bradscholars

Performance evaluation of metamodelling methods for engineering problems: towards a practitioner guide

Item Type	Article
Authors	Kianifar, Mohammed R.;Campean, Felician
Citation	Kianifar MR and Campean F (2020) Performance evaluation of metamodelling methods for engineering problems: towards a practitioner guide. Structural and Multidisciplinary Optimization. 61: 159-186.
DOI	https://doi.org/10.1007/s00158-019-02352-1
Publisher	Springer
Rights	© The Author(s) 2019. Open Access. This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.
Download date	2025-05-28 03:29:34
Link to Item	http://hdl.handle.net/10454/17192

RESEARCH PAPER

Performance evaluation of metamodelling methods for engineering problems: towards a practitioner guide

Mohammed Reza Kianifar[1] · **Felician Campean**[1]

Abstract

Metamodelling or surrogate modelling techniques are frequently used across the engineering disciplines in conjunction with expensive simulation models or physical experiments. With the proliferation of metamodeling techniques developed to provide enhanced performance for specific problems, and the wide availability of a diverse choice of tools in engineering software packages, the engineering task of selecting a robust metamodeling technique for practical problems is still a challenge. This research introduces a framework for describing the typology of engineering problems, in terms of dimensionality and complexity, and the modelling conditions, reflecting the noisiness of the signals and the affordability of sample sizes, and on this basis presents a systematic evaluation of the performance of frequently used metamodeling techniques. A set of metamodeling techniques, selected based on their reported use for engineering problems (i.e. Polynomial, Radial Basis Function, and Kriging), were systematically evaluated in terms of accuracy and robustness against a carefully assembled set of 18 test functions covering different types of problems, sampling conditions and noise conditions. A set of four real-world engineering case studies covering both computer simulation and physical experiments were also analysed as validation tests for the proposed guidelines. The main conclusions drawn from the study are that Kriging model with Matérn 5/2 correlation function performs consistently well across different problem types with smooth (i.e. not noisy) data, while Kriging model with Matérn 3/2 correlation function provides robust performance under noisy conditions, except for the very high noise conditions, where the Kriging model with nugget appears to provide better models. These results provide engineering practitioners with a guide for the choice of a metamodeling technique for problem types and modelling conditions represented in the study, whereas the evaluation framework and benchmarking problems set will be useful for researchers conducting similar studies.

Keywords Metamodelling · Kriging · Radial basis functions · Polynomials · Correlation function · Kernel functions · Response surfaces

1 Introduction

1.1 Background

High fidelity simulation modelling of complex engineered systems, based on tools such as finite element analysis (FEA) and computational fluid dynamics (CFD), plays an increasingly important role as they enable detailed analysis and optimisation of the design at an early stage, enabling

significant cost and time compression in product development. However, complex engineering simulations are computationally expensive, which often precludes an efficient design exploration through parametric studies and optimisation, or the integration of multiple simulations in multi-physics models to study complex multidisciplinary systems. Metamodelling or surrogate modelling (Box and Draper 1987), underpinned by response surface modelling techniques originally introduced to develop prediction models for expensive physical experimental responses (Simpson et al. 2001b), is increasingly and extensively employed to replace complex simulation-based models across the engineering disciplines, (e.g. Fang et al. 2005; Zhu et al. 2009). Metamodelling techniques do not only provide relatively cheap and accurate response models to replace the expensive analysis tools, but also provide a filtering method to handle noisy data (Box and Draper 1987; Forrester et al. 2008).

✉ Felician Campean
   F.Campean@bradford.ac.uk

1   Faculty of Engineering & Informatics, University of Bradford, Bradford, UK

Metamodelling techniques are commonly classified into parametric and non-parametric models (Rango et al. 2013). Parametric models (such as polynomials (Khan 2011) are explicitly dependent on the underlying model structure, whereas non-parametric methods (such as Radial Basis Function (RBF) (Khan 2011), Neural network (Hagan and Demuth 1999) and Kriging models (Sacks et al. 1989) do not require explicit model assumptions and use the experimental measurements to define the underlying relationship among the parameters.

Despite the availability of different software packages that enable the application of various metamodelling methods, the engineering task of selecting the most robust modelling technique, given the practical engineering problem, is still a challenge for the practitioner, especially when the system behaviour is unknown (Didcock et al. 2014). Several studies have been reported over the years evaluating the performance of metamodelling techniques; such studies have focussed either on identifying the most suitable type of model for a particular engineering application or modelling conditions (e.g. relating to the affordable sample size or uncertainty in the underpinning experiment) or, more general, the study of performance in relation to the type or characteristics of the problem (such as scale and nonlinearity). However, no study has systematically considered the complete typology of the modelling problems to drive the evaluation study of metamodelling techniques, and thus, a generic guideline for the engineering practitioners is still not available.

The research presented in this paper aims to address this issue, with a study of a selection of the metamodelling techniques frequently used in practice, systematically evaluated in a comprehensive experiment designed to replicate the characteristics of the metamodelling problems commonly encountered in product design and development of engineered systems in industries such as automotive and aerospace. Metamodelling problems associated with both physical experiments (such as steady-state engine calibration tests) and computer-based experiments (e.g. based on FEA, CFD, or multi-physics simulation experiments) are within the scope for this study. An analysis of literature relating to the evaluation of metamodelling techniques will be considered first in order to facilitate the identification of the key characteristics of engineering metamodelling problems, supporting the development of a coherent methodology for the study.

1.2 Review of related work

The review of the studies reported in literature that have evaluated the performance of metamodelling techniques for engineering applications is focussed on two aspects:

(i) The metamodelling methods considered in each study

(ii) The characteristics of the engineering problems considered in the study

The review and associated analysis are organised in chronological order to provide a view on both the use of metamodelling techniques in engineering as well as the expanding nature of problems tackled and issues.

Simpson et al. (1998) have analysed the performance of the Kriging method against Polynomials for a three-dimensional multidisciplinary optimisation problem of an aerospike nozzle design based on FE and CFD, arguing that Kriging based on a space-filling design can provide competitive modelling performance for the engineering design problem considered. Giunta et al. (1998) compared Polynomials with Kriging for three problems with different numbers of parameters (1, 5 and 10 parameters), based on mathematical benchmark problems. Over the specific set of problems considered, they observed that a quadratic polynomial consistently delivers highly accurate prediction models. Yang et al. (2000) compared four metamodelling techniques (i.e. Moving Least Square, Neural Network (NN), Stepwise Regression, and Multivariate Adaptive Regression Splines (MARS)) to predict a low-dimensional safety function in an automotive crash model. Varadarajan et al. (2000) compared the performance of Neural Networks and Polynomial methods for an engine combustion modelling application. The research by Jin et al. (2001) reports the first systematic comparative approach to study the performance of four metamodelling methods (i.e. Radial Basis Function (RBF), Kriging, Polynomial and Regression Splines), over different sampling sizes (i.e. scarce, small and large), and for different problem types in terms of scale and nonlinearity. Thirteen case studies were employed to develop a standard procedure to compare the metamodelling methods, including an extra noisy test function to briefly investigate how the performance of metamodels is affected by the noisy conditions, concluding that RBF with Gaussian kernel function performs the best across the case studies considered. However, the test functions considered are unduly biased towards only two problem dimensionality cases: small (with less than 4 variables) and large (with 10 and more parameters), not covering the medium size test functions. In a different study, Jin et al. (2003) compared the performance of Kriging, RBF and Polynomials for modelling a structural engineering problem, considering uncertainty in the values of parameters, over two sample sizes (small and large). They observed that the Kriging method with a Gaussian correlation matrix outperforms the other techniques and that the accuracy of metamodels was not affected by the sample size in the range considered (in particular for the RBF method). Also, the second-order polynomial was observed to be incapable of capturing the nonlinearity of the performance variations for the case study problem. Seabrook et al. (2003) compared NN, Kriging and RBFs for engine calibration experiments where the results are

affected by experimental noise, concluding that Kriging offers the most robust modelling technique. Forsberg and Nilsson (2004) compared the performance of Kriging and Polynomial methods in modelling crashworthiness of a vehicle structure, for which Kriging was found to be more accurate. Fang et al. (2005) also compared RBF (using five different kernel functions) and polynomial techniques for a crashworthiness application. They reported that both modelling methods are accurate for the case study (especially the RBF with Multiquadric kernel function), while RBF outperforms Polynomial for smaller sampling sizes. In the comparative study presented by Mullur and Messac (2005), four metamodelling techniques were studied (Polynomials, RBF, Kriging and a RBF-based method named extended RBF) over 8 case studies selected from the ones used in the Jin et al. study. The study focused on the effects of different problem scales (small and large) and sample sizes (small, medium and large) using 2 sampling strategies (Latin Hypercube and random sampling); the extended RBF was found to outperform the other methods. Chen et al. (2006) studied the performance of RBF, Kriging, Polynomial, Neural Network and MARS methods over three case studies (two 2-dimensional test functions and one 10-dimensional engineering problem). For the limited number of case studies considered, they also investigated the effects of different sampling strategies. Ben-Ari et al. (2007) presented a comparative study of the performance of three metamodelling strategies (Splines, Kriging and projection pursuit regression) for three simulation-based case studies, concluding that while Kriging provides the best modelling technique, it is also the most computationally expensive. Kim et al. (2009) studied the performance of four metamodelling techniques (RBF, Kriging, Moving Least Square (MLS) and Support Vector Regression (SVR)) over six low-dimensional case studies, using two sample sizes (small and large). They reported that MLS and Kriging provided superior results compared to the other methods; however, these methods were found ineffective when the sample size is too small. In the study reported by Zhu et al. (2009), RBF, Kriging, SVR and NN methods were used to model a real-world engineering case study in the automotive engineering field, for which SVR outperformed the other techniques. Paiva et al. (2009) compared the performance of Kriging, quadratic polynomial and Neural Network over three aircraft applications. They observed that both Kriging and Neural Network delivered high fidelity models for large-scale problems. Zhao and Xue (2010) applied four metamodelling strategies (RBF, Kriging, linear polynomials and Bayesian neural network) to model six test functions (three small-scale and three large-scale problems), selected from Jin et al. (2001). This research also considered the effect of different sample sizes and the influence of different levels of noise conditions. In the study reported by Campean et al. (2010), Polynomial, Kriging and RBF (with different kernel functions including Thinplate,

Multiquadric and Linear) metamodels were evaluated for engine fuel consumption and gaseous emissions modelling, based on noisy engine test data measurement (with different level of noise associated with measurement of different responses), concluding that Kriging outperforms the other modelling techniques. Li et al. (2010) have reported an evaluation of the performance of SVR against NN, RBF, Kriging and MARS for a range of benchmark engineering problems with between 2 and 8 variables, and also considering the effect homogeneous and heterogeneous stochastic error, induced via simulation. They have concluded that SVR performs best, closely matched by Kriging, while RBFs provide computational advantage for larger samples. They have also validated the performance of SVR on several simulation-based process optimisation case studies. Similarly, Wang et al. (2011) have compared the performance of SVR against polynomial regression, Kriging, RBF, NN and MARS for several nonlinear benchmark problems, concluding that SVR provides the most robust modelling technique. However, both studies relied on large sample sizes for the model fitting. Van Gelder et al. (2014) carried out a comparative study of metamodelling techniques (Polynomial, MARS, Kriging, RBF and NN) for large-scale building simulation, considering probabilistic input parameters, and the effect of the training sample size on the reliability of the metamodelling strategy. Their aim was to derive a guideline for practitioners and have concluded that NN and Kriging performed best, with Kriging requiring lower training set. They have however pointed out that Kriging models are harder to interpret. Liu et al. (2016) considered the impact of problem dimensionality (nine engineering benchmark problems with between 4 and 51 variables) and the space filling sampling strategy and size (between $10\times$ and $40\times$) on the performance of metamodelling strategies (Polynomial, Kriging, RBF and RBF-HDMR—i.e. high dimensional model representation), concluding that RBF performed best. Kroetz et al. (2017) contrasted performance of Kriging models against Neural Networks and Polynomial Chaos Expansion (PCE) metamodelling techniques for a range of structural reliability problems and concluded that Kriging and NN models converge efficiently compared to PCE. Chen et al. (2019) studied the problem of efficiency of metamodelling in high dimensionality problems considering a set of nine test functions, the influence of sample size and presence of random noise, for a selection of metamodelling techniques including Kriging, RBF, SVR and a range of high dimensional model representation derivatives of these methods. They have found that Kriging (with Gaussian kernel) does not provide satisfactory global modelling results even with large sample sets. Østergård et al. (2018) evaluated the performance of six metamodelling techniques (Linear Regression, Random Forest (RF), SVR, MARS, Kriging, and NN) in relation to building performance simulation, focussing both on performance (accuracy, efficiency and robustness) and user-

focussed metrics such as ease of use and interpretability. Their review considered 13 benchmark problems, with varying dimensionality and complexity (but not including "noisy" problems as these were not relevant to the application domain), and including both discrete and continuous variable problems. They have attempted to provide some guidance for metamodelling choice in relation to the level of expertise of the analyst (expert/ non-expert / automatic metamodelling) and the time/cost afforded for the experiment (large / limited / minimal). They have concluded that Kriging offers the most accurate metamodelling approach, although not the most computationally efficient, and providing models that are not easy to interpret by the user in the same way in which Polynomial regression models can be interpreted.

