Smart Structures and Systems, *Vol. 17, No. 4 (2016) 611-629*
DOI: http://dx.doi.org/10.12989/sss.2016.17.4.611

Self-adaptive sampling for sequential surrogate modeling of time-consuming finite element analysis

Seung-Seop Jin[a] and Hyung-Jo Jung[*]

Department of Civil and Environmental Engineering, Korea Advanced Institute of Science and Technology, 291 Daehak-ro, Yuseong-gu, Daejeon 305-701, Republic of Korea

(Received December 28, 2015, Revised February 19, 2016, Accepted February 23, 2016)

Abstract. This study presents a new approach of surrogate modeling for time-consuming finite element analysis. A surrogate model is widely used to reduce the computational cost under an iterative computational analysis. Although a variety of the methods have been widely investigated, there are still difficulties in surrogate modeling from a practical point of view: (1) How to derive optimal design of experiments (i.e., the number of training samples and their locations); and (2) diagnostics of the surrogate model. To overcome these difficulties, we propose a sequential surrogate modeling based on Gaussian process model (GPM) with self-adaptive sampling. The proposed approach not only enables further sampling to make GPM more accurate, but also evaluates the model adequacy within a sequential framework. The applicability of the proposed approach is first demonstrated by using mathematical test functions. Then, it is applied as a substitute of the iterative finite element analysis to Monte Carlo simulation for a response uncertainty analysis under correlated input uncertainties. In all numerical studies, it is successful to build GPM automatically with the minimal user intervention. The proposed approach can be customized for the various response surfaces and help a less experienced user save his/her efforts.

Keywords: surrogate modeling; Gaussian process model; self-adaptive sampling; sequential Bayesian framework; time-consuming FE analysis

1. Introduction

Various engineering problems (e.g., reliability analysis and uncertainty analysis) require a large number of the iterative finite element analysis. Especially in civil engineering, structures such as bridges and dams are massive and complex, so that their finite element analysis (FEA) may be sometimes expensive in terms of the computing time for a single run.

In this context, a surrogate model has been gaining a considerable attention as a cost-effective substitute for a time-consuming FEA (Jones 2001, Queipo *et al*. 2005, Forrester *et al*. 2008, Forrester and Keane 2009). The surrogate model is a way to emulate an FEA in the form of a mathematical/statistical approximation by using an input/output (I/O) of the FEA. The input/output (I/O) of the FEA is typically referred to as response surface. The surrogate model is categorized as follows: (1) regression-based model such as polynomial function (Bucher and Bourgund 1990) and

∗Corresponding author, Professor, E-mail: hjung@kaist.ac.kr
[a] Ph.D. Candidate, E-mail: seungsab@kaist.ac.kr

http://www.techno-press.org/?journal=sss&subpage=8 ISSN: 1738-1584 (Print), 1738-1991 (Online)

(2) interpolation-based model such as radial basis function (Zhang *et al.* 2014) and Gaussian process model (DiazDelaO and Adhikari 2011, Dubourg *et al*. 2013). It is known that the interpolation-based surrogate model can approximate a response surface more accurately than the regression-based one, as the new training samples are added (Jones 2001).

Conventional surrogate modeling uses training samples at once to construct a response-surface (i.e., model construction phase). As a conventional model construction, various design of experiments (DOEs) were investigated: (1) classic DOEs such as central composite designs (Box *et al*. 1978), Box-Behnken design (Box and Behnken, 1960) and optimal design (Kiefer and Wolfowitz 1959); and (2) space-filling design such as uniform design (Fang *et al*. 2000), Latin hypercube design (LHD) (Mckay *et al.* 1979), Orthogonal LHD (Ye 1998) and Generalized LHD (Dette and Pepelyshev, 2010). Once a surrogate model is constructed, a user should check the adequacy of the surrogate model by using additional validation samples (i.e., model diagnostic phase) (Bastos and O'Hagan 2009).

Although a variety of the methods have been widely investigated in surrogate modeling, the conventional surrogate modeling has two difficulties from a practical point of view: (1) How to derive optimal design of experiments in the model construction phase (i.e., the number of training samples and their locations); and (2) Diagnostics of the surrogate model adequacy before using it as a substitute of FEA (i.e., model diagnostic phase).

Firstly, a key element of surrogate modeling is how to generate the training samples. However, optimal number of samples and their location are not known, since a response surface is not known beforehand. In practice, a response surface may respond much more rapidly to changes at some regions of the input space than others (Xiong *et al.* 2007). Moreover, it is getting more challenging to determine the optimal number of the training samples and their locations, as the dimension of the input increases. Because no single method works best for every problem (Bucher and Most 2008, Goel *et al*. 2008), the conventional surrogate modeling performs the model construction process by trial-and-error method with user intervention. The efficiency of the conventional surrogate modeling is depending on the user's knowledge and experience.

Secondly, it is the most important to validate and assess the adequacy of the surrogate model. Diagnostics of the surrogate model can be found as follows: (1) independent validation data set (Bastos and O'Hagan 2009); and (2) resampling methods (Jackknifing (Kleijnen and van Beers, 2004) and leave-one-out cross-validation (Rougier *et al*. 2009)). Typically, additional validation samples are typically required to evaluate the adequacy of the surrogate model (Bastos and O'Hagan 2009). Especially in interpolation-based model, they can reproduce target outputs of the training samples, additional validation samples are essential. Therefore, it requires computational efforts and user-intervention additionally in model diagnostic phase.

To overcome the abovementioned difficulties, this paper presents a sequential surrogate modeling based on Gaussian process model (GPM) with self-adaptive sampling. The proposed method starts with initial and small training samples. Once surrogate model is constructed, three infill criteria are used in parallel to find the infill samples (i.e., self-adaptive sampling). Based on these infill criteria, sequential surrogate modeling enable not only further sampling to make GPM more accurate, but also evaluates the model adequacy within a sequential framework.

The main contributions are summarized as follows: (1) the proposed method provides the optimal samples and their location automatically; (2) the proposed method integrates the model diagnostic phase into the model construction phase, so that the proposed method does not require the model diagnostic phase; and (3) the proposed method minimizes the user-intervention in both model construction and diagnostic phases.

The paper is organized as follows. Section 2 describes the background of the GPM and practical difficulties in surrogate modeling. In Section 3, the proposed approach is presented with the infill criteria and self-adaptive sampling. Section 4 verifies the proposed method by using mathematical test functions. The proposed method is also applied as engineering example to the Monte Carlo simulation (MCS) for a response uncertainty. Finally, it is concluded with Section 5. Hereafter, the bold-face letters indicates vectors or matrices.

2. Gaussian process model

The proposed method in this study uses GPM, also known as Kriging from Geostatistics (Krige 1994). Since theoretical backgrounds of GPM is well established in the literatures (Jones *et al.* 1998, Jones 2001, Kennedy and O'Hagan 2001, Kleijnen and van Beers 2004, Forrester *et al.* 2008, Forrester and Keane 2009), this section briefly introduces GPM.

2.1 Background of Gaussian process model

GPM assumes that an unknown output $\hat{y}(\boldsymbol{x}_{new})$ as the Gaussian process random variable conditional on training samples (Kennedy and O'Hagan 2001) as shown in Fig. 1. GPM is modeled as

$$\hat{y}(\boldsymbol{x}_{new}) = \mu(\mathbf{x}) + \mathrm{Z}(\mathbf{x}) \;\; with \;\; \mathrm{Z}(\mathbf{x}) \sim \mathrm{GP}\left(0, \mathrm{C}(\mathbf{x}, \mathbf{x}')\right) \tag{1}$$

where $\boldsymbol{x}_{new}$ is an unknown output of interest; $\mu(\boldsymbol{x})$ is a deterministic component to capture the global trend (i.e., a constant or polynomial function); $\mathrm{Z}(\boldsymbol{x})$ denotes a stochastic component of a zero mean stationary Gaussian process with a covariance function $\boldsymbol{C}(\boldsymbol{x}, \boldsymbol{x}')$.

