Journal of the Taiwan Institute of Chemical Engineers 000 (2016) 1–11

Contents lists available at ScienceDirect

Journal of the Taiwan Institute of Chemical Engineers

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

Meta-modelling in chemical process system engineering

Olumayowa T. Kajero[a], Tao Chen[a], Yuan Yao[b], Yao-Chen Chuang[b], David Shan Hill Wong[b,*]

[a] Department of Chemical and Process Engineering, University of Surrey, Guildford GU2 7XH, United Kingdom
[b] Department of Chemical Engineering, National Tsing Hua University, Hsinchu, Taiwan, ROC

ARTICLE INFO

Article history:
Received 29 April 2016
Revised 12 October 2016
Accepted 13 October 2016
Available online xxx

Keywords:
Metal-modelling
Chemical process system engineering
Surrogate modelling

ABSTRACT

Use of computational fluid dynamics to model chemical process system has received much attention in recent years. However, even with state-of-the-art computing, it is still difficult to perform simulations with many physical factors taken into account. Hence, translation of such models into computationally easy surrogate models is necessary for successful applications of such high fidelity models to process design optimization, scale-up and model predictive control. In this work, the methodology, statistical background and past applications to chemical processes of meta-model development were reviewed. The objective is to help interested researchers be familiarized with the work that has been carried out and problems that remain to be investigated.

1. Introduction

Systematic accumulation of knowledge and drive towards optimal design is the key to intelligent and rapid development of chemical processes and products. Traditionally this is done in two distinct approaches: the first-principle approach and data-driven or black-box approach.

First principles, or physical approach requires scientific understandings of the workings of process and integrates them into deterministic input-output simulation models. Simulation models can be developed at different physical scales, *e.g.,* steady state and dynamic process simulations model the operation performance of a plant, computer fluid dynamic (CFD) simulations model the momentum, material and heat transfer in an equipment, and molecular simulations model the relation between molecular structure as input and material property as output.

The black-box or data-driven approach rooted on the statistical theory of design of experiment (DOE) to direct experiments. DOE can be divided into two categories, the exploration of design space, *e.g.* screening designs, and finding the optimum, *e.g.,* response surface method. Traditional DOE theory were based on the assumptions that the input-output relations are relatively simple, consisting of linear, interaction, quadratic effects etc.

With the development of powerful computers, we can include more and more details into first-principle simulation models so as to improve the fidelity of the model. For example, we can model a continuous stirred tank reactor (CSTR) by assuming that it is a well-mix reactor. The mixing and heat transfer can be modelled using a CFD simulator and their effects can be integrated to the well-mix reactor through residence time distribution and heat transfer rate. Alternatively, one can take into account reactions and change in physical properties with change in composition and temperature in a CFD simulation. Even with a given simulation model, the fidelity can increase by including more mesh into the solver. As the fidelity of the physical model increases, the number of parameters needed to be estimated, *i.e.*, costs of calibrating the first-principle model increase. However, the computer time required for simulation also increase and the high cost the first-principle model becomes difficult to use.

This dilemma leads to a continuous effort to develop meta-models (models of model, or surrogate models) so that knowledge accumulated in such high fidelity models can be used efficiently in design, optimization and control. There have already been many useful reviews and books in the development and application of metamodel and design of computer experiments, a brief list is provided here [1–8]. However, to the best of our knowledge, there is no such review attempt specifically from the perspective of a chemical process system engineer. Specifically, we shall attempt to address the following important issues:

(1) model representation,
(2) model construction and evolution, and
(3) applications in chemical process system engineering.

* Corresponding author. Fax: +886 35715408.
E-mail address: dshwong@che.nthu.edu.tw (D.S.H. Wong).

http://dx.doi.org/10.1016/j.jtice.2016.10.042
1876-1070/

2. Models representation

2.1. Input–output relation

Consider an actual process with input $\boldsymbol{x} = [x_1 \cdots x_{K_x}]$ and output $\boldsymbol{y} = [y_1 \cdots y_{K_y}]$.

$$\boldsymbol{y} = \boldsymbol{\Phi}(\boldsymbol{x}). \tag{1}$$

Let $\boldsymbol{\Psi}$ be a high-fidelity physical model that predict an output $\boldsymbol{y}$ at a given set of input

$$\boldsymbol{y} = \boldsymbol{\Psi}(\boldsymbol{x}). \tag{2}$$

A meta-model is any model $\boldsymbol{\Omega}(\boldsymbol{x}, \boldsymbol{\beta})$ that approximate this high fidelity model

$$\boldsymbol{\Psi}(\boldsymbol{x}) = \boldsymbol{\Omega}\left(\boldsymbol{x}, \boldsymbol{\beta}\right) + \boldsymbol{\Gamma}(\boldsymbol{x}), \tag{3}$$

