Applications of Multi-fidelity Multi-Output Kriging to Engineering Design Optimization

David J.J. Toal[1]

[1]Faculty of Engineering & Physical Sciences, University of Southampton, Boldrewood Innovation Campus, Southampton, SO16 7QF, Hampshire, U.K..

Contributing authors: djjt@soton.ac.uk;

Abstract

Surrogate modelling is a popular approach for reducing the number of high fidelity simulations required within an engineering design optimization. Multi-fidelity surrogate modelling can further reduce this effort by exploiting low fidelity simulation data. Multi-output surrogate modelling techniques offer a way for categorical variables e.g. the choice of material, to be included within such models. While multi-fidelity multi-output surrogate modelling strategies have been proposed, to date only their predictive performance rather than optimization performance has been assessed. This paper considers three different multi-fidelity multi-output Kriging based surrogate modelling approaches and compares them to ordinary Kriging and multi-fidelity Kriging. The first approach modifies multi-fidelity Kriging to include multiple outputs whereas the second and third approaches model the different levels of simulation fidelity as different outputs within a multi-output Kriging model. Each of these techniques is assessed using three engineering design problems including the optimization of a gas turbine combustor in the presence of a topological variation, the optimization of a vibrating truss where the material can vary and finally, the parallel optimization of a family of airfoils.

Keywords: Multi-output, Multi-fidelity, Kriging

1 Introduction

Engineering design optimization typically involves the use of expensive computational simulations in order to accurately predict the performance of a design. The expense of these simulations, which can be of the order of days even with modern computing hardware resources, makes their direct application within an optimization algorithm completely infeasible. Instead, it has become the norm within the design optimization community to employ some sort of surrogate modelling-based approach whereby a surrogate (otherwise known as a metamodel or response surface) is constructed using a relatively small amount of simulation data and then used to guide a subsequent design search.

This process of surrogate construction, followed by sequential updating and improvement in regions of interest, was popularized by authors such as Jones et al (1998) almost 25 years ago. Since then, whilst the general philosophy has changed little, there have been numerous advances in terms of the surrogate model formulation (Stork et al, 2020; Alizadeh et al, 2020), sampling plans (Garud et al, 2017; Alizadeh et al, 2020), updating and infilling strategies (Haftka et al, 2016; Liu et al, 2017; Zhan and Xing, 2020) to name but a few. While the application of surrogate based approaches to robust design (Chatterjee et al,

2017), multi-objective (Tabatabaei et al, 2015; Deb et al, 2020) as well as traditional single objective optimization problems has grown throughout engineering.

Most surrogate-based optimization applications within the literature tend to focus on problems where the design variables are continuous in nature. While this makes the problem tractable for most surrogate modelling techniques it is perhaps somewhat removed from the realities of real engineering design where designers are often faced with discrete choices as part of the design process, for example, the choice of optimal feature or material. As discussed by Allmendinger et al (2017) there is a considerable business case in favour of being able to efficiently tackle such problems within industry.

A potential solution to this problem comes in the form of surrogate modelling strategies for the prediction of multiple outputs (also referred to as multi-task prediction/learning (Bonilla et al, 2007), categorical modelling or multi-variate modelling (Kleijnen and Mehdad, 2014) within the literature). However, these methods appear to have been applied to relatively few engineering design optimization problems. Notable exceptions include the design optimization of metamaterials (Tran et al, 2019), multi-scale topology optimization (Wang et al, 2020), sounding rocket engine (Pelamatti et al, 2018) and centrifugal pump blade optimization (Zhang and Zhao, 2020). Beyond design optimization, the predictive performance advantages of multi-output models have been demonstrated on a number of occasions (Liu et al, 2018b, 2014; Wang and Chen, 2015). It is worth noting that multi-output Kriging models have also been recently extended to include gradient information (Quan et al, 2022).