The covariance between any samples $\boldsymbol{C}(\boldsymbol{x}, \boldsymbol{x}')$ is derived as

$$\mathbf{C}(\mathbf{x}, \mathbf{x}') = \sigma^2 \boldsymbol{\psi}(\mathbf{x}, \mathbf{x}') \tag{2}$$

where the constant variance σ^2 provides overall dispersion relative to the mean of the Gaussian process and $\boldsymbol{\psi}(\cdot,\cdot)$ denotes the spatial correlation matrix. A typical choice of the correlation matrix $\boldsymbol{\psi}(\cdot,\cdot)$ is the k-dimensional Gaussian correlation function as

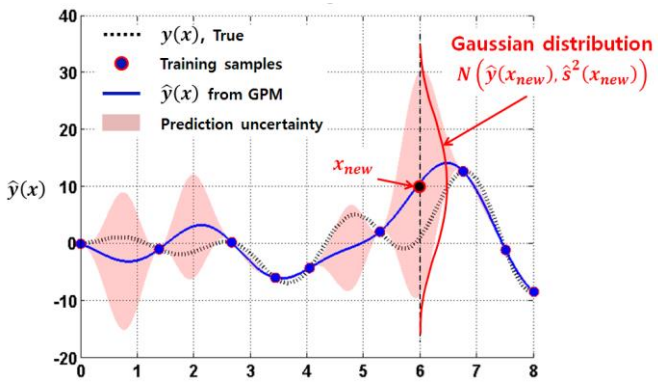


Fig. 1 Prediction of Gaussian process model conditional on training samples

$$\psi(\mathbf{x}^{\mathrm{i}}, \mathbf{x}^{\mathrm{j}}) = \exp\left(-\sum_{\mathrm{p}=1}^{\mathrm{k}} \theta_{\mathrm{p}} \left\| \mathrm{x}_{\mathrm{p}}^{\mathrm{i}} - \mathrm{x}_{\mathrm{p}}^{\mathrm{j}} \right\|^2\right) \tag{3}$$

where the subscript p denotes the dimension of $\boldsymbol{x}$, the superscript i and j indicate the i-th and j-th training sample, respectively, and $\left\| \mathrm{x}_{\mathrm{p}}^{\mathrm{i}} - \mathrm{x}_{\mathrm{p}}^{\mathrm{j}} \right\|^2$ is the Euclidean distance measure between two samples. θ_{p} is a vector to scale the correlation length in each dimension. It is known that the correlation length reflects the relative significance of each input.

Gaussian process model determines the parameters $\mu(\boldsymbol{x})$, σ^2, and θ_p by maximizing the log-likelihood function. $\mu(\boldsymbol{x})$ is hereafter represented by μ for simplicity. To model GPM, μ and σ^2 are estimated by maximum likelihood estimation. The estimators of $\hat{\mu}$ and $\hat{\sigma}^2$ are given in a closed form as

$$\hat{\mu} = \frac{1^T \boldsymbol{\psi}^{-1} \boldsymbol{Y}}{1^T \boldsymbol{\psi}^{-1} \mathbf{1}} \tag{4}$$

$$\hat{\sigma}^2 = \frac{(\boldsymbol{Y} - \mathbf{1}\hat{\mu})^T \boldsymbol{\psi}^{-1} (\boldsymbol{Y} - \mathbf{1}\hat{\mu})}{n} \tag{5}$$

where $\boldsymbol{Y} = [\mathrm{y}(\boldsymbol{x}^1), \quad \ldots \quad , \mathrm{y}(\boldsymbol{x}^n)]^T$, n is the number of the training samples, $\mathbf{1}$ is the n-by-1 unit vector, and $\boldsymbol{\psi}$ is the n-by-n symmetric matrix of ψ^{ij}. With the estimates of $\hat{\mu}$ and $\hat{\sigma}^2$, θ_p is estimated by optimization by maximizing the log-likelihood function (Eq. (6))

$$ln\left(L(\hat{\mu}, \hat{\sigma}, \theta_p)\right) = -\frac{n}{2} ln(\hat{\sigma}^2) - \frac{1}{2} ln|\boldsymbol{\psi}| - \frac{(\boldsymbol{Y} - \mathbf{1}\hat{\mu})^T \boldsymbol{\psi}^{-1} (\boldsymbol{Y} - \mathbf{1}\hat{\mu})}{2\sigma^2} \tag{6}$$

Once $\hat{\mu}$, $\hat{\sigma}^2$ and θ_p are estimated, unknown output $\hat{y}(\boldsymbol{x}_{new})$ is estimated by a weighted linear combination of all output already observed. Thus, Eq. (1) is rewritten *a*

$$\hat{y}(x_{new}) = \hat{\boldsymbol{\mu}} + \hat{\boldsymbol{\psi}}^T \boldsymbol{\psi}^{-1} (\boldsymbol{Y} - \mathbf{1}\hat{\boldsymbol{\mu}}) \tag{7}$$

where $\hat{\boldsymbol{\psi}}$ is an n-by-1 vector between training samples and an unknown sample ($\boldsymbol{\psi}(\mathbf{x}^{\mathrm{i}}, \boldsymbol{x}_{new})$). A detailed derivation of Eqs. (4), (5) and (7) is given in Jones (2001).

2.2 Practical difficulties in surrogate modeling

Although GPM is very flexible to represent a smooth hyper-surface, it is invalid correctly to predict an output in some situations (Xiong *et al*. 2007, Bastos and O'Hagan 2009, Gramacy and Lee 2012). From a practical point of view, these difficulties of surrogate modeling are as follows: (1) invalid assumption of GPM (in a non-stationary response surface); (2) insufficient training samples to capture a response surface; and (3) additional computational costs due to diagnostics of model adequacy.

Firstly, the assumption of a stationary Gaussian process may be inappropriate. In practice, FEA output may respond much more rapidly to changes in the particular region than others. To illustrate this problem, the mathematical test function is adopted from Gramacy and Lee (2012). As shown in Fig. 2(a), the test function is a one-dimensional non-stationary function as given in

$$f(x) = \frac{sin(10\pi x)}{2x} + (x-1)^4, \qquad 0.5 \le \mathrm{x} \le 2.5 \tag{8}$$

where Eq. (8) has a chirp component which the frequency of the output decreases along input x. Due to the assumption of stationary covariance, GPM approximate the smooth shape and fails to

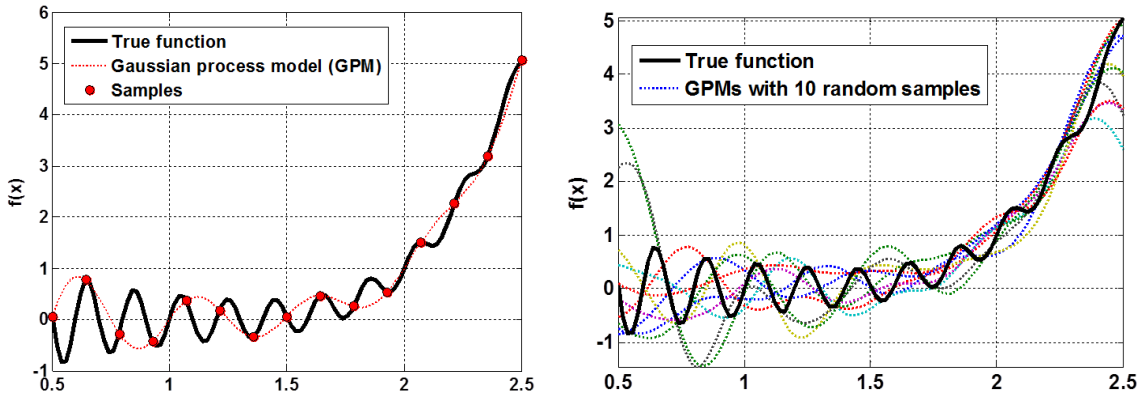
approximate the true function.

Secondly, insufficient training samples may construct an inappropriate GPM, even if the stationary assumption is reasonable. As the dimension of the input space increases, the number of training samples increases exponentially (i.e., curse of dimensionality (Bellman 2003)). Therefore, the training samples sometimes may be sparse. To show the concept of sparseness in higher-dimensional space, 10 samples are randomly generated by Eq. (8) and GPM is constructed by 10 times with the different sample configurations. As shown in Fig. 2(b), it is clear that the 10 training samples are not sufficient for GPM to represent this function accurately. In general, it is impossible to understand how many samples are required to represent a response surface in high-dimensional input space, since the hyper-surface of the response surface is not known in advance.

Lastly, it is very important to validate and assess the model adequacy before using it as a fast surrogate of FEA. Due to their flexibility, the outputs on the training samples are exactly the same as the corresponding predictions of GPM. Generally, the diagnostic is commonly based on comparison between FEA outputs and GPM prediction for some test data, known as the validation data (Bastos and O'Hagan 2009). Other diagnostics are also found based on resampling methods (Kleijnen and van Beers, 2004, Rougier *et al*. 2009). To perform these diagnostic methods, it requires computational efforts additionally.