This review of related literature studies, summarised in Table 1, shows that the effectiveness of metamodelling techniques has been intensely studied from both theoretical (i.e. through synthetic test cases) and practical (i.e. application to specific engineering problems) perspectives. The review reveals the complexity of the research and application landscape, illustrated by the complex evolution and development of metamodelling techniques, interconnected with the evolution of the types of problems tackled. The chronological analysis of comparative studies of metamodeling techniques summarised in Table 1 illustrates well this trend, showing the evolution from polynomial based response surface methods introduced by Box and Draper (1987), to the prevalence of Kriging and RBF, and more recently the increasing use of techniques like SVR (Wang et al. 2009; Pan et al. 2010; Wang et al. 2011). Also, the emerging interest in tackling high-dimensional engineering problems led to increased use of HDMR techniques (Huang et al. 2015; Dey et al. 2015; Liu et al. 2016; Chen et al. 2018).

In relation to computer-based experiments, the improvements in computation power and speed have enabled metamodelling studies for significantly increased dimensionality of problems and based on larger samples, which in turn supported the development and application of an increasingly diverse range of metamodeling techniques. Several studies compared the effect of different mathematical options for a particular modelling technique (such as correlation function options for Kriging (Toal and Keane 2013; Kleijnen and van Beers 2005; Ulaganathan et al. 2014, 2015) or kernel function options for RBF (Campean et al. 2010; Fang et al. 2005) over different applications. The impact of the hyperparameters associated with the modelling techniques and model selection criteria has been also increasingly considered in comparative studies—see for example Østergård et al. (2018). The continued research interest for comparative metamodelling studies over time reflects the fact that this problem is still of significant real-world impact. The fact that different practical test problems and applications areas appear to point to different

optimal metamodelling choice conflicts the practical engineering interest, which usually revolves around prescriptive guidelines for a robust choice of a metamodelling technique for a specific type of engineering problem. Efforts to develop such guidelines are well illustrated by studies such as Van Gelder et al. (2014) and Østergård et al. (2018) for high dimensional problems of building performance simulations, or Blondet et al. (2019) for mechanical engineering simulations, underpinned by a knowledge-based system. Examples of generic recommendations for the choice of metamodelling techniques for physical experiments associated with engine tests are illustrated by the work of Seabrook et al. (2003) and Berger (2012).

Widely available commercial software, such as Matlab (e.g. the Model-Based Calibration toolbox), provide easy access for engineers to a very large choice of metamodeling techniques, including model fitting options (hyperparameters and model selection settings). However, this raises modelling productivity and effectiveness issues for practical application, for reasons discussed by Østergård et al. (2018), which justifies the need for further studies to support practical metamodeling choices of methods.

1.3 Research objectives and contribution

Similar to Van Gelder (2014), instead of focussing on the theory of metamodelling and developing new techniques and algorithms, the aim of this paper is to guide users in choosing the appropriate metamodelling technique (implemented in readily available algorithms) for their specific engineering problem. A deeper analysis of the studies in the literature review (summarised in Table 1) from the point of view of the characteristics of the metamodeling problems considered in each study, suggests the following classification:

1. Characteristics of the problem type:

   (a) *Dimensionality* or scale of the problem—typically indicated by the number of independent variables involved
   (b) Complexity of the problem defined in relation to the relationship among factors for approximation and usually quantified in terms of *nonlinearity*, interaction between variables and importance of terms (Shan and Wang 2010)

2. Modelling choices and conditions:

   (a) The s*ample size* practically afforded for the experiment
   (b) Uncertainty or *noise*, relating to the experimental measurement uncertainty/noisiness

Table 1 Summary of literature cases: metamodelling techniques and problem characteristics considered

	Metamodelling technique									Problem characteristic				Test problem		Engineering case study
	Polynomial	Kriging	RBF	MLS	NN	MARS	Splines	SVR	Other (e.g. HDMR)	Dimensionality	Nonlinearity	Sample size	Noise	Benchmark functions	Engineering problem	
Simpson et al. (1998)	x	x								x					x	Aerospike nozzle design
Giunta and Watson (1998)	x	x								x				x		
Yang et al. (2000)	x		x	x	x	x				x					x	Crashworthiness simulation
Varadarajan et al. (200)	x				x										x	Engine combustion modelling
Jin et al. (2001)	x	x	x				x			x	x	x	x	x		
Jin et al. (2003)	x	x	x									x	x		x	Structural design
Seabrook et al. (2003)		x	x		x								x		x	Engine mapping experiments
Forsberg and Nilsson (2004)	x	x													x	Crash simulation
Fang et al. (2005)	x		x									x			x	
Mullur and Messac (2006)	x	x	x		x	x				x		x			x	Crash simulation
Chen et al. (2006)	x	x	x							x		x		x	x	Wastewater treatment
Ben-Ari and Steinberg (2007)		x					x		x	x				x	x	Piston dynamics, electric circuit
Kim et al. (2009)		x	x	x				x				x		x		
Zhu et al. (2009)		x	x		x			x						x	x	Automotive - body structure
Paiva et al. (2009)		x	x		x					x					x	Aircraft wing design
Zhao and Xue (2010)		x	x		x					x			x			
Campean et al. (2010a, b)		x	x							x		x	x	x	x	Engine modelling experiments
Li et al. (2010)		x	x		x	x		x		x				x	x	Process simulation
Wang et al. (2011)	x	x	x			x		x		x	x		x	x	x	Crashworthiness optimisation
Van Gelder et al. (2014)	x	x	x		x					x		x			x	Building simulation
Liu et al. (2016)	x	x				x		x	x	x		x			x	9 engineering problems
Kroetz et al. (2017)		x			x				x	x	x	x	x	x	x	Structural reliability
Chen et al. (2018, 2019)		x	x					x	x	x	x	x	x	x		
Ostergard et al. (2018)		x	x		x				x	x		x		x	x	

The objective of the research presented in this paper was to carry out a systematic and comprehensive study of metamodelling techniques with the aim of developing a framework for the robust choice of method and model based on both the type of the engineering problem (scale and nonlinearity) and the modelling conditions or constraints (uncertainty and size of the metamodelling experiment). The scope of the study was set on three of the most popular metamodelling techniques in the engineering field, namely Polynomial, RBF and Kriging. This choice is justified both by the analysis of previous studies considered (these three methods are seen as the most popular in the analysis in Table 1), availability in common engineering software packages, and also the authors' extensive experience and observation of practice in automotive product development. The key points observed in relation to practical engineering metamodelling problems, in particular early in product development, relate to (i) affordability of larger sample sizes for both computer-based experiments (including crash simulation, CFD and structural simulations, as well as increasingly used multi-physics simulations) and physical experiments (typified by the expensive engine calibration experiments) and (ii) uncertainty about the variables and parameters involved in the modelling experiments. While Neural Networks are also commonly used metamodelling techniques, most of the studies considered in Table 1 point to the fact that limitation with sample sizes severely affect the performance of NNs for engineering modelling problems in product development. For this reason, the NNs were not included in the study. While most of the recent research, in particular in relation to large-scale nonlinear metamodeling problems, points out the limitations of Polynomial regression models, from a practical point of view, they are still important because they have the advantage that as models are easier to interpret by the engineering analysts (van Gelder et al. 2014). They are still widely seen in practice as the first preference of engineers, and for this reason, they were included in this study.

The approach taken for this study, aiming to reflect the dual theoretical—practical nature of the interest for the metamodelling performance evaluation, includes two parts:

1. An extensive theoretical (i.e. based on known mathematical functions) study of the performance of metamodelling techniques across a range of problems of different scale and nonlinearity, and with consideration of the size of experimental training set for model fitting and sensitivity to noise in the response variables. In order to ensure the validity of the study, a set of 18 benchmark problems extracted from previous studies was assembled to provide a representative and balanced cover across the problem parameters. Based

on the findings from this study, a guideline for preferred metamodelling technique based on problem criteria is proposed.

2. A set of validation studies based on real-world engineering problems with experimental data from both physical and computer-based simulation experiments/tests. These studies are used as empirical validation for the proposed guideline.

Thus, the paper aims to make a contribution to the research of metamodelling techniques for engineering problems by:

(i) Presenting a comprehensive and complete evaluation study of common metamodeling techniques in relation to the characteristics of engineering problem types (dimensionality and complexity) and modelling conditions (experimental sample size and noisiness);

(ii) Defining a set of benchmarking problems with systematic coverage of criteria for metamodelling problems; and

(iii) Introducing a framework for guiding selection of metamodelling technique based on the problem classification criteria, with relevance to the typology of problems commonly seen in product design and development.

The impact of the paper to the research community is defined by the first two points, while the latter point should provide the basis for the impact of the study to practitioners in the engineering design, modelling and simulation community.

The organisation of the paper is as follows: first, the mathematical properties of the three metamodelling techniques considered in this study are outlined, followed by a detailed description of the research methodology, including the benchmark test functions and the procedure to evaluate the properties of the constructed models. The results from different modelling techniques are presented next and analysed for different problem types and modelling conditions to provide a sound guideline to choose the best metamodelling technique for different classes of problems. The proposed guideline is validated on four real-life engineering problems. The paper ends with a discussion of the results and directions for further work.

2 Metamodelling techniques

Three types of metamodelling techniques are included in this study: Polynomial, Radial Basis Function and Kriging. The principal features of these techniques and the corresponding mathematical representations are described in the following sections.

2.1 Polynomial

Polynomials have been frequently used by engineers to predict the behaviour of complex engineering systems (Lach et al. 2007; Schlober et al. 2007). The general form of a polynomial model can be expressed as (Morris and Mitchell 1995; Myers et al. 1989):

$$y(x) = f^T.a = \sum_{p=1}^{P} a_p.f_p(x) \tag{1}$$

where $y(x)$ predicts the response value at design parameters x by linear combination of a set of base functions f. In this equation, P indicates the number of base functions corresponding to a polynomial order, and a defines the regression coefficients for each of the base functions. These coefficients are usually retrieved through a linear regression based on least squares estimation (Chen et al. 2006). The main reasons for the popularity of polynomials compared to the non-parametric metamodelling techniques are:

- Polynomials have a simple structure, which makes them easy to understand and manipulate (Hartmann et al. 2013).
- Polynomials require low computational effort due to linearity in the unknown parameters (Rango et al. 2013).
- Low-order polynomials show less possibility of overfitting (i.e. better smoothing capability (Jin et al. 2001)), which is particularly important for modelling noisy measurements (Khan 2011).

Regardless of these advantages, there are some important limitations in using polynomial metamodels for complex responses (i.e. highly nonlinear or large-scale problems) since the number of model terms increases significantly, increasing the number of experimental measurements required to calculate the coefficients (Khan, 2011). In this research, the second- and third-order polynomials were used, while the maximum order of parameter interactions was limited to 2.

2.2 Radial Basis Function (RBF)

Radial Basis Function is a non-parametric metamodelling technique which was introduced as a variant of Neural Network models in the late 1980s (Chen et al. 1996). This interpolation-based modelling method uses linear combinations of radially symmetric functions (called kernel functions) (Jin et al. 2001). A general form of a RBF model can be described by (Khan 2011):

$$y(x^i) = \sum_{j=1}^{n} w_i \phi(\|x^i - x^j\|) = t^i \quad \text{where} \quad i \neq j \tag{2}$$

x^i is the i'th measured point in the training data, n is the number of measured points, $\Phi()$ indicates the kernel function (i.e. the shape of the radially symmetric function), w_i is the weight given to the i'th kernel function, t^i is the measured output at x^i and $x^i - x^j$ is the distance between the measured points i and j. RBF models can use different kernel functions, such as Gaussian, Multiquadric, Thinplate, Spline, Cubic and Linear functions (Khan 2011).

The main advantages of RBF models are as follows:

- RBF can learn the complex and nonlinear relationship between the input and output parameters (Hagan and Demuth 1999).
- RBF can confront missing and noisy data with a good generalisation capability (Hagan and Demuth 1999).
- RBF models can be more efficient (in relation to the number of measurements required) than polynomials for large-scale nonlinear problems, since the approximation model is purely driven from the collected data (through learning), rather than assuming a fixed model type in advance (He and Rutland 2004).
- RBF is very fast in learning the relationship between the input and output variables because of using two-stage network training. The first stage is to determine the weights from the 'Input' to the 'Hidden' layer, and the second stage from the 'Hidden' layer to the 'Output' layer (Khan 2011).

Although the RBF method has proved to be an efficient modelling technique for complex problems, such as modelling of internal combustion engines (Howlett et al. 1999), there is always the possibility of over-fitting (Forrester et al. 2008), particularly for noisy measurements.

In this paper, the performance of the RBF method is investigated by applying three of the frequently used kernel functions: Gaussian, Multiquadric and Thinplate (Kim et al. 2009; Rango et al. 2013).

2.3 Kriging

The theoretical basis of the Kriging modelling was developed by Matheron (Matheron, 1976), based on the original work by Krige (Forrester et al., 2008). Kriging metamodels are derived from (3) (Sacks et al. 1989; Simpson et al. 2001a), which postulates a combination of a global model (f) and localised departures (G).

$$y(x) = f(x) + G(x) \tag{3}$$

In this equation, the response y is predicted at a point x using a global approximation function f, which is usually a polynomial, and 'localized' deviations G, which are calculated by interpolation of the measured sample points. $G(x)$

denotes the realisation of a weakly stationary stochastic process with mean 0, process variance σ^2 and nonzero covariance function, as given by (4) (Khan, 2011):

$$Cov\left(G\left(x^i\right), G\left(x^j\right)\right) = \sigma^2 R\left(x^i, x^j\right) \tag{4}$$

$R(x^i, x^j)$ represents the correlation between any two of the measured sample points, x^i and x^j. R is assumed to be a function of a small set of parameters, which are estimated based on the Likelihood function (Forrester et al. 2008). The Likelihood function defines the probability of the measured outputs, given a specific set of parameter values (Cressie 1990).