The performance of a surrogate model-based optimization strategy is largely a function of the predictive accuracy of the underlying surrogate used. Generally speaking, the more a surrogate model is representative of the trends in the true response the greater the rate of convergence towards an optimal design. That is to say, the surrogate model does not have to be accurate everywhere in the design space but in order for the surrogate to be useful, from an optimization point of view, it should help guide our searches effectively towards an optimum.

In engineering design optimization applications, where the data is generated from expensive simulations or experiments, often it is not feasible to construct a surrogate model from a large amount of data. Multi-fidelity surrogate modelling methods offer an alternative to this by enabling multiple sources of data to be exploited within a single model. These methods enable inexpensive but large low fidelity data sets to provide an accurate global trend throughout the design space which is then corrected by a much smaller subset of high fidelity data. Recent advances in this field include, for example, the use of radial basis functions (Serani et al, 2019; Eddine et al, 2022), support vector regression (Shi et al, 2020), neural networks (Li et al, 2022), conditional generative adversarial networks (Zhang et al, 2022), graph neural networks (Black and Najafi, 2022) and deep neural networks (Li et al, 2022). For the interested reader the article by Peherstorfer et al (2018) provides a useful review of multi-fidelity surrogate modelling approaches for optimization and uncertainty quantification.

The literature contains numerous examples of multi-fidelity surrogate modelling applied to the design optimization of engineering systems demonstrating both an improvement in surrogate model predictive, and subsequent optimization, performance. Examples, focusing in particular on the application of Kriging or Gaussian process models, include the design of airfoils (Liu et al, 2018a,b; Phiboon et al, 2021), gas turbine compressor blades (Brooks et al, 2011), combustion systems (Zhang et al, 2015; Toal et al, 2021), engine structures (Yong et al, 2019), composite structures (Guo et al, 2020) and hull forms (Liu et al, 2022).

In summary, multi-output surrogate models enable designers to model problems with integer or categorical variables. The correlation of responses across outputs, in such an approach, can produce a more accurate surrogate model than if a single independent surrogate were constructed for each output. This assumes, of course, that the outputs are indeed correlated. Multi-fidelity surrogate models, by fusing different sources of data together, can potentially reduce the amount of high fidelity data required, improve surrogate model accuracy and accelerate convergence towards an optimum.

Given the advantages of multi-output and multi-fidelity surrogate models, it makes perfect sense to attempt to combine these into a single modelling approach. To date, the work of Lin et al (2021) represents the only serious attempt to achieve this. Here the autoregressive model of Kennedy and O'Hagan (2000) is combined with a correlation matrix parameterized using the hypersphere approach of Zhou et al (2011) to model the interactions between outputs. However, while Lin et al (2021) demonstrated the predictive performance advantage of such a model for airfoil aerodynamic coefficient prediction, the optimization performance of this approach was never assessed. One could also argue that their choice of engineering problem (where both lift and drag come from the same simulation) resulted in an isotropic dataset which has been demonstrated to offer poorer performance on multi-output prediction problems than a heterotopic dataset (Liu et al, 2018b; Lin et al, 2021). Since this publication, Nasti et al (2022) have applied the modelling approach of Lin et al (2021) to the design optimization of the secondary air seals within an aero engine. However, this study only compared the resulting optimization performance to that of a genetic algorithm and not alternative surrogate modelling strategies.

This paper aims to address a number of the observed limitations from the literature with a focus on Kriging based surrogate modelling approaches. The multi-fidelity multi-output Kriging approach of Lin et al (2021) is compared to competing surrogate modelling strategies with respect to their optimization performance. This includes a comparison to standard single and multi-fidelity, single-output modelling approaches. But also assesses an alternative approach to embedding multi-fidelity data within a multi-output model by including it as an additional output of the model. While this approach has been suggested within the literature (Kleijnen and Mehdad, 2014) it is yet to be demonstrated on multi-fidelity problems with discrete design variables.

Optimization of the hyperparameters is often an extremely important, and often overlooked, step in the creation of any surrogate model. To date, little emphasis within the literature has been given to the solution of this problem for multi-output Kriging models in general, whether they are single or multi-fidelity in nature. This paper goes some way to addressing this by deriving an analytical expression for the adjoint of the likelihood of both single and multi-fidelity multi-output Kriging models which can be used to accelerate hyperparameter optimization.