3. Sequential surrogate modelling with self-adaptive sampling

In this section, we present a sequential surrogate modelling with self-adaptive sampling to address the difficulties abovementioned in Section 2.2. The proposed approach is based on a Bayesian framework, since candidate samples for infill are inferred by a current GPM. The proposed approach not only enables further sampling to make GPM more accurate, but also evaluates the model adequacy within a sequential sampling framework. First, we introduce the infill criteria for further sampling to improve the GPM. Then, we present the proposed sequential framework with the toy problem (i.e., Eq. (8)).



(a) non-stationary response surface and (b) sparseness of training samples

Fig. 2 Practical challenges with an unknown response surface

3.1 Infill criteria for self-adaptive sampling

By virtue of the treatment of an unknown output as a realization of Gaussian process, the infill criteria are developed to evaluate the prediction uncertainty and improvement of a current best value in global optimization (Jones 2001, Forrester *et al*. 2008, Forrester and Keane 2009). In the proposed method, three infill criteria are used for self-adaptive sampling: (1) Mean squared error, (2) Expected improvement on minimum, and (3) Expected improvement on maximum.

Mean squared error (MSE) at an unknown output is interpreted as prediction uncertainty and it is estimated as

$$\hat{s}^2(\boldsymbol{x}) = \hat{\sigma}^2\left[\mathbf{1} - \hat{\boldsymbol{\psi}}^{\boldsymbol{T}}\boldsymbol{\psi}^{-\mathbf{1}}\hat{\boldsymbol{\psi}}\right] \tag{9}$$

As shown in Fig. 1, MSE is portrayed by the red shaded area. It is easily found that the prediction uncertainty at an unknown sample is generally large at the region which the training samples are sparse. MSE can be used as a measure of the sparseness of the training samples in the input space as shown in Fig. 3(a). The full derivation of MSE is presented in Sacks *et al*. (1983).

Expected improvement (EI) is an infill criterion to evaluate how much improvement of the current GPM is expected if a new sample is obtained. It is developed in computationally intensive optimization. The reference value is required to compute EI. Usually, the minimum in the training samples (y_{min}) is used. As illustrated in Fig. 3(b), the EI value corresponds to the probability that GPM prediction ($\hat{y}(\boldsymbol{x}_{new})$) may be better than y_{min} (i.e., shaded area). At any unknown sample with a current GPM, the probability of an improvement upon y_{min} is quantified by the integrating the Gaussian distribution with mean $\hat{y}(\boldsymbol{x}_{new})$ and variance $\hat{s}^2(\boldsymbol{x}_{new})$. The EI on minimum is derived in the closed form as

$$\mathrm{EI}_{min}(\boldsymbol{x}_{new}) = \left(\mathrm{y}_{\min} - \hat{\mathrm{y}}(\boldsymbol{x}_{new})\right)\Phi\left(\frac{\mathrm{y}_{\min} - \hat{\mathrm{y}}(\boldsymbol{x}_{new})}{\hat{\mathrm{s}}(\boldsymbol{x}_{new})}\right) + \hat{\mathrm{s}}(\boldsymbol{x}_{new})\phi\left(\frac{\mathrm{y}_{\min} - \hat{\mathrm{y}}(\boldsymbol{x}_{new})}{\hat{\mathrm{s}}(\boldsymbol{x}_{new})}\right) \tag{10}$$

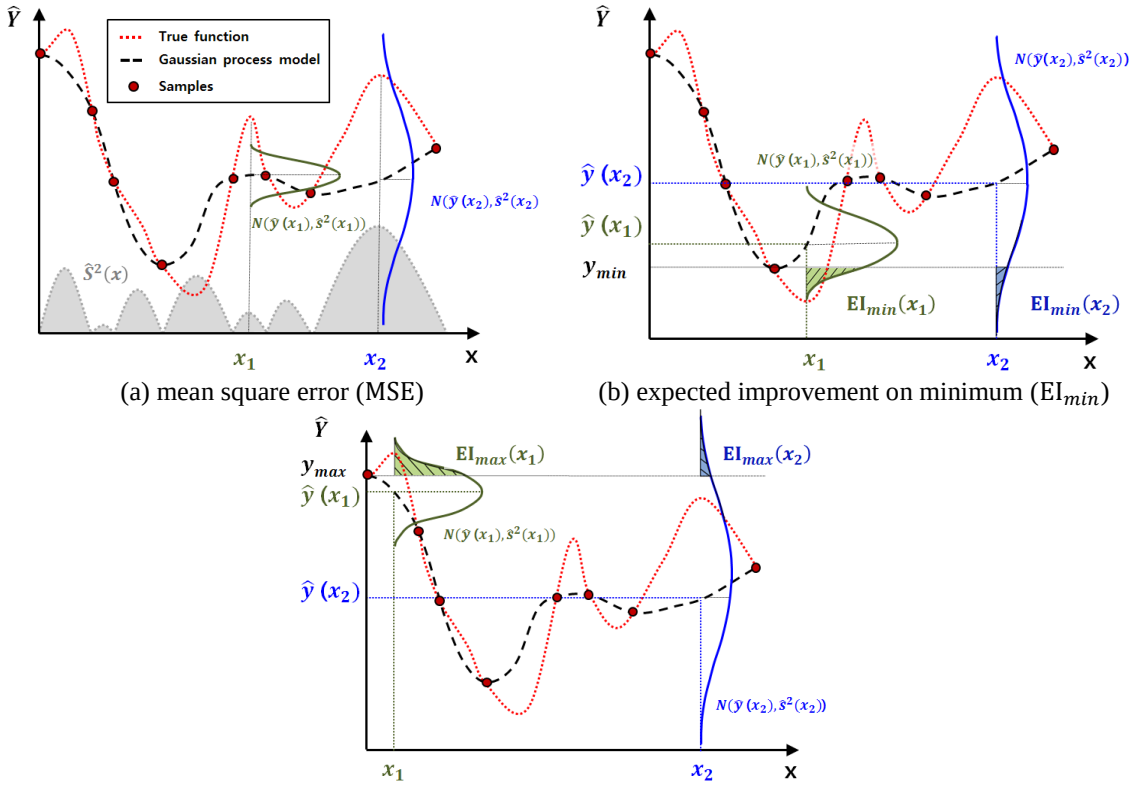
where $\Phi(\mathrm{x})$ and $\phi(\mathrm{x})$ are the Gaussian cumulative distribution function (using error function) and probability density function respectively. The derivation of EI can be found in literatures of global optimization, which has been given in Jones *et al*. (1998) and Forrester *et al*. (2008). Expected improvement on a maximum (EI_{min}) is also derived in the same way and it is

$$\mathrm{EI}_{max}(\boldsymbol{x}_{new}) = \left(\hat{\mathrm{y}}(\boldsymbol{x}_{new}) - \mathrm{y}_{\max}\right)\Phi\left(\frac{\hat{\mathrm{y}}(\boldsymbol{x}_{new}) - \mathrm{y}_{\max}}{\hat{\mathrm{s}}(\boldsymbol{x}_{new})}\right) + \hat{\mathrm{s}}(\boldsymbol{x}_{new})\phi\left(\frac{\hat{\mathrm{y}}(\boldsymbol{x}_{new}) - \mathrm{y}_{\max}}{\hat{\mathrm{s}}(\boldsymbol{x}_{new})}\right) \tag{11}$$

where $\mathrm{y}_{\max}$ is the maximum in the training samples.

3.2 Sequential surrogate modeling with self-adaptive sampling

Sequential surrogate modeling with self-adaptive sampling is based on a Bayesian framework, since candidate samples for infill are inferred by a current GPM. Three infill criteria are used as follows: MSE, EI_{min}, and EI_{max}. The inferred sample from MSE is served only to reduce the regions which are likely higher uncertain (i.e., global exploration), while inferred samples from EI_{min} and EI_{max} balance the local exploitation of current best values and global exploration. As the proposed approach goes along sequentially, the inferred samples improve the current GPM to approximate a response surface more accurately.



(a) mean square error (MSE)

(b) expected improvement on minimum (EI_{min})

(c) expected improvement on maximum (EI_{max})

Fig. 3 Infill criteria for self-adaptive sampling

The inferred samples are also used as a diagnostic to evaluate the adequacy of the current GPM. As shown in Fig. 4(a), there is the deviation between true outputs (Y) and corresponding GPM predictions ($\hat{Y}$) under an inadequate GPM. Through sequential model updates by the inferred samples, the predictions of inferred samples are getting close to true outputs as shown in Fig. 4(b).