A variety of correlation functions are available in literature to fit a Kriging model (Couckuyt et al. 2012; Fang et al. 2005; Kaymaz 2005; Kersaudy et al. 2015; Kleijnen 2009; Kleijnen and van Beers 2005; Moyeed and Papritz 2002; Passos et al. 2015; Picheny et al. 2013a; Ulaganathan et al. 2014, 2015; Wang and Shan 2005). For this study, a selection of commonly used correlation functions, i.e. Gaussian (de Oliveira et al. 2013; Kleijnen 2009; Passos et al. 2015; Picheny et al. 2010), Matérn 3/2 (Picheny et al. 2013a; Ulaganathan et al. 2014, 2015) and Matérn 5/2 (Ulaganathan et al. 2014, 2015), is considered for the study of the performance of the Kriging metamodels. The common choice for the functional form for the global function f (see (3)) which gives the overall trend is a polynomial (linear or quadratic). However, for this study, a constant term (zero order polynomial) has been chosen, similar to the approach commonly used in engineering practice—see for example Forsberg and Nilsson (2004) and van Gelder (2014). This approach is referred to as ordinary Kriging, a particular case of universal Kriging—which uses a polynomial global trend function.

Furthermore, Kriging metamodels can be formulated to deal with noise by introducing an extra variation term (σ_e), which measures the amount of random variation over the n-dimensional design space, through a so-called nugget effect (Kleijnen, 2009; Kleijnen and van Beers 2005; Picheny et al. 2013a; Chen et al. 2018, 2019). In a 'nugget effect' Kriging the nonzero covariance function can be defined as:

$$\text{Cov}\left(G\left(x^i\right), G\left(x^j\right)\right) = \sigma^2 R\left(x^i, x^j\right) + \sigma_e^2 I\left(x^i, x^j\right) \tag{5}$$

where I is an $n \times n$ identity matrix. The performance of Kriging with a nugget effect is not well documented in the existing literature.

Some of the main advantages of Kriging models can be summarised as (Khan, 2011; Seabrook et al., 2005; Welch et al., 1992a):

- Kriging models are highly flexible due to the wide range of correlation functions.

- Kriging models require fewer measurements due to the strong interpolation among the measured sample points.
- Kriging models can either 'honour the data' by providing an exact interpolation of the data, or 'smooth the data' by providing an inexact interpolation.

A potential pitfall of using Gaussian Kriging models is the *curse of over-fitting* (Forrester et al., 2008), particularly with noisy sets of measurements. Also, implementation of Kriging models is a relatively time-consuming process since determining the maximum likelihood parameters is a complex optimisation problem (Khan, 2011).

It is noteworthy that the performance of Kriging can be endangered by the type of designed experiment used to collect the experimental data. This method has shown some difficulty in fitting response models for Full Factorial and Central Composite Design of Experiments (DoE) methods (Meckesheimer et al., 2001), since the correlation matrix becomes almost singular when multiple sample points are located close to each other.

3 Research methodology

3.1 Test functions and test cases

To investigate the performance of different metamodelling strategies, 18 test functions were selected from literature to be representative of the range of dimensionality and complexity (nonlinearity) seen across the different engineering problems discussed in related literature, and seen by the authors in engineering practice. The mathematical definitions of selected test functions are listed in Appendix A.

Table 2 summarizes the balanced structure of the test cases for the experimental study based on the test function used, showing that each combination of problem scale and complexity is tested with three problems.

In terms of problem dimensionality (i.e. the number of function variables k), the 18 functions can be classified into three categories:

– Small-scale ($k \leq 4$) (6 test functions)
– Medium-scale ($5 \leq k \leq 8$) (6 test functions)
– Large-scale ($k \geq 9$) (6 test functions)

In terms of problem complexity, the 18 functions represent two levels:

– Low-order nonlinear problems (9 test functions)
– High-order nonlinear problems (9 test functions)

Table 2 Experimental matrix for the study—summary of test functions and their characteristics

Test function	No of variables (n)	Problem type		Modelling conditions				
		Scale	Nonlinearity	DoE size (test points)				Noise
				MBs (10n)	MBl (30n)	Validation		
Test 1	2	Small	Low	20	60	1000		0–15%
Test 2	3	Small	Low	30	90	1000		0–15%
Test 3	4	Small	Low	40	120	1000		0–15%
Test 4	2	Small	High	20	60	1000		0–15%
Test 5	3	Small	High	30	90	1000		0–15%
Test 6	4	Small	High	40	120	1000		0–15%
Test 7	5	Medium	Low	50	150	1500		0–15%
Test 8	6	Medium	Low	60	180	1500		0–15%
Test 9	8	Medium	Low	80	240	1500		0–15%
Test 10	6	Medium	High	60	180	1500		0–15%
Test 11	7	Medium	High	70	210	1500		0–15%
Test 12	8	Medium	High	80	240	1500		0–15%
Test 13	10	Large	Low	100	300	2000		0–15%
Test 14	16	Large	Low	160	480	2000		0–15%
Test 15	20	Large	Low	200	600	2000		0–15%
Test 16	10	Large	High	100	300	2000		0–15%
Test 17	14	Large	High	140	420	2000		0–15%
Test 18	18	Large	High	180	540	2000		0–15%

For the classification of complexity, the test functions for which the value of R^2 is more than 0.99, when constructing a second-order polynomial metamodel, are considered low-order nonlinear, otherwise they are considered as high-order nonlinear (Jin et al. 2001; Mullur and Messac 2005; Zhao and Xue 2010).

The performance of the chosen metamodelling techniques was also examined in relation to the modelling conditions. To reflect the common choices of sample sizes for Model Building (MB) engineering experiments, two DoE sizes were considered:

- Small DoE sample size (denoted as MBs in Table 2): an Optimal Latin Hypercube (OLH) design with a number of test points '10×number of parameters' (Jin et al. 2001; Loeppky et al. 2009).
- Large DoE sample size (MBl in Table 2): an OLH design a with number of test points '30×number of parameters' (Jin et al. 2001; Mullur and Messac 2005; Zhao and Xue 2010).

The main reason of using an OLH design to generate the test points is the ability of this sampling strategy to deliver uniformly distributed test points within the range of parameters and also the flexibility on the number of test points (Fang et al. 2010; Sacks et al. 1989). In this paper, a Permutation Genetic Algorithm is used to generate the Optimal Latin

Hypercube test point based on the Audze-Eglais objective function (Kianifar et al. 2015).

The validation sets designed to check the metamodels' prediction accuracy were also OLH DoE. Two different validation DoE sizes were considered, ranging between 1000 and 2000 validation tests, based on the problem dimensionality, as shown in Table 2.

Table 2 summarizes the base experiment which includes the evaluation of the effect of problem dimensionality and complexity, and the scale of the experimental DoE used for metamodel building. In order to evaluate the impact of experimental uncertainty/noisiness of the response on the performance of different metamodelling techniques, the experiment was repeated 3 times with different noise conditions. Noise was introduced in the experiment by adding a random term to the function response, generated from a standard Gaussian distribution multiplied by set values of noise-to-signal ratio (expressed/referred to as a "% noise"), as follows:

- Base test—'0%' noise condition (or smooth data);
- Repeat 1—'5%' noise condition;
- Repeat 2—'10%' noise condition;
- Repeat 3—'15%' noise condition.

The choice of these level of noise was based on the comprehensive experimental studies such as those reported in Berger (2012) and Kianifar et al. (2014), which systematically

considered the experimental variability associated with modelling different engine responses.

3.2 Selection of Metamodelling techniques

As discussed, the metamodelling techniques considered within the scope of this study include Polynomial, RBFs and Kriging. Several options, selected based on the common choices in engineering practice, were considered for each type, as follows:

- Two polynomial metamodels: order 2 ('Poly 2') and three ('Poly 3_c2'—i.e. cubic polynomial with interaction order 2);
- Three RBF metamodels: using Gaussian, Thinplate and Multiquadric kernel functions (shown as 'RBF_G', 'RBF_TP' and 'RBF_MQ');
- Three Kriging metamodels: using Gaussian ('Krig_G'), Matérn 3/2 ('Krig_M32') and Matérn 5/2 ('Krig_M52') correlation functions;
- Three Kriging metamodels with nugget factor, using Gaussian ('KrigN_G'), Matérn 3/2 ('KrigN_M32') and Matérn 5/2 ('KrigN_M52') correlation functions;

Thus, the base experiments included a total of 11 metamodels fitted for each of the 18 test functions considered. In order to account for the impact of the modelling conditions (sample size and noisiness of the response), the base experiment was repeated for "Small" and "Large" DoE sizes and different noise conditions (5%, 10% and 15%, in addition to the base experiments which considered a smooth signal—0% noise).

3.3 Performance metrics and analysis procedure

The metrics used to investigate the performance of the constructed metamodels are as follows:

(a) Normalised Root Mean Squared Error (NRMSE) of the validation set, as the indicator of a metamodel's prediction accuracy over the design range (Mullur and Messac 2005; Ulaganathan et al. 2014, 2015), expressed by (6).

$$NRMSE = \frac{\sqrt{\frac{1}{v} \cdot \Sigma_{i=1}^{v} \left(y_i - \hat{y}_i\right)^2}}{max(y) - min(y)} \qquad (6)$$

NRMSE is the normalised discrepancy between the real values (y) of the v validation sample points and the corresponding prediction values ($\hat{y}$) (Hartmann et al. 2013). The smaller the value of NRMSE, the more accurately the metamodel predicts.

The main reason for using an external validation metric (based on additional validation samples) to define the prediction accuracy, instead of statistical properties of the constructed model such as root mean squared error (RMSE) (Forsberg and Nilsson 2004; Kaymaz, 2005; Mullur and Messac, 2005; Zhu et al. 2009) and R^2 (Ben-Ari and Steinberg 2007; Fang et al. 2005; Jin et al. 2003; Vicente-Serrano et al. 2003), is that non-parametric metamodels are susceptible to over-fitting, especially in noisy data conditions. Thus, by measuring the validation NRMSE, an evaluation is provided for the capability of a metamodel to predict the response behaviour over the design range rather than the ability to construct a model which follows the sample points (Fang et al. 2005; Vicente-Serrano et al. 2003).

(b) The variance of NRMSEs for a problem type; this provides an indicator of modelling robustness, coherent with Jin et al. (2001) and Chen et al. (2018, 2019), which relates to the capability of achieving good accuracy for different problems, thus providing an indicator whether a modelling technique is highly problem dependent. In the context of this study, the variance of NRMSE was calculated across the test-problem clusters defined through the experimental matrix, illustrated in Table 2. A smaller variance of NRMSE value indicates consistent performance of the methodology across the type of problem, which can be taken as an indication of robustness.

(c) The computational effort was taken as an indicator of efficiency. The required time to construct a metamodel is used to define the computational efficiency of different metamodelling techniques for each of the problem types.

3.4 Validation test cases

The analysis of the experimental data generated from the study planned for this research will facilitate the articulation of a set of practitioner guidelines for the robust choice of a metamodelling technique for a problem type and modelling conditions. In order to validate the guidelines provided, an empirical validation exercise was included in the research methodology, by choosing a set of four real-world engineering problems, summarised in Table 3. The test problems were selected from reported case studies (including authors' previous work—test problems V1–V3) where actual DoE data was available/accessible, with the aim of covering a good mix of modelling conditions—the size of the data set and the level of noise.

Table 3 Summary of the validation experiments

Type of experiment	Test problem	Problem characteristics
Physical	V1: Modelling of fuel consumption response for a diesel engine based on dynamometer testing data (Kianifar et al. 2014)	•Medium scale •High-order nonlinearity •*Medium/large data set* •*Low noise*
	V2: Modelling of NOx response for a diesel engine based on dynamometer testing data (Kianifar et al. 2014)	•Medium scale •High-order nonlinearity •*Medium/large data set* •*High noise*
Computer-based/CAE	V3: Modelling the head injury criteria (HIC) response for a tall man during a car-pedestrian impact based on dynamic CAE experiments (Zhao et al. 2010)	•Medium scale •High-order nonlinearity •*Small sample data set* •*Medium noise (induced in the experiment)*
	V4: Modelling of contaminant concentrations response in the upper aquifer of a waste disposal site in Moscow based on numerical simulation data (Marrel et al. 2008; Volkova et al. 2006).	•Large scale •High-order nonlinearity •*Small sample data set* •*No noise*

4 Results and analysis

Based on the designed research methodology, for each of the 18 case studies listed in Table 2, 11 model types are created under four noise conditions and two sampling sizes. Thus, a total of 1584 metamodels were constructed.

This section provides a systematic analysis of the results of the study. Taking the engineering practitioner viewpoint, the analysis of the results focusses on the modelling choices and conditions across the types of problems. Thus, the analysis will consider the comparative performance (based on the criteria outlined in Section 3.3) of the metamodelling techniques for different scale and nonlinearity of test problems, following the sequence below:

1. Small sample DoE
2. Large sample DoE
3. Sensitivity to noise

The efficiency of the metamodelling techniques (in terms of computation cost) will be also evaluated, before articulating a set of practical guidelines for the choice of metamodelling techniques.

4.1 Evaluation of metamodelling techniques for small sample DoEs

Figure 1 summarises the performance of modelling techniques aggregated based on the results across the fitted models for all the 18 test functions for small sampling sizes. This figure includes both NRMSE (showing the average across all fitted models and whiskers extending to the minimum and maximum NRMSE across the fitted models) and variance of NRMSE, to give an overall simultaneous measure of accuracy and robustness, respectively. As discussed, a smaller value for the mean of NRMSE indicates that the modelling strategy is more accurate, and a smaller range and variance of NRMSE across the fitted models indicates a higher robustness of the modelling strategy. In line with the literature, considering the limited number of test functions to calculate the variance, if the variance of NRMSE is less than 0.05 (i.e. less than 5% of the average response value), the modelling methodology can be accepted as robust.

