Simulation Modelling Practice and Theory 49 (2014) 245–257

Contents lists available at ScienceDirect

Simulation Modelling Practice and Theory

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

Comparative study of metamodelling techniques in building energy simulation: Guidelines for practitioners

Liesje Van Gelder [a,*], Payel Das [b], Hans Janssen [a], Staf Roels [a]

[a] Building Physics Section, Department of Civil Engineering, KU Leuven, Kasteelpark Arenberg 40 Bus 2447, 3001 Heverlee, Belgium
[b] The Bartlett School of Graduate Studies, University College London, Central House, 14 Upper Woburn Place, London WC1H 0NN, UK

ARTICLE INFO

Article history:
Received 30 July 2014
Received in revised form 1 September 2014
Accepted 6 October 2014

Keywords:
Metamodelling strategy
Polynomial regression
Multivariate adaptive regression splines
Kriging
Radial basis function network
Neural network

ABSTRACT

Computer simulation of real system behaviour is increasingly used in research and development. As simulation models become more reliable, they also often become more complex to capture the progressive complexity of the real system. Calculation time can be a limiting factor for using simulation models in optimisation studies, for example, which generally require multiple simulations. Instead of using these time-consuming simulation models, the use of metamodels can be considered. A metamodel approximates the original simulation model with high confidence via a simplified mathematical model. A series of simulations then only takes a fraction of the original simulation time, hence allowing significant computational savings.

In this paper, a strategy that is both reliable and time-efficient is provided in order to guide users in their metamodelling problems. Furthermore, polynomial regression (PR), multivariate adaptive regression splines (MARS), kriging (KR), radial basis function networks (RBF), and neural networks (NN) are compared on a building energy simulation problem. We find that for the outputs of this example and based on Root Mean Squared Error (RMSE), coefficient of determination (R^2), and Maximal Absolute Error (MAE), KR and NN are the overall best techniques. Although MARS perform slightly worse than KR and NN, it is preferred because of its simplicity. For different applications, other techniques might be optimal.

1. Introduction

Computational models are commonly used in various fields such as engineering and economics to simulate system behaviour. Due to the increasing reliability of these models, simulation is usually more useful and less expensive than real-time experiments, which may not be feasible. Depending on the model complexity, simulations can take from only a second to several days, weeks, or even months. Despite the huge potential of these simulations, excessive calculation time might be a limiting factor, especially in optimisation problems. Moreover, lowering calculation time by reducing the model complexity is not an option as this might lead to less reliable results. To counter this computational barrier, metamodels – also known as surrogate models – have been introduced to replace potentially time-consuming models [1,2]. Metamodels aim at mimicking the original complex simulation model via a simplified mathematical model, statistically determined based on original model realisations. The simulation then only takes a fraction of the original simulation time, allowing

* Corresponding author. Tel.: +32 16 321344; fax: +32 16 321980.
E-mail address: liesje.vangelder@bwk.kuleuven.be (L. Van Gelder).

http://dx.doi.org/10.1016/j.simpat.2014.10.004
1569-190X/

significant computational savings, without compromising the reliability. Both stationary and dynamic metamodels can be constructed, however the focus of this paper is limited to the former.

Several metamodelling techniques were developed to address this increasing interest, and different applications appeared in the literature during the last few decades. Based on both mathematical and engineering test problems with a varying number of inputs, number of samples, degree of non-linearity, noisy behaviour, and applied fitting algorithm, several metamodelling techniques were compared [1,3–8]. Amongst others, mainly polynomial regression, multivariate adaptive splines, kriging, radial basis function networks, and neural networks were explored. In the above-mentioned literature, polynomial regression is considered as the worst performing technique, while neural networks are advised for problems with many inputs, and kriging is recommended for highly non-linear problems. When the considered problem is noisy however, kriging typically performs the worst. Depending on the number of samples, the degree of non-linearity and noisy behaviour of the particular problem and also the employed algorithm, yet other techniques may provide better models.

Instead of focussing on the theory of metamodelling and developing new techniques and algorithms, this paper aims at guiding users in metamodelling more complex engineering problems based on readily available algorithms. This was already intended by Simpson et al. [1] and Wang and Shan [2] for design optimisation, however no concrete guidelines were presented. This paper therefore wants to take this one step further by proposing a practical metamodelling strategy. To illustrate this, a building energy simulation example is chosen as metamodelling was only very recently introduced in this field to overcome the time barrier in optimisation [9,10]. The time-efficiency of metamodels enable them to also be used in robust optimisation such as in [11,12], which generally requires significantly more simulations. Although the metamodels themselves are computationally inexpensive to run, they are not always constructed in the most time-efficient way. Eisenhower et al. [9] and Ferreira et al. [10], for example, both used about 5000 samples to fit their metamodels. As this is not feasible for very computationally-expensive simulation models, this paper proposes a metamodelling strategy dealing with both time-efficiency and reliability: a well-performing metamodel trained on as few samples as possible will be preferred.

Section 2 first describes the theory concerning fitting and validating metamodels. Five groups of metamodelling techniques are described, with a focus on the selection of algorithms: polynomial regression, multivariate adaptive regression splines, kriging, radial basis function networks, and neural networks. Section 3 then proposes a metamodelling strategy, using as little data as possible without compromising the reliability of the metamodel. The building energy simulation application is described in Section 4, and the results for this are presented in Section 5.

2. Metamodelling theory

Wang and Shan [2] and Kleijnen and Sargent [13] emphasise the importance of both fitting and validating metamodels in view of model reliability. This is indeed one of the major concerns as metamodelling aims to replace a model without becoming unreliable. Therefore, this section first describes the fitting of a metamodel followed by the validation. Furthermore, a selection of five metamodelling techniques is described as these are thought to be most useful based on literature: polynomial regression (PR), multivariate adaptive regression splines (MARS), kriging (KR), radial basis function networks (RBF), and sigmoidal neural networks (NN).

2.1. Fitting

A metamodel is a mathematical function for which the coefficients are determined based on a limited number of input/output combinations. To create n samples of the p inputs, a Monte-Carlo based sampling technique is used in this paper. This is described briefly in Section 3.1 and in more detail in [14]. The original simulation model is run for these samples to obtain the n corresponding values for each of the outputs. These data will be referred to as *training data*. In general, these training data are standardised (zero mean, unit variance) to overcome influences from parameter units. Each output is then modelled separately or together with a metamodelling technique. With these techniques, the training process results in an independent model to estimate new input/output combinations within the range of the sampled combinations. The more training data are used, the better the metamodel can perform in general. It is however possible that the training data are perfectly fit, while unseen data are not approximated well at all. This phenomenon is called *overfitting* and can be avoided by employing generalisation methods that reduce the complexity of the model by

- *regularisation:* limiting the Euclidean norm of the coefficients vector, in order to avoid unnecessarily large coefficients, or
- *pruning:* reducing the number of coefficients before or after fitting them, in order to avoid too many coefficients.

Amongst the variety of readily available algorithms for each of the metamodelling groups described in Section 2.3, algorithms employing these generalisation methods are selected. In this paper, all algorithms are provided by MATLAB toolboxes and are referred to when the techniques are described.

Each of the selected algorithms contain several settings that have to be defined by the user. Different settings might result in differing metamodels, of which only the best is retained, selected via a model selection criterion. Such a criterion indicates the trade off between the goodness of fit and model complexity in order to avoid more coefficients than needed. The Akaike information criterion (AIC) [15] is commonly used for that purpose and is given by