As the validation measures for the diagnostic, R-squared value (R^2) and root mean square error (RMSE) are used in the proposed approach. R-squared value is also known as the coefficient of determination. It measures how well the plot between observed outputs and their prediction fits the 1:1 line. It is the proportion of total variation of outputs explained by the prediction model and expressed as

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\sum_{i=1}^{3}\left(y(\mathbf{x}_i)-\hat{y}(\mathbf{x}_i)\right)^2}{\sum_{i=1}^{3}(y(\mathbf{x}_i)-\bar{y})^2}, \qquad 0 \le R^2 \le 1 \tag{12}$$

where SSE is the sum of squares of residuals, SST is the total sum of squares. $\hat{y}(\boldsymbol{x}_i)$ and $y(\mathbf{x}_i)$ denote the prediction value from the current GPM and true output at the inferred samples. $\bar{y}$ is the average value of y. R^2 ranges from 0 to 1. The larger value close to one indicates the more accurate prediction model; See Fig. 4(b).

Another measure is RMSE that aggregates the magnitude of the residuals into a single measure of the prediction accuracy. The smaller value close to zero means the higher prediction accuracy.

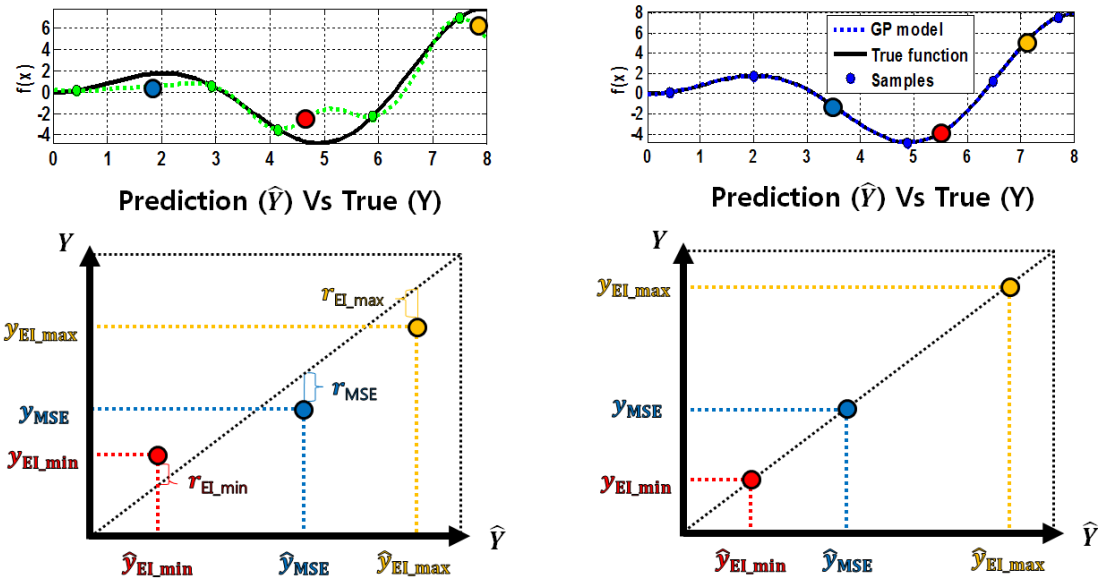
$$\mathrm{RMSE} = \sqrt{\frac{\sum_1^{\mathrm{n_{val}}}\left(\mathrm{y}(\mathbf{x}_{\mathrm{i,val}}) - \hat{\mathrm{y}}(\mathbf{x}_{\mathrm{i,val}})\right)^2}{\mathrm{n_{val}}}} \tag{13}$$

In practice, it is impossible to obtain the ideal values of R^2 (i.e., one) and RMSE (i.e., zero) due to numerical truncation error and the nature of approximation (i.e., small deviation also known as code uncertainty (Kennedy and O'Hagan 2001)). Therefore, thresholds are adopted to evaluate whether the current GPM approximates a response surface accurately or not. If both R_k^2 and RMSE_k are satisfied with predefined thresholds, the current GPM is considered to be adequate. Especially, RMSE may fluctuate greatly at the early stages. So there is a risk that the proposed approach stops prematurely. To avoid such an undesired stopping, stall generation (*# stall*) with two stopping criteria are used as given in

$$R_k^2 \geq R_{Thresh}^2 \tag{14}$$

$$|\mathrm{RMSE}_k - \mathrm{RMSE}_{k+1}| \leq \Delta\mathrm{RMSE}_{Thresh}(1 + |\mathrm{RMSE}_k|) \tag{15}$$

where the subscript k indicates the k-th infill stage and *Thresh* indicates the predefined threshold. Especially, Eq. (15) is a lower bound on the relative change of the RMSE value as a measure of the convergence. Since two stopping criteria are very unstable at the initial phase with an inaccurate GPM, two stopping criteria (R^2 and relative change of RMSE) are complementary measures at the initial phase of the propose method. In addition, the stall generation also works for a diagnostic measure to evaluate whether GPM is adequate or not, once adequate GPM meet the two stopping criteria sequentially within the stall generation.



(a) under inadequate GPM (b) under adequate GPM

Fig. 4 Convergence to adequate Gaussian process model with the proposed approach

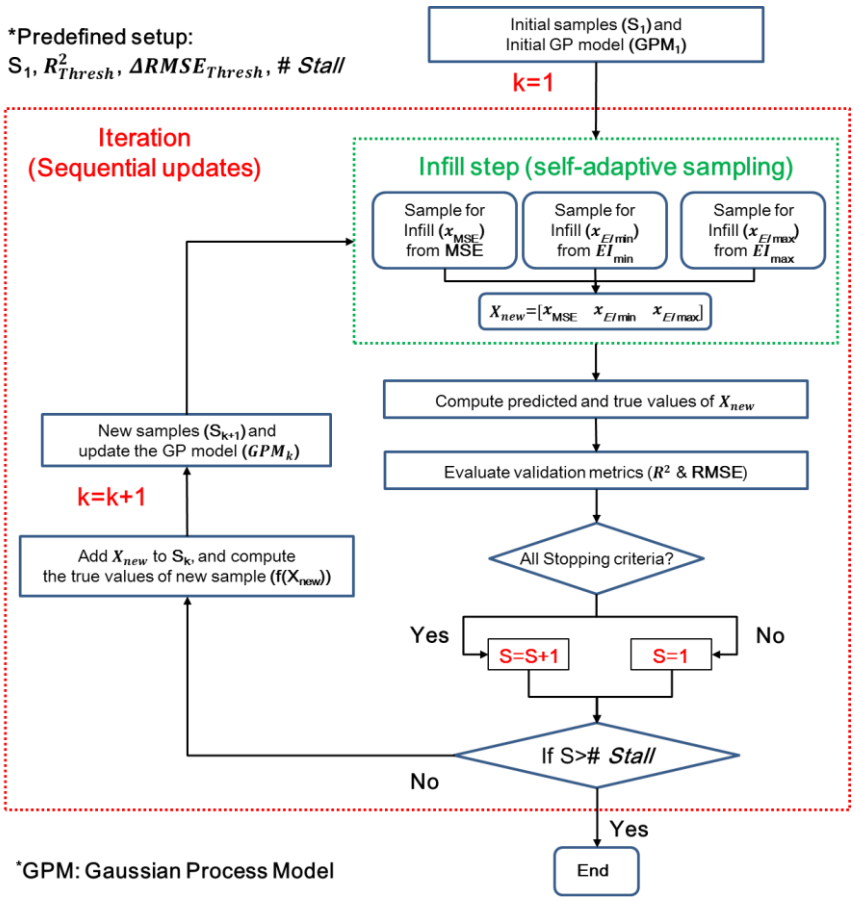
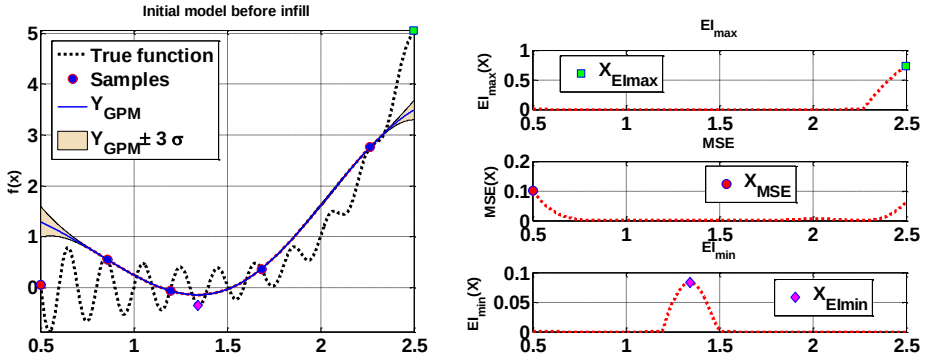


