic and aleatory
237 uncertainty. Fig. 1 presents the flow chart for the proposed NDK method. Starting from the
238 identification of input variables, three major processes in the method are (1) correlation matrix
239 calculation, (2) locally weighted regression process for aleatory uncertainty estimation, and (3)
240 epistemic variance estimation. Individual outputs from these processes are later used for the final
241 outputs of the NDK model, which includes the estimated mean response and prediction variance
242 including both epistemic and aleatory uncertainties. The detailed steps of the model are presented
243 in the following subsections. For clarity in symbols, bold fonts were used for vectors and matrices.

244 Consider $f(\boldsymbol{x}, \boldsymbol{z})$ is a probabilistic black-box function that is defined between an input $\boldsymbol{v}(\boldsymbol{x}, \boldsymbol{z}) \in \mathbb{V} \subset \mathbb{R}_{n+q}$ and an output, $y \in \mathbf{Y} \subset \mathbb{R}$, where n and q represent the number of continuous and
246 discrete variables respectively, $\mathbb{V}$ is the matrix of input data locations, and $\mathbf{Y}$ is the vector of
247 responses at input data locations. According to the proposed method, developing an NDK model
248 for this function starts with determining the input variables and their characteristics. Consider that
249 $\boldsymbol{x}$ is defined as a vector containing continuous inputs ($\boldsymbol{x} = \{x_1, x_2, x_3, .. x_k \ldots .. x_r\}$) while $\boldsymbol{z}$ is a
250 vector that consists of discrete inputs ($\boldsymbol{z} = \{z_1, z_2, z_3, .., z_k, \ldots .. z_q\}$). Each discrete variable z_k has

b_k number of possible values, also known as levels, resulting in a total of $\prod_{k=1}^{k=q} b_k$ combinations of categorical inputs.

In general, a Kriging model works like weighted regression that processes a training data set or a design of experiments (D), of N_s samples $\{\boldsymbol{v}_i, y_i\}$, where i = 1, 2, …, N_s. For a given training data point, the deterministic Kriging model is represented as follows;

$$y(\boldsymbol{v}) = m(\boldsymbol{v}) + Z_0(\boldsymbol{v}) \text{ ---------- Eq (1)}$$

where $m(\boldsymbol{v})$ represents the global trend. $Z_0(\boldsymbol{v})$ is a zero mean stochastic process defined as

$$Cov(Z_0(\boldsymbol{v_i}), Z_0(\boldsymbol{v_j})) = \sigma^2 R(\boldsymbol{v_i}, \boldsymbol{v_j}; \boldsymbol{\theta}) \text{ ---------- Eq (2)}$$

where σ^2 is the process variance, $R(\boldsymbol{v_i}, \boldsymbol{v_j}; \boldsymbol{\theta})$ is the correlation among data points, and $\boldsymbol{\theta}$ are hyperparameters that need to be calibrated. The stochastic nature of Z_0 refers to the extrinsic (epistemic) uncertainty since it is imposed on the problem to assist in developing the model. This stochastic process accounts for the modeling error in the Kriging model. The conventional Kriging method is more suitable for data without noise since the estimation only represents the modeling uncertainty and does not consider the natural randomness present in the response variable of the data set. This highlights the need for a Kriging model which accounts for both epistemic uncertainty (due to modeling error) and aleatory uncertainty (due to natural randomness of data).

As a solution for this need, the Non-Deterministic Kriging (NDK) method includes both epistemic and aleatory uncertainties within its predictions by adding aleatory uncertainty as a separate stochastic process into the deterministic Kriging model given in Eq (1). Accordingly, taking the above-described deterministic Kriging derivation as a basis, the NDK model is represented as follows:

$$y_{nd}(\boldsymbol{v}) = m(\boldsymbol{v}) + Z_E(\boldsymbol{v}) + Z_A(\boldsymbol{v}) \text{ ---------- Eq (3)}$$

where $m(\boldsymbol{v})$ represents the global trend and $Z_E(\boldsymbol{v})$ and $Z_A(\boldsymbol{v})$ are the stochastic processes of epistemic and aleatory uncertainties respectively (Clark, 2019). The mean estimation (y_{nd}) can be formulated using a linear predictor as;

$$y_{nd}(\boldsymbol{v}) = \boldsymbol{c}^T\boldsymbol{Y} \text{ ----------Eq (4)}$$

where $\boldsymbol{c} = c(\boldsymbol{v}) \in \mathbb{R}_{N_s}$. With the use of unbiasedness condition $\boldsymbol{c}^T\boldsymbol{F} - \boldsymbol{f} = 0$, where $\boldsymbol{f}$ and $\boldsymbol{F}$ are the assumed basis function vector at the unsampled point and regression design matrix at training data points respectively, the prediction variance (σ_{nd}^2) at an unsampled point ($\boldsymbol{v}$) is obtained as:

$$\sigma_{nd}^2(\boldsymbol{v}) = E\left[\left(y_{nd}(\boldsymbol{v}) - Y(\boldsymbol{v})\right)^2\right] = E[\left(\boldsymbol{c}^T(\boldsymbol{Z}_E + \boldsymbol{Z}_A) - (z_E + z_A)\right)^2] \text{ ----------Eq (5)}$$

where $\boldsymbol{Z}_E$ and $\boldsymbol{Z}_A$ are true epistemic and aleatory uncertainty vectors from training data points and z_E and z_A are true epistemic and aleatory uncertainties at the unsampled point. In general, epistemic and aleatory uncertainties are different and independent since their sources are different according to the definitions. Therefore, it is assumed that the aleatory uncertainty and model uncertainty are not correlated, thus there is no correlation between z_E and z_A. Based on this assumption, the above equation for prediction variance of the NDK model can be expanded as;

$$\sigma_{nd}^2(\boldsymbol{v}) = E[\boldsymbol{c}^T\boldsymbol{Z}_E\boldsymbol{Z}_E^T\boldsymbol{c}] + E[\boldsymbol{c}^T\boldsymbol{Z}_A\boldsymbol{Z}_A^T\boldsymbol{c}] + E[z_E z_E^T]$$

$$+E[z_A z_A^T] - 2E[\boldsymbol{c}^T\boldsymbol{Z}_A z_A] - 2E[\boldsymbol{c}^T\boldsymbol{Z}_E z_E] \text{ ---------- Eq (6)}$$

Using variance (σ_E^2, σ_A^2), correlation ($\boldsymbol{R}_E$), and covariance ($\boldsymbol{V}_A$) terms this can be further simplified as;

$$\sigma_{nd}^2 = \sigma_E^2(1 + \boldsymbol{c}^T\boldsymbol{R}_E\boldsymbol{c} - 2\boldsymbol{c}^T\boldsymbol{r}_E) + (\boldsymbol{c}^T\boldsymbol{V}_A\boldsymbol{c} - 2\boldsymbol{c}^T\mathbf{v}_A + \sigma_A^2) \text{ ----------Eq (7)}$$

where σ_E^2 and σ_A^2 are epistemic and aleatory variances, $\boldsymbol{R_E}$ is the correlation matrix among training data samples, $\boldsymbol{r_E}$ represents the correlation vector between the unsampled point and the already sampled training data points. $\boldsymbol{V_A}$ and $\mathbf{v_A}$ are the aleatory covariance matrix and vector respectively.

By minimizing the prediction variance of the estimate, the mean estimation (y_{nd}) is solved as,

$$y_{nd}(\boldsymbol{v}) = \boldsymbol{f}^T(x)\boldsymbol{\beta}_{nd} + \mathbf{v}_{nd}^T\boldsymbol{V}_{nd}^{-1}(\boldsymbol{Y} - \boldsymbol{F}\boldsymbol{\beta}_{nd}) \text{ ---------- Eq (8)}$$

where $\boldsymbol{\beta}_{nd} = [\boldsymbol{F}^T\boldsymbol{V}_{nd}^{-1}\boldsymbol{F}]^{-1}\boldsymbol{F}^T\boldsymbol{V}_{nd}^{-1}\boldsymbol{F}$, $\boldsymbol{V}_{nd} = \sigma_E^2\boldsymbol{R}_E + \boldsymbol{V}_A$, and $\mathbf{v}_{nd} = \sigma_E^2\boldsymbol{r}_E + \mathbf{v}_A$.

In practice, aleatory uncertainty is considered independent across the design space. It is assumed that there is no correlation between that affect each other's aleatory uncertainty. Due to this independence and the definition of covariance matrix within the zero-mean stochastic process of Kriging according to Eq. 2, $\mathbf{v}_A$ can be neglected and $\boldsymbol{V}_A$ becomes a diagonal matrix. Even if the aleatory uncertainty is correlated among design points the Locally Weighted Regression process presented in Section 3.2 captures that effect approximately. The derivation of $\boldsymbol{V}_A$ is discussed in the next section of the paper. The prediction variance of the NDK model, representing a total measurement of epistemic and aleatory uncertainty, can be formulated as,

$$\sigma_{nd}^2(\boldsymbol{v}) = \sigma_E^2 + \sigma_A^2(\boldsymbol{v}) + \boldsymbol{u}(\boldsymbol{v})^T[\boldsymbol{F}^T{\boldsymbol{V}_{nd}}^{-1}\boldsymbol{F}]^{-1}\boldsymbol{u}(\boldsymbol{v}) - \mathbf{v}_{nd}(\boldsymbol{v})^T{\boldsymbol{V}_{nd}}^{-1}\mathbf{v}_{nd}(\boldsymbol{v}) \text{ ---------- Eq (9)}$$

where $\boldsymbol{u}(\boldsymbol{v}) = \boldsymbol{F}^T{\boldsymbol{V}_{nd}}^{-1}\mathbf{v}_{nd}(\boldsymbol{v}) - \boldsymbol{f}(\boldsymbol{v})$.

3.1 Determining the correlation matrix $\boldsymbol{R}_E$

After identifying the characteristics of the input variables, the correlation matrix $\boldsymbol{R}_E$ should be determined. $\boldsymbol{R}_E$ is a square matrix with the size $N_s \times N_s$ and contains correlation values between each sample of the training data set. It can be calculated as;

$$\boldsymbol{R}_{E\,i,j} = k(\boldsymbol{v}_i, \boldsymbol{v}_j) \text{ ---------- Eq (10)}$$

where i, j = 1, 2, …, N_s and $k(\cdot)$ is a user-defined kernel function, which is symmetric and positive semi-definite over the input space (Lanckriet et al., 2004; Mohammadi, 2016). In this proposed method for NDK, the correlation function was adapted to capture the influence of discrete variables in addition to continuous variables. Two kernel operators were used in this study to meet the requirements for defining discrete kernels and combining them with continuous kernels. These selected operators ensure that the resulting correlation functions are also symmetric and positive semi-definite over the respective input spaces. The following kernel operations were used in this proposed NDK methodology.

1. Product - Consider two continuous or discrete input subspaces, F_1 and F_2. If k_1 and k_2 are two kernels defined over input subspaces F_1 and F_2, and $'\beta'$ is a real number, $k = k_1 \times k_2$ and $k' = \beta k_1$ become valid kernels over the input spaces $F_1 \times F_2$ and F_1 respectively (John Shawe-Taylor and Nello Cristianini, 2004).
2. Mapping – If k is a kernel in the space F_1, $F_1{}'$ is a set and $g(.)$ is a mapping function from $F_1{}'$ to F_1, $k'(x, x') = k(g(x), g(x'))$ becomes a valid kernel over $F_1{}'$ (Steinwart and Christmann, 2008).

Consider an input ***v (x, z)***, which includes both continuous and discrete variables. In combining discrete and continuous kernels for the input ***v (x, z)***, Li and Racine (2003) proposed to use the product of kernels. Furthermore, the Schur product theorem proves that the Hadamard product between two positive semi-definite matrices results in a positive semi-definite matrix. Therefore, to define a valid kernel function for the NDK model by kernel combination, the product operation of kernels was used in this study. The other kernel operation, mapping was used later for defining

a discrete kernel. Accordingly, two different kernels were introduced for continuous and discrete variables separately and combined as follows.

$$k(\boldsymbol{v}_i, \boldsymbol{v}_j) = k_x(\boldsymbol{x}_i, \boldsymbol{x}_j) \times k_z(\boldsymbol{z}_i, \boldsymbol{z}_j) \text{ ---------- Eq (11)}$$

Here k_x and k_z are kernels for continuous and discrete variables. As shown in Fig. 1, the 'Selection of Correlation Kernels' stands as an important task within the proposed methodology.

3.1.1 Continuous kernels

For continuous variables, the kernel function can be formulated as:

$$k_x(\boldsymbol{x}_i, \boldsymbol{x}_j) = \prod_{l=1}^{n} k(\theta_l, d_l) \text{ ---------- Eq (12)}$$

where n is the number of continuous variables, $d_l = |x_{i,l} - x_{j,l}|$ is the difference between the i^{th} and j^{th} data points in the l^{th} dimension, and θ_l is a hyperparameter that determines the rate at which correlation decreases with the difference between two points in the l^{th} dimension (Clark, 2019). Table 1 presents a set of commonly used correlation functions for continuous variables.