The paper begins by presenting the mathematics behind the construction of each of the Kriging based surrogate models employed within the remainder of the paper. This includes details of ordinary single output Kriging (Section 2.1), its extension to multi-output Kriging (Section 2.2), multi-fidelity Kriging (Section 2.3) and its extension to multi-fidelity multi-output Kriging (Section 2.4). The formulation of the adjoint of the likelihood function with respect to the hyperparameters is presented for each model.

With the model formulations defined the paper then proceeds to explore the potential of single and multi-fidelity multi-output Kriging models through a combination of analytical and engineering test problems. In Section 3 the one-dimensional Forrester function is used to illustrate graphically the predictive power of these methods before progressing to the engineering problems.

Each of the three engineering design optimization problems is novel in its formulation and showcases the benefits of a multi-output approach to multi-fidelity design optimization. Section 4 considers a multi-point airfoil design optimization problem where four airfoils from the same family are optimized simultaneously. Section 5 considers the multi-fidelity optimization of a vibrating truss where the choice of material can vary. Finally, Section 6 considers the optimization of a gas turbine combustor sub-system where a multi-fidelity multi-output approach is used to optimize the positions of two competing topologies of dilution port in parallel.

Section 7 concludes the paper with a summary of the results from each study and a discussion on the application of these surrogate modelling techniques to future engineering problems.

2 Surrogate Model Formulations

The single and multi-fidelity multi-output models employed within the current paper have their origins in the popular surrogate modelling approach

known as Kriging. The following section captures the mathematics of these models commencing from the basics of ordinary Kriging.

2.1 Ordinary Kriging

A Kriging model is based on the assumption that points within a design space close to each-other will have a similar function value. Given a d dimensional hypercube, it is assumed that the correlation between points is given by,

$$\boldsymbol{R}_{ij} = \exp\left(-\sum_{l=1}^{d} 10^{\boldsymbol{\theta}^{(l)}} \|\boldsymbol{x}_i^{(l)} - \boldsymbol{x}_j^{(l)}\|^{\boldsymbol{p}^{(l)}}\right) \quad (1)$$

where θ defines the rate of correlation decrease and p defines the 'smoothness' of the model in each dimension. These parameters, along with a regularization term, 10^λ, if applied to the diagonal of the correlation matrix (Forrester et al, 2006), can be determined by optimizing the concentrated log-likelihood function (Jones et al, 1998),

$$\phi = -\frac{n}{2}\ln(\hat{\sigma}^2) - \frac{1}{2}\ln(|\boldsymbol{R}|). \quad (2)$$

Here the variance is given by,

$$\hat{\sigma}^2 = \frac{1}{n}(\boldsymbol{y} - \boldsymbol{1}\hat{\mu})^T \boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{1}\hat{\mu}), \quad (3)$$

and the mean of the observations by,

$$\hat{\mu} = \frac{\boldsymbol{1}^T \boldsymbol{R}^{-1}\boldsymbol{y}}{\boldsymbol{1}^T \boldsymbol{R}^{-1}\boldsymbol{1}}, \quad (4)$$

where $\boldsymbol{1}$ is a vector of n ones. n is equal to the number of data points contained within the sampling plan upon which the model is constructed. With the hyperparameters determined, the predictor,

$$y(\boldsymbol{x}^*) = \hat{\mu} + \boldsymbol{r}^T \boldsymbol{R}^{-1}(\boldsymbol{y} - \boldsymbol{1}\hat{\mu}), \quad (5)$$

provides the output of the model at an unsampled location.