Fig. 5 Flowchart of the proposed approach

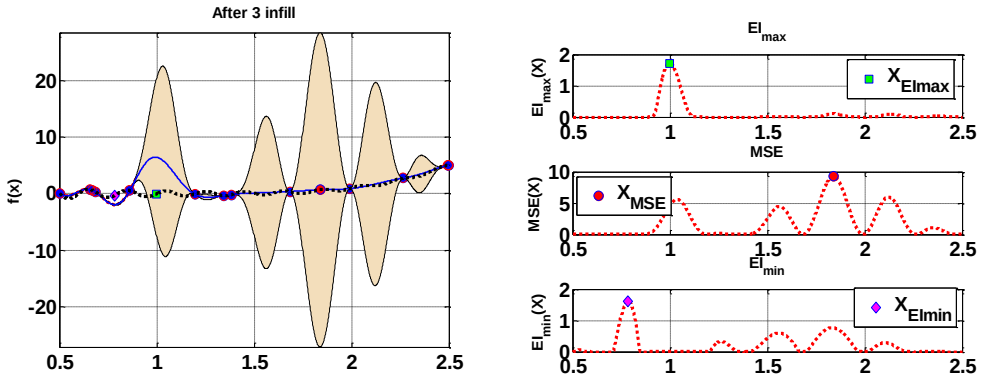
The procedure of the proposed approach is illustrated in Fig. 5. At initialization, there are four setups to be predefined by user: initial DOE ($\mathbf{X}_1$); thresholds (R^2_{Thresh} & $\Delta RMSE_{Thresh}$); and stall generation (*# stall*). After, the rest of the algorithm is performed automatically.

To illustrate the proposed approach, Eq. (8) is revisited as the toy example, which is a non-stationary 1D-function. To represent the concept of sparseness in high-dimensional space, four initial training samples are used. R^2_{Thresh} and $\Delta RMSE_{Thresh}$ are set to 0.98 and 0.05 respectively. Stall generation (*# stall*) is set to 3.

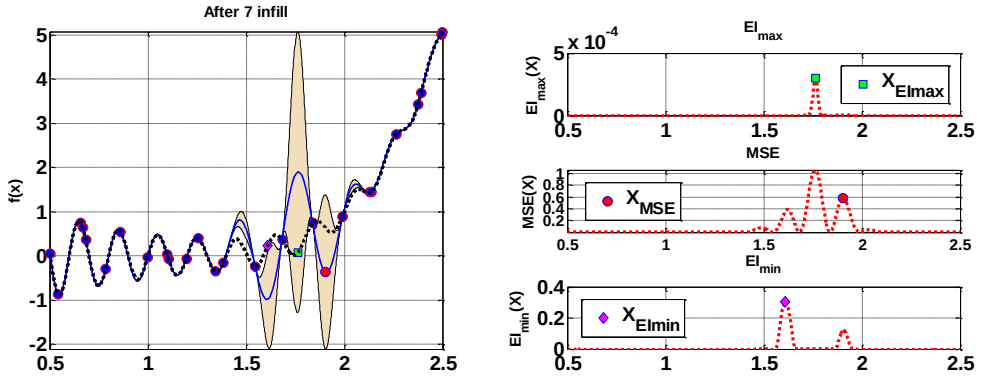
The progress of the proposed approach is shown in Fig. 6. The left side describes the convergence history to the true function, while the right side shows the infill criteria with the inferred samples. Shaded area of the right side expresses the uncertainty of each prediction with 3 sigma-level. MSE allocates the candidate (red marker) to the higher uncertainty region (i.e., large standard error), while EI_{min} and EI_{max} allocate the candidates (magenta and green markers) to the regions which is likely to improve the best values of the current GPM. It is observed that GPM is getting converged to the true function by the sequential GPM updates.



(a) GPM before candidates for infill



(b) updated GPM after 3 infill stage (adding 9 samples for infill)



(c) updated GPM after 7 infill stage (adding 21 samples for infill)

Fig. 6 Progress in the approximation of updated Gaussian process model

After the 11 infill stage, the proposed approach is stopped with a total of 37 samples: 4 initial samples and additional 33 samples for infill. Although the initial DOE is insufficient to approximate the true function (See Fig. 6(a)), the differences between the true values and the final GPM predictions are not recognized and the corresponding residuals are almost zero (See bottom plot of Fig. 7).

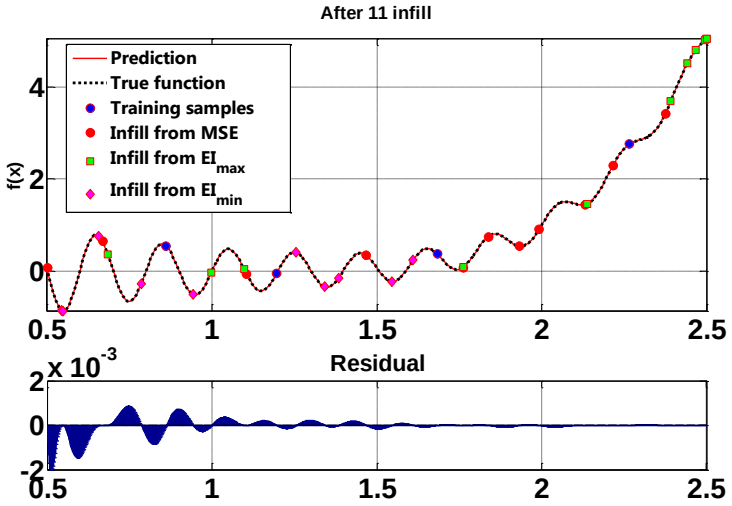


Fig. 7 Final updated Gaussian process model by the proposed approach (After the 11 infill stage)

4. Numerical study

In this section, we evaluate the efficiency and effectiveness of the proposed approach with mathematical test functions and an engineering application. To simulate the sparseness of training samples, small initial samples are generated randomly by LHD. In all examples, R^2_{Thresh} and $\Delta RMSE_{Thresh}$ are set to 0.98 and 0.05, respectively.

For the mathematical test functions, stall generation is set to 3 and a total of 10,000 validation samples are generated by independently and identically distributed uniform distributions. On the other hand, in the engineering application, a total of 30,000 random samples are generated under various input uncertainties, and the generated samples are used for Monte Carlo simulation (MCS). The response uncertainty from both MCS, based on direct FEA and surrogate model, are compared to evaluate the performance of the proposed approach.

4.1 5-dimensional test function

The test function with five input variables is given in Eq. (16) (Friedman 1991). It is apparent that there is a strong interaction effect that decays at different rates due to $10\sin(\pi x_1 x_2)$ in the test function. It is evaluated on a hypercube with $x_{i=1,..,5} \in [0,\ 1]$. As the training samples, 20 samples are randomly generated and the proposed approach is performed. It is repeated by 10 times.

$$f(x) = 10\sin(\pi x_1 x_2) + 20(x_3 - 0.5)^2 + 10x_4 + 5x_5 \tag{16}$$

The total number of the required samples ranges 68 from 134, since each repetition starts with the different subset of the initial samples. Mean and standard deviation of the training samples is 102.8 and 20.75, respectively (See Fig. 8(c)).

10,000 validation samples are used to evaluate how well final GPMs approximate the true function. Under the identical validation samples, 10 GPMs are used to estimate the predictions. To remove the scale-dependent problem in RMSE (Hyndman and Koehler 2006), RMSE is normalized by the range of the true outputs, $\text{NRMSE} = \text{RMSE}/(y_{max} - y_{min})$.