3.1.2 Discrete kernel functions

In a similar format, the kernel functions for discrete variables were derived in this study as follows:

$$k_z(\boldsymbol{z}_i, \boldsymbol{z}_j) = \prod_{m=1}^{q} k(\theta_l, d_m) \text{---------- Eq (13)}$$

where d_m represents the difference between i^{th} and j^{th} data points in the m^{th} dimension. In order to represent the discrete difference between data points, the study employed several kernel types as described below.

Weighted Gower distance kernel

Hutter et al., (2011) proposed a new kernel function, that engages weighted Hamming distances, to capture the influence of categorical variables in Gaussian process models. Lately, Halstrup, (2016) further improved this concept by using Gower distance for both continuous and discrete variables. The Gower distance between two data points can be represented as follows (Gower, 1971);

$$d_{gow}(\boldsymbol{v}_i, \boldsymbol{v}_j) = \frac{\sum_{k=1}^{k=n} \frac{|x_k^i - x_k^j|}{\Delta x_k}}{n+q} + \frac{\sum_{k=1}^{k=q} \partial\left(z_k^i, z_k^j\right)}{n+q} \quad \text{---------- Eq (14)}$$

where Δx_k is the range in the k-th dimension and $\partial\left(z_k^i, z_k^j\right)$ is defined as follows;

$$\partial\left(z_k^i, z_k^j\right) = \begin{cases} 0 & if\ z_k^i = z_k^j \\ 1 & otherwise \end{cases} \text{---------- Eq (15)}$$

Halstrup, (2016) proposed the use of Gower distance as the difference between two data points, which corresponds to d_l and d_m in Eqs. 12 and 13 respectively. Since Gower distance calculates the difference between data points by considering both continuous and discrete input variables together, Eqs. 12 and 13 were combined as;

$$k\left(\boldsymbol{v}_i, \boldsymbol{v}_j\right) = \prod_{l=1}^{n+q} k(\theta_l, d_{gow}^l) \text{---------- Eq (16)}$$

where θ_l is the hyperparameter corresponding to the l^{th} dimension and d_{gow}^l represents the Gower distance in the l^{th} dimension between i^{th} and j^{th} data points in the system. For example, the Gower distance is used to define the mixed variable kernel function in the form of p^{th} exponential correlation as;

$$k\left(\boldsymbol{v}_i, \boldsymbol{v}_j\right) = \exp\{-\sum_{l=1}^{l=n+q} \theta_l\ [d_{gow}^l(\boldsymbol{v}_i, \boldsymbol{v}_j)]^{p^l}\} \text{---------- Eq (17)}$$

where p^l represents p^{th} exponential in the l^{th} dimension.

Latent variable kernel

Recently in metamodels, discrete variables were mapped into non-observed Latent variables and considered continuous variables (Y. Zhang et al., 2020). It was proposed to map each level of discrete variables into a vector of continuous variables in an l-dimensional hyperspace. For instance, for the m^{th} level of a categorical input, there exists in particular a Latent variable, $\mathbf{t} = [t_1,\ldots,t_h] \in \mathbf{T} \subset \mathbb{R}_1 \times\cdots\times\mathbb{R}_h$ where $\mathbf{T}$ represents the Latent space and h is the number of dimensions in the Latent space (Cuesta Ramirez et al., 2022). In the present study, the proposed Latent variable kernel maps each level of discrete variables into a 2-dimensional space as follows;

$$\varphi(z);\ \boldsymbol{F}_z \rightarrow \mathbb{R}^2$$

$$\varphi(z) = (\vartheta_{m,1}, \vartheta_{m,2})$$

where $\vartheta_{m,1}$ and $\vartheta_{m,2}$ are hyperparameters that represent the coordinates in the Latent variable space corresponding to discrete variable level m. The fact that discrete levels are already represented by hyperparameters eliminates the need for using separate hyperparameters for the distance between discrete inputs. According to the kernel operation of mapping, the Latent variable kernel function can be formulated as;

$$k(\mathbf{z}_i, \mathbf{z}_j) = k'(\varphi(\mathbf{z}_i), \varphi(\mathbf{z}_j)) \text{ ---------- Eq (18)}$$

In this proposed methodology, the distance between coordinates in the Latent space was used as the measure of the difference between discrete inputs. For instance, the Gaussian kernel function for the q-th discrete variable can be presented as;

$$k_z(\mathbf{z}_i, \mathbf{z}_j) = \exp(-\theta_q \left\| \varphi(\mathbf{z}_i) - \varphi(\mathbf{z}_j) \right\|^2) \text{ ---------- Eq (19)}$$

where θ_q is the hyperparameter in the q^{th} dimension. However, θ_q can be removed since Latent variables already depend on hyperparameter values. According to suggestions made by Zhang et al. (2020), one of the Latent variable coordinates was set to the origin of the 2-dimensional plane (0,0). Another coordinate was set on one of the axes (0, $\vartheta_{m,1}$) to reduce the number of hyperparameters used in the kernel function.

3.2 Aleatory uncertainty estimation through Locally Weighted Regression (LWR)

In this proposed NDK for mixed input variables, aleatory uncertainty quantification stands as the third major process as shown in Fig. 1. Accordingly, this process depends on three main activities namely, the selection of kernels for the weight matrix in LWR, determining the smoothening parameter, and covariance matrix calculation. In this process, it was assumed that aleatory uncertainty in the data set varies with the location of data in the design space. Therefore, to determine the aleatory uncertainty, locally weighted regression is used (Sam and Ker, 2006). The general regression model for locally weighted regression is given as;

$$y_l(\boldsymbol{v}) = \widetilde{m}_l(\boldsymbol{v}) + e_l(\boldsymbol{v}) \text{ ---------- Eq (20)}$$

where l denotes the samples located within the l^{th} local neighborhood, $\widetilde{m}_l$ is the local estimator, and $e_l(v)$ is uncorrelated random errors or deviations from measurements with zero mean and finite variance. The standard Nadaraya–Watson estimator was used to determine the local estimator as;

$$\widetilde{m}_l(\boldsymbol{v}) = \frac{\sum_1^{N_S} Y_i W_{i,i}^l(v_l,v_i)}{\sum_1^{N_S} W_{i,i}^l(v_l,v_i)} \text{----------Eq (21)}$$

where $\boldsymbol{Y}$ is a vector of the mean responses at the data points, and $\boldsymbol{W}^l$ is a diagonal matrix of weights or degrees of membership to the local neighborhood (Aljuhani and Turk, 2014). The data points

were weighted via kernel functions within the size of the neighborhood (kl). Gaussian, quadratic, and sigmoid functions are a few such kernel functions used for the calculation of weights. In this study, kernel functions were combined as described in Section 3.1 to capture the influence of discrete variables on LWR. The weight matrix is formulated as a diagonal matrix as shown in Eq. 22. Eq. 23 presents the combined kernel function for individual weights.

$$\boldsymbol{W}^{l} = \begin{bmatrix} w_{1,1}^{l} & 0 & \rightarrow & 0 \\ 0 & w_{2,2}^{l} & 0 & \downarrow \\ 0 & 0 & \ddots & 0 \\ 0 & \rightarrow & 0 & w_{Ns,Ns}^{l} \end{bmatrix} \text{---------- Eq (22)}$$

$$w_{i,i}^{l}(\boldsymbol{v}_l, \boldsymbol{v}_i) = w_x(\boldsymbol{x}_l, \boldsymbol{x}_i) \times w_z(\boldsymbol{z}_l, \boldsymbol{z}_i) \text{ ---------- Eq (23)}$$

The selection of kernels for the weight matrix in LWR is the next activity of the proposed methodology. In this study, different kernel combinations were used to represent Eq 23. Table 2 presents a set of kernel combinations used for the weight calculations.

The size of the neighborhood (kl), also called the smoothening parameter was determined through log-likelihood maximization presented in Section 3.3. After that, the aleatory uncertainty at $\boldsymbol{v}_l$ was estimated as a locally weighted mean squared error using Eq,24,

$$\sigma_l^2(\boldsymbol{v}_l) = \frac{\sum_1^{N_s} e_i^2 W_{i,i}^{l}(\boldsymbol{v}_l, \boldsymbol{v}_i)}{\sum_1^{N_s} W_{i,i}^{l}(\boldsymbol{v}_l, \boldsymbol{v}_i)} \text{ ---------- Eq (24)}$$

where $e_i = Y_i - \widetilde{m}_i(x)$ and l = 1, 2, 3, …., N_s. This LWR variance estimation is performed at each sample point and the estimated variance is used to form the diagonal matrix $\boldsymbol{V_A}$ as shown in Eq 25.

$$\boldsymbol{V_A} = \begin{bmatrix} \sigma_1^2(\boldsymbol{v}_1) & 0 & \rightarrow & 0 \\ 0 & \sigma_2^2(\boldsymbol{v}_2) & 0 & \downarrow \\ 0 & 0 & \ddots & 0 \\ 0 & \rightarrow & 0 & \sigma^2(\boldsymbol{v}_{Ns}) \end{bmatrix} \text{---------- Eq (25)}$$

Accordingly, using the locally weighted regression, the aleatory uncertainty estimation in the design space can be formulated as;

$$\sigma_A^2(\boldsymbol{v}) = \frac{\sum_1^{N_s} W_{i,i}(v,\boldsymbol{v}_i) \times V_{A(i,i)}}{\sum_1^{N_s} W_{i,i}(v,\boldsymbol{v}_i)} \text{ ---------- Eq (26)}$$

3.3 Epistemic Uncertainty Estimation

Epistemic uncertainty quantification comes after the initial formulation of the correlation and aleatory covariance matrices in the proposed framework as shown in Fig. 1. The optimum values for hyperparameters (θ), neighborhood size of LWR process (kl), and epistemic uncertainty term (σ_E^2) of the model are calculated through the log-likelihood maximization using Eq. 27 (Rasmussen and Williams, 2006):

$$max_{\theta,\sigma_E^2} L = -\frac{1}{2}(N_s\, ln(2\pi) + ln(|\boldsymbol{V_{nd}}(\boldsymbol{\theta}, \sigma_E^2, kl)|) + \boldsymbol{Y}_\beta^T \boldsymbol{V}^{-1}(\boldsymbol{\theta}, \sigma_E^2, kl)\boldsymbol{Y}_\beta) \text{ ---------- Eq (27)}$$

Here $\boldsymbol{Y}_\beta = \boldsymbol{Y} - \boldsymbol{F}\boldsymbol{\beta_{nd}}$, and $\boldsymbol{\beta_{nd}}$ is also a function of $\boldsymbol{\theta}$ and σ_E^2. It was possible in conventional deterministic Kriging to reduce the optimization problem into multiple one-dimensional problems for individual $\theta_i, i = 1,2, \ldots\ldots, N_{n+q}$. However, in NDK due to the non-proportional scale of σ_E^2 in the likelihood function, the maximization needs to be performed across all model parameters simultaneously. The optimum values obtained for hyperparameters, epistemic uncertainty parameter, and neighborhood size are then used for calculating correlation matrix R and parameters related to aleatory covariance.

As the final step, the values obtained from sections 3.1, 3.2, and 3.3 can be used in Eqs 8 and 9 to obtain the estimated mean (y_{nd}) and prediction variance (σ_{nd}^2) respectively at any point in the design space. The prediction variance obtained through Eq. 9 consists of information for both aleatory and epistemic uncertainties of the model. In NDK it is expected that with enough training data points, the epistemic uncertainty of the NDK model converges to zero. Subsequently, all the uncertainty present in the NDK prediction variance (σ_{nd}^2) converges to the aleatory uncertainty in the response variable.