The optimization of the hyperparameters is an important step in the construction of any Kriging model and has implications on the accuracy of the resulting model. Unfortunately due to the $O(n^3)$ cost of inverting the correlation matrix $\boldsymbol{R}$ this can be rather expensive if the size of the training set is large. To reduce this burden and accelerate the optimization process Toal et al (2011) presented an adjoint formulation of the likelihood function which can efficiently calculate the derivatives of the likelihood function with respect to each of the hyperparameters. Using the collection of adjoints provided by Giles (2008), it can be shown that the adjoint of the correlation matrix, $\boldsymbol{R}$, is given by,

$$\bar{\boldsymbol{R}} = \frac{1}{2\hat{\sigma}^2}\boldsymbol{R}^{-T}(\boldsymbol{y}-\boldsymbol{1}\hat{\mu})^T(\boldsymbol{y}-\boldsymbol{1}\hat{\mu})^T\boldsymbol{R}^{-T} - \frac{1}{2}\boldsymbol{R}^{-T}. \quad (6)$$

The derivatives of the log-likelihood function with respect to θ are given by,

$$\frac{\partial\phi}{\partial\theta_l} = \ln 10 \sum_{ij} -10^{\theta_l} \|\boldsymbol{x}_{i_l} - \boldsymbol{x}_{j_l}\|^{p_l} \boldsymbol{R}_{ij}\bar{\boldsymbol{R}}_{ij} \quad (7)$$

and with respect to p by,

$$\frac{\partial\phi}{\partial p_l} = \sum_{ij} -10^{\theta_l} \|\boldsymbol{x}_{i_l} - \boldsymbol{x}_{j_l}\|^{p_l} \ln \|\boldsymbol{x}_{i_l} - \boldsymbol{x}_{j_l}\| \boldsymbol{R}_{ij}\bar{\boldsymbol{R}}_{ij}. \quad (8)$$

If a regularization term has been included to account for noise within the training data its derivative can also be found,

$$\frac{\partial\phi}{\partial\lambda} = 10^\lambda \ln 10 \sum_i \bar{\boldsymbol{R}}_{ii}. \quad (9)$$

In order to solve the hyperparameter optimization problem the hybridized particle swarm algorithm of Toal et al (2011) is employed. A modified version of the particle swarm developed by Eberhart and Kennedy (1995), this search algorithm employs a particle initialization based on an optimal Latin hypercube with a rank-based selection scheme used to select a particle in each generation for local improvement via sequential quadratic programming (SQP). In addition to this diversity in the swarm is preserved via re-initialization process which aims to explore regions of the space not previously searched through a maxmin distance criterion. A terminal local SQP search is applied to the global optimum indicated by the swarm.

This optimization algorithm along with all of the surrogate models presented within this paper have been implemented within the proprietary Rolls-Royce optimization toolkit, CM02, formally known as OPTIMATv2. In all cases, the particle swarm used in the hyperparameter optimization employs a population size of 50 points for 100 generations.

2.2 Multi-Output Kriging

In order to extend ordinary Kriging to predict multiple outputs we follow the approach of Bonilla et al (2007) and recently Lin et al (2021). Bonilla et al. proposed that a single Gaussian process model could be constructed which would be capable of modelling multiple tasks or outputs. They achieved this through the construction of a 'free-form' correlation matrix which would 'learn' the correlations between the multiple tasks within the model. Furthermore, they drew inspiration from Cholesky decomposition in the parameterization of this correlation matrix to ensure that is was positive semi-definite. Lin et al. took the approach of Bonilla et al. a step further and introduced this 'free-form' correlation matrix approach into the multi-fidelity framework of Kennedy and O'Hagan (2000) to enable multiple levels of fidelity to enhance the prediction of multiple tasks or outputs.

In the following notation it is assumed that in addition to the training data $\boldsymbol{X}$ comprising of n training points within a d dimensional hypercube, there exists a vector of categorical values, $\boldsymbol{q}$. The vector, $\boldsymbol{q}$, contains n entries where each entry corresponds to the location of a training point, $\boldsymbol{x}$. Each entry within $\boldsymbol{q}$ can have one of m different levels. As will be discussed later there is no reason why the model cannot have multiple categories each with a differing number of levels but for simplicity, it is assumed for now that the model has only one categorical variable with m different levels.