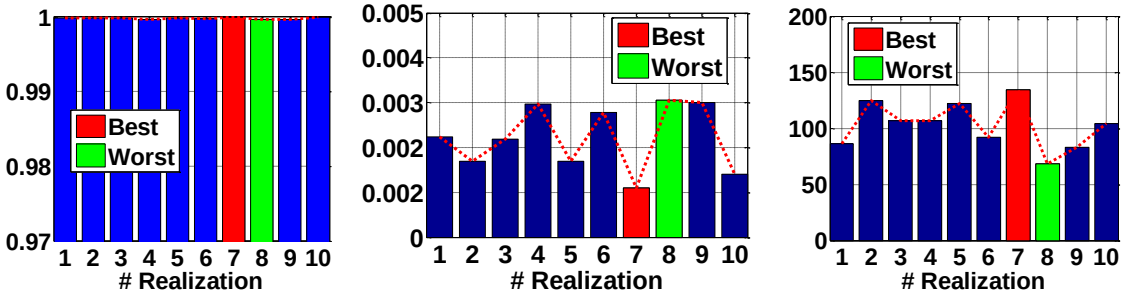
Fig. 8 shows that 10 GPMs exhibit the highly accurate approximation of the true function. Among the final GPMs, the best and worst GPMs are selected and used to plot the true outputs versus predictions as shown in Fig. 9. It is shown that two GPMs reproduce the true output with the negligible deviations.

4.2 8-dimensional test function

The test function with eight input variables is given in

$$f(x) = 4(x_1 - 2 + 8x_2 - 8x_2^2)^2 + (3 - 4x_2)^2 + 16\sqrt{x_3 + 1}(2x_3 - 1)^2 + \sum_{i=4}^{8} i \ln\left(1 + \sum_{j=3}^{i} x_j\right) \quad (17)$$

It is highly curved in some variables and is less curved in other variables. It is evaluated on a hypercube with $x_{i=1,\ldots,8} \in [0,\ 1]$. As the training samples, 80 samples are randomly generated and the proposed approach is performed. It is also repeated by 10 times.



(a) R-squared value (b) normalized RMSE (c) number of final samples

Fig. 8 Validation of 10 Gaussian process models (10,000 validation samples)

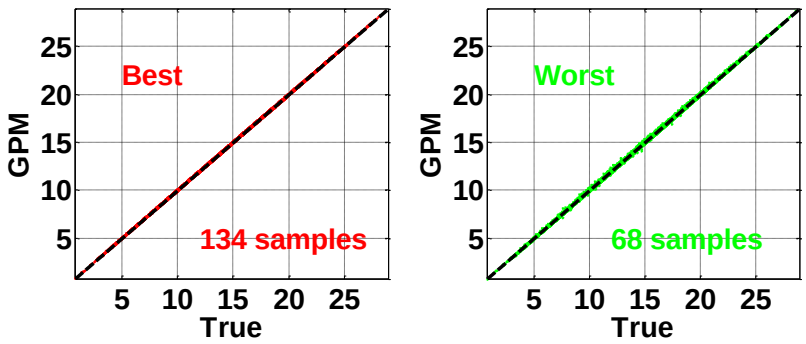


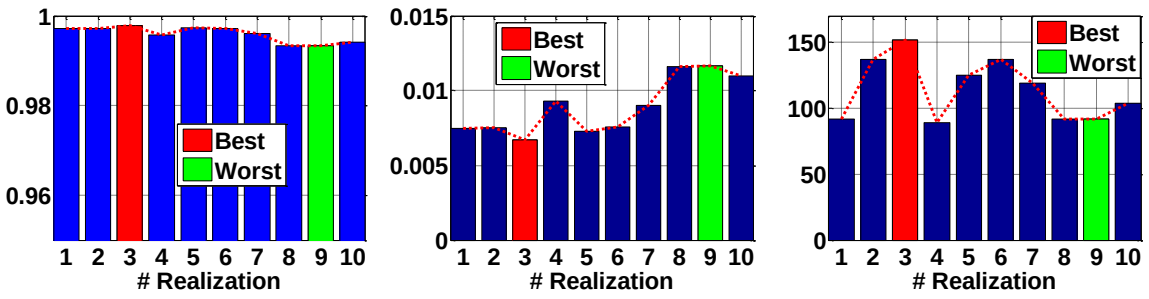
Fig. 9 Prediction versus true outputs

It is observed that the computed NRMSE is larger than those in Fig. 8(b) and the number of final samples is also quite increased. The total number of the required samples ranges 92 from 152, since each repetition starts with the different subset of the initial samples. Mean and standard deviation of the training samples is 113.9 and 23.15 respectively (See Fig. 9(c)).

As in the previous test, all GPMs can reproduce the true outputs within the highly acceptable deviations as shown in Fig. 11, although the input dimension increases and the true function is much complicated than Eq. (16).

4.3 Response uncertainty analysis based on MCS with correlated input parameters

As an engineering application, a three-bay-five-story frame structure is investigated. The geometric representation of the structure is illustrated in Fig. 12. Using SAP2000, a finite element model (FEM) is modeled by the frame element and fixed supports. Frame elements are grouped into four groups (C1, C2, B1, and B2). Such grouping is chosen to make the response surface complicated. In Fig. 12, Groups are represented by different colors. Based on this grouping, there are 13 input parameters: three point loads (P1~P3); two Young's moduli (E4, E5); four moments of inertia (I6~ I8); and sectional areas (A10~A14).



(a) R-squared value (b) normalized RMSE (c) number of final samples

Fig. 10 Validation of 10 Gaussian process models (10,000 validation samples)

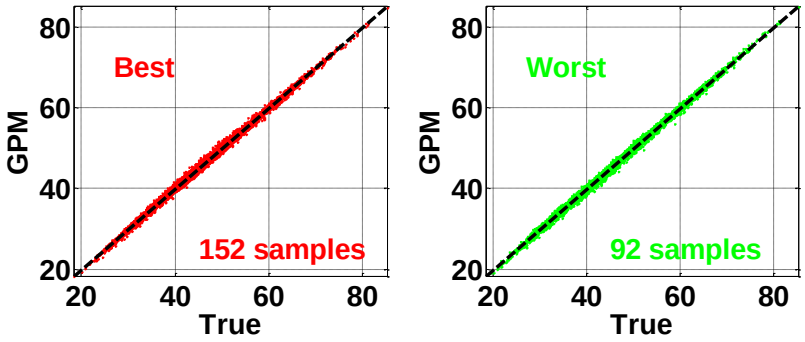


Fig. 11 Prediction versus true outputs

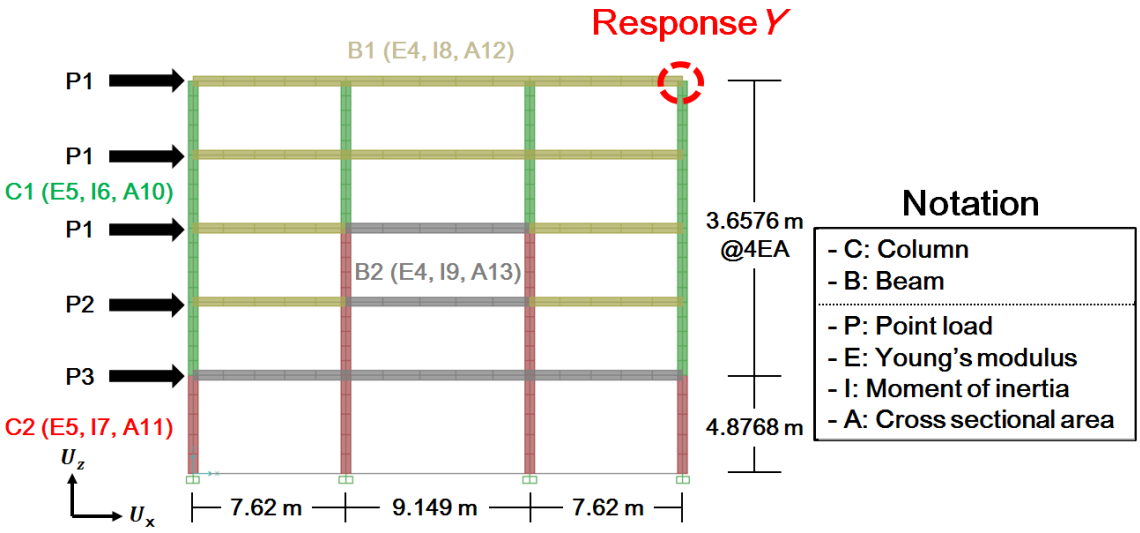


Fig. 12 Geometric representation

Table 1 Statistical information of input uncertainty

Description	Notation	Distribution	Mean, μ	St dev, σ	COV, %	Interval, [min max]
Load (KN)	P1	Rayleigh	133.45	40.04	30%	[53.38 333.63]
	P2	Rayleigh	88.97	26.69	30%	[35.59 222.43]
	P3	Rayleigh	71.18	21.35	30%	[28.47 177.95]
Young's Modulus (MPa)	E4	Normal	21,738	1,956	9%	[10,869 32,606]
	E5	Normal	23,796	2,142	9%	[11,898 35,695]
Moment of inertia (m^4)	I6	Normal	0.0081	0.0008	10%	[0.0041 0.0122]
	I7	Normal	0.0115	0.0012	10%	[0.0058 0.0173]
	I8	Normal	0.0232	0.0023	10%	[0.0116 0.0348]
	I9	Normal	0.0259	0.0026	10%	[0.0130 0.0389]
Sectional area (m^2)	A10	Normal	0.0312	0.0025	8%	[0.0156 0.0468]
	A11	Normal	0.3716	0.0297	8%	[0.1858 0.5574]
	A12	Normal	0.3725	0.0298	8%	[0.1863 0.5588]
	A13	Normal	0.4181	0.0334	8%	[0.2091 0.6272]