The inclusion of repeated training data points can make the correlation matrix singular. This is considered one of the drawbacks of deterministic Kriging models. In physical experiments, the test cases are repeated at the same experimental conditions to identify the effect of natural randomness. The proposed NDK method accommodates such repeats and estimates the mean response and uncertainties. For this purpose, the design of experiments ($\boldsymbol{D}$) was modified to eliminate the effect of repeats if there are any. Accordingly, the matrix of location vectors of input data $\mathbb{V}$, was rearranged to $\mathbb{V}_{\boldsymbol{mod}}$ by including only one entry for the repeated training data points. Subsequently, the response vector, $\boldsymbol{Y}$ was also modified to $\boldsymbol{Y}_{\boldsymbol{mod}}$. For example, consider there are repeats at the design point ($\boldsymbol{v}_0$) such that;

$$\boldsymbol{v}_0 = \boldsymbol{v}_j \text{ and } f(\boldsymbol{v}_j) = Y_j \; ; \boldsymbol{v}_j \in \boldsymbol{V},\; Y_j \in \boldsymbol{Y}, \text{ and } j = 1, 2, 3, \dots, p$$

where p is the number of repeats at $\boldsymbol{v}_0$. Out of all $\boldsymbol{v}_j$ data points, only $\boldsymbol{v}_0$ was included in the modified input variable vector $\mathbb{V}_{\boldsymbol{mod}}$. The corresponding response, Y_0 at the design point $\boldsymbol{v}_0$, was defined as;

$$Y_0 = \frac{\sum_{j=1}^{p} Y_j}{p} \; ; Y_0 \in \boldsymbol{Y}_{\boldsymbol{mod}} \text{ ------ Eq (28)}$$

Accordingly, the modified design of experiments ($\boldsymbol{D}_{mod} = \{\mathbb{V}_{mod}, \boldsymbol{Y}_{\boldsymbol{mod}}\}$) was used to calculate the correlation matrix, $\boldsymbol{R}$. However, for the aleatory uncertainty quantification, the design of experiments ($\boldsymbol{D}$) was used without any modification within the locally weighted regression process.

Elaborating on the limitation of the study, there are other discrete kernels that can be used in NDK models with mixed input variables in addition to the discussed kernels in this study. However, those kernels required a larger number of hyperparameters. The required number of hyperparameters can increase exponentially with the number of discrete variables and levels. Such increments result in higher computational time, thus increasing the cost of the model ultimately. However, there can be instances where the use of these discrete kernels becomes useful. For example, when there is a large discrete design space or the available training data is limited, choosing kernels such as the coregionalization matrix kernel (Pelamatti et al., 2021) can generate better predictions from the model. Even in the continuous design space, the requirement for a higher number of hyperparameters is a continuously investigated issue in the domain of Kriging models.

This study did not focus on the efficiency of the proposed NDK method for mixed input variables with discrete kernels that employ a larger number of hyperparameters and methods of reducing the impact of larger numbers of hyperparameters. Furthermore, the study did not investigate the performance of the proposed NDK model with systems that are classified as high-dimensional problems. Therefore, further research is needed to develop the proposed method with such high-dimensional problems. Also, this study did not focus on how the performance of the proposed model varies with the number of input variables in the system although test cases consisted of different numbers of input variables and combinations of variables.

4. Application of the proposed NDK method to analytical examples

This section presents the application of the proposed NDK methodology on several pre-defined probabilistic functions. The aim was to assess how the NDK for mixed variables performs on multi-dimensional data samples with a larger number of categorical combinations. For this purpose, two analytical functions with 4 and 8 dimensions were considered. In evaluating the performance, the estimated mean response at the unsampled points was compared with the actual values of the functions using a set of goodness-of-fit measures. To demonstrate the proposed method's capability of capturing the aleatory uncertainty in the noisy response, the predicted standard deviation given by the NDK model was compared with the actual standard deviation of the response due to noise (aleatory uncertainty). Furthermore, the predicted mean $\pm$ 3*standard deviation bounds were also compared to demonstrate the performance of the NDK model.

Accordingly, the R^2 value (Chicco et al., 2021), Normalized Root Mean Squared Error (NRMSE) (Patel and Ramachandran, 2015), and Normalized Maximum Absolute Error (NMAE) (Xu et al., 2020), were determined for each example using Eqs 29, 30, and 31, respectively.

$$R^2 = 1 - \frac{\frac{\sum_{i=1}^{n}(y_i'-y_i)^2}{n}}{\sigma_y^2} \text{ ------ Eq (29)}$$

$$NRMSE = \frac{\sqrt{\frac{\sum_{i=1}^{n}(y_i'-y_i)^2}{n}}}{y_{max}-y_{min}} \text{ ------ Eq (30)}$$

$$NMAE = \sqrt{\frac{max\{(y_i'-y_i)^2, i=1,..n\}}{\sigma_y^2}} \text{ ------ Eq (31)}$$

where y_i' and y_i are the predicted response and actual response at the *i*-th testing data point. In common practice, an R^2 value closer to 1 indicates high accuracy. Both NRMSE and NMAE values

512 can range from 0 to $+\infty$, while values closer to 0 in both metrics account for high accuracy in
513 predictions.

514 The evaluation of the goodness-of-fit measures for the NDK model was repeated with
515 randomly selected training and test data sets for each analytical case. These repetitions were done
516 to quantify and compensate for the effect of the random nature in the initial design of experiments.
517 In previous similar work, different numbers of repetitions were used for this purpose; 10 repetitions
518 (Pelamatti et al., 2019) and 20 repetitions (Pelamatti et al., 2021) to eliminate the possible
519 imprecision in the results of accuracy metrics. Following these previous work, 20 repetitions were
520 conducted in this study. At each iteration, the developed NDK model was used to obtain
521 predictions at randomly selected 2000 unsampled points from the input space. These random
522 training and test data sets were sampled by combining the continuous Latin Hypercube Sampling
523 (LHS) method (Mckay et al., 2000) and a sampling over a uniform discrete distribution in the
524 discrete input space (Pelamatti et al., 2021). When defining the training data sample, at each
525 training data point, 3 replications were considered to obtain three different response values at the
526 same data location as a result of random perturbation in the functions. A MATLAB script written
527 by the authors was used for building the NDK model for these two analytical examples. This script
528 was modified according to the kernel functions used for correlation matrix calculation and LWR
529 process in each example.

530 **4.1 Modified Branin function**

531 A modified version of the Branin function that consists of two continuous variables and
532 two discrete variables, each with two levels, was taken as the first analytical example. Due to the
533 two levels for each discrete variable, this four-dimensional function consists of four categorical
534 combinations. The modified version of the Branin function can be presented as follows;

$$f(x_1, x_2, z_1, z_2) = \begin{matrix} g(x_1, x_2) + h(x_1, x_2)\varepsilon & z_1 = 0 \; and \; z_2 = 0 \\ 0.5g(x_1, x_2) + 1 + 1.4h(x_1, x_2)\varepsilon & z_1 = 0 \; and \; z_2 = 1 \\ -0.6g(x_1, x_2) + 6.2 + 0.8h(x_1, x_2)\varepsilon & z_1 = 1 \; and \; z_2 = 0 \\ -0.4g(x_1, x_2) - 1.5 + 0.6h(x_1, x_2)\varepsilon & z_1 = 1 \; and \; z_2 = 1 \end{matrix} \text{---------- Eq (32)}$$

$$g(x_1, x_2) = \left[\left(\left(15x_2 - \frac{5}{4\pi^2}(15x_1 - 5)^2 + \frac{5}{\pi}(15x_1 - 5) - 6\right)^2 + 10\left(1 - \frac{1}{8\pi}\right)\cos(15x_1 - 5) + 10\right) - 54.8104\right]\frac{1}{51.9496} \text{ ---------- Eq (33)}$$

$$h(x_1, x_2) = |x_1 + x_2| \text{ ---------- Eq (34)}$$

where $x_1 \in [0,1], x_2 \in [0,1], \quad x_3 \in \{0,1\}, x_4 \in \{0,1\}$, and ε is a standard normal variable N (0,1). Fig. 2 visualizes the actual mean of the modified Branin function.

In this 4-D mixed variable function, both discrete variables are binary categorical variables. Therefore, out of the discrete kernels described in section 2, the weighted Gower distance kernel was used to develop the correlation matrix in the NDK model for this 4-D function. Furthermore, the Gaussian kernel coupled with the Racine and Li estimator was used for the locally weighted regression step in covariance matrix calculations. Both of these kernels consist of the same hyperparameter value for non-identical levels of a discrete variable and have the lowest number of hyperparameters compared to other kernels. With nonbinary discrete variables, these two kernels do not consider correlations between different discrete levels. This approach is similar to a category-wise approach. The use of such approaches requires a larger number of training data samples when the numbers of discrete variables and discrete levels increase. On the other hand, in this 4-D example, each discrete input variable has only two discrete levels. Given that, even the Latent variable kernel employs a similar category-wise approach based on how the Latent space is defined in this proposed methodology using a cartesian plane. However, since the weighted Gower distance kernel and Racine Li estimator are generally used with binary variables, these two kernels were used for calculations with discrete input variables in this 4-D function.

In this analytical example, 3 replications at each training data point were considered. Starting from a training data sample size of 60, the training sample size was gradually increased while evaluating the goodness-of-fit of estimations at 2000 unsampled data points. Out of the goodness-of-fit measures used, Figs. 3(a) and 3(b) present how the mean and median of R^2 values corresponding to the estimated mean and standard deviation of the function varied with the sample size of training data points. A summary of mean values obtained for the other goodness-of-fit measures is shown in Table 3.

The estimated mean showed a higher goodness-of-fit with a mean R^2 value exceeding 0.93 even with 60 training data samples. Thereafter R^2 values further increased gradually with increasing training data sample size while gaining goodness-of-fit for the estimated mean. However, the prediction of standard deviation in the developed NDK model required a higher number of training data samples to show a higher goodness of fit. For instance, at the training data sample size of 160, the mean R^2 value of the predicted standard deviation reached 0.8 for the first time. After that mean R^2 value of predicted standard deviation showed no significant variation at training sample sizes 180 and 200. Also, at a training sample size of 160, the mean R^2 value of the estimated mean response of the function reached a value over 0.96 and remained almost unchanged with increasing training data sample sizes. As shown in Table 3, a similar pattern can be observed in goodness-of-fit measures for mean $\pm$ 3*standard deviation bounds as well. Therefore, the results indicate that a training sample size between 180-200 provides accurate predictions of both the mean response and standard deviation for this 4-D probabilistic function.

4.2 Eight-Dimensional Powell Function

The next analytical case was set up with the Powell equation which consists of eight dimensions. Accordingly, the selected modified version of the Powell equation is characterized by

four continuous variables and four discrete variables, each with three different levels. Since this function consists of 81 categorical combinations with 4 continuous variables, it was aimed to evaluate the performance of the proposed NDK method, in systems with a larger number of dimensions and categorical combinations. The used 8-D function in the third analytical case can be presented as;

$$f(x_1, \dots, x_8) = g(x_1, \dots, x_8) + h(x_1, \dots, x_8) * \varepsilon \text{ ----------Eq (35)}$$

$$g(x_1, \dots, x_8) = \sum_{i=1}^{2} \begin{matrix}[(x_{4i-3} + 10x_{4i-2})^2 + 5(x_{4i-1} - x_{4i})^2 \\ +(x_{4i-2} - 2x_{4i-1})^4 + 10(x_{4i-3} - x_{4i})^4]\end{matrix} \text{----------Eq (36)}$$

$$h(x_1, \dots, x_8) = \sum_{i=1}^{2} \begin{matrix}[0.8(x_{4i-3} + 10x_{4i-2})^2 + 4.5(x_{4i-1} - x_{4i})^2 \\ +0.7(x_{4i-2} - 2x_{4i-1})^4 + 9(x_{4i-3} - x_{4i})^4]\end{matrix} \text{----------Eq (37)}$$

where $x_1, x_2, x_3, x_4 \in \left[-\frac{\pi}{20}, \frac{\pi}{20}\right]$, $(x_5, x_6, x_7, x_8) \in \{\pi/60, \pi/30, \pi/20\}$, and ε is a standard normal variable N (0,1).

There are 3 discrete levels in each of the discrete levels resulting in 81 categorical combinations. Employing the weighted Gower distance kernel and Racine and Li estimator for this analytical function requires a larger number of training data points. Therefore, the Latent variable kernel was used for discrete variables in this function since Latent variable kernels consider correlations between non-identical discrete levels as well. The initial training data sample consisted of 50 randomly selected samples from the above-introduced input space. Using the developed NDK method, mean values were estimated at 2000 testing data points. Considering the higher number of discrete variables and categorical combinations available in the test case, the size of the training data sample was increased by 50 at each time and the goodness-of-fit measures were calculated for predictions. Fig. 4 presents the variation of calculated mean R^2 values for the

599 estimated mean and standard deviation with the training data sample size. Table 4 presents a
600 summary of other goodness-of-fit measures used for evaluating the performance of the NDK
601 model for 8-D function. The results obtained for the goodness-of-fit measures related to
602 predictions of the 8-D function also showed a similar pattern compared to the previous analytical
603 case. After a training data sample size of 200, no significant variation was observed in the
604 goodness-of-fit measures for both the estimated mean and standard deviation of the 8-D function.