The correlation between each level is defined by a second correlation matrix, $\boldsymbol{K}$, which is $m \times m$ in size and should be unit diagonal, symmetrical and positive definite. This matrix is then used to modify the Kriging correlation matrix, $\boldsymbol{R}$, defined by equation 1, such that the i, j^{th} entry of the updated correlation matrix, $\boldsymbol{R}_K$, is given by,

$$\boldsymbol{R}_{K_{ij}} = \boldsymbol{K}_{q_i, q_j} \boldsymbol{R}_{ij}. \tag{10}$$

In simple terms, each entry in the original correlation matrix, $\boldsymbol{R}$, is multiplied by an entry in the correlation matrix, $\boldsymbol{K}$ where the index of the entry in $\boldsymbol{K}$ depends on the categorical level indicated by the corresponding pair of values in the vector, $\boldsymbol{q}$.

To aid with the definition of the adjoint of the likelihood function we define an additional piece of notation, a subscript n, to represent the mapping from an $m \times m$ matrix to an $n \times n$ matrix using the indices of the categorical variables stored within $\boldsymbol{q}$. In this notation, the i, j^{th} entry of the matrix $\boldsymbol{K}_n$ is given by,

$$\boldsymbol{K}_{n_{ij}} = \boldsymbol{K}_{q_i, q_j} \tag{11}$$

where $\boldsymbol{R}_K$ can now be written as,

$$\boldsymbol{R}_K = \boldsymbol{K}_n \boldsymbol{R}. \tag{12}$$

In order for $\boldsymbol{R}_K$ to be symmetric positive definite, $\boldsymbol{K}$ must also be symmetric and positive definite. This requirement on $\boldsymbol{K}$ is difficult to achieve with a direct parameterization. Instead the approach of Zhou et al (2011) is used whereby the matrix $\boldsymbol{L}$, which is the lower triangular matrix resulting from the Cholesky decomposition of $\boldsymbol{K}$,

$$\boldsymbol{K} = \boldsymbol{L}\boldsymbol{L}^T, \tag{13}$$

is parameterized using a spherical coordinate system. If each entry of $\boldsymbol{L}$ is defined as $l_{r,s}$ then letting $l_{1,1} = 1$, for $r = 2, \ldots, m$ the below spherical coordinate system (Zhou et al, 2011) is used,

$$
\begin{aligned}
l_{r,1} &= \cos(\alpha_{r,1}) \\
l_{r,s} &= \cos(\alpha_{r,s}) \prod_{i=1}^{s-1} \sin(\alpha_{r,i}) \\
&\quad \text{for} \quad s = 2, \ldots, r-1 \\
l_{r,r} &= \prod_{i=1}^{r-1} sin(\alpha_{r,i}),
\end{aligned}
\tag{14}
$$

where $\alpha_{r,s} \in (0, \pi)$.

For a model with m levels this parameterization results in $m(m-1)/2$ additional hyperparameters to be determined in addition to those of the original ordinary Kriging model. The number of parameters therefore increases quite rapidly with the number of levels in the model.

Optimization of the hyperparameters is performed in an identical manner to that of an ordinary Kriging model. The likelihood function defined in equation 2 is again minimized but with $\boldsymbol{R}$ now replaced by $\boldsymbol{R}_K$. With the additional hyperparameters the presence of an efficient derivative calculation is of even greater importance.

The adjoint of the likelihood with respect to each of the hyperparameters of the spherical coordinate system can be derived by extending the approach outlined in the previous section. As noted previously, the formulation of the likelihood function is identical for both single and multi-output Kriging with $\boldsymbol{R}_K$ replacing $\boldsymbol{R}$. It therefore follows that the adjoint of the likelihood function with respect to $\boldsymbol{R}_K$ is given by,

$$\bar{\boldsymbol{R}}_K = \frac{1}{2\hat{\sigma}^2}\boldsymbol{R}_K^{-T}(\boldsymbol{y}-\boldsymbol{1}\hat{\mu})^T(\boldsymbol{y}-\boldsymbol{1}\hat{\mu})^T\boldsymbol{R}_K^{-T} - \frac{1}{2}\boldsymbol{R}_K^{-T} \tag{15}$$

where $\boldsymbol{R}_K$ has directly replaced $\boldsymbol{R}$, but this time in Equation 6.