Table 1 shows the 13 input parameters and their statistical uncertainty. Additionally, it is assumed that some parameters are correlated as follows: Young's modulus ($\rho_{P_i,P_j} = 0.9$ with $i \neq j$); sectional properties of the identical element ($\rho_{I_i,A_i} = 0.95$); and all other section properties ($\rho_{I_i,I_j} = \rho_{I_i,A_j} = \rho_{A_i,A_j} = 0.13$ with $i \neq j$). To generate the correlated random samples, the Gaussian copula model is used. In order to make the response surface complicated, the upper and lower bounds are chosen more than 3 standard deviations from the mean.

With the 13 input parameters, Monte Carlo simulation (MCS) with the 30,000 random samples is performed to estimate the uncertainty of the horizontal displacement at the top floor. The structural analysis on generated random samples is performed by the open application programming interface (OAPI) of SAP2000 linked with MATLAB 2014a.

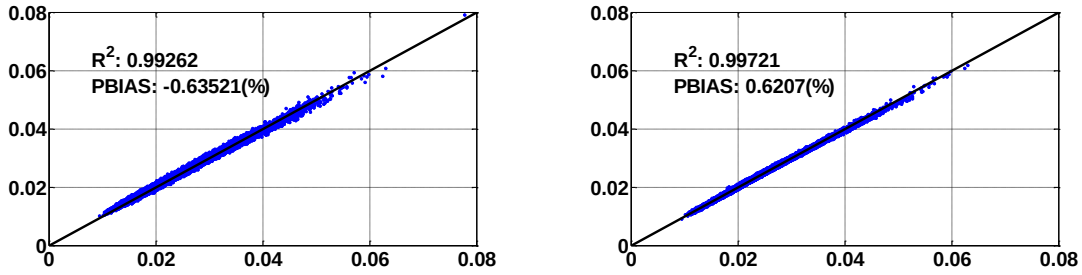
For surrogate modeling, two cases with different sample size are considered. The first case with 50 initial samples is referred to "Case A", while the other case starts with 100 initial samples and it is referred to "Case B". Initial samples of both cases are generated within the upper and lower bounds as tabulated in Table 1 (See 7th and 8th columns). The proposed approach is performed to each case.

The number of infill samples and total training samples are tabulated in Table 2. Although both cases start with different initial sample, the proposed approach is finished with the similar number of training sample: 237 (case A) and 196 (case B). It implies that GPM requires around 200 of the number of training sample to approximate the true response surface with sufficient accuracy.

As shown in Fig. 13, the responses of FEA (Y_{True}) is plotted against to those of the final GPM (Y_{GPM}). As given in Eq. (18), Percent bias (PBIAS) is the average tendency of the simulated data ($\hat{Y}$) to their observed data (Y) (Moriasi *et al*. 2007). Positive values indicate the underestimation bias of the prediction, and negative values indicate the overestimation bias. It is found that both GPMs can produce the target response accurately. In both cases, the good agreements between Y_{True} and Y_{GPM} is observed with 0.99 of the R^2 and 0.64% of PBIAS.

$$\text{PBIAS}(\%) = \frac{\sum(\hat{Y}-Y)}{\sum Y} \times 100 \tag{18}$$

Fig. 14 shows the empirical distributions of the response. It is observed that both empirical estimates from Y_{True} and Y_{GPM} are matched well. Lastly, the statistical parameters of the response are computed as tabulated in Table 3. The estimates of statistical parameters are also matched well with maximum 3.5% of the relative error. Based on these comparisons, the final GPMs from the proposed approach reproduce the true outputs with the acceptable deviations automatically.



(a) Case A: 50 initial samples (b) Case B: 100 initial samples

Fig. 13 Prediction versus true outputs (X-axis: true output, Y-axis: GPM prediction)

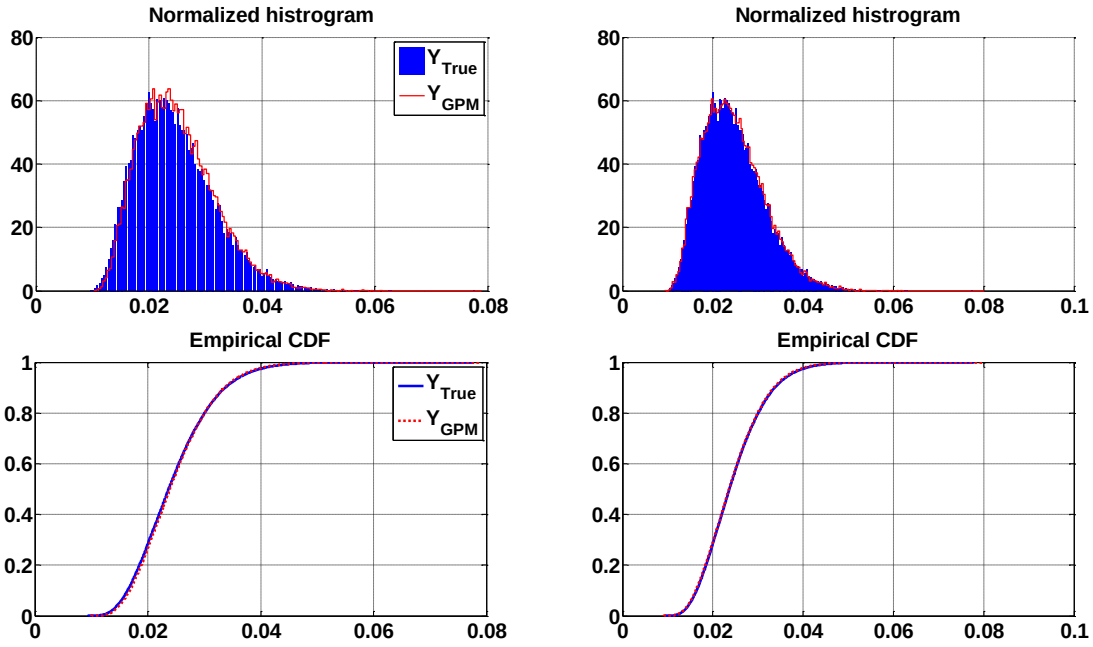
Table 2 Total number of the sample evaluation

	Initial samples size	Infill samples size	Total samples size
Case A	50	177 (59 infill stages)	237
Case B	100	96 (32 infill stages)	196

Table 3 Comparison of the statistical parameters

	FEA	Case A (50 initial samples)		Case B (100 initial samples)	
		GPM	Relative error*	GPM	Relative error*
Mean	0.0245	0.0247	-0.64 %	0.0244	0.62 %
Standard deviation	0.0069	0.0066	3.35 %	0.0068	0.98 %
Skewness	0.7308	0.7059	3.40 %	0.7118	2.60 %
Kurtosis	3.6875	3.6638	0.64 %	3.5961	2.48 %

*$\left(\frac{\text{FEA}-\text{GPM}}{FEA}\times 100\right)$



(a) Case A: 50 initial samples (b) Case B: 100 initial samples

Fig. 14 Frequency distributions of responses

Table 4 Summary statistics of computational prediction time

Computation time (sec)	Direct finite element analysis	Gaussian process model
Mean	10.41	0.0036
Standard deviation	0.26	0.0057

Table 5 Comparison for the computational efficiency

	Surrogate modeling	Uncertainty analysis	Estimated total time
Direct FEA	0 sec (0 FEA × 10.41 sec)	312,300 sec (30,000 FEA × 10.41 sec)	312300 sec (86.75 hours)
GPM by proposed method (Case A)	2467.17 sec (237 FEA × 10.41 sec)	108 sec (30,000 GPM × 0.0036 sec)	2575.17 sec (0.72 hours)

In order to show the computational efficiency of the proposed method over the direct FE analysis, the total time of performing uncertainty analysis is estimated. Under the identical computational environment (Intel(R) Xeon(R) E5-2660 v2 processor running with 2.20 and 2.90 GHz, 16 GB RAM, and Windows7 OS), the average prediction times of both FEA and GPM are estimated by a total of 30 simulations and tabulated in Table 4. The prediction time needed for FEA is 10.41 seconds on average. On the other hand, the prediction time of GPM is around 0.0036 seconds.