605 In addition to that, the mean R^2 value for the predicted mean -3*standard deviation bound
606 showed considerably low values in this 8-D function although it improved when the size of training
607 data points was increased. Furthermore, this bound showed large values for the NMAE metric as
608 well. Although low R^2 values and large NMAE values indicated a low prediction accuracy,
609 NRMSE value showed considerably satisfactory values for the mean -3*standard deviation
610 bound. Delving into the obtained results showed that the variance of the mean -3*standard
611 deviation bound is low compared to the mean + 3* standard deviation bound. Therefore, based on
612 how R^2 is calculated, corresponding R^2 values for predicted mean -3*standard deviation bound
613 show lower values with a relatively smaller variance value, irrespective of the goodness-of-fit in
614 predictions. Due to the same reason, the NMAE metric also yields large mean values.

615 **5. Testing the proposed NDK model with engineering applications**

616 In order to evaluate the applicability of the proposed NDK for mixed input variables
617 methodology in actual engineering applications, it was applied to two engineering problems; 1) a
618 physical experiment for wave forces on an elevated coastal structure and 2) numerical modeling
619 for the performance of bridges under earthquake excitations. The following sub-sections present
620 the results of applying the proposed NDK method to the two engineering cases.

5.1 A physical experiment for wave forces on an elevated coastal structure

As the first engineering application, the NDK model was applied to estimate wave forces on an elevated coastal structure obtained using physical experiments conducted by Park et al. (2017). During this experiment, data was collected by measuring the wave-induced forces on an idealized coastal structure. Three different types of waves were used for the experiment with varying significant wave heights (H) and peak periods (T). Furthermore, the experiment included the air gap (a) between the idealized elevated structure and the initial water level as a variable. However, this experiment used a fixed water height (h) for all the test cases. Fig. 5 presents the experimental setup used for test cases. Accordingly, the experiment consisted of four variables namely, wave type, wave height, peak period, and air gap between the structure and water level. The type of waves was the discrete variable with wave types namely, regular, irregular, and transient (tsunami type) waves. The other three variables were considered continuous variables.

Table 5 presents the combinations of test cases used for the experiments based on wave heights and peak periods. For the test cases presented in Table 5, the air gap between the structure and water level was selected from different values (a_0, a_1, …., a_9) given in Table 6. For some combinations in Table 5, all the air gap values were used while some combinations were performed only with a selected number of air gap values. With the given details above, 236 experimental cases were conducted. For each test case vertical and horizontal forces on the idealized structure were measured using load cells f_y and f_x respectively as shown in Fig. 5. Typically, forces induced by these extreme events are modeled using the lognormal distribution (Melchers and Beck, 2018). Therefore, the logarithms of maximum loads on the elevated structure are assumed to be following a random Gaussian process. In this test case, the logarithm of maximum horizontal force was taken as the dependent variable, which was predicted herein.

During the application of the proposed NDK methodology in the above-explained experiment, 150 data points were considered for training while 70 data points were taken for testing. Since there are three discrete levels in the discrete input variable of this test case, the use of a category-wise approach demands a higher number of training data samples. However, there are only 150 data points to be used to train the NDK model. Furthermore, the Gower distance kernel combined with the Racine and Li estimator neglects any possible correlations between non-identical discrete levels (the wave types in this example). In such cases, the performance of the model can be affected due to the scarcity of data. Therefore, in developing the NDK model for this problem, the Latent variable kernel was used for the discrete variable during both correlation matrix calculation and the LWR process since it considers the correlations between non-identical discrete levels as well.

Following the approach used in previous numerical examples, 20 repetitions were performed on randomly selected training and testing data points. The previously described goodness-of-fit measures were used to evaluate the prediction accuracy. At each iteration, randomly selected 150 data points were used for training and 70 data points were used for testing. A mean R^2 value of 0.80 was obtained for the estimated horizontal forces at the testing data points. Since this was an experimental case, actual standard deviations were not available to calculate the goodness-of-fit measures for predicted standard deviation values. However, the values obtained for the goodness-of-fit measures corresponding to predicted standard deviations in previous numerical examples suggest that predictions for standard deviations in this experimental case would have been reasonable. Furthermore, the mean values obtained for NRMSE and NMAE metrics corresponding to the predicted logarithm of horizontal forces were 0.099 and 2.1908 respectively.

5.2 Numerical modeling of bridge performance

As the second engineering application for the proposed NDK method, a numerical simulation of bridge performance under earthquake conditions was selected. The data on the responses and bridge parameters was obtained from Kameshwar et al. (2019). Non-linear time history analysis for seismic response was conducted in OpenSees, a finite element software. The bridge components such as columns and bent were modeled using fiber-based elements. The bridge deck was modeled using a grillage; and the bearings, abutments, and foundations were modeled using non-linear springs. The response of these bridge components was recorded. During these simulations, a wide range of bridge material and geometric properties were varied, resulting in 45 variables as shown in Table 7. A total of 1044 and 108 bridge simulations were conducted to obtain training and testing datasets, respectively. Depending on the degrees of freedom in the model, each of these simulations required 3 to 24 hours on a single CPU. Since these simulations require a lot of computational time, the seismic response was used to demonstrate the usefulness of the proposed NDK method.

Although the original data set consisted of 45 variables, the NDK model developed for this problem considered 17 continuous variables and 3 discrete variables resulting in only 20 input variables altogether. These 20 input variables are marked with a (*) in Table 7. The remaining set of variables were considered uniform random variables in the problem. Out of the 20 input variables for the NDK mode, values for continuous input variables were normalized in their ranges, except for discrete variables and peak ground acceleration (PGA). In this example, the discrete input variables have 8, 5, and 18 discrete levels respectively. In such conditions, using both kernels with a category-wise approach and Latent variable kernels has disadvantages. These category-wise approaches demand a larger number of training data while the use of Latent variable kernels

690 increases the number of hyperparameters in the model, thus increasing the computational cost. In
691 this model, two separate NDK models were built using both types of kernels to compare their
692 effectiveness in this type of condition. Accordingly, one NDK model consisted of the weighted
693 Gower distance kernel in correlation matrix calculation and the Racine and Li estimator in the
694 LWR process for discrete input variables. The other NDK model consisted of Latent variable
695 kernels for discrete input variables in both correlation matrix calculation and LWR process.

696 This simulation consisted of 956 data points for the training purpose and 95 data points for
697 the testing purpose out of all the simulations. Some of the seismic simulations did not converge;
698 therefore, the number of usable data points in the test and training set is lower than in the original
699 testing and training data sets. No repetitions were done in this test case since the data set already
700 consisted of designated training and test data sets. In a similar way to the previous engineering
701 application, the dependent variable of this system, maximum relative drifts of the bridge columns,
702 were assumed to be in a lognormal distribution and their logarithm values follow a random
703 Gaussian process. Therefore, goodness-of-fit measures were calculated for the prediction of
704 logarithm values of the maximum relative drift in bridge columns at the 95 testing data points.
705 Since the training and test data points were already designated, the predictions using the NDK
706 model were not repeated like in previous test cases.

707 NDK models that included kernels with a category-wise approach and Latent variable
708 kernels indicated R^2 values of 0.77 and 0.68 respectively. The results did not show an improvement
709 in using Latent variable kernels over the weighted Gower distance kernel and Racine and Li
710 estimator. This feature was also observed in NRMSE values as well. The two models yielded
711 values of 0.11 and 0.13 for the NRMSE metric respectively. However, for the NMAE metric, both
712 NDK models yielded a value of 1.4. The results suggested that there was enough training data

713 within this specific data set for the category-wise approach to satisfactorily capture the effect of
714 discrete input variables. The NDK model with Latent variable kernels produced a smaller R^2 value
715 and a larger NRMSE indicating less accurate predictions compared to the category-wise kernels.
716 The limited capacity of Latent variables to effectively represent relationships between a higher
717 number of discrete levels in a 2-D Latent space can be a possible reason for this difference. Also,
718 the higher computational cost due to the large number of hyperparameters indicated that the use
719 of Latent variable kernels in this test case is not cost-effective. However, it is also important to
720 note that this may vary if the NDK model is developed over a different set of data even from the
721 same numerical model. Similar to the first engineering applications, there are no actual values for
722 the system uncertainty at design points to determine the accuracy of predicted uncertainties.
723 However, the results from the analytical functions suggest that reasonable values would have been
724 estimated for the standard deviations as well.

725 The proposed NDK model showed satisfactory performances in predicting the seismic
726 response of the bridges. However, it is important to note that predictions from the NDK model do
727 not support detailed bridge designs under seismic loads. The use of such an NDK model is more
728 suitable for identifying uncertainties associated with bridges when carrying out risk assessments
729 and mitigation measures in a holistic manner. Since the NDK model provides separate estimates
730 of both epistemic and aleatory uncertainties as a function of input variables, the homoscedasticity
731 assumption, which is commonly used in seismic fragility models of bridges can be avoided.
732 Therefore, it leads to better fragility models, thus resulting in better risk assessments. To
733 demonstrate the usefulness of the proposed NDK model outcomes in practical applications, the
734 developed model was employed in a case study for the seismic risk assessment of a bridge.

735 Accordingly, the annual probability of seismic failure was estimated for a bridge in Charleston,
736 South Carolina using the methodology presented by Kameshwar and Padgett (2014).

737 Data obtained from USGS (Petersen et al., 2008) for Charleston (32.8N 79.9W), were used
738 for the seismic hazard curve. The failure probability of the bridge at any given Peak Ground
739 Acceleration (PGA) value was predicted using the trained NDK model which consisted of the
740 Gower distance kernel and Racine and Li estimator. A maximum of 1.2g was taken for the PGA
741 in the region of interest (Fernandez and Rix, 2012). The characteristics of the bridge employed in
742 the case study are given in Table 9. The values of the set of variables marked with (*) were obtained
743 from (Kameshwar and Padgett, 2014). The other parameter values were assumed from the ranges
744 of corresponding input variables in the data set used for the NDK model. More details of the
745 employed method and the considered case study can be found in (Kameshwar and Padgett, 2014).
746 The complete damage state was considered for the relative drift capacity of the columns. The limit
747 state was defined with a mean value of 5.0 and 0.35 logarithmic standard deviation (Xie et al.,
748 2019; Yi et al., 2007). For a given PGA, the annual failure probability was estimated using the
749 Monte Carlo Simulation with 1.0E+06 samples. Finally, the annual failure probability of the
750 selected bridge was approximated as 3.86E-05 in the range of 0.28-1.2g for PGA. This value aligns
751 with the seismic risk of the particular bridge, 5.57E-05 which is based on column curvature
752 ductility (Kameshwar and Padgett 2014).

753 **6. Discussion of results for analytical cases**

754 In the previous two sections, the proposed NDK method for mixed variables was tested on
755 a set of analytical functions and engineering problems. Overall, the results from numerical
756 examples showed that mixed variable NDK provided accurate predictions for mean and standard

757 deviation including both aleatory and epistemic uncertainties in probabilistic systems. All the
758 metrics calculated for predictions on each numerical case showed satisfactory values confirming
759 goodness-of-fit in both mean response and standard deviation. However, to achieve a satisfactory
760 level of the goodness-of-fit for the standard deviation of analytical cases, a larger number of
761 training data points were required in comparison to mean responses. Although goodness-of-fit
762 measures were not calculated for standard deviations in the two engineering applications,
763 predictions of mean responses matched the actual value at a satisfactory level.

764 In this proposed NDK method for mixed input variables, the selection of discrete kernel
765 functions in the correlation matrix and LWR plays a major role. Table 8 shows the discrete kernels
766 used for these two processes in each of the numerical examples and the engineering application.
767 The first numerical example consisted of the weighted Gower distance kernel for the correlation
768 matrix. Racine and Li estimator was chosen to use in LWR for aleatory uncertainty calculation.
769 Both of these kernels consist of the same hyperparameter value for non-identical levels of a
770 discrete variable, thus employing a category-wise approach. The number of training data points
771 required by a Kriging model with this type of kernel increases when the number of discrete
772 variables and discrete categories in the problem increases. The first numerical example is
773 characterized by binary categorical discrete variables and contains only four categorical
774 combinations. This is a special case where even Latent variable kernels also employ a category-
775 wise approach ultimately since both discrete variables are binary categorical variables. However,
776 the weighted Gower distance kernel and Racine and Li estimator are used with binary input
777 variables in general. Therefore, these two kernels were used in the first numerical example while
778 keeping the number of hyperparameters and computational cost as low as possible.