From this point the adjoint of the likelihood with respect to both $\boldsymbol{R}$ and $\boldsymbol{K}_m$ can be found to be,

$$\bar{\boldsymbol{R}} = \bar{\boldsymbol{R}}_K \boldsymbol{K}_n \tag{16}$$

and

$$\bar{\boldsymbol{K}}_n = \bar{\boldsymbol{R}}_K \boldsymbol{R}. \tag{17}$$

With $\bar{\boldsymbol{R}}$ calculated, the derivatives of the likelihood with respect to the hyperparameters θ, p and λ can be found as before via Equations 7, 8 and 9 respectively.

The derivative of the likelihood with respect to a hypersphere parameter α, is given by,

$$\frac{\partial\phi}{\partial\alpha} = \sum_{ij}\bar{\boldsymbol{K}}_n\dot{\boldsymbol{K}}_n \tag{18}$$

where $\dot{\boldsymbol{K}}_n$ is an $n\times n$ matrix where each element represents the derivative of the corresponding element of $\boldsymbol{K}_n$ with respect to the α hypersphere parameter of interest. This is effectively an expansion of the $m\times m$ matrix of derivatives $\dot{\boldsymbol{K}}$ to a $n\times n$ representation using the vector of categorical indices, $\boldsymbol{q}$. As $\boldsymbol{K} = \boldsymbol{L}\boldsymbol{L}^T$ then,

$$\dot{\boldsymbol{K}} = \dot{\boldsymbol{L}}\boldsymbol{L}^T + \boldsymbol{L}\dot{\boldsymbol{L}}^T, \tag{19}$$

with $\dot{\boldsymbol{L}}$ representing the derivative of the matrix $\boldsymbol{L}$ with respect to α.

The definition of $\boldsymbol{L}$, as a function of each α, is outlined in Equation 14. $\dot{\boldsymbol{L}}$, like $\boldsymbol{L}$, is lower diagonal in nature and as $l_{1,1}$ is a constant $\dot{l}_{1,1}=0$. For the remaining components,

$$\frac{\partial l_{r,1}}{\partial\alpha_{r,i}} = -\sin(\alpha_{r,i}), \tag{20}$$

when $i=1$ and 0 when $i\neq 1$. For the diagonal terms,

$$\frac{\partial l_{r,r}}{\partial\alpha_{r,i}} = \frac{l_{r,r}\cos(\alpha_{r,i})}{\sin(\alpha_{r,i})} \tag{21}$$

while for the remaining terms,

$$\frac{\partial l_{r,s}}{\partial\alpha_{r,i}} = \frac{l_{r,s}\cos(\alpha_{r,i})}{\sin(\alpha_{r,i})} \tag{22}$$

for $i=1,\ldots,s-1$ and finally,

$$\frac{\partial l_{r,s}}{\partial\alpha_{r,s}} = -\frac{l_{r,s}\sin(\alpha_{r,s})}{\cos(\alpha_{r,s})}. \tag{23}$$

With the above adjoint formulation providing a relatively efficient calculation of the derivatives of the likelihood function, the hybridized particle swarm algorithm of Toal et al (2011) can again be used to perform the optimization.

With the hyperparameters determined, the predictor given in equation 5 can be modified to predict any of the m levels,

$$y(\boldsymbol{x}^*) = \hat{\mu} + \boldsymbol{r}_K^T\boldsymbol{R}_K^{-1}(\boldsymbol{y}-\boldsymbol{1}\hat{\mu}). \tag{24}$$

Here the vector of correlations between the new point $\boldsymbol{x}^*$ and the existing points, $\boldsymbol{r}_K$, has also been modified by the correlation matrix $\boldsymbol{K}$ where the correlation used is dependent on the categorical level of $\boldsymbol{x}^*$ and that of the training data.