The results in Fig. 1 show that for small DoE sizes, Kriging without a nugget factor outperforms other modelling strategies in terms of robustness—judged based on the average and range of NRMSE across the fitted models and the variance of NRMSE. In particular, Kriging with the Matérn 5/2 correlation function (shown as 'Krig_M52') performs slightly better than other correlation functions. It should be noted that the variance of NRMSE for all the modelling strategies is within an acceptable range (less than 0.05). Polynomial models can have very good performance (in some cases NMRSE is 0), but their robustness is quite poor—as shown by the large range and variance of NRMSE. RBFs do not perform very well with small sample sizes; the Gaussian kernel appears to provide more robustness, but the worst accuracy (highest average NRMSE).

In order to better understand the behaviour of different metamodelling techniques, a deeper analysis is provided by evaluating the performance in relation to the problem characteristics—nonlinearity and scale.

Fig. 1 Summary of performance of metamodelling techniques for small DoE size
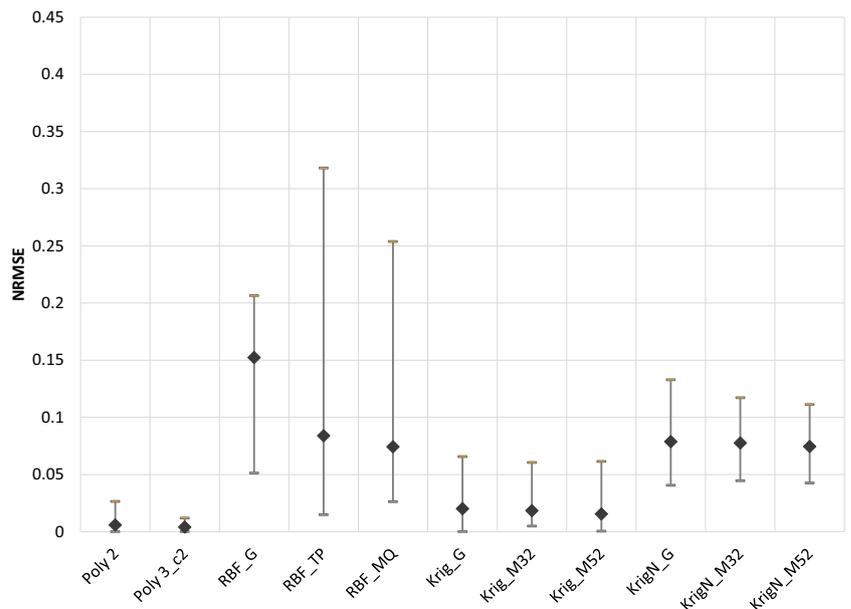

Therefore, Kriging modelling technique with Matérn 5/2 correlation function is shown to be the best method in terms of both accuracy and robustness over all types of problems for small sampling sizes.
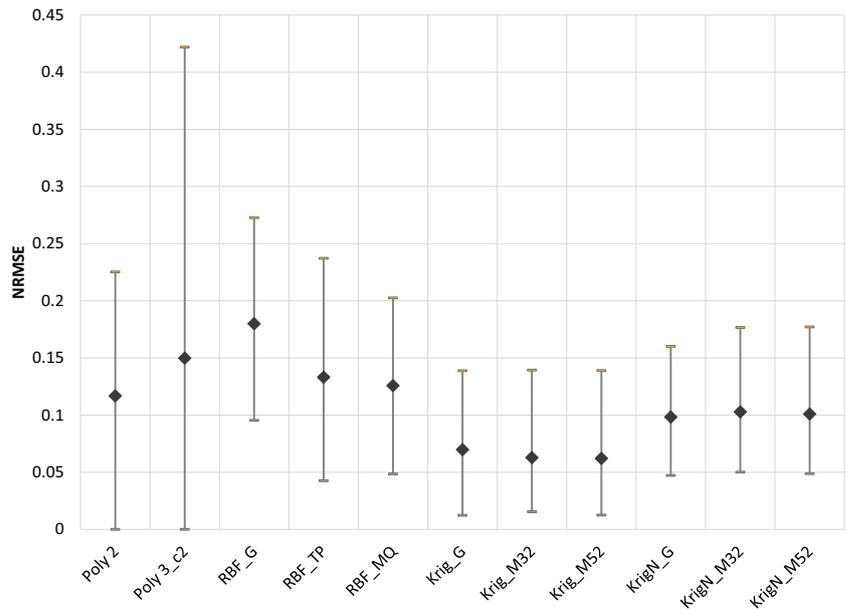
Figures 2 and 3 illustrate the mean of NRMSE for each of the nonlinearity order categories, i.e. low and high nonlinearity, respectively. Figure 2 indicates that with small sampling size, for low-order nonlinear problems, the Polynomials appear to perform better than all other methods. Kriging models (without nugget) perform reasonably well in terms of both average NRMSE and robustness (measured by the range of

NRMSE across the fitted models)—although not as good as polynomials. RBF models do not perform robustly—as shown by the large range of NRMSE. The results in Fig. 3 show that for high-order nonlinear problems, Kriging metamodels outperform all the others in terms of accuracy and robustness. While polynomial models can still deliver good models in some cases (the whiskers extend to 0), in general, they do not have enough flexibility to model nonlinearity, which gives poor robustness (the range of NRMSE is quite large). RBFs perform relatively better in terms of robustness, although they have inferior average NRMSE.

Fig. 2 Analysis of NRMSE performance for low nonlinearity/ small DoE size

Fig. 3 Analysis of NRMSE performance for high nonlinearity/small DoE size

Figures 4, 5 and 6 summarise the analysis of the performance of metamodelling techniques in relation to problem dimensionality. The average and range of NRMSE across the fitted models are illustrated for small (Fig. 4), medium (Fig. 5), and large (Fig. 6) scale test functions. The results across show that Kriging performs best in terms of model accuracy and robustness across the problem scales. In particular, Kriging with correlation function Matérn 5/2 provides the best performance across the tests. Thus, it can be concluded that the problem scale has a negligible effect on the performance of the modelling techniques, which was also reported by Jin et al. (2001).

Other notable observations from the results in Figs. 4 to 6 include:

– Polynomial models can deliver good models (the NRMSE whiskers extend to 0), but their robustness is worse than most other models. In particular, the cubic polynomial model shows a very wide NRMSE range for the large-scale test functions, which can be explained by the number of coefficients needed for the larger scale test problems in relation to the DoE size.

Fig. 4 Analysis of NRMSE performance for small-scale problems/small DoE size

Fig. 5 Analysis of NRMSE performance for medium-scale problems/small DoE size

– The performance of the RBF models improves relative to polynomials for the large-scale problems, in particular in terms of robustness.

– Kriging with nugget metamodelling delivers robust performance, but inferior to the standard Kriging in terms of accuracy (average NRMSE), which is not surprising given that there is no noise in the response.

To provide an even more granular view of the results, Figs. 7 and 8 show the performance of modelling techniques in terms of the average NRMSE for different problem scales

with low nonlinearity (Fig. 7) and high nonlinearity (Fig. 8), respectively.

The results in Fig. 7 confirm that for low nonlinearity problems and small sample DoEs, the polynomials (in particular the cubic polynomials) deliver the best average NRMSE performance. The Kriging models (without nugget) provide very good models with NRMSE below 0.05 in all cases. RBFs can provide competitive NRMSE performance, but not in all cases, so not a robust choice. For high-nonlinearity problems, the results in Fig. 8 show that the Kriging method with both Matérn 3/2

Fig. 6 Analysis of NRMSE performance for large-scale problems/small DoE size

Fig. 7 Detailed analysis of NRMSE performance for low nonlinearity/small DoE size

and Matérn 5/2 correlation functions perform the best in terms of average NRMSE across the test functions considered.

It is worth reflecting that if the problem scale is large and the sample size is small, using a non-parametric model like Kriging can be preferable since it uses fewer model parameters (i.e. while the Polynomial will require many model coefficients in relation to the Poly order). In other words, the performance of Polynomials is very 'problem-type' and 'sample-size' dependent (Jin et al., 2001).

4.2 Evaluation of metamodelling techniques for large sample DoEs

The analysis of the results for the large sample DoE follows the same sequence as the analysis of small sample DoE results presented in the previous section.

Figure 9 summarises the performance of modelling techniques aggregated based on the results across the fitted models for all the 18 test functions for large sampling sizes, in terms of NRMSE mean and range (minimum/maximum across the

Fig. 8 Detailed analysis of NRMSE performance for high nonlinearity/small DoE size

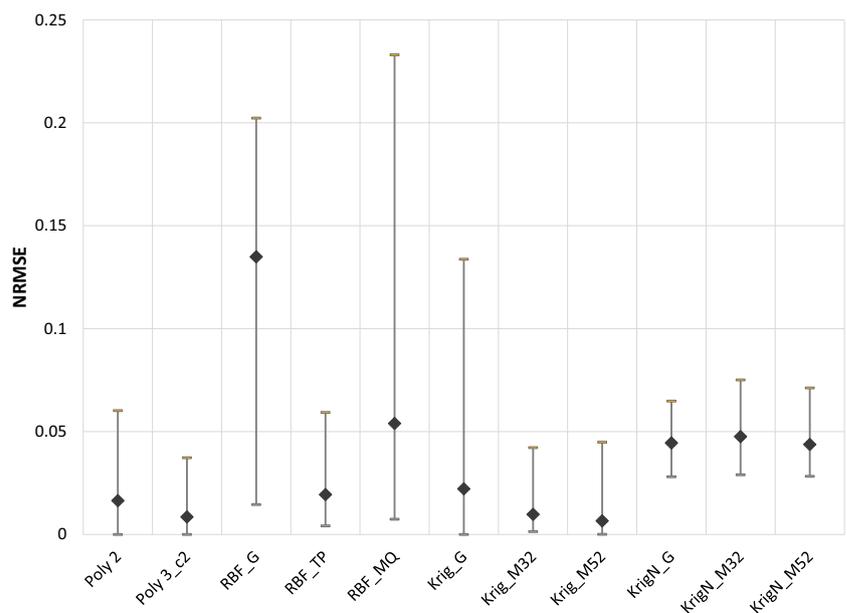**Fig. 9** Summary of performance of metamodelling techniques for large DoE size

fitted models) and variance of NRMSE. The results show that Kriging, in particular Kriging with Matérn 5/2, perform over-all better (more robust) than the other metamodelling techniques, results similar to the conclusion drawn from the results seen in Fig. 1. It can also be noted that, unlike the results in Fig. 1 for small size DoE, Polynomial order 3 (i.e. cubic model) performs marginally better than Poly order 2 with larger sampling DoE test plans. The performance of the RBF model with Thinplate kernel function is also improved significantly by increasing the number of sampling points. Moreover, Fig. 9 shows that in comparison with the results in Fig. 1 for small sample DoE, the robustness of all modelling strategies is

improved noticeably with the large DoE sizes (as shown by both NRMSE and variance of NRMSE values).

Therefore, Kriging modelling technique with Matérn 5/2 correlation function is shown to be the best method in terms of both accuracy and robustness over all types of problems for not only the small but also large sampling sizes.

Figures 10 and 11 show that when a large sampling data is accessible, Kriging, and in particular Kriging with Matérn 5/2, delivers the best accuracy and robustness across both low and high nonlinearity test functions. As expected, and similar to what can be observed in Figs. 2 and 3 for small sample DoE, the NRMSE is lower for low-nonlinearity problems compared

Fig. 10 Analysis of NRMSE performance for low nonlinearity/large DoE size

Fig. 11 Analysis of NRMSE performance for high nonlinearity/large DoE size
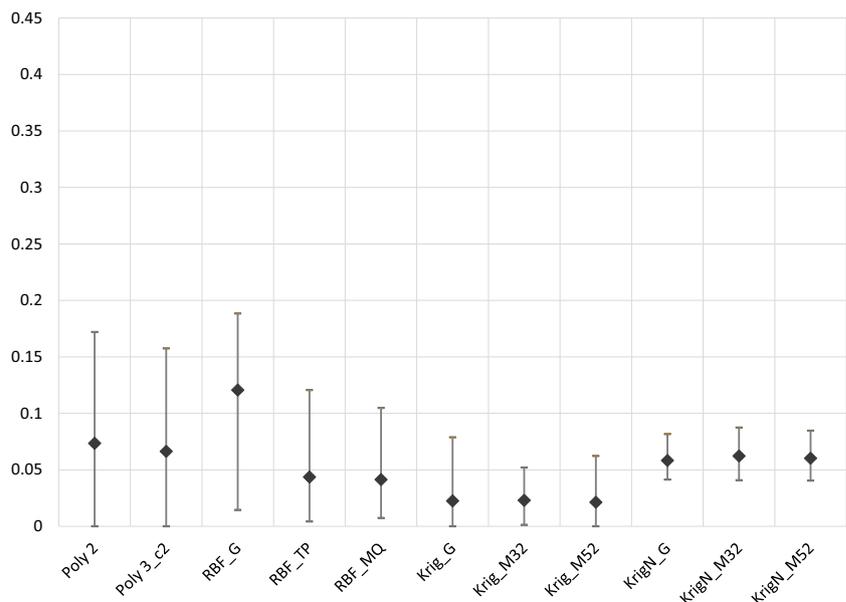
to high- nonlinearity problems, for all modelling techniques. Similarly, contrasting the range of NRMSE (visually illustrated by the whiskers) seen in Figs. 10 and 11 for a large DoE sample, with Figs. 2 and 3 (small DoE size), it is noted that the robustness improves with the DoE sample size for all modelling techniques. RBF models, in particular with Gaussian and Multiquadric kernel functions, show less robustness compared to the other methods.