However, the second numerical example and engineering application I consisted of discrete variables that have more than 2 discrete levels (non-binary variables). The modified 8-D Powell function in the second numerical example consisted of 3 discrete variables with 3 levels in each of them. The first engineering application also had a discrete variable with 3 discrete levels. The weighted Gower distance kernel and Racine and Li estimator demand a larger number of training data points in these types of systems because these kernels do not consider correlations between non-identical discrete levels. Therefore, the Latent variable kernel was used in the two NDK models for numerical example II and engineering application I since this kernel considers correlations between non-identical levels in a non-binary discrete variable. In these two examples, the Latent variable kernel was used for both correlation matrix and aleatory uncertainty calculations. However, the second engineering application led to a situation where a decision had to be made between the accuracy of estimations and the computational cost of the NDK model when choosing the types of kernels for the NDK model. Therefore, a comparison was made between the accuracy of NDK models developed with both types of kernels. The results show that the NDK model with the weighted Gower distance kernel and Racine and Li estimator has rendered satisfactory accuracy in estimations given that 956 data points were available for training. The NDK model with Latent variable kernels indicated the need for Latent variables in higher dimensional spaces which ultimately increases its large demand for computational power due to the large number of hyperparameters in the NDK model. Therefore, altogether the results from both numerical examples and engineering applications suggest that the selection of the discrete kernels depends on the size of the available training data sample, the number of discrete variables, and their discrete levels.

6 Conclusion

In this paper, a Non-Deterministic Kriging methodology was proposed for probabilistic systems with mixed continuous and discrete variables. The proposed methodology has the capability of predicting both epistemic and aleatory uncertainties along with the mean response. Furthermore, this method considers the effect of aleatory uncertainty in the computations for mean response estimation as well. Compared to Stochastic Kriging, which considers the effect of aleatory uncertainty in predictions, the proposed method does not require a larger number of replications at the same design point. During the construction of the correlation matrix, both continuous and discrete kernels were combined to incorporate the influence of mixed variables. Within this proposed model, aleatory uncertainty, which is an additional estimation compared to conventional Kriging, is estimated using locally weighted regression. To capture the effect of discrete variables on the local weights, continuous and discrete kernels were combined together. Furthermore, the proposed methodology can process replications at the same training data point, which is a common condition, especially with physical experiments due to inherent randomness.

This approach was tested on two probabilistic numerical examples and two engineering examples using three goodness-of-fit measures to determine its effectiveness in the prediction of both mean response and standard deviation. According to the results, the proposed NDK methodology shows better goodness-of-fit in both mean response and standard deviation predictions. Two engineering applications demonstrate the usefulness of the proposed method with large-scale physical experiments and computer simulations which are common in engineering fields. Furthermore, in the second engineering application, the trained NDK model was further used for a risk assessment of a bridge to demonstrate one of the potential uses of the proposed NDK method. However, it is important to note that the performance of the model depends on the nature of the problem and the selection of discrete kernels as well. The selection of discrete kernels

825 can be generally done based on the number of discrete variables and levels in the system and the
826 number of available training data points. The computational time and cost of the model depend on
827 the discrete kernel used since it largely contributes to the number of hyperparameters in the model.
828 This study did not focus on other discrete kernels and handling a larger number of hyperparameters
829 in the proposed NDK model. With adequate research, the proposed method can be assisted as a
830 supervised learning tool in machine learning methods for optimization problems under the
831 complexity of mixed input variables and natural stochasticity.

832 **7. Data availability statement**

833 Some or all data, models, or codes that support the findings of this study are available from the
834 corresponding author upon reasonable request.

835 **8. Acknowledgment**

836 The authors would like to gratefully acknowledge the support of this research by the National
837 Science Foundation (NSF) under Award No. 2203116. Any opinions, findings, conclusions, or
838 recommendations expressed in this material are those of the authors and do not necessarily reflect
839 the views of the National Science Foundation.

840 **List of Symbols**

$\boldsymbol{c}$ - vector of linear predictor

$\boldsymbol{D}$ - design of experiments

$\boldsymbol{D}_{mod}$ – modified design of experiments without repeats

d_{gow}- Gower distance

$\mathbb{V}_{mod}$ - modified matrix of input data locations without repeats

$\boldsymbol{V_A}$ - aleatory covariance matrix

$\mathbf{v_A}$ - aleatory covariance vector.

$\boldsymbol{v}$ – location vector of an input data point

e_i- uncorrelated correlation error in locally weighted regression (LWR)

$\boldsymbol{f}$- basis function vector at the unsampled point

$\boldsymbol{F}$ - regression design matrix at training data points

kl - size of neighborhood in LWR

$\widetilde{m}_l$- standard Nadaraya-Watson estimator in LWR

N_s – number of training data points

N_{r+q} – number of all input variables

q – number of discrete input variables

R - correlation among data points

$\boldsymbol{R_E}$ - correlation matrix

$\boldsymbol{r_E}$ - correlation vector

n – number of continuous input variables

t - Latent variable

T - Latent space

$\mathbb{V}$ - matrix of input data locations

$\boldsymbol{W}^l$- diagonal matrix of the local polynomial regression weights

$\boldsymbol{x}$ – vector of discrete input variables

y_{nd} – Non-Deterministic Kriging (NDK) prediction

$\boldsymbol{Y}$ – response vector in the design of experiments

$\boldsymbol{Y_{mod}}$-modified response vector without repeats

$\boldsymbol{z}$ – vector of discrete input variables

Z_0 - zero-mean stochastic process

Z_E - stochastic process of epistemic uncertainty

Z_A - stochastic process of aleatory uncertainty

$\boldsymbol{\beta}_{nd}$- regression coefficient vector

θ_l- hyperparameter in the *l* th dimension

σ_{nd}^2- prediction variance of NDK model

σ_E^2 - epistemic variance

σ_A^2 - aleatory variance

References

Abbas, A.E.H.A., Khattab, M.R., Abdel Azeem, M.M., 2018. Natural radionuclides distribution and environmental impacts of ferruginous sand-siltstone (raw material) and their manufactured Ahmer oxide used as wall paints. Environmental Forensics 19, 217–224. https://doi.org/10.1080/15275922.2018.1485791

Aljuhani, K.H., Turk, L.I.A., 2014. Modification of the adaptive Nadaraya-Watson kernel regression estimator. SRE 9, 966–971. https://doi.org/10.5897/SRE2014.6121

An, H., Youn, B.D., Kim, H.S., 2022a. Optimal placement of non-redundant sensors for structural health monitoring under model uncertainty and measurement noise. Measurement 204, 112102. https://doi.org/10.1016/j.measurement.2022.112102

An, H., Youn, B.D., Kim, H.S., 2022b. A methodology for sensor number and placement optimization for vibration-based damage detection of composite structures under model uncertainty. Composite Structures 279, 114863. https://doi.org/10.1016/j.compstruct.2021.114863

An, H., Youn, B.D., Kim, H.S., 2021. Reliability-based Design Optimization of Laminated Composite Structures under Delamination and Material Property Uncertainties. International Journal of Mechanical Sciences 205, 106561. https://doi.org/10.1016/j.ijmecsci.2021.106561

Ankenman, B., Nelson, B.L., Staum, J., 2008. Stochastic kriging for simulation metamodeling, in: 2008 Winter Simulation Conference. Presented at the 2008 Winter Simulation Conference, pp. 362–370. https://doi.org/10.1109/WSC.2008.4736089

Asher, M.J., Croke, B.F.W., Jakeman, A.J., Peeters, L.J.M., 2015. A review of surrogate models and their application to groundwater modeling. Water Resources Research 51, 5957–5973. https://doi.org/10.1002/2015WR016967

Bae, H., Clark, D.L., Forster, E.E., 2019. Nondeterministic Kriging for Engineering Design Exploration. AIAA Journal 57, 1659–1670. https://doi.org/10.2514/1.J057364

Boser, B.E., Guyon, I.M., Vapnik, V.N., 1992. A training algorithm for optimal margin classifiers, in: Proceedings of the Fifth Annual Workshop on Computational Learning Theory. pp. 144–152.

Broomhead, D., Lowe, D., 1988. Multivariable Functional Interpolation and Adaptive Networks. Complex Systems 2, 321–355.

Chicco, D., Warrens, M.J., Jurman, G., 2021. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Comput Sci 7, e623. https://doi.org/10.7717/peerj-cs.623

Choi, S.-K., Grandhi, R., Robert A. Canfield, 2006. Reliability-based Structural Design, 2007th edition. ed. Springer, London.

Clark, D.L., 2019. Non-Deterministic Metamodeling for Multidisciplinary Design Optimization of Aircraft Systems Under Uncertainty. Wright State University.

Cuesta Ramirez, J., Le Riche, R., Roustant, O., Perrin, G., Durantin, C., Glière, A., 2022. A comparison of mixed-variables Bayesian optimization approaches. Adv. Model. and Simul. in Eng. Sci. 9, 6. https://doi.org/10.1186/s40323-022-00218-8

Delage, T., Zannane, S., Neveux, T., 2022. Metamodeling of chemical engineering unit operations using Kriging and prediction error estimation, in: Montastruc, L., Negny, S. (Eds.), Computer Aided Chemical Engineering, 32 European Symposium on Computer Aided Process Engineering. Elsevier, pp. 535–540. https://doi.org/10.1016/B978-0-323-95879-0.50090-4

Di Maio, F., Belotti, M., Volpe, M., Selva, J., Zio, E., 2022. Parallel density scanned adaptive Kriging to improve local tsunami hazard assessment for coastal infrastructures. Reliability Engineering & System Safety 222, 108441. https://doi.org/10.1016/j.ress.2022.108441

Dunn, P.F., 2019. Measurement, Data Analysis, and Sensor Fundamentals for Engineering and Science. CRC Press.

Fernandez, J.A., Rix, G.J., 2012. Seismic Hazard Analysis and Probabilistic Ground Motions in the Upper Mississippi Embayment 1–10. https://doi.org/10.1061/40975(318)8

Friedman, J.H., 1991. Multivariate Adaptive Regression Splines. The Annals of Statistics 19, 1–67. https://doi.org/10.1214/aos/1176347963

Gajewicz-Skretna, A., Kar, S., Piotrowska, M., Leszczynski, J., 2021. The kernel-weighted local polynomial regression (KwLPR) approach: an efficient, novel tool for development of QSAR/QSAAR toxicity extrapolation models. Journal of Cheminformatics 13, 9. https://doi.org/10.1186/s13321-021-00484-5

Gower, J.C., 1971. A General Coefficient of Similarity and Some of Its Properties. Biometrics 27, 857–871. https://doi.org/10.2307/2528823

Halstrup, M., 2016. Black-box optimization of mixed discrete-continuous optimization problems. Technische Universität Dortmund.

Hao, W., Shaoping, W., Tomovic, M.M., 2010. Modified Sequential Kriging Optimization for Multidisciplinary Complex Product Simulation. Chinese Journal of Aeronautics 23, 616–622. https://doi.org/10.1016/S1000-9361(09)60262-4

Hengl, T., Heuvelink, G.B.M., Stein, A., 2004. A generic framework for spatial prediction of soil variables based on regression-kriging. Geoderma 120, 75–93. https://doi.org/10.1016/j.geoderma.2003.08.018

Hong, L., Li, H., Fu, J., 2022. Novel Kriging-Based Variance Reduction Sampling Method for Hybrid Reliability Analysis with Small Failure Probability. ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering 8, 04022017. https://doi.org/10.1061/AJRUA6.0001231

Huang, T., Bisram, M., Li, Y., Xu, H., Zeng, D., Su, X., Cao, J., Chen, W., 2023. Mixed-Variable Concurrent Material, Geometry, and Process Design in Integrated Computational Materials Engineering, in: Rabczuk, T., Bathe, K.-J. (Eds.), Machine Learning in Modeling and Simulation: Methods and Applications, Computational Methods in Engineering & the Sciences. Springer International Publishing, Cham, pp. 395–426. https://doi.org/10.1007/978-3-031-36644-4_11

Hutter, F., Hoos, H.H., Leyton-Brown, K., 2011. Sequential Model-Based Optimization for General Algorithm Configuration, in: Coello, C.A.C. (Ed.), Learning and Intelligent Optimization, Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 507–523. https://doi.org/10.1007/978-3-642-25566-3_40

Jiang, P., Zhang, Y., Zhou, Q., Shao, X., Hu, J., Shu, L., 2018. An adaptive sampling strategy for Kriging metamodel based on Delaunay triangulation and TOPSIS. Appl Intell 48, 1644–1656. https://doi.org/10.1007/s10489-017-1031-z

John Shawe-Taylor, Nello Cristianini, 2004. Kernel Methods for Pattern Analysis. Cambridge University Press, Cambridge, UK.

