Computers and Chemical Engineering 92 (2016) 143–159

Contents lists available at ScienceDirect

Computers and Chemical Engineering

journal homepage: www.elsevier.com/locate/compchemeng

Estimation of percentiles using the Kriging method for uncertainty propagation

Frits Byron Soepyan[a], Selen Cremaschi[a,*], Cem Sarica[b], Hariprasad J. Subramani[c], Gene E. Kouba[c], Haijing Gao[c]

[a] The University of Tulsa, Russell School of Chemical Engineering, 800 South Tucker Drive, Tulsa, OK 74104, USA
[b] The University of Tulsa, McDougall School of Petroleum Engineering, 800 South Tucker Drive, Tulsa, OK 74104, USA
[c] Chevron Energy Technology Company, 1400 Smith Street, Houston, TX 77002, USA

ARTICLE INFO

Article history:
Received 10 December 2015
Received in revised form 24 May 2016
Accepted 25 May 2016
Available online 27 May 2016

Keywords:
Uncertainty propagation
Model output uncertainty
Kriging surrogate model
Solid particle transport
Threshold velocity

ABSTRACT

The use of simulation-based uncertainty propagation approaches (e.g., Monte Carlo simulation method) can be computationally expensive if evaluating the function requires a relatively large computation time. To reduce the computation time, uncertainty propagation methods that use surrogate models (e.g., the Kriging method) may be used. In this paper, we extend the Kriging method to propagate the uncertainties from multiple sources, and for cases where the distribution of the prediction is produced at each trial (replication) of the simulation-based uncertainty propagation approach (i.e., at each sample point). The outputs of the methodology are the approximate percentiles of the output distribution. The capability of the methodology is tested using a Case Study involving the transport of solid particles in pipelines to prevent solid particle deposition and improve pipeline efficiency. Statistical comparisons suggest that our methodology successfully replicates the outputs from the Monte Carlo simulation method with a 94% reduction in computational cost.

1. Introduction

Uncertainty propagation methods are used in cases where a function $f(\underline{x})$ is needed to predict the value of a dependent variable y, where $f(\underline{x})$ may be an analytical function, or a “black-box-type performance function” (Lee and Chen, 2009). Here, y is dependent on $\underline{x}$, with $\underline{x}$ representing a $d \times 1$ vector whose components contain the values of the independent variables, and d the number of independent variables. In this case, at least one of the components of $\underline{x}$ is a random variable with a probability density function or marginal distribution function and correlations (Lee and Chen, 2009).

In many applications, simulation-based uncertainty propagation methods have been used because of their simplicity and robustness (Roy and Oberkampf, 2011). Examples of simulation-based uncertainty propagation methods include the Monte Carlo simulation method (Dieck, 2007), the Latin hypercube sampling method (Hora and Helton, 2003; McKay et al., 2000), and the stratified sampling method (Giunta et al., 2006; McKay et al., 2000). To propagate the uncertainties of the random independent variables using the Monte Carlo simulation method (assuming the random input or independent variables are statistically independent), first, a random number generator selects random values of the independent variables from the distributions of these variables (Dieck, 2007). Then, the value of $f(\underline{x})$ is computed using the randomly-selected values of the independent variables. This process is repeated hundreds or thousands of times, and the output of the Monte Carlo simulation method is an estimate of the distribution of $f(\underline{x})$ (i.e., the approximate distribution of the simulation response y).

The procedure for implementing the Latin hypercube sampling method is similar to the Monte Carlo simulation method. However, prior to selecting random values of the random independent variables, the interval of possible values for each random independent variable is divided into n_{int} number of subintervals of equal probability (Hora and Helton, 2003; McKay et al., 2000). Using a random number generator, in each interval, a random value of the independent variable is selected from the distribution of the variable. Then, the n_{int} random values of the independent variable are grouped randomly with the other random values of the remaining independent variables, without replacement. The value of $f(\underline{x})$ is then quantified for each combination of the

* Corresponding author. Present address: Auburn University, Department of Chemical Engineering, 212 Ross Hall, Auburn, AL 36849-5127, USA.
E-mail addresses: selen-cremaschi@auburn.edu, selen.aydogan@gmail.com (S. Cremaschi).