Figures 12, 13 and 14 summarise the analysis of the performance of metamodelling techniques in relation to problem scale. These results show that the Kriging 5/2 outperforms all the other modelling techniques *regardless of problem scale*,
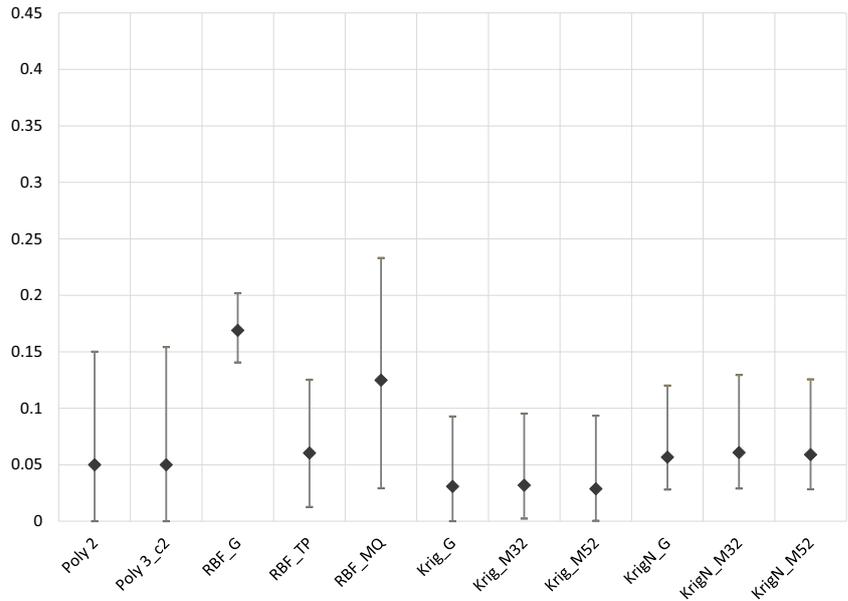
similar to what was seen for small DoE sizes. Kriging models with Matérn 3/2 correlation function are also competitive.

Figures 15 and 16 provide further granularity for the analysis, showing the performance of metamodelling techniques in terms of the average NRMSE for different problem scales with low nonlinearity (Fig. 15) and high nonlinearity (Fig. 16), respectively. Figure 6 shows that for large sampling sizes, for high-order nonlinear problems the Kriging method with both Matérn 3/2 and Matérn 5/2 correlation functions perform the best in terms of accuracy (Kriging 5/2 is the best) across problem scales. However, for low-order nonlinear problems, it was noted that Polynomials (in particular cubic polynomials)

Fig. 12 Analysis of NRMSE performance for small scale problems/large DoE size

Fig. 13 Analysis of NRMSE performance for medium-scale problems/large DoE size

perform as well as Kriging models. It is also noted that for small and medium-scale problems Kriging with Gaussian correlation function performs particularly well. Of the RBF models, the RBFs with Thinplate kernel appear to perform best.
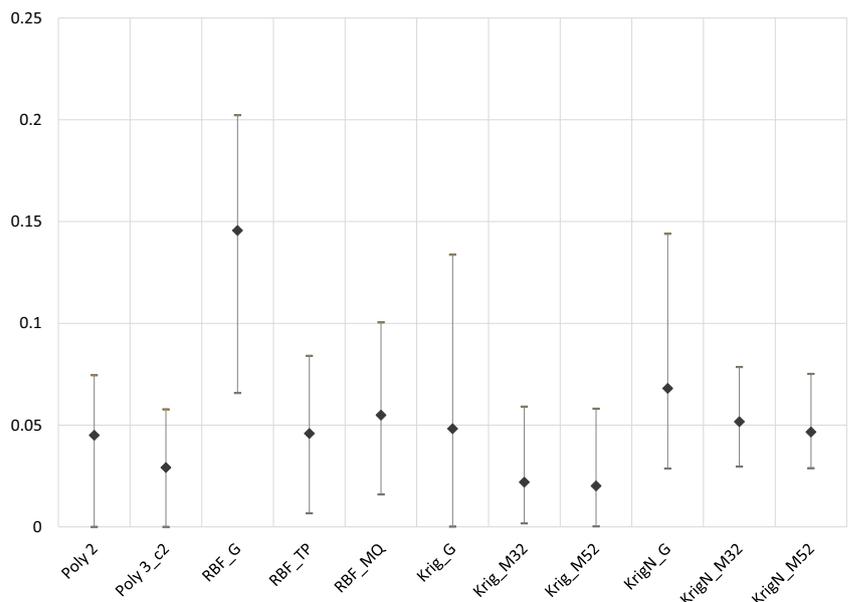
4.3 Sensitivity to noise

In order to investigate the effects of noise in the response variable measurement on the performance of metamodelling techniques, in terms of both accuracy and robustness, all the 18 test case functions in Table 1 were modelled under three noise conditions, i.e. 5%, 10% and 15% Gaussian random noise artificially added to the
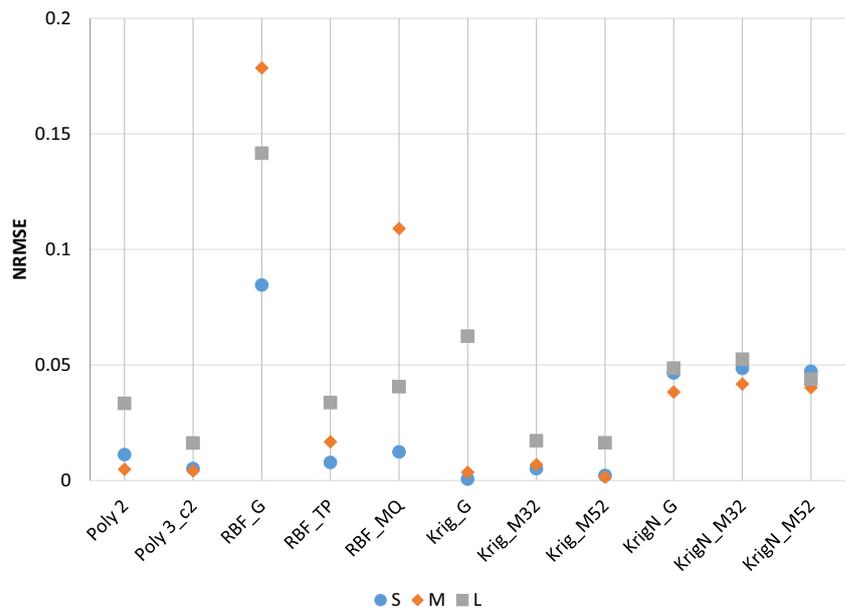
response data. The whole set of experiments (including both small and large DoE sampling size) was repeated with the three levels of noise.

Figs. 17 and 18 summarise the prediction accuracy of different modelling techniques under four different noise conditions for the small and large DoE sizes, respectively. These figures show that while Kriging with Matern 5/2 correlation function overall performance outperforms other modelling techniques for smooth data (i.e. 0% noise) regardless of the DoE size (as also seen in Figs. 1 and 9), Kriging with Matern 3/2 correlation function appears to perform marginally better under noisy data conditions, in terms of average NRMSE.

Fig. 14 Analysis of NRMSE performance for large-scale problems/large DoE size
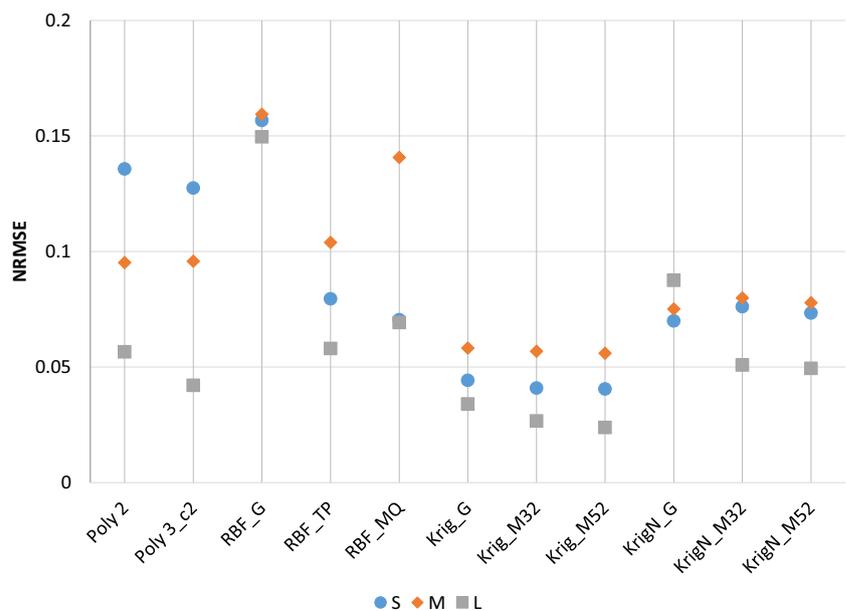
Fig. 15 Detailed analysis of NRMSE performance for low nonlinearity/large DoE size
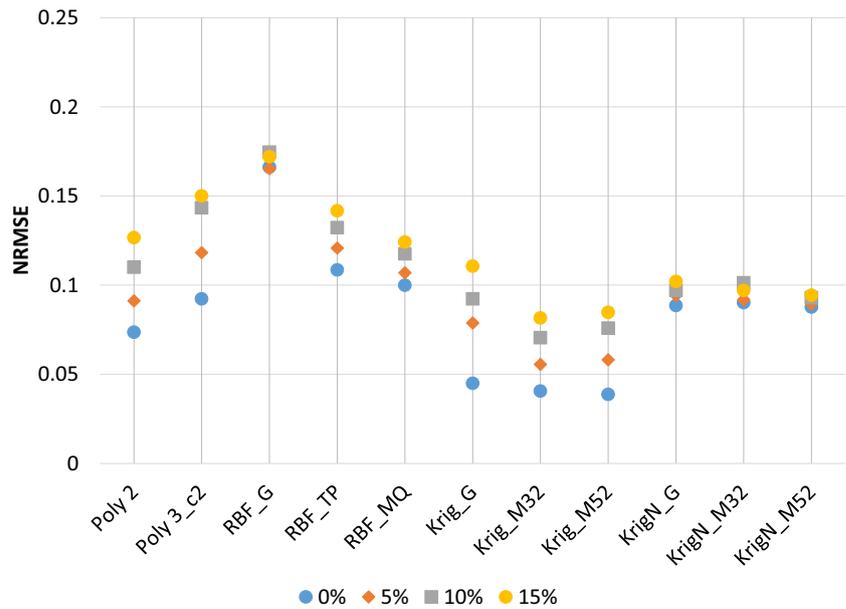

Other important observations from these results can be summarised as follows:

– The performance of the Kriging with nugget metamodelling technique is remarkably quite insensitive to noise levels within the range considered in the study. However, the average NRMSE of Kriging with nugget (with Matern 5/2 correlation function) outperforms the standard Kriging under highest levels of noise (15%) and with a large DoE sample size.

– As expected, a general trend of worsening NRMSE with higher noise is observed across the metamodelling techniques, with the exception of Kriging with nugget. The polynomial models also show good stability in relation to increase noise for large DoE sample sizes.

– It is also seen that the accuracy of modelling improves with the large DoE sample size, except for the RBF with Gaussian Kernel. In general, RBF metamodels show the greatest sensitivity to noise.

Fig. 16 Detailed analysis of NRMSE performance for high nonlinearity/large DoE size

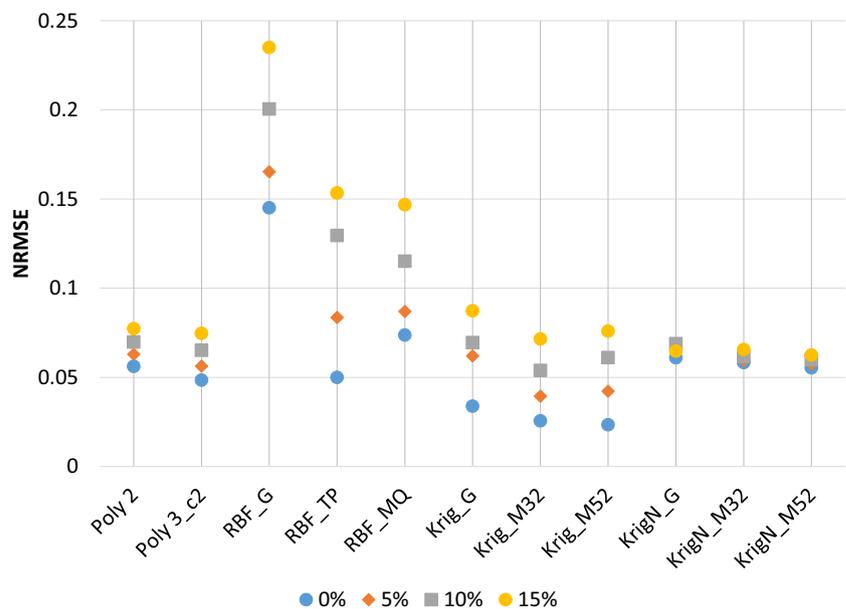**Fig. 17** Average NRMSE for different levels of noise/small DoE size

4.4 Evaluation of computational efficiency of Metamodelling techniques

The efficiency of metamodelling techniques was evaluated by the time required to fit a model on an Intel i5, 2.60 GHz computer (with 8GB RAM), as summarised in Table 2 for each problem scale and sample size. This table indicates that fitting a model using either Polynomial or RBF methods is quite fast

regardless of the problem dimensionality (i.e. took less than 1 s to fit a model). Therefore, relatively speaking, while Polynomial is observed to be the most time-efficient method, Kriging method is relatively very time-consuming. This result was expected since the Kriging modelling technique without nugget factor requires a k-dimensional optimisation, regardless of the correlation function, to find the optimum value of the maximum likelihood function parameters. The optimisation

Fig. 18 Average NRMSE for different levels of noise/large DoE size

Table 4 Time needed (in second) to fit a model

	DoE size					
	MBs	MBs	MBs	MBl	MBl	MBl
Model/scale	Small	Medium	Large	Small	Medium	Large
Poly 2	0.01	0.01	0.46	0.00	0.02	0.20
Poly 3_c2	0.00	0.00	0.05	0.00	0.15	0.43
RBF_G	0.00	0.00	0.06	0.01	0.01	0.20
RBF_TP	0.00	0.00	0.02	0.01	0.02	0.09
RBF_MQ	0.00	0.00	0.02	0.00	0.03	0.06
Krig_G	1.81	8.61	102.22	5.88	55.03	653.34
Krig_M32	1.96	8.85	131.79	8.43	82.31	645.37
Krig_M52	1.92	8.09	109.66	7.12	48.38	1365.52
KrigN_G	1.82	7.39	58.17	6.59	56.17	724.76
KrigN_M32	1.50	5.44	71.21	6.70	50.34	603.08
KrigN_M52	1.81	8.08	109.49	5.24	55.88	781.47

problem is even more complex for Kriging with nugget factor which requires a $(k+1)$-dimensional optimisation problem, having an extra parameter to optimise for the nugget.