Jones, D.R., Schonlau, M., Welch, W.J., 1998. Efficient Global Optimization of Expensive Black-Box Functions. Journal of Global Optimization 13, 455–492. https://doi.org/10.1023/A:1008306431147

Kabir, A.M., Langsfeld, J.D., Zhuang, C., Kaipa, K.N., Gupta, S.K., 2017. A systematic approach for minimizing physical experiments to identify optimal trajectory parameters for robots, in: 2017 IEEE International Conference on Robotics and Automation (ICRA). Presented at the 2017 IEEE International Conference on Robotics and Automation (ICRA), pp. 351–357. https://doi.org/10.1109/ICRA.2017.7989045

Kahrizi, E., Rajaee, T., Sedighi, M., 2022. Probabilistic and Experimental Investigation of the Effect of Mineral Adsorbents on Porous Concrete Using Kriging, PRSM, and RBF

939 Methods. ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A:
940 Civil Engineering 8, 04022047. https://doi.org/10.1061/AJRUA6.0001258
941 Kameshwar, S., Padgett, J.E., 2014. Multi-hazard risk assessment of highway bridges subjected to
942 earthquake and hurricane hazards. Engineering Structures, Performance Based
943 Engineering: Current Advances and Applications 78, 154–166.
944 https://doi.org/10.1016/j.engstruct.2014.05.016
945 Kameshwar, S., Vishnu, N., Padgett, J., 2019. Response and Fragility Modeling of Aging Bridges
946 Subjected to Earthquakes and Truck Loads. https://doi.org/10.17603/ds2-5tzv-qz91
947 Khan, M.A.Z., 2011. Transient engine model for calibration using two-stage regression approach
948 (thesis). Loughborough University.
949 Kianifar, M.R., Campean, F., 2020. Performance evaluation of metamodelling methods for
950 engineering problems: towards a practitioner guide. Struct Multidisc Optim 61, 159–186.
951 https://doi.org/10.1007/s00158-019-02352-1
952 Kitahara, M., Bi, S., Broggi, M., Beer, M., 2021. Bayesian Model Updating in Time Domain with
953 Metamodel-Based Reliability Method. ASCE-ASME Journal of Risk and Uncertainty in
954 Engineering Systems, Part A: Civil Engineering 7, 04021030.
955 https://doi.org/10.1061/AJRUA6.0001149
956 Kohonen, T., 1982. Self-organized formation of topologically correct feature maps. Biological
957 cybernetics 43, 59–69.
958 Koziel, S., Pietrenko-Dabrowska, A., 2022. Performance-Driven Yield Optimization of High-
959 Frequency Structures by Kriging Surrogates. Applied Sciences 12, 3697.
960 https://doi.org/10.3390/app12073697
961 Krige, D., 1951. A statistical approach to some basic mine valuation problems on the
962 Witwatersrand 2, 119–139.
963 Kumar, P., Rao, B., Burman, A., Kumar, S., Samui, P., 2023. Spatial variation of permeability and
964 consolidation behaviors of soil using ordinary kriging method. Groundwater for
965 Sustainable Development 20, 100856. https://doi.org/10.1016/j.gsd.2022.100856
966 Lanckriet, G., Cristianini, N., Cristianini, P., Ghaoui, L., Jordan, M., 2004. Learning the Kernel
967 Matrix with Semidefinite Programming. Journal of Machine Learning Research 5, 27–72.
968 Li, C., Lu, Z., Ma, T., Zhu, X., 2009. A simple kriging method incorporating multiscale
969 measurements in geochemical survey. Journal of Geochemical Exploration 101, 147–154.
970 https://doi.org/10.1016/j.gexplo.2008.06.003
971 Li, Q., Racine, J., 2014. Cross-Validated Local Linear Nonparametric Regression 14, 485–512.
972 Li, Q., Racine, J., 2003. Nonparametric estimation of distributions with categorical and continuous
973 data. Journal of Multivariate Analysis 86, 266–292. https://doi.org/10.1016/S0047-
974 259X(02)00025-8
975 Lopez, R.H., Bismut, E., Straub, D., 2022. Stochastic efficient global optimization with high noise
976 variance and mixed design variables. J Braz. Soc. Mech. Sci. Eng. 45, 7.
977 https://doi.org/10.1007/s40430-022-03920-1
978 Loquin, K., Dubois, D., 2010. Kriging and Epistemic Uncertainty: A Critical Discussion, in:
979 Jeansoulin, R., Papini, O., Prade, H., Schockaert, S. (Eds.), Methods for Handling
980 Imperfect Spatial Information, Studies in Fuzziness and Soft Computing. Springer, Berlin,
981 Heidelberg, pp. 269–305. https://doi.org/10.1007/978-3-642-14755-5_11
982 Matheron, G., 1962. Traité de géostatistique appliquée. Editions Technip.

Mckay, M.D., Beckman, R.J., Conover, W.J., 2000. A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output From a Computer Code. Technometrics 42, 55–61. https://doi.org/10.1080/00401706.2000.10485979

Melchers, R., Beck, A., 2018. Structural Reliability Analysis and Prediction, 1st ed. John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119266105

Mohammadi, H., 2016. Kriging-based black-box global optimization : analysis and new algorithms (phdthesis). Université de Lyon.

Moore, R.A., Romero, D.A., Paredis, C.J.J., 2014. Value-Based Global Optimization. Journal of Mechanical Design 136. https://doi.org/10.1115/1.4026281

Mortazavi, M., Kuczera, G., Cui, L., 2012. Multiobjective optimization of urban water resources: Moving toward more practical solutions. Water Resources Research 48. https://doi.org/10.1029/2011WR010866

Moustapha, M., Galimshina, A., Habert, G., Sudret, B., 2022. Multi-objective robust optimization using adaptive surrogate models for problems with mixed continuous-categorical parameters. Struct Multidisc Optim 65, 357. https://doi.org/10.1007/s00158-022-03457-w

Mukhopadhyay, T., Dey, T.K., Chowdhury, R., Chakrabarti, A., 2015. Structural Damage Identification Using Response Surface-Based Multi-objective Optimization: A Comparative Study. Arab J Sci Eng 40, 1027–1044. https://doi.org/10.1007/s13369-015-1591-3

Musiol, G., 1997. An Introduction to Categorical Data Analysis. Computational Statistics & Data Analysis 23, 565–563.

Nazeeh, K.M., Dilip, D.M., Sivakumar Babu, G.L., 2023. Quantile-Based Design and Optimization of Shallow Foundation on Cohesionless Soil Using Adaptive Kriging Surrogates. International Journal of Geomechanics 23, 06023014. https://doi.org/10.1061/IJGNAI.GMENG-8226

Palar, P.S., Liem, R.P., Zuhal, L.R., Shimoyama, K., 2019. On the use of surrogate models in engineering design optimization and exploration: the key issues, in: Proceedings of the Genetic and Evolutionary Computation Conference Companion, GECCO '19. Association for Computing Machinery, New York, NY, USA, pp. 1592–1602. https://doi.org/10.1145/3319619.3326813

Park, H., Tomiczek, T., Cox, D.T., van de Lindt, J.W., Lomonaco, P., 2017. Experimental modeling of horizontal and vertical wave forces on an elevated coastal structure. Coastal Engineering 128, 58–74. https://doi.org/10.1016/j.coastaleng.2017.08.001

Patel, S.S., Ramachandran, P., 2015. A Comparison of Machine Learning Techniques for Modeling River Flow Time Series: The Case of Upper Cauvery River Basin. Water Resour Manage 29, 589–602. https://doi.org/10.1007/s11269-014-0705-0

Pelamatti, J., Brevault, L., Balesdent, M., Talbi, E.-G., Guerin, Y., 2021. Mixed Variable Gaussian Process-Based Surrogate Modeling Techniques: Application to Aerospace Design. Journal of Aerospace Information Systems 18, 813–837. https://doi.org/10.2514/1.I010965

Pelamatti, J., Brevault, L., Balesdent, M., Talbi, E.-G., Guerin, Y., 2020. Overview and Comparison of Gaussian Process-Based Surrogate Models for Mixed Continuous and Discrete Variables: Application on Aerospace Design Problems, in: Bartz-Beielstein, T., Filipič, B., Korošec, P., Talbi, E.-G. (Eds.), High-Performance Simulation-Based Optimization, Studies in Computational Intelligence. Springer International Publishing, Cham, pp. 189–224. https://doi.org/10.1007/978-3-030-18764-4_9

Pelamatti, J., Brevault, L., Balesdent, M., Talbi, E.-G., Guerin, Y., 2019. Efficient global optimization of constrained mixed variable problems. J Glob Optim 73, 583–613. https://doi.org/10.1007/s10898-018-0715-1

Petersen, M., Frankel, A., Harmsen, S., Mueller, C., Haller, K., Wheeler, R., Wesson, R., Zeng, Y., Boyd, O., Perkins, D., Luco, N., Field, E., Wills, C., Rukstales, K., 2008. Documentation for the 2008 Update of the United States National Seismic Hazard Maps. USGS Open-File Report 2008–1128.

Picheny, V., Wagner, T., Ginsbourger, D., 2013. A benchmark of kriging-based infill criteria for noisy optimization. Struct Multidisc Optim 48, 607–626. https://doi.org/10.1007/s00158-013-0919-4

Qian, J., Yi, J., Cheng, Y., Liu, J., Zhou, Q., 2020. A sequential constraints updating approach for Kriging surrogate model-assisted engineering optimization design problem. Engineering with Computers 36, 993–1009. https://doi.org/10.1007/s00366-019-00745-w

Rasmussen, C.E., Williams, C.K.I., 2006. Gaussian processes for machine learning, Adaptive computation and machine learning. MIT Press, Cambridge, Mass.

Rosness, R., 1993. Limits to Analsyis and Verification, in: Verification and Validation of Complex Systems: Human Factors Issues.

Roustant, O., Padonou, E., Deville, Y., Clément, A., Perrin, G., Giorla, J., Wynn, H., 2020. Group Kernels for Gaussian Process Metamodels with Categorical Inputs. SIAM/ASA J. Uncertainty Quantification 8, 775–806. https://doi.org/10.1137/18M1209386

Sam, A.G., Ker, A.P., 2006. Nonparametric regression under alternative data environments. Statistics & Probability Letters 76, 1037–1046. https://doi.org/10.1016/j.spl.2005.12.002

Saves, P., Bartoli, N., Diouane, Y., Lefebvre, T., Morlier, J., David, C., Van, E.N., Defoort, S., 2021. Bayesian optimization for mixed variables using an adaptive dimension reduction process: applications to aircraft design, in: AIAA SCITECH 2022 Forum. American Institute of Aeronautics and Astronautics. https://doi.org/10.2514/6.2022-0082

Saves, P., Diouane, Y., Bartoli, N., Lefebvre, T., Morlier, J., 2022. A general square exponential kernel to handle mixed-categorical variables for Gaussian process, in: AIAA AVIATION Forum. Presented at the AIAA AVIATION 2022 Forum.

Sriver, T., Chrissis, J., 2004. Framework for Mixed-Variable Optimization Under Uncertainty Using Surrogates and Statistical Selection, in: 10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference. American Institute of Aeronautics and Astronautics. https://doi.org/10.2514/6.2004-4591

Steinwart, I., Christmann, A., 2008. Support Vector Machines, 1st ed. Springer Publishing Company, Incorporated.

Stuckner, J., Piekenbrock, M., Arnold, S.M., Ricks, T.M., 2021. Optimal experimental design with fast neural network surrogate models. Computational Materials Science 200, 110747. https://doi.org/10.1016/j.commatsci.2021.110747

Su, D.-X., Zhao, J., Wang, Y., Qu, M.-J., 2019. Kriging-based orthotropic closure for flow-induced fiber orientation and the part stiffness predictions with experimental investigation. Polymer Composites 40, 3844–3856. https://doi.org/10.1002/pc.25243

Tao, S., Apley, D.W., Plumlee, M., Chen, W., 2021. Latent variable Gaussian process models: A rank-based analysis and an alternative approach. International Journal for Numerical Methods in Engineering 122, 4007–4026. https://doi.org/10.1002/nme.6690

Trochu, F., Vernet, N., Sun, Y., Echaabi, J., Makradi, A., Belouettar, S., 2022. Hybrid twin models of fiber compaction for composite manufacturing based on dual kriging. Int J Mater Form 15, 36. https://doi.org/10.1007/s12289-022-01679-3

Vapnik, V., 1999. The nature of statistical learning theory. Springer science & business media.

Wang, D., Zeng, X.-J., Keane, J.A., 2007. Hierarchical hybrid fuzzy-neural networks for approximation with mixed input variables. Neurocomputing, Neural Network Applications in Electrical Engineering 70, 3019–3033. https://doi.org/10.1016/j.neucom.2006.07.015

Wang, G., Dong, Z., Aitchison, P., 2001. Adaptive Response Surface Method - a Global Optimization Scheme for Approximation-Based Design Problems. Engineering Optimization 33, 707–733. https://doi.org/10.1080/03052150108940940

Welch, W.J., Buck, Robert.J., Sacks, J., Wynn, H.P., Mitchell, T.J., Morris, M.D., 1992. Screening, Predicting, and Computer Experiments. Technometrics 34, 15–25. https://doi.org/10.2307/1269548

Widrow, B., Hoff, M.E., 1960. Adaptive switching circuits (No. TR-1553-1). Stanford Univ Ca Stanford Electronics Labs.