with $\boldsymbol{\Gamma}(\boldsymbol{x})$ being the error of the meta model at a specific input configuration. The high fidelity model $\boldsymbol{\Psi}(\boldsymbol{x})$ is usually a deterministic first principle model. Common types of meta-models $\boldsymbol{\Omega}(\boldsymbol{x}, \boldsymbol{\beta})$ include polynomial, kriging, radial basis function and artificial neural network, etc. The coefficients $\boldsymbol{\beta}$ are regression results that usually are not susceptible to any physical interpretation. We should bear in mind that the definition of input $\boldsymbol{x}$ and output $\boldsymbol{y}$ may be different in different applications. For example, let us consider the CFD model of a heat exchanger with fixed geometry. We can try to construct a meta-model that only apply to hot and cold streams with specific physical properties. The input parameters $\boldsymbol{x}$ are the inlet flowrates and temperatures of the inlet hot and cold streams. However, if we want to construct a more general meta-model that can be applied to different fluids, then physical properties such as viscosities and thermo-conductivities will also be classified as inputs. Similarly we can define various sets of $\boldsymbol{y}$. A simple version of $\boldsymbol{y}$ will be the average temperature of outlet streams. A detailed version of $\boldsymbol{y}$ can be the temperatures and the velocities at different points inside the heat exchanger.

2.2. Types of meta-models

2.2.1. Polynomial

Polynomial meta-model is perhaps the simplest form of model presentation used in meta-modelling research. Some studies in the literature on this method are: Simpson et al. [9], Palmer and Realff [10], Dutournie et al. [11], Chen et al. [12]. Simplicity implies ease in construction and application but also inability to describe complex input–output relationships.

2.2.2. Kriging, Gaussian process model, and radial basis function

The work of Krige [13] was widely used in geostatistics [14] and spatial statistics [15]. Kriging assumes some form of correlation between points in the multi-dimensional input space, with the correlation being used to predict response values between observed points. A brief introduction to the formulation and construction of Kriging model [16,17] are described as follows.

Let $\hat{\boldsymbol{X}} = [\hat{\boldsymbol{x}}_1, \ldots, \hat{\boldsymbol{x}}_N]^T$ be a set of training data points (sites) and $\hat{\boldsymbol{Y}} = [\hat{\boldsymbol{y}}_1, \ldots, \hat{\boldsymbol{y}}_N]^T$ be the corresponding response variables for development of a Kriging model. The prediction for a new data point, $\boldsymbol{x}$ is given by

$$\boldsymbol{y}(\boldsymbol{x}) = \boldsymbol{f}^T(\boldsymbol{x})\boldsymbol{\beta} + \boldsymbol{r}^T(\boldsymbol{x})\boldsymbol{\Sigma}\left(\hat{\boldsymbol{X}}\right)^{-1}\left(\hat{\boldsymbol{Y}} - \boldsymbol{F}\left(\hat{\boldsymbol{X}}\right)\boldsymbol{\beta}\right), \tag{4}$$

where $\boldsymbol{f}(\boldsymbol{x})$ contains a set of regression functions of the input variables, and $\boldsymbol{\beta}$ is the corresponding regression coefficients to be estimated. $\boldsymbol{F}(\hat{\boldsymbol{X}}) = [\boldsymbol{f}(\hat{\boldsymbol{x}}_1), \ldots, \boldsymbol{f}(\hat{\boldsymbol{x}}_N))]^T$ is a matrix containing the regression functions calculated for all the training data points. $\sum(\hat{\boldsymbol{X}})$ is the correlation matrix which is obtained from correlation functions evaluated at each pair of the training points:

$$\boldsymbol{\Sigma}\left(\hat{\boldsymbol{X}}\right) = \begin{bmatrix} \rho\left(\hat{\boldsymbol{x}}_1, \hat{\boldsymbol{x}}_1\right) & \cdots & \rho\left(\hat{\boldsymbol{x}}_1, \hat{\boldsymbol{x}}_N\right) \\ \vdots & \ddots & \vdots \\ \rho\left(\hat{\boldsymbol{x}}_N, \hat{\boldsymbol{x}}_1\right) & \cdots & \rho\left(\hat{\boldsymbol{x}}_N, \hat{\boldsymbol{x}}_N\right) \end{bmatrix}. \tag{5}$$