Table 4 illustrate that the time required to fit a model increases by the increase of the problem dimensionality, especially for the Kriging method. It is also noted that modelling using a larger sampling data (i.e. MBl) is less time-efficient than a small sampling data (i.e. MBs) which matches well with intuition considering that the optimisation process becomes computationally more expensive.

4.5 Overall guideline for choosing a robust metamodelling method

Summarising all the presented results, in order to recommend on the selection of the most robust metamodelling technique, it can be concluded that for smooth response data:

– Kriging with Matérn 5/2 correlation function outperforms all the other modelling methods, regardless of problem scale and DoE size.
– Kriging with Matérn 5/2 correlation function performs best for highly nonlinear functions, regardless of the problem scale and DoE size, while also performing reasonably accurate for low nonlinear problems. Polynomials (especially Cubic models) are also providing competitive results to the Krig_M52 method for low-order non-linear functions, however; a Polynomial might require a larger DoE sample to calculate the model coefficients.

And for *noisy data*:

– Kriging with Matérn 3/2 correlation function excels other modelling techniques in terms of accuracy and robustness.
– Kriging with nugget factor (Matérn 5/2 correlation function) can be a better modelling option for highly noisy data (15% noise) with large DoE samples.

5 Validation case studies

In order to provide empirical validation for the guidelines provided in section 4.5, a modelling analysis was performed for four real-world engineering test problems, as outlined in the methodology section 3.4. For each of these four validation test case problems, all the considered metamodelling techniques were deployed, and the performance of the techniques measured in terms of NRMSE of validation test points was analysed. The results are discussed in the following sections.
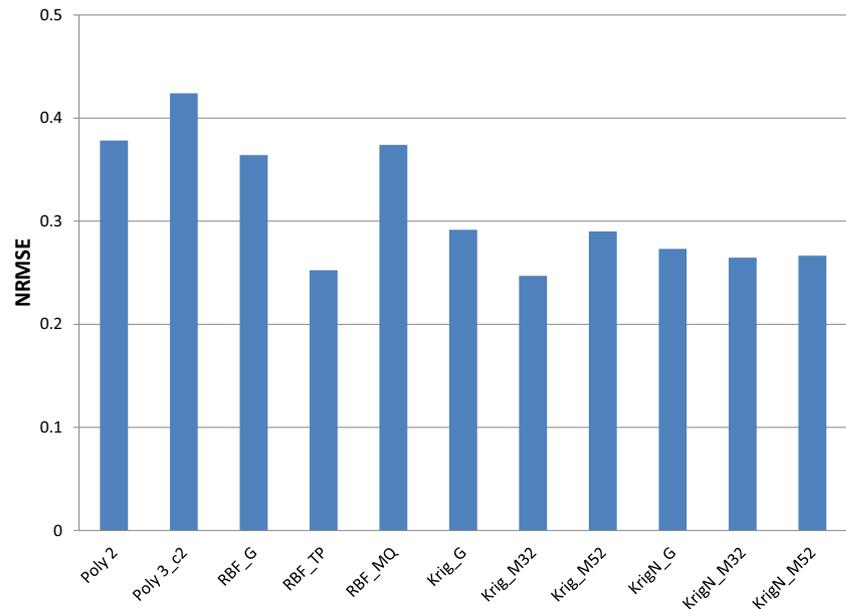
5.1 Diesel engine fuel consumption

The aim of this case study was to predict the fuel consumption response of a Diesel engine on an engine calibration test. The test data was collected using 127 test points scheduled based on an adaptive D-Optimal DoE, covering the range of 6 design parameters, with 15 additional space-filling test points for validation (for more details see (Kianifar et al. 2014)).

This case study can be categorised as a problem of:

– Medium scale: 6 design parameters;
– Medium- to high-order nonlinearity: due to the expected physical behaviour of the fuel consumption response (Kianifar et al. 2014);
– Medium DoE sample: 127 Model Building points for 6 parameters, which is bigger than a small DoE of 60 (i.e. 10×6) points, but smaller that the "large" DoE (i.e. $30 \times 6 = 180$);
– Low noise condition ($< 5\%$): measurement of fuel flow for a Diesel engine on an engine dynamometer test setting is reasonably accurate—though not 0% noise.

Figure 19 illustrates the performance of different metamodelling techniques for the case study in terms of NRMSE. This figure indicates that Kriging model with Matérn 3/2 correlation function performs best, while RBF with Thinplate kernel function is also providing competitive accuracy.

Fig. 19 NRMSE for the diesel engine fuel consumption

This result is consistent with the guideline provided—which would have suggested the Kriging model with Matérn 3/2 correlation function—given the presence of noise.
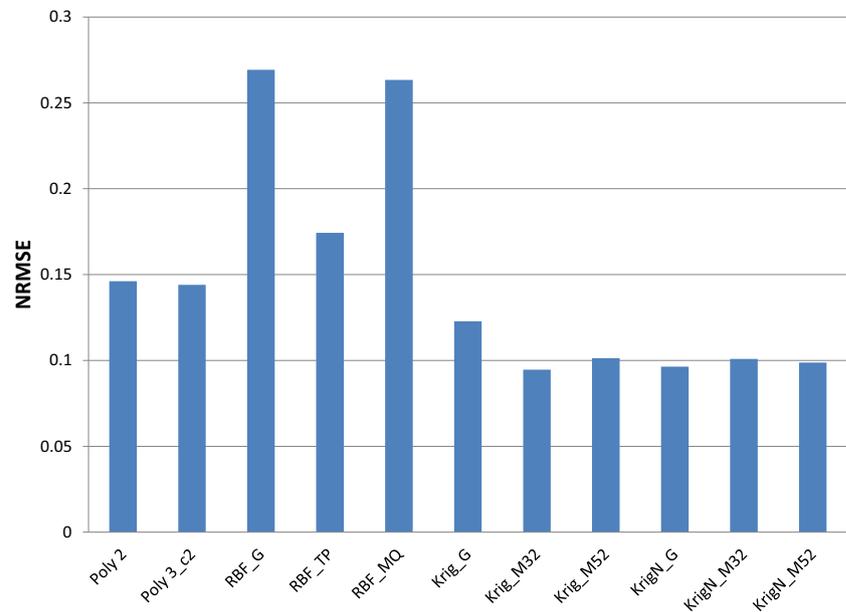
5.2 Diesel engine NOx emissions

This test case aims to predict the NOx emission response of a Diesel engine, based on data collected from the same engine dynamometer testing experiment as described in the previous section. The main difference is that measurement of NOx is

characterised by higher uncertainty—hence higher noise modelling conditions. Accordingly, this problem type is also a: medium-scale, high-order nonlinear with large sample size, but it is under 15% noise condition. The accuracy of metamodelling techniques is again compared based on the NRMSE of the 15 space-filling validation test points.

Figure 20 indicates that Kriging with Matérn 3/2 outperforms other modelling techniques, validating the results provided in the guideline. Moreover, similar to the trends observed in section 4.3, for such a problem, which is under a

Fig. 20 NRMSE for the diesel engine NOx

large noise condition with a large sampling data, Kriging models with nugget factor are also providing highly accurate response models.

5.3 Metamodelling for crashworthiness/pedestrian impact simulation

This case study aims to use an efficient metamodelling technique to model the head injury criteria as simulation response. The study focusses on the vehicle impact event simulation with a tall male pedestrian, based on a Optimal Latin Hypercube design of 100 test points; 70 points for constructing the models and 30 points for validation, over the range of 7 design parameters (for more details see (Zhao et al. 2010)).

This problem can be categorised as:

– Medium scale: with 7 design parameters
– High-order nonlinear: due to the highly nonlinear behaviour of head injury response in relation to the design parameters (Zhao et al. 2010)
– Small sampling data: 70 points for 7 parameters (i.e. 10×7)
– About 10% noise condition, due to the applied error range during the simulation process

Figure 21 illustrates the accuracy of different modelling techniques. It is noted that Kriging model with Matérn 3/2 correlation function performs best (i.e. slightly better than Krig_M52). Therefore, the proposed modelling method by the guideline for this type of problem excels other techniques. This figure also shows that the RBF models are also

performing comparatively well, matching the author's decision in the original study (Zhao et al. 2010) that they preferred RBFs to the polynomials.
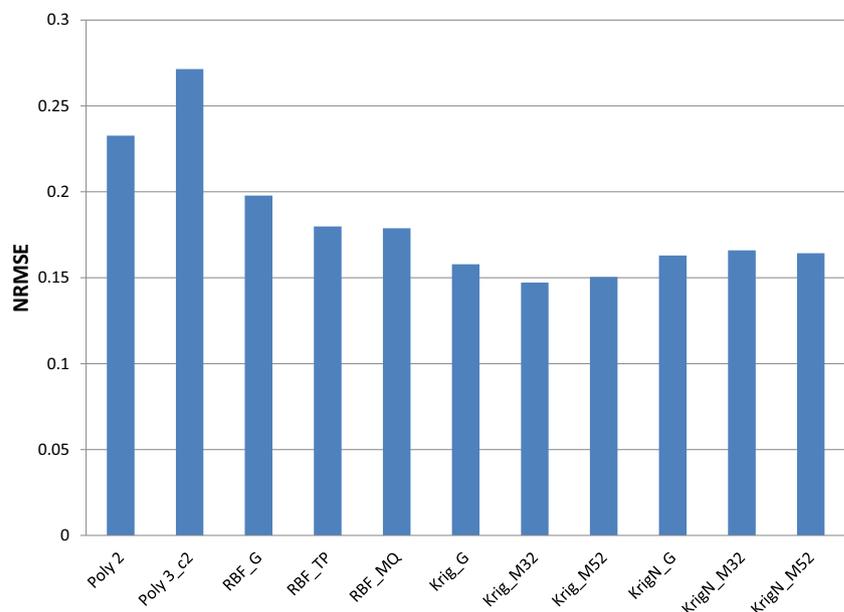
5.4 Metamodelling for the MARTHE simulation data

The MARTHE data problem has been used in literature to test the performance of metamodels (Marrel et al. 2008; Volkova et al. 2006). For this problem, 300 data points are available from the MARTHE code, which provides a numerical simulation of Strontium-90 transport in the upper aquifer of a waste disposal site in Moscow (Volkova et al. 2006). The data were collected over the range of 20 design parameters to model the contaminant concentrations. In this paper, 200 test points were randomly selected to construct the response model, and the remaining 100 points used as the validation set. Accordingly, the test problem type can be described as:
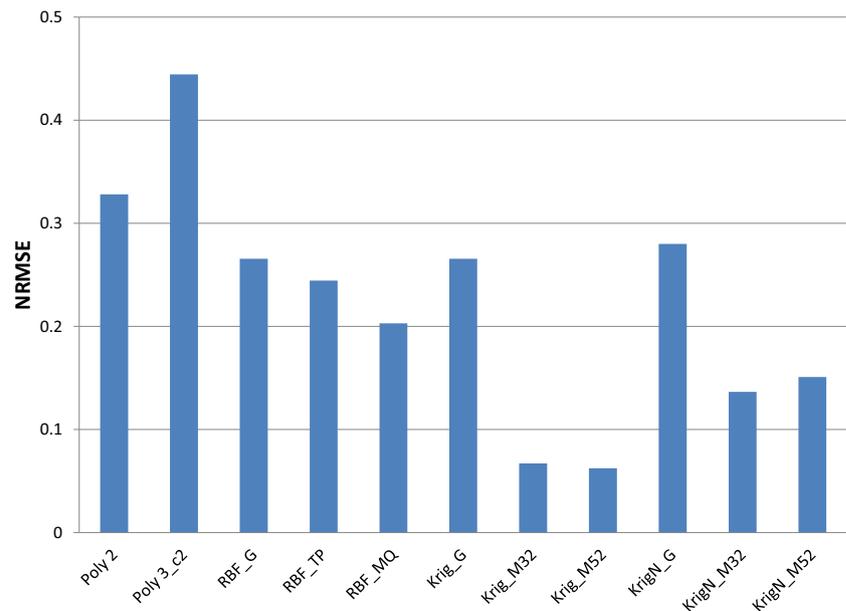
– Large scale: with 20 design parameters
– High-order nonlinear: due to the high-order nonlinear behaviour of the response (Volkova et al. 2006)
– Small sampling data: 200 points for 20 parameters (i.e. 10×20)
– 0% noise condition: collected from the MARTHE simulation code

Figure 22 shows the NRMSE of the validation set for different metamodelling techniques. This figure indicates that Kriging with Matérn 5/2 correlation function outperforms

Fig. 21 NRMSE for the crashworthiness model

Fig. 22 NRMSE for the
MARTHE model

other methods (i.e. slightly better than Krig_M32), validating the guideline provided in this research for such problem types.