Xie, Y., Zhang, J., DesRoches, R., Padgett, J.E., 2019. Seismic fragilities of single-column highway bridges with rocking column-footing. Earthquake Engineering & Structural Dynamics 48, 843–864. https://doi.org/10.1002/eqe.3164

Xu, G., Kareem, A., Shen, L., 2020. Surrogate Modeling with Sequential Updating: Applications to Bridge Deck–Wave and Bridge Deck–Wind Interactions. Journal of Computing in Civil Engineering 34, 04020023. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000904

Yi, J.-H., Kim, S.-H., Kushiyama, S., 2007. PDF interpolation technique for seismic fragility analysis of bridges. Engineering Structures 29, 1312–1322. https://doi.org/10.1016/j.engstruct.2006.08.019

Yi, S., Taflanidis, A.A., 2023. Computationally Efficient Adaptive Design of Experiments for Global Metamodeling through Integrated Error Approximation and Multicriteria Search Strategies. Journal of Engineering Mechanics 149, 04023050. https://doi.org/10.1061/JENMDT.EMENG-7083

Yin, H., Fang, H., Wen, G., Gutowski, M., Xiao, Y., 2018. On the ensemble of metamodels with multiple regional optimized weight factors. Structural and Multidisciplinary Optimization 58, 245–263.

Zhang, J., Chowdhury, S., Messac, A., 2012. An adaptive hybrid surrogate model. Structural and Multidisciplinary Optimization 46, 223–238.

Zhang, X., Wang, L., Sørensen, J.D., 2020. AKOIS: An adaptive Kriging oriented importance sampling method for structural system reliability analysis. Structural Safety 82, 101876. https://doi.org/10.1016/j.strusafe.2019.101876

Zhang, Y., Tao, S., Chen, W., Apley, D.W., 2020. A Latent Variable Approach to Gaussian Process Modeling with Qualitative and Quantitative Factors. Technometrics 62, 291–302. https://doi.org/10.1080/00401706.2019.1638834

Zhou, Q., Wang, Y., Choi, S.-K., Jiang, P., Shao, X., Hu, J., Shu, L., 2018. A robust optimization approach based on multi-fidelity metamodel. Struct Multidisc Optim 57, 775–797. https://doi.org/10.1007/s00158-017-1783-4

Zhuang, X., Pan, R., 2012. Epistemic uncertainty in reliability-based design optimization, in: 2012 Proceedings Annual Reliability and Maintainability Symposium. Presented at the 2012 Proceedings Annual Reliability and Maintainability Symposium, pp. 1–6. https://doi.org/10.1109/RAMS.2012.6175496

Table 1. Correlation functions for continuous variables

Name	$k(\theta_l, d_l)$	Parameter limits
Exponential	$\exp(-\theta_l \lvert d_l \rvert)$	-
Gaussian	$\exp(-\theta_l d_l^2)$	-
Linear	$\max(0, 1-\theta_l d_l^2)$	-
Spherical	$1-1.5\varepsilon_l+0.5\varepsilon_l^3$	$\varepsilon_l = \min(1, \theta_l \lvert d_l \rvert)$
Cubic	$1-3\varepsilon_l^2+2\varepsilon_l^3$	$\varepsilon_l = \min(1, \theta_l \lvert d_l \rvert)$

Table 2. Combinations of kernel functions for weight matrix

Comb. Number	Continuous variable kernel	Discrete variable kernel
01	Gaussian kernel function (Gajewicz-Skretna et al., 2021) $w_x(x_k^i, x_k^j) = \frac{1}{c\sqrt{2\pi}}\exp(-\frac{(x_k^i - x_k^j)^2}{2kl^2})$	Racine and Li estimator (Li and Racine, 2014) $w_z(z_k^i, z_k^j) = \begin{cases} 1 & if\ z_k^i = z_k^j \\ \gamma_k & otherwise \end{cases}$
02	Weighted Gower Distance Kernel $w(\boldsymbol{v}^i, \boldsymbol{v}^j) = \exp\{-\sum_{k=1}^{k=n+q} \theta_k\ [d_{gow}^k(v_k^i, v_k^j)]^{p^k}\}$	
03	Gaussian kernel function (Gajewicz-Skretna et al., 2021) $w_x(x_k^i, x_k^j) = \frac{1}{c\sqrt{2\pi}}\exp(-\frac{(x_k^i - x_k^j)^2}{2kl^2})$	Latent variable kernel $w_z(z_k^i, z_k^j) = \exp(-\theta_q\left\|\varphi(z_i) - \varphi(z_j)\right\|^2)$

Note: v_k^i represents the i-th variable in the k-th dimension

Table 3. Summary of mean values of the Goodness-of-fit (GOF) measures in 4D function

Sample Size	GOF measures for the estimated mean			GOF measures for the estimated standard deviation			GOF measures for the estimated mean − 3*standard deviation			GOF measures for the estimated mean + 3*standard deviation		
	R^2 value	NRMSE	NMAE	R^2 value	NRMSE	NMAE	R^2 value	NRMSE	NMAE	R^2 value	NRMSE	NMAE
60	0.93	0.083	1.40	0.62	0.151	2.25	0.81	0.115	1.90	0.83	0.097	1.35
80	0.94	0.071	1.24	0.61	0.155	2.33	0.81	0.116	1.83	0.81	0.103	1.40
100	0.94	0.068	1.40	0.72	0.131	2.14	0.80	0.117	1.97	0.84	0.093	1.32
120	0.95	0.065	1.22	0.73	0.127	2.08	0.81	0.117	1.92	0.84	0.094	1.30
140	0.95	0.059	1.19	0.77	0.117	1.87	0.82	0.114	1.83	0.85	0.092	1.27
160	0.96	0.059	1.16	0.81	0.107	1.71	0.84	0.107	1.75	0.86	0.087	1.19
180	0.96	0.060	1.29	0.80	0.110	1.87	0.81	0.115	1.99	0.85	0.092	1.26
200	0.95	0.059	1.20	0.80	0.110	1.93	0.82	0.113	1.87	0.86	0.088	1.18

Table 4. Summary of mean values of the goodness-of-fit measures in 8D function

Sample Size	Goodness-of-fit measures for the estimated mean			GOF measures for estimated standard deviation			GOF measures for the estimated mean − 3*standard deviation			GOF measures for the estimated mean + 3*standard deviation		
	R^2 value	NRMSE	NMAE	R^2 value	NRMSE	NMAE	R^2 value	NRMSE	NMAE	R^2 value	NRMSE	NMAE
50	0.82	0.095	1.69	0.70	0.2	1.72	0.31	0.201	3.06	0.82	0.123	1.55
100	0.93	0.056	1.04	0.75	0.177	1.97	0.46	0.177	3.07	0.87	0.102	1.50
150	0.95	0.050	0.98	0.69	0.203	2.09	0.34	0.196	3.12	0.83	0.118	1.58
200	0.95	0.048	0.98	0.75	0.183	1.91	0.45	0.180	3.02	0.87	0.104	1.42
250	0.95	0.050	1.06	0.76	0.177	1.82	0.47	0.176	2.84	0.88	0.102	1.38

1127 **Table 5**. Experimental wave conditions

Exp.	Regular waves (h =2.15m)		Irregular waves (h =2.15m)		Transient waves (h =2.0m)	
	H(m)	T(s)	H(m)	T(s)	A(m)	T(s)
X1	0.1	4.10	0.10	3.72	0.51	36.4
X2	0.21	4.10	0.19	3.86	0.34	51.0
X3	0.29	4.10	0.29	4.10	0.28	87.2
X4	0.40	4.10	0.40	4.10	0.21	109
X5	0.50	4.10	0.50	3.86	0.18	117
X6	0.16	2.52	0.16	2.52	0.16	120
X7	0.23	2.98	0.21	2.98	0.14	154
X8	0.26	3.64	0.25	3.28	0.13	162
X9	0.35	4.68	0.34	4.68		
X10	0.42	5.04	0.39	5.04		

1128

Table 6. Air gag conditions for experimental cases

Air Gap cases	*a*(m)	
	Regular and Irregular Waves	Transient Waves
a_0	-0.40	-0.25
a_1	-0.30	-0.15
a_2	-0.20	-0.05
a_3	-0.10	0.05
a_4	-0.05	0.10
a_5	0.00	0.15
a_6	0.05	0.20
a_7	0.10	0.25
a_8	0.20	0.30
a_9	0.28	0.43

1131 **Table 7**. Details of input variables used in Engineering Application II

Variable	Category	Range	Unit
*Log of Peak Ground Acceleration	Continuous	[-3.22 0.67]	
*Deck width	Continuous	[150.03- 1096.50]	Inches
*Concrete compressive strength	Continuous	[3.00- 8.00]	ksi
*Steel yield strength	Continuous	[39.50 – 95.00]	ksi
*Dowel strength	Continuous	[10.40 – 15.60]	kips
*Number of spans	Discrete	2-9	-
*Span length	Continuous	[276.08 – 1102.21]	Inches
*Number of columns	Discrete	2-6	-
*Column height	Continuous	[130.00 – 354.33]	Inches
*Column diameter	Continuous	[23.65 – 59.10]	Inches
*Concrete cover depth	Continuous	[0.50 – 4.50]	Inches
*Number of girders along the width of the deck	Discrete	2-19	-
*Girder spacing	Continuous	[53.05 – 210.25]	Inches
*Column spacing	Continuous	[150.03 – 300.00]	Inches
*Slab weight per girder	Continuous	[0.03- 0.28]	Kips/inch
*Bearing pad area	Continuous	[90.11 – 654.64]	Sq. inches
*Bearing pad thickness	Continuous	[0.20 – 1.20]	Inches
*Decrease in rebar diameter	Continuous	[0.00 – 0.70]	Inches
*Stiffness factor to account for oxidation of elastomeric bearings	Continuous	[0.90 – 2.00]	-

*Decrease in bearing dowel diameter	Continuous	[0.00 – 0.70]	Inches
Coefficient of friction for bearing pad	Continuous	[0.50 - 2.5]	-
Stiffness of bearing pad	Continuous	[0.04 - 0.80]	ksi
Dowel gap	Continuous	[0.00 – 2.00]	Inches
Abutment passive stiffness	Continuous	[1.46- 4.39]	kip/in/in
Abutment active stiffness	Continuous	[20.00 – 60.00]	kip/in/pile
Foundation vertical stiffness	Continuous	[500.41 – 1500.00]	kip/in
Foundation transverse stiffness	Continuous	[20.00 – 60.00]	kip/in/pile
Mass participation ratio	Continuous	[0.90 – 1.10]	-
Damping ratio	Continuous	[0.02 – 0.08]	-
PGA – geometric mean of the two ground motion components	Continuous	[0.03 – 1.90]	g
Gap 1 (used for bearing model)	Continuous	[1.41 – 1.57]	Inches
Gap 2 (used for bearing model)	Continuous	[1.41 – 1.57]	Inches
Gap 3 (used for bearing model)	Continuous	[0.78 – 1.22]	Inches
Gap 4 (used for bearing model)	Continuous	[0.78 – 1.22]	Inches
Longitudinal steel reinforcement ratio	Continuous	[0.01 - 0.04]	-
Transverse steel reinforcement ratio	Continuous	[0.00 – 0.02]	-
Deck slab c/s area	Continuous	[362.31 – 3412.81]	$inches^2$
Girder steel area	Continuous	[754.04 – 4282.90]	$inches^2$
Girder concrete strength	Continuous	[7.00 – 11.00]	ksi
I_x of deck slab	Continuous	[1458.26 – 96455.30]	$Inches^4$

I_z of deck slab	Continuous	[95243.05-11934069.40]	Inches^4
I_x of girder	Continuous	[94262.079-1392317.95]	Inches^4
I_z of girder	Continuous	[86750.26-1392317.95]	Inches^4
Earthquake direction	Continuous	[0.14 – 360.00]	degree
Weight of one AASHTO prestressed girder	Continuous	[0.02 – 0.1]	kip/in

1132

Table 8. Summary of discrete kernels used in examples

Example	No. of Categorical combinations (No. of discrete variables)	Discrete Kernels		Total number of Hyperparameters in Discrete kernels
		Correlation Matrix	LWR	
4-D function	4(2)	Weighted Gower Distance	Racine and Li Estimator	4
8-D function	81 (4)	Latent variable kernel	Latent variable kernel	12
Engineering Application I	3 (1)	Latent variable kernel	Latent variable kernel	3
Engineering Application II	760 (3)	Weighted Gower Distance	Racine and Li Estimator	6