Based on these prediction times, a comparison between computation costs for the direct FEA and the GPM by proposed method is tabulated in Table 5. The total number of the required FEA in the proposed method is 237 (i.e., case A), as tabulated in Table 2. Once GPM is constructed, additional FEA is not required. In contrast, direct FEA only requires 30,000 times of FEA in uncertainty analysis. Uncertainty analysis with direct FEA takes around 86.75 hours to complete, while uncertainty analysis with GPM takes around 0.72 hours. It implies that GPM by the proposed method can reduce computational times and resources dramatically with highly acceptable predictive capability.

5. Conclusions

A variety of the methods have been widely investigated in surrogate modeling to reduce the computational cost under an iterative codes analysis. Although the various methods are available, there are still difficulties in surrogate modeling from a practical point of view: (1) How to generate the training samples; and (2) diagnostics of the surrogate model. Because prior knowledge on a true response surface behavior is not available, the trial-and-error approach with the user intervention is required.

In this regard, we propose the sequential surrogate modeling with self-adaptive sampling for time-consuming finite element analysis. The novelty of the proposed approach is the generative

learning ability about a true response surface in a Bayesian framework. The proposed approach uses infill criteria in parallel to infer the candidate samples to improve the surrogate model.

To demonstrate the applicability of the proposed approach, it is numerically tested with the mathematical test functions and engineering application: (1) the non-stationary 1-D test function; (2) 5-D test function with the strong interaction effect; (3) the highly non-linear 8-D test function; and (4) Monte Carlo simulation for the response uncertainty under the 13 correlated random inputs. Good agreement between the true outputs and predictions is observed in all numerical examples. The proposed approach provides not only further sampling for the better approximation, but also diagnostics of model adequacy under the insufficient samples and complicated response surface.

The proposed approach seems to be promising to build a surrogate model automatically with the minimal user intervention. The generative learning ability about a true response surface can enable the proposed approach to be customized for the various response surfaces. Therefore, it is expected that less experienced user with surrogate modeling can significantly save the time and efforts.

Acknowledgments

This research was partially supported by a grant (13SCIPA01) from Smart Civil Infrastructure Research Program funded by Ministry of Land, Infrastructure and Transport (MOLIT) of Korea government and Korea Agency for Infrastructure Technology Advancement (KAIA) and financially supported by Korea Ministry of Land, Infrastructure and Transport (MOLIT) as (U-City Master and Doctor Course Grant Program).

References

Bastos, L.S. and O'Hagan, A. (2009), "Diagnostics for Gaussian process emulators", *Technometrics*, **51**(4), 425-438.

Bellman, R. (2003), *Dynamic Programming*, Dover Publications, Mineola, N.Y.

Box, G.E.P. and Behnken, D.W. (1960), "Some new three level designs for the study of quantitative variables", *Technometrics*, **2**(4), 455-475.

Box, G.E.P., Hunter, W.G. and Hunter, J.S. (1978), *Statistics for Experimenters : An Introduction to Design, Data Analysis, and Model Building*, Wiley, New York.

Bucher, C. and Most, T. (2008), "A comparison of approximate response functions in structural reliability analysis", *Probabilist. Eng. Mech.*, **23**(2-3), 154-163.

Bucher, C.G. and Bourgund, U. (1990), "A fast and efficient response-surface approach for structural reliability problems", *Struct. Safety*, **7**(1), 57-66.

Dette, H. and Pepelyshev, A. (2010), "Generalized latin hypercube design for computer experiments", *Technometrics*, **52**(4), 421-429.

DiazDelaO, F.A. and Adhikari, S. (2011), "Gaussian process emulators for the stochastic finite element method", *Int. J. Numerical Meth. Eng.*, **87**(6), 521-540.

Dubourg, V., Sudret, B. and Deheeger, F. (2013), "Metamodel-based importance sampling for structural reliability analysis", *Probabilist. Eng. Mech.*, **33**, 47-57.

Fang, K.T., Lin, D.K.J., Winker, P. and Zhang, Y. (2000), "Uniform design: Theory and application", *Technometrics*, **42**(3), 237-248.

Forrester, A.I.J. and Keane, A.J. (2009), "Recent advances in surrogate-based optimization", *Prog.*

Aerospace Sci., **45**(1-3), 50-79.
Forrester, A.I.J., Sóbester, A.S. and Keane, A.J. (2008), *Engineering Design Via Surrogate Modelling : A Practical Guide*, J. Wiley, Chichester, West Sussex, England ; Hoboken, NJ.
Friedman, J. H. (1991), "Multivariate adaptive regression splines", *Annals of Statistics*, **19**(1), 1-67.
Goel, T., Haftka, R.T., Shyy, W. and Watson, L.T. (2008), "Pitfalls of using a single criterion for selecting experimental designs", *Int. J. Numer. Meth. Eng.*, **75**(2), 127-155.
Gramacy, R.B. and Lee, H.K.H. (2012), "Cases for the nugget in modeling computer experiments", *Statist. Comput.*, **22**(3), 713-722.
Hyndman, R.J. and Koehler, A.B. (2006), "Another look at measures of forecast Accuracy", *Int. J. Forecast.*, **22**(4), 679-688.
Jones, D.R. (2001), "A taxonomy of global optimization methods based on response surfaces", *J. Global Optim.*, **21**(4), 345-383.
Jones, D.R., Schonlau, M. and Welch, W.J. (1998), "Efficient global optimization of expensive black-box functions", *J. Global Optim.*, **13**(4), 455-492.
Kennedy, M. C. and O'Hagan, A. (2001), "Bayesian calibration of computer models", *J. Roy. Statist. Soc. Series B-Statistical Methodology*, **63**, 425-450.
Kiefer, J. and Wolfowitz, J. (1959), "Optimum designs in regression problems", *Annals Mathematical Statist.*, **30**(2), 271-294.
Kleijnen, J.P.C. and van Beers, W.C.M. (2004), "Application-driven sequential designs for simulation experiments: Kriging metamodelling", *J. Operational Res. Soc.*, **55**(8), 876-883.
Krige, D.G. (1994), "A Statistical approach to some basic mine valuation problems on the witwatersrand", *J. South African Institute of Mining and Metallurgy*, **94**(3), 95-111.
Mckay, M.D., Beckman, R.J. and Conover, W.J. (1979), "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code", *Technometrics*, **21**(2), 239-245.
Moriasi, D.N., Arnold, J.G., Van Liew, M.W., Bingner, R.L., Harmel, R.D. and Veith, T.L. (2007), "Model evaluation guidelines for systematic auantification of accuracy in watershed simulations", *Transactions of the Asabe*, **50**(3), 885-900.
Queipo, N.V., Haftka, R.T., Shyy, W., Goel, T., Vaidyanathan, R. and Tucker, P.K. (2005), "Surrogate-based analysis and optimization", *Prog. Aerospace Sci.*, **41**(1), 1-28.
Rougier, J., Sexton, D.M.H., Murphy, J.M. and Stainforth, D. (2009), "Analyzing the climate sensitivity of the Hadsm3 climate model using ensembles from different but related experiments", *J. Climate*, **22**(13), 3540-3557.
Sacks, J., Welch, W.J., Mitchell, T.J. and Wynn, H.P. (1983), "Design and analysis of computer experiments", *Statist. Sci.*, **4**(4), 409-423.
Xiong, Y., Chen, W., Apley, D. and Ding, X.R. (2007), "A non-stationary covariance-based kriging method for metamodelling in engineering design", *Int. J. Numer. Meth. Eng.*, **71**(6), 733-756.
Ye, K.Q. (1998), "Orthogonal column latin hypercubes and their application in computer experiments", *J. American Statistical Association*, **93**(444), 1430-1439.
Zhang, Z., Jiang, C., Han, X., Hu, D. and Yu, S. (2014), "A response surface approach for structural reliability analysis using evidence theory", *Adv. Eng. Software*, **69**, 37-45.