6 Discussion, conclusions and future work

The main aim of this study was to carry out a comprehensive and systematic evaluation of the performance of several metamodelling techniques frequently utilised by engineers, considering the characteristics of the problems and the modelling choices and conditions—i.e. sample size and noise conditions. The motivation for this research was given by the need to provide guidelines for engineers who might seek to select a robust metamodelling strategy (from a diverse choice available) for real-world engineering problems.

The paper introduced a framework for the study of the metamodeling techniques designed to provide a balanced evaluation against all the characteristics that define the metamodeling typology—dimensionality, nonlinearity, sample size and noise. The careful selection of synthetic benchmark problems/test functions ensured a balanced set is available for each problem type. The systematic consideration of noise as a factor in metamodeling is important because it projects the validity of the study onto both physical and computer-based experiments.

The choice of techniques to be included in the evaluation was driven by the review of related studies from literature, and the scope/focus of the study on providing guidance for practical product development engineering metamodeling

problems, where affordability of large sample sizes for the experiments to generate the data for model building is often a limiting factor. Therefore, a range of techniques that in literature have shown promising results in relation to high dimensional problems and large sample sizes (e.g. HDMR) have not been included in this study. The study was also limited in relation to the selection of correlation functions for Kriging and kernels for RBF, confined to the choice of a set that was seen from previous studies to give consistently good results for engineering problems. For the same reason, we have not considered global trend functions for either Kriging (i.e. Universal Kriging) or RBF (i.e. hybrid RBF models). We have also not considered in this study the effect of hyperparameters that can be used to further improve the fit that can be achieved with a specific metamodeling technique. The main reason for this rests with engineering practice considerations: real-world modelling problems often require many models to be developed as part of a larger modelling exercise. For example, modelling engine performance—either through physical tests or computer-based experiments (for example using CFD models for combustion) requires many models (10 s and often 100 s) to be developed at different conditions (e.g. based on data collected at a range of engine speed/load setpoints). Through hyperparameters tuning and choosing from a wider pool of correlation functions or kernels, slightly better models fits could perhaps be achieved for each individual data set. However, this might mean that overall we end up with 100 s of different model structures—for the same fundamental engineering problem modelled. This will not only be confusing for the analyst, but will raise issues of confidence in the

robustness of the models and modelling, as well as the interpretability of the models. Such issues have been also discussed by other researchers that sought to develop guidelines for the practitioners, see for example Van Gelder et al. (2014). Therefore, our strategy has been to focus on the development of a consistent framework for describing engineering modelling problem types, and to design a systematic experiment to evaluate a set of candidates that have already been proven as good modelling choices, leading to practitioner guidelines for choosing a metamodeling technique for their specific problem type.

Considering the results from the systematic evaluation of metamodels experiments, it was observed that:

- Under 0% noise condition, Kriging with Matérn 5/2 correlation function outperforms the other methods for high-order nonlinear problems; however, for low-order problems, Polynomials are providing competitive response models in terms of accuracy, and have the advantage that are easier to interpret;
- Problem scale has an insignificant effect on the performance of modelling techniques in terms of accuracy, and Kriging with both Matérn 5/2 and Matérn 3/2 correlation functions provide highly accurate response models;
- Increasing the number of sampling points usually results in an enhancement in the performance of modelling techniques, in terms of accuracy and robustness; however, this is also dependent on the noise condition;
- For both small and large DoE sample sizes, Kriging with Matérn 5/2 correlation function excels in terms of accuracy and robustness;
- For engineering problems under noise condition, Kriging modelling technique with Matérn 3/2 correlation function performs better than other techniques; however, its performance might deteriorate under very high-noise conditions (i.e. 15% noise or more), where using a Kriging model with nugget factor might provide better models;
- In terms of computation-efficiency, both Polynomials and RBFs take a trivial amount of time to construct the models, while the Kriging can be very time-consuming, in particular for large samples; however, from a practical engineering point of view, the cost of computation is perhaps less significant compared to the benefits gained from a better quality model.

These findings, based on the results from the evaluation against the 18 test problems, have supported the development of a set of modelling guidelines for engineers, which in essence can be summarised as follows: (i) for engineering problems with smooth-data (i.e. 0% noise), Kriging with Matérn 5/2 correlation function is the most dependable metamodeling method, and (ii) when experimental noise conditions are significant, Kriging with Matérn 3/2 correlation function provides a robust metamodeling choice. The capability of Kriging with the Matérn class of correlation functions is due to the flexibility and realistic smoothness assumptions in modelling many physical processes (Stein 2012). Within our study, the proposed guidelines have been validated with four engineering case studies, which showed that the suggested modelling technique based on the problem type confirmed the performance expectations.

The framework developed in this paper should provide a basis for further research on evaluation of metamodeling techniques. The structure of the experiment, including the balanced set of benchmark functions structured on problem types, provides a good testbed for further studies, which could include a broader set of metamodelling techniques as well as the effect of hyperparameters. The role of dynamic approaches for optimal metamodel selection, e.g. as discussed by Zhao et al. (2011), could also be evaluated within the context of this framework.

Most importantly, we see the impact of this paper onto the engineering practitioners dealing with modelling problems. The principle of the guidelines developed in this paper is to identify a robust metamodeling choice for a problem type. This encourages the engineer to start by analysing the problem in relation to its characteristics, firstly in relation to dimensionality and complexity (either nonlinearity, or in terms of interactions and importance of terms, which could be evaluated through smaller screening experiments), and then in relation to the sample size affordability and the level of noise affecting the variables and the experiment (which could be assessed with standard repeatability and reproducibility tests, which are common practice in engineering). This should not only provide a more robust basis for metamodeling, but also increase the confidence and ultimately the take-up of metamodeling in engineering practice.

7 Replication of results

The test functions are listed in the Appendix, and the hyperparameters of the metamodelling techniques implemented in the paper are discussed in Section 3; therefore, the results should be fully reproducible by other researchers. The implementation was in Matlab, and the scripts can be made available by request to interested researchers via the corresponding author.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Appendix

Table 5 Test functions

Test	Equation	Reference
1	$f(x) = sin(x_1 + x_2) + (x_1 - x_2)^2 - 1.5x_1 + 2.5x_2 + 1$	Jin et al. 2001
2	$f(x) = 5.357854x_2^2 + 0.835689xx_3 + 37.293239x_1 - 40792.141$	Jin et al. 2001
3	$f(x) = \sum_{i=1}^{4}\left[\left(\sum_{j=1}^{4}x_j^i\right) - b_i\right]^2$	Global Optimization Test Problems n.d.
4	$f(x) = -a\,exp\left(-b\sqrt{\frac{1}{2}\sum_{i=1}^{2}x_i^2}\right) - exp\left(\frac{1}{2}\sum_{i=1}^{2}cos(cx_i)\right) + a + exp(1)$	Bäck 1996
5	$f(x) = -\sum_{i=1}^{4}a_i\,exp\left(-\sum_{j=1}^{3}A_{ij}\left(x_j - P_{ij}\right)^2\right)$	Global Optimization Test Problems n.d.
6	$f(x) = \frac{1}{0.839}\left[1.1 - \sum_{i=1}^{4}a_i\,exp\left(-\sum_{j=1}^{4}A_{ij}\left(x_j - P_{ij}\right)^2\right)\right]$	Picheny et al. 2013b
7	$f(x) = \sum_{i=1}^{5}(x_i - 1)^2 - \sum_{i=2}^{5}x_i x_{i-1}$	Global Optimization Test Problems n.d.
8	$V_m(x) = \frac{(V_{b1} + 0.74)\beta(R_{c2} + 9)}{\beta(R_{c2} + 9) + R_f} + \frac{11.35R_f}{\beta(R_{c2} + 9) + R_f} + \frac{0.74R_f\beta(R_{c2} + 9)}{\left(\beta(R_{c2} + 9) + R_f\right)R_{c1}}$	Ben-Ari and Steinberg 2007; Moon et al. 2012
9	$f(x) = \frac{2piT_u(H_u - H_l)}{ln\left(\frac{r}{r_w}\right)\left(1 + \frac{2LT_u}{ln\left(\frac{r}{r_w}\right)r_w^2 K_w} + \frac{T_u}{T_l}\right)}$	Gramacy and Lian 2012; Xiong et al. 2013; Zhou et al. 2012
10	$f(x) = -\sum_{i=1}^{4}a_i\,\exp\left(-\sum_{j=1}^{6}A_{ij}\left(x_j - P_{ij}\right)^2\right)$	Picheny et al. 2013b
11	$C(x) = 2pi\sqrt{\frac{M}{k + S^2\frac{P_0 V_0 T_0}{T_0\ V^2}}}$	Ben-Ari and Steinberg 2007; Moon et al. 2012
12	$f(x) = \left(\sum_{i=1}^{4}L_i cos\left(\sum_{j=1}^{i}\theta_j\right) + \sum_{i=1}^{4}L_i sin\left(\sum_{j=1}^{i}\theta_j\right)\right)^{0.5}$	An and Owen 2001
13	$f(x) = x_1^2 + x_2^2 + x_1 x_2 - 14x_1 - 16x_2 + (x_3 - 10)^2 + 4(x_4 - 5)^2 + (x_5 - 3)^2$ $\quad + 2(x_6 - 1)^2 + 5x_7^2 + 7(x_8 - 11)^2 + 2(x_9 - 10)^2 + (x_{10} - 7)^2 + 45$	Jin et al. 2001
14	$f(x) = \sum_{i=1}^{16}\sum_{j=1}^{16}a_{ij}\left(x_i^2 + x_i + 1\right)\left(x_j^2 + x_j + 1\right)$	Jin et al. 2001
15	$f(x) = \frac{5x_{12}}{1 + x_1} + 5(x_4 - x_{20})^2 + x_5 + 40x_{19}^3 - 5x_{19} + 0.05x_2 + 0.08x_3 - 0.03x_6$ $\quad + 0.03x_7 - 0.09x_9 - 0.01x_{10} - 0.07x_{11} + 0.25x_{13}^2 - 0.04x_{14} + 0.06x_{15}$ $\quad - 0.01x_{17} - 0.03x_{18}$	Welch et al. 1992b
16	$f(x) = \sum_{i=1}^{10}x_i^2 + \left(\sum_{i=1}^{10}0.5ix_i\right)^2 + \left(\sum_{i=1}^{10}0.5ix_i\right)^4$	Global Optimization Test Problems n.d.
17	$f(x) = (x_1 - 1)^2 + \sum_{i=2}^{14}i\left(2x_i^2 - x_{i-1}\right)^2$	Global Optimization Test Problems n.d.
18	$f(x) = \sum_{i=1}^{17}\left[100\left(x_{i+1} - x_i^2\right)^2 + (x_i - 1)^2\right]$	Dixon 1978; Picheny et al. 2013b

References

An J, Owen A (2001) Quasi-regression. J Complex 17(4):588–607

Bäck T (1996) Evolutionary algorithms in theory and practice: evolution strategies, evolutionary programming, genetic algorithms. Oxford University Press

Ben-Ari EN and Steinberg DM (2007) Modeling Data from Computer Experiments: An Empirical Comparison of Kriging with MARS and Projection Pursuit Regression, Quality Engineering, 19:4, 327-338, https://doi.org/10.1080/08982110701580930

Berger B (2012) Modelling and optimisation for Stationary Base engine calibration, PhD thesis, TU Munich

Blondet G, Duigou J Le and Boudaoud N (2019) A knowledge-based system for numerical design of experiments processes in mechanical engineering. Expert Syst Appl 122. Pergamon: 289–302

Box GEP, Draper NR (1987) Empirical model-building and response surfaces. John Wiley & Sons, Oxford, England

Campean F, Ananta K, Yin XF, Grove D (2010) Evaluation of modelling techniques for engine response models, proc Int congress on automotive and transport engineering CONAT 2010, vol 2, pp 95–102 ISSN 2069-0428

Chen S, Chng ES and Alkadhimi K (1996) Regularized orthogonal least squares algorithm for constructing radial basis function networks. International journal of control 64(5). Taylor & Francis: 829–837

Chen VCP, Tsui K-L, Barton RR, et al. (2006) A review on design, modeling and applications of computer experiments. IIE transactions 38(4). Taylor & Francis Group: 273–291

Chen L, Qiu H, Jiang C, et al. (2018) Support vector enhanced kriging for metamodeling with noisy data. Struct Multidiscip Optim 57(4): 1611–1623

Chen, L., Wang, H., Ye, F., & Hu, W. (2019). Comparative study of HDMRs and other popular metamodeling techniques for high dimensional problems. Struct Multidiscip Optim 59(1):21–42

Couckuyt I, Forrester A, Gorissen D et al (2012) Blind kriging: implementation and performance analysis. Adv Eng Softw 49:1–13. https://doi.org/10.1016/j.advengsoft.2012.03.002

Cressie N (1990) The origins of kriging. Math Geol 22(3):239–252

Dey S, Mukhopadhyay T, Adhikari S (2015) Stochastic free vibration analysis of angle-ply composite plates-a RS-HDMR approach. Compos Struct 122:526–536