Table 9. Input values for the predictor variables

Variable	Input value	Unit
*Deck width	600	Inches
*Concrete compressive strength	4.35	ksi
*Steel yield strength	66.7	ksi
Dowel strength	11	kips
*Number of spans	4	-
*Span length	393.7	Inches
*Number of columns	3	-
*Column height	157.48	Inches
*Column diameter	35.83	Inches
Concrete cover depth	3	Inches
Number of girders along the width of the deck	6	-
Girder spacing	120	Inches
*Column spacing	300	Inches
Slab weight per girder	0.24	Kips/inch
Bearing pad area	372.58	Sq. inches
Bearing pad thickness	0.7	Inches
Decrease in rebar diameter	0.35	Inches
Stiffness factor to account for oxidation of elastomeric bearings	1.4	-
Decrease in bearing dowel diameter	0.35	Inches

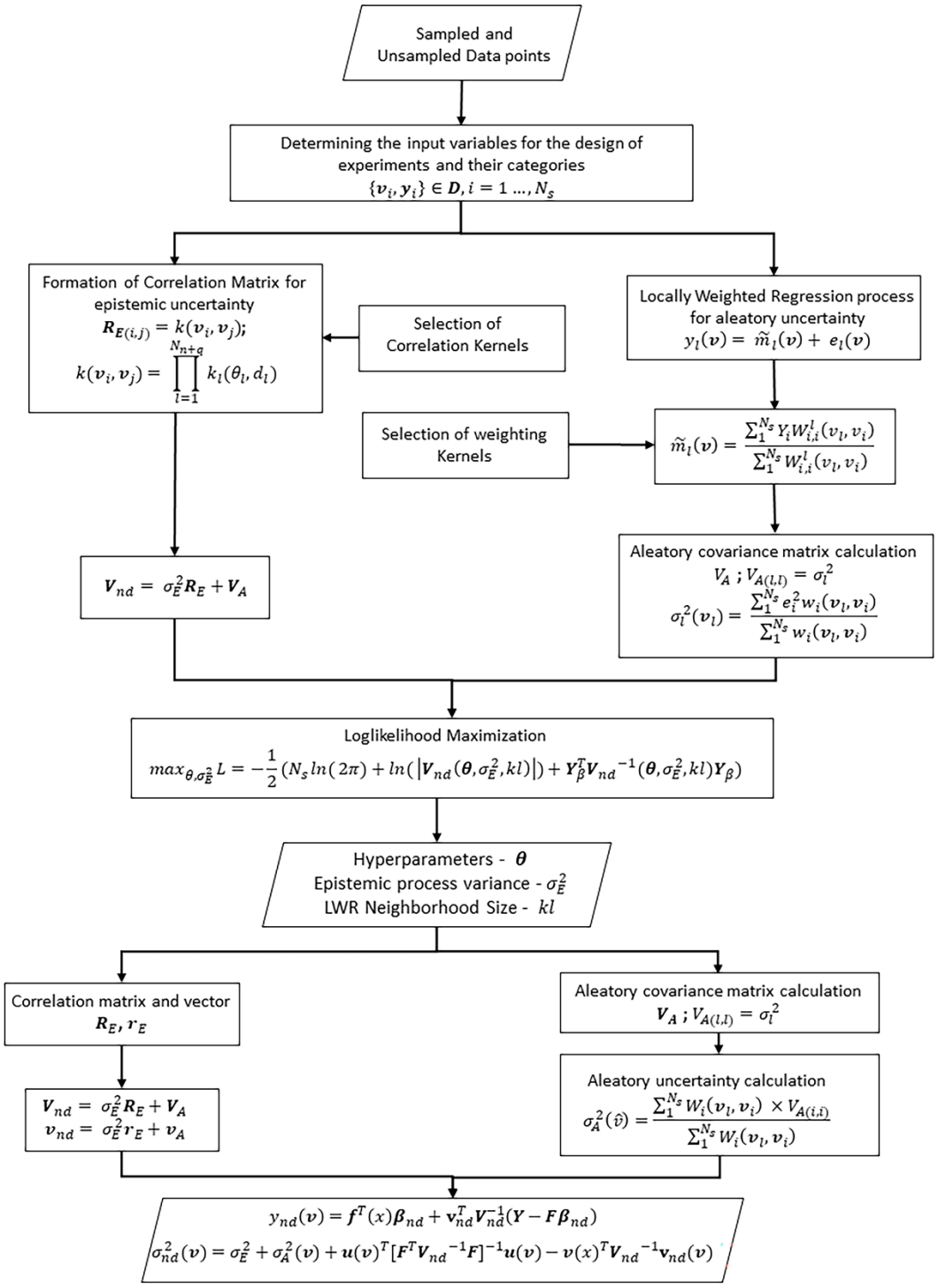
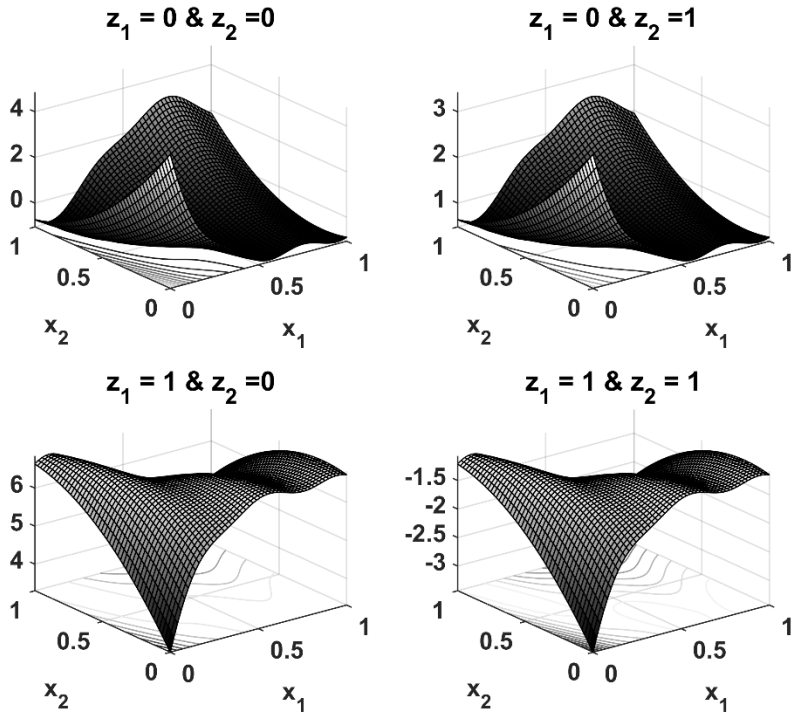
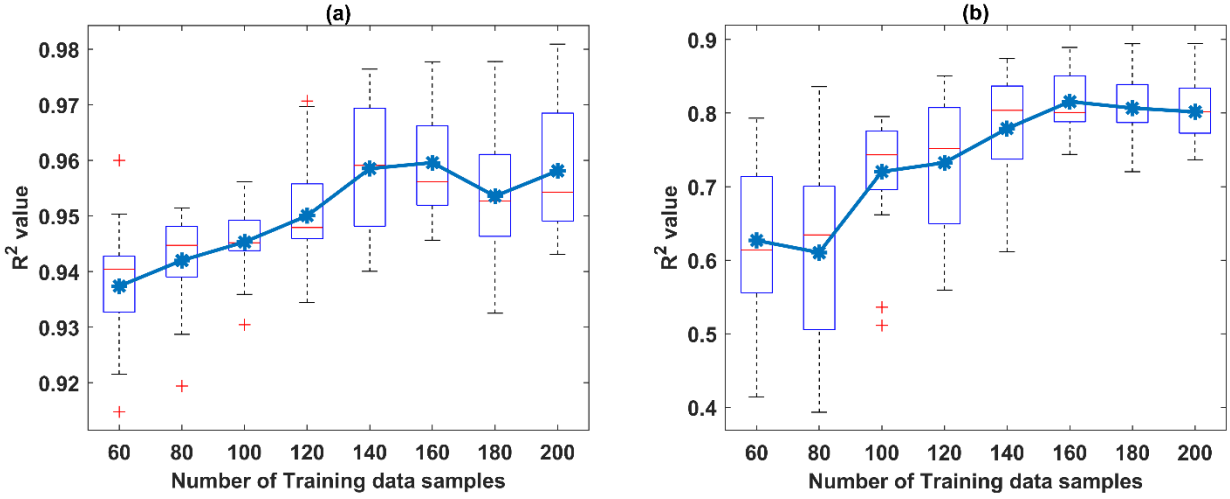


Fig. 1: Flowchart for the proposed NDK method



1138

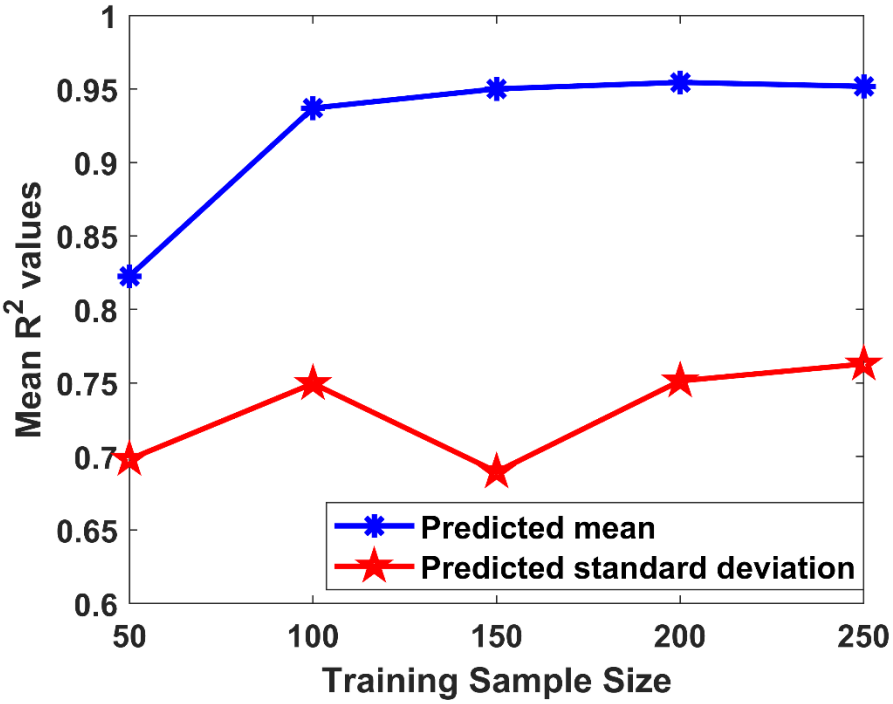
1139 **Fig. 2:** Actual mean of the 4-D function



1140

1141 **Fig. 3: (a)** Variation of R^2 values of the estimated mean in the 4D function **(b)** Variation of R^2

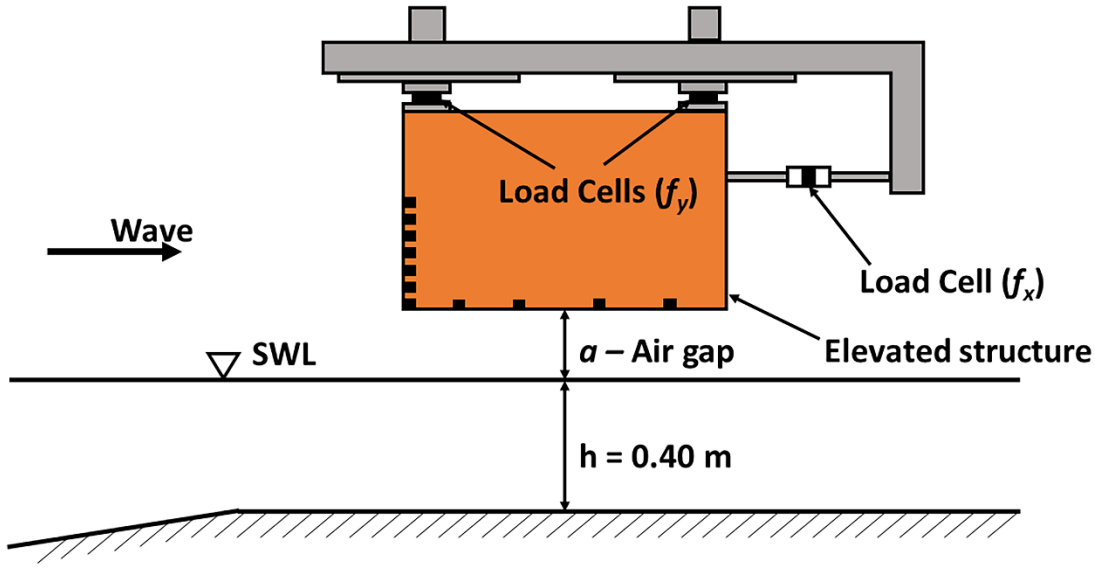
1142 values of the estimated standard deviation in the 4D function



1143

1144 **Fig. 4:** Variation of mean R^2 values for the estimated mean and predicted standard

1145 deviation with the training data sample size in 8-D function



1146

1147 **Fig. 5:** Side view of the experimental setup

1148