A widely used correlation function is the Gaussian function

$$\rho(\boldsymbol{x}, \boldsymbol{x}') = exp\left[-(\boldsymbol{x} - \boldsymbol{x}')^T diag[\theta_1 \ldots \theta_{K_x}](\boldsymbol{x} - \boldsymbol{x}')\right], \tag{6}$$

while

$$\boldsymbol{r}(\boldsymbol{x}) = \left[\rho\left(\boldsymbol{x}, \hat{\boldsymbol{x}}_N\right), \ldots \rho\left(\boldsymbol{x}, \hat{\boldsymbol{x}}_N\right)\right] \tag{7}$$

is vector of the correlation between a general point in the input space and the training sites. The parameters of the Kriging model are the parameters in the correlation function $\boldsymbol{\theta} = [\theta_1 \cdots \theta_{K_x}]$ and the regression coefficient $\boldsymbol{\beta}$. They can be estimated by the following iterative procedure. First assume a value for $\boldsymbol{\theta}$, estimate the regression coefficient $\boldsymbol{\beta}$ by

$$\tilde{\boldsymbol{\beta}} = \left(\boldsymbol{F}\left(\hat{\boldsymbol{X}}\right)^T \boldsymbol{\Sigma}\left(\hat{\boldsymbol{X}}\right)^{-1} \boldsymbol{F}\left(\hat{\boldsymbol{X}}\right)\right)^{-1} \boldsymbol{F}\left(\hat{\boldsymbol{X}}\right)^T \boldsymbol{\Sigma}\left(\hat{\boldsymbol{X}}\right)^{-1} \boldsymbol{Y}. \tag{8}$$

The process variance can be calculated as

$$\sigma_p^2 = \frac{1}{N}\left(\boldsymbol{Y} - \boldsymbol{F}\left(\hat{\boldsymbol{X}}\right)\right)^T \boldsymbol{\Sigma}\left(\hat{\boldsymbol{X}}\right)^{-1}\left(\boldsymbol{Y} - \boldsymbol{F}\left(\hat{\boldsymbol{X}}\right)\right). \tag{9}$$

A new set of correlation parameters $\boldsymbol{\theta}$ can be estimated by

$$\tilde{\boldsymbol{\theta}} = \min_{\boldsymbol{\theta}}\left(\left|\boldsymbol{\Sigma}\left(\hat{\boldsymbol{X}}\right)\right|^{1/N} \sigma_p^2\right)$$

The above procedure is repeated until values of $\tilde{\boldsymbol{\theta}}$ and $\tilde{\boldsymbol{\beta}}$ converge.

Kriging is also termed Gaussian process in the literature with slightly different formulation [18–20].

Applications of Kriging, Gaussian process models in meta-model development have been extensively investigated by many authors. A chronological, but far from exhaustive, list is provided here [18,21–47]. A review was provided by Kleijin [4] in 2009.

2.2.3. Support vector machine

Support vector machine (SVM) [48] was originally developed as a supervised learning classifier so that data in a (high dimensional) input space can be classified into groups according to their locations. The SVM can also be formulated into an input–output known as support vector regression (SVR) [49]. For a specific dimension in the output space, given a training data set $\hat{\boldsymbol{X}} = [\hat{\boldsymbol{x}}_1, \ldots \hat{\boldsymbol{x}}_N]^T$ and $[\hat{y}_1, \ldots \hat{y}_N]^T$; a nonlinear SVR can be expressed as

$$y(\boldsymbol{x}) = \alpha_o + \sum_{i=1}^{N}\left(\alpha_i - \alpha_i^*\right)K\left(x, \hat{\boldsymbol{x}}_i\right), \tag{10}$$

where $\boldsymbol{\alpha}_i, \boldsymbol{\alpha}_i^*$ can be obtained by solving optimization problem

$$\max\left\{\begin{array}{l} -\frac{1}{2}\sum_{n=1}^{N}\sum_{n'=1}^{N}\left[(\alpha_n - \alpha_n^*)\left(\alpha_{n'} - \alpha_{n'}^*\right)K\left(\hat{\boldsymbol{x}}_n, \hat{\boldsymbol{x}}_{n'}\right)\right] \\ \quad + \varepsilon\sum_{n=1}^{N}(\alpha_n + \alpha_n^*) + \sum_{n=1}^{m}\hat{y}_n(\alpha_n - \alpha_n^*) \end{array}\right\}, \quad \text{subject to:} \begin{cases} \sum_{n=1}^{N}(\alpha_n - \alpha_n^*) = 0, \\ 0 < \alpha_n, \alpha_n^* < C. \end{cases} \tag{11}$$

The off-set parameter α_o, tolerance parameter ε and constraint parameter C are parameters to be chosen in training [50]. The solution of the above problem can be determined using a least square approach that uniquely determined by the input–output training data; the resulting model is known as known as least square support vector machine (LS–SVM) [51]. Several reports of using SVR in meta-modelling are given here [52–56].

Please cite this article as: O.T. Kajero et al., Meta-modelling in chemical process system engineering, Journal of the Taiwan Institute of Chemical Engineers (2016), http://dx.doi.org/10.1016/j.jtice.2016.10.042