Didcock N, Rainer A and Jakubek S (2014) Optimization and optimal control in automotive systems. Waschl H, Kolmanovsky I, Steinbuch M, et al. (eds). Lecture Notes in Control and Information Sciences. Cham: Springer International Publishing

Dixon LCW (1978) The global optimization problem: An introduction. 2. Available at: https://www.researchgate.net/publication/247892177_The_global_optimization_problem_An_introduction (accessed 24 March 2016)

Fang H, Rais-Rohani M, Liu Z et al (2005) A comparative study of metamodeling methods for multiobjective crashworthiness optimization. Comput Struct 83(25–26):2121–2136

Fang K, Li R, Sudjianto A (2010) Design and modeling for computer experiments. CRC Press

Forrester A, Sobester DA, Keane A (2008) Engineering design via surrogate modelling: a practical guide. John Wiley & Sons

Forsberg J, Nilsson L (2004) On polynomial response surfaces and kriging for use in structural optimization of crashworthiness. Struct Multidiscip Optim 29(3):232–243

Giunta AA and Watson LT (1998) A Comparison of Approximation Modeling Techniques: Polynomial Versus Interpolating Models, 7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, AIAA 98–4758, Vol. 1, pp. 392–404. https://doi.org/10.2514/6.1998-4758

Global Optimization Test Problems (n.d.) Available at: http://www.mat.univie.ac.at/~neum/glopt/test.html (accessed 24 March 2016)

Gramacy RB and Lian H (2012) Gaussian process single-index models as emulators for computer experiments. Technometrics. Taylor & Francis Group 54(1):30–41

Hagan MT and Demuth HB (1999) Neural networks for control. In: Proceedings of the 1999 American control conference (cat. No. 99CH36251), pp. 1642–1656. IEEE

Hartmann B, Baumann W, Nelles O (2013) Axes-Oblique Partitioning of Local Model Networks for Engine Calibration. Design of Experiments (DoE) in Engine Development, Expert-Verlag GmbH: 92–106

He Y, Rutland CJ (2004) Application of artificial neural networks in engine modelling. Int J Engine Res 5(4):281–296

Howlett RJ, Zoysa MM, Walters SD, et al. (1999) Neural Network techniques for monitoring and control of internal combustion engines'. In: Int. Symposium on Intelligent Industrial Automation, Genova, Italy, 1999

Huang Z, Qiu H, Zhao M, Cai X, Gao L (2015) An adaptive SVR-HDMR model for approximating high dimensional problems. Eng Comput 32(3):643–667

Jin R, Chen W, Simpson TW (2001) Comparative studies of metamodelling techniques under multiple modelling criteria. Struct Multidiscip Optim 23(1):1–13

Jin R, Du X, Chen W (2003) The use of metamodeling techniques for optimization under uncertainty. Struct Multidiscip Optim 25(2):99–116

Kaymaz I (2005) Application of kriging method to structural reliability problems. Struct Saf 27(2):133–151

Kersaudy P, Sudret B, Varsier N et al (2015) A new surrogate modeling technique combining kriging and polynomial chaos expansions – application to uncertainty analysis in computational dosimetry. J Comput Phys 286:103–117

Khan MAZ (2011) Transient engine model for calibration using two-stage regression approach. PhD Thesis, Loughborough University. Available at: https://dspace.lboro.ac.uk/dspace-jspui/handle/2134/8456 (accessed 18 December 2013)

Kianifar MR, Campean F, Beattie T, et al. (2014) Analytical target cascading framework for diesel engine calibration optimisation. In: 13 October 2014

Kianifar MR, Campean F, Wood A (2015) Application of permutation genetic algorithm for sequential model building–model validation design of experiments. Soft Comput:1–22

Kim B-S, Lee Y-B, Choi D-H (2009) Comparison study on the accuracy of metamodeling technique for non-convex functions. J Mech Sci Technol 23(4):1175–1181

Kleijnen JPC (2009) Kriging metamodeling in simulation: a review. Eur J Oper Res 192(3):707–716

Kleijnen JPC, van Beers WCM (2005) Robustness of Kriging when interpolating in random simulation with heterogeneous variances: some experiments. Eur J Oper Res 165(3):826–834

Kroetz HM, Tessari RK and Beck AT (2017) Performance of global metamodeling techniques in solution of structural reliability problems. Adv Eng Softw 114. Elsevier: 394–404

Lach R, Weber F, Siebertz K, et al. (2007) DoE application within the base engine design. In: Design of Experiments in Engine Development (ed. K Roepke), 2007, pp. 249–260. Expert Verlag

Li YF, Ng SH, Xie M, Goh TN (2010) A systematic comparison of metamodeling techniques for simulation optimization in decision support systems. Appl Soft Comput 10(4):1257–1273

Liu H, Xu S and Wang X (2016) Sampling strategies and metamodeling techniques for engineering design: comparison and application. In: Volume 2C: Turbomachinery, 13 June 2016, p. V02CT45A019. ASME

Loeppky JL, Sacks J and Welch WJ (2009) Choosing the sample size of a computer experiment: a practical guide. Technometrics 51(4). Taylor & Francis: 366–376

Marrel A, Iooss B, Van Dorpe F et al (2008) An efficient methodology for modeling complex computer codes with Gaussian processes. Compu Statist Data Anal 52(10):4731–4744

Matheron G (1976) Advanced geostatistics in the mining industry. Guarascio M, David M, and Huijbregts C (eds). Dordrecht: Springer Netherlands

Meckesheimer M, Barton RR, Simpson T et al (2001) Metamodeling of combined discrete/continuous responses. AIAA J 39(10):1950–1959

Moon H, Dean AM and Santner TJ (2012) Two-stage sensitivity-based group screening in computer experiments. Technometrics. Taylor & Francis Group 54(4):376–387

Morris MD, Mitchell TJ (1995) Exploratory designs for computational experiments. J Stat Plan Inference 43(3):381–402

Moyeed RA, Papritz A (2002) An Empirical Comparison of Kriging Methods for Nonlinear Spatial Point prediction. Mathematical Geology 34(4):365–386. https://doi.org/10.1023/A:1015085810154

Mullur AA, Messac A (2005) Metamodeling using extended radial basis functions: a comparative approach. Eng Comput 21(3):203–217

Myers RH, Khuri AI and Carter WH (1989) Response surface methodology. Technometrics 31(2). Taylor & Francis: 137–157

de Oliveira MA, Possamai O, Dalla Valentina LVO et al (2013) Modeling the leadership – project performance relation: radial basis function, Gaussian and Kriging methods as alternatives to linear regression. Expert Syst Appl 40(1):272–280

Østergård T, Jensen RL and Maagaard SE (2018) A comparison of six metamodeling techniques applied to building performance simulations. Appl Energy 211. Elsevier: 89–103

Paiva R, Crawford C, Suleman A, et al. (2009) A comparison of surrogate models in the framework of an MDO tool for wing design. In: 50th AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics, and materials conference, Reston, Virigina, 2009. American Institute of Aeronautics and Astronautics

Pan F, Zhu P, Zhang Y (2010) Metamodel-based lightweight design of B-pillar with TWB structure via support vector regression. Comput Struct 88(1–2):36–44

Passos F, Gonzalez-Echevarria R, Roca E, et al. (2015) Surrogate modeling and optimization of inductor performances using Kriging functions. In: 2015 International Conference on Synthesis, Modeling, Analysis and Simulation Methods and Applications to Circuit Design (SMACD), 2015, pp. 1–4

Picheny V, Ginsbourger D, Roustant O, et al. (2010) Adaptive Designs of Experiments for Accurate Approximation of a Target Region. J Mech Des 132(7):9. https://doi.org/10.1115/1.4001873

Picheny V, Wagner T, Ginsbourger D (2013a) A benchmark of kriging-based infill criteria for noisy optimization. Struct Multidiscip Optim 48(3):607–626

Picheny V, Wagner T, Ginsbourger D (2013b) A benchmark of kriging-based infill criteria for noisy optimization. Struct Multidiscip Optim 48(3):607–626

Rango J, Schnorbus T, Kwee H et al (2013) Comparison of different approaches for global modeling of combustion engines. In: Design of Experiments (DoE) in engine development, vol 2013, pp 70–91

Sacks J, Welch WJ, Mitchell TJ etal (1989) Design and analysis of computer experiments (with discussion), Statistical Science, 4:409–435

Sacks J, Welch WJ, Mitchell TJ et al (1989) Design and analysis of computer experiments. Statistical science. Institute of Mathematical Statistics 4(4):409–423

Schlober A, Linssen R and Bozelie P (2007) Model-based calibration of SI-engines: map optimization. In: Design of Experiments in Engine Development (ed. K Roepke), 2007, pp. 145–164. Expert Verlag

Seabrook J, Salamon T, Edwards S et al (2003) A comparison of neural networks, stochastic process methods and radial basis functions for the optimisation of engine control parameters. Tagung, Haus der Technik, Design of Experiments in engine development. Berlin

Seabrook J, Collins J and Edwards S (2005) Application of advanced modelling techniques to the calibration of gasoline engines with direct injection and variable valve timing. In: Design of Experiments in Engine Development (ed. K Roepke), 2005, pp. 235–245. Expert Verlag

Shan S and Wang GG (2010) Survey of modeling and optimization strategies to solve high-dimensional design problems with computationally-expensive black-box functions. Structural and multidisciplinary optimization 41(2). Springer-Verlag: 219–241

Simpson TW, Korte JJ, Mauery TM, et al. (1998) Comparison of response surface and kriging models for multidisciplinary design optimization. In: 7th AIAA/USAF/ NASA/ISSMO symposium on multidisciplinary analysis & optimization

Simpson TW, Mauery TM, Korte JJ et al (2001a) Kriging models for global approximation in simulation-based multidisciplinary design optimization. AIAA J 39(12):2233–2241

Simpson TW, Poplinski JD, Koch PN et al (2001b) Metamodels for computer-based engineering design: survey and recommendations. Eng Comput 17(2):129–150

Stein ML (2012) Interpolation of spatial data: some theory for kriging. Springer Science & Business Media, New York

Toal J, DJ, Keane AJ (2013) Performance of an ensemble of ordinary, universal, non-stationary and limit Kriging predictors. Struct Multidiscip Optim 47(6):893–903

Ulaganathan S, Couckuyt I, Ferranti F et al (2014) Performance study of multi-fidelity gradient enhanced kriging. Struct Multidiscip Optim 51(5):1017–1033

Ulaganathan S, Couckuyt I, Dhaene T et al (2015) Performance study of gradient-enhanced Kriging. Eng Comput 32(1):15–34

Van Gelder L, Das P, Janssen H, et al. (2014) Comparative study of metamodelling techniques in building energy simulation: guidelines for practitioners. Simul Mod Prac and Theory 49:245–257, https://doi.org/10.1016/j.simpat.2014.10.004

Varadarajan S, Chen W, Pelka CJ (2000) Robust concept exploration of propulsion systems with enhanced model approximation capabilities. Engineering Optimization 32(3). Taylor & Francis. Group:309–334

Vicente-Serrano S, Saz-Sánchez M, Cuadrat J (2003) Comparative analysis of interpolation methods in the middle Ebro Valley (Spain): application to annual precipitation and temperature. Clim Res 24:161–180

Volkova E, Iooss B, Van Dorpe F (2006) Global sensitivity analysis for a numerical model of radionuclide migration from the RRC "Kurchatov Institute" radwaste disposal site. Stoch Env Res Risk A 22(1):17–31

Wang G and Shan S (2005) Review of metamodeling techniques for product design with computation-intensive processes. In: Proceedings of the Canadian Engineering Education Association

Wang H, Li E, Li GY (2009) The least square support vector regression coupled with parallel sampling scheme metamodeling technique and application in sheet forming optimization. Mater Des 30(5):1468–1479

Wang H, Shan S, Wang GG, Li G (2011) Integrating least square support vector regression and mode pursuing sampling optimization for crashworthiness design. J Mech Des 133(4)

Welch WJ, Buck RJ, Sacks J, et al. (1992a) Screening, predicting, and computer experiments. Technometrics 34(1). Taylor & Francis roup

Welch WJ, Buck RJ, Sacks J, et al. (1992b) Screening, predicting, and computer experiments. Technometrics 34(1). Taylor & Francis Group: 15–25

Xiong S, Qian PZG and Wu CFJ (2013) Sequential design and analysis of high-accuracy and low-accuracy computer codes. Technometrics. Taylor & Francis Group 55(1):37–46

Yang RJ, Gu L, et al. (2000) Approximations for safety optimization of large systems. In: Proceeding of the 2000 ASME design engineering technical conferences and computers and information in engineering conference, Baltimore, Maryland, 2000, pp. 763–772

Zhao D, Xue D (2010) A comparative study of metamodeling methods considering sample quality merits. Struct Multidiscip Optim 42(6):923–938

Zhao Y, Rosala GF, Campean IF, et al. (2010) A response surface approach to front-car optimisation for minimising pedestrian head injury levels. International journal of crashworthiness. Taylor & Francis Group 15(2):143–150

Zhao L, Choi KK, Lee I (2011) Metamodeling method using dynamic kriging for design optimization. AIAA J 49(9):2034–2046. https://doi.org/10.2514/1.J0551017

Zhou Q, Qian PZG and Zhou S (2012) A simple approach to emulation for computer models with qualitative and quantitative factors. Technometrics. Taylor & Francis

Zhu P, Zhang Y, Chen G-L (2009) Metamodel-based lightweight design of an automotive front-body structure using robust optimization. Proc Inst Mech Eng, Part D: Journal of Automobile Engineering 223(9):1133–1147, https://doi.org/10.1243/09544070JAUTO1045