http://dx.doi.org/10.1016/j.compchemeng.2016.05.015
0098-1354/

Nomenclature

A_F	Flow area of the fluid in the pipe
d	Total number of independent variables
D_h	Hydraulic diameter of the pipe
$\hat{f}$	Approximate value or approximate percentile value of y generated using the Kriging function
$\underline{\hat{f}_{MC}}$	$n \times 1$ vector whose elements are the values of the model's prediction at each sample point
$\underline{\hat{f}_{MC,p}}$	$n \times 1$ vector whose elements are the values of the p^{th} percentile of the simulation output y at each sample point
$\underline{\hat{f}_{MC,95}}$	$n \times 1$ vector whose elements are the values of the 95^{th} percentile of the simulation output y at each sample point
$f(\underline{x})$	Function that predicts the value of a dependent variable
$F(z)$	Cumulative distribution function of z
$F_1(x)$	Distribution function of a set of numbers
$F_2(x)$	Distribution function of a set of numbers
H_a	Alternative hypothesis
H_0	Null hypothesis
i	Index used for the sample points
j	Index used for the sample points
k	Index used for the independent variables
m	Slope of the best-fitting line in the parity plot of the model
m_0	Slope of the best-fitting line in the parity plot of the model, with the intercept forced to be at the origin
n	Total number of sample points used to generate the Kriging function
n_{data}	Number of data points that populate the reduced database that contains the most representative experimental data points for the user-defined input combination
n_{int}	Number of intervals of equal probability
$n_{Kriging}$	Number of trials (replications) for the Monte Carlo simulation method that are used to generate the Kriging function
n_{MC}	Number of trials (replications) for the Monte Carlo (MC) simulation method when the Kriging method is not used
n_{sample}	Sample size
p	Percentile value
P_W	Wetted perimeter of the pipe
$\underline{\hat{R}}$	$n \times n$ matrix of correlations of all n number of sample points
$\hat{r}_i$	i^{th} component of the vector $\underline{\hat{r}}(\underline{x_0})$
$\hat{R}_{ij}$	Element of the ith row and jth column of $\underline{\hat{R}}$
$\underline{\hat{r}}(\underline{x_0})$	$n \times 1$ vector of correlations between $\underline{x_0}$ and all n number of sample points
$R^2{}_{adj}$	Value of modified adjusted-R^2 statistic of the model
t	Index used for the experimental data points in the reduced database
u	$n \times 1$ vector whose components have values of one
v_{calc}	Threshold velocity prediction of the model
$v_{5,all}$	5^{th} percentile value of the threshold velocity prediction of the model produced from propagating all uncertainties
$v_{5,input}$	5^{th} percentile value of the threshold velocity prediction of the model produced from propagating only the uncertainty of the user-defined input combination
v_{95}	95^{th} percentile value of the threshold velocity prediction of the model
$v_{95,all}$	95^{th} percentile value of the threshold velocity prediction of the model produced from propagating all uncertainties
$v_{95,input}$	95^{th} percentile value of the threshold velocity prediction of the model produced from propagating only the uncertainty of the user-defined input combination
x	Independent variable
$\underline{x}$	$d \times 1$ vector whose components contain the values of the independent variables
$x_{k,i}$	Value of the k^{th} independent variable at the i^{th} sample point
$\overline{x_{k,i}}$	Value of $x_{k,i}$ upon normalization
$x_{k,j}$	Value of the k^{th} independent variable at the j^{th} sample point
$\underline{x_0 x_0}$	$d \times 1$ vector whose components contain the average values of the independent variables of the user-defined input combination
x_1	First independent variable
x_2	Second independent variable
y	Dependent variable
z	Variable whose value has a stochastic nature
z_{calc}	Model's prediction for the user-defined input combination
$z_{calc,avg}$	Average value of the predictions at the user-defined input combination of all the models that were not discarded during the model screening process
z_{dev}	Absolute deviation of z_{calc} from $z_{calc,avg}$
$z_{exp,t}$	Observed value of the dependent variable at experimental datum point t
α	Level of significance for the Smirnov statistical test
$\hat{\beta}$	Generalized least squares estimate of the mean response of $f(\underline{x})$
ε_l	Percentage that represents the acceptable lower limit of the deviation of the model predictions from the experimental observations
ε_u	Percentage that represents the acceptable upper limit of the deviation of the model predictions from the experimental observations
η	p^{th} percentile of a set of number
θ_k	Correlation parameter of the k^{th} independent variable used in the Kriging function
$\hat{\sigma}$	Approximate standard deviation of $\hat{f}$
$\hat{\sigma}_5$	Approximate standard deviation of the 5^{th} percentile estimate from the Kriging method
$\hat{\sigma}_{95}$	Approximate standard deviation of the 95^{th} percentile estimate from the Kriging method

independent variables. Similar to the Latin hypercube sampling method, the stratified sampling approach partitions the d-dimensional space of the independent variables into "a grid of equal-probability bins" (Giunta et al., 2006). Then, in each bin, random values of the d independent variables are selected from the distributions of these variables, and $f(\underline{x})$ is evaluated using these randomly-selected values of the independent variables. Similar to the Monte Carlo simulation method, the approximate distribution of $f(\underline{x})$ becomes the output of both the Latin hypercube sampling method and the stratified sampling method. References to additional variants of the Latin hypercube sampling method are given in Pages 202–203 of Kleijnen (2015).