The matrix, $\boldsymbol{K}$ describing correlations between levels in a multi-output Kriging model enables one level to influence another. This is quite different to a multi-fidelity Kriging model (described in the following section) where the inexpensive data influences the expensive prediction but not vice versa. This mutual influence between levels leads to some important considerations regarding the definition of the sampling plan used in the construction of the model. Employing an identical sampling plan for each level (an isotropic dataset) has been demonstrated to result in a poorer performing multi-output model as the correlation between outputs adds no useful information to the prediction of each output (Liu et al, 2018b; Lin et al, 2021). A heterotopic approach (where the sampling plans differ between levels) allows the predictor to benefit from the potential correlation between levels.

The above multi-output Kriging formulation makes the assumption of a common set of θ and

p hyperparameters for the d input variables. For the case when the correlation between each of the levels is high this makes sense. A high level of correlation implies a similar level of activity and smoothness in each response and therefore a similar optimal set of θ and p hyperparameters. However, when the correlation between levels is poor, the optimization of the hyperparameters naturally leads to a compromise in the parameter values and therefore a set of parameters which is sub-optimal for either level. This results in a reduction in predictive performance (Kleijnen and Mehdad, 2014). This is similar to the construction of a Kriging model of a non-stationary response where a single value of each hyperparameter is inappropriate at all locations in the space (Toal and Keane, 2012). This effect can also be observed when the amount of data from which to construct the model is sufficiently large. In this case, separate ordinary Kriging models for each output have better predictive performance than a single multi-output model combining all of the data (Liu et al, 2018b).

As noted above the number of hyperparameters can increase relatively quickly with an increasing number of outputs. This is the case if a categorical variable has many different levels but also if multiple categorical variables are to be captured within a single matrix. A problem with two categorical variables each with two levels, for example, results in four possible combinations of these variables and therefore a total of six additional hyperparameters to optimize. For such a problem the number of hyperparameters can be reduced significantly by assigning each category to a separate correlation matrix, $\boldsymbol{K}$. Instead of a single 4×4 correlation matrix in the above example, one would employ two separate 2×2 matrices resulting in two additional hyperparameters instead of six. However, this will naturally come at the cost of some flexibility within the model.

2.3 Multi-fidelity Kriging

Ordinary Kriging can be extended into multi-fidelity Kriging using the autoregressive approach of Kennedy and O'Hagan (2000). Here it is assumed that the predicted output of an expensive simulation is equal to the scaled prediction of an inexpensive simulation plus a prediction of the difference between the two,

$$Z_e(\boldsymbol{x}) = \rho Z_c(\boldsymbol{x}) + Z_d(\boldsymbol{x}). \qquad (25)$$

Here ρ is a scaling parameter determined during the subsequent hyperparameter optimization. The sub-scripts c, e and d are used to denote, inexpensive, expensive and the difference between inexpensive and expensive respectively.

The correlation matrix for such a model is defined as,

$$\boldsymbol{C} = \begin{pmatrix} \sigma_c^2 \boldsymbol{R}_c(\boldsymbol{X}_c, \boldsymbol{X}_c) & \rho \sigma_c^2 \boldsymbol{R}_c(\boldsymbol{X}_c, \boldsymbol{X}_e) \\ \rho \sigma_c^2 \boldsymbol{R}_c(\boldsymbol{X}_e, \boldsymbol{X}_c) & \rho^2 \sigma_c^2 \boldsymbol{R}_c(\boldsymbol{X}_e, \boldsymbol{X}_e) + \sigma_d^2 \boldsymbol{R}_d(\boldsymbol{X}_e, \boldsymbol{X}_e) \end{pmatrix} \quad (26)$$

where $\boldsymbol{R}_c(\boldsymbol{X}_c, \boldsymbol{X}_c)$ defines the correlation matrix between the known inexpensive sample points and $\boldsymbol{R}_c(\boldsymbol{X}_c, \boldsymbol{X}_e)$ defines the correlation between expensive and inexpensive sample points both using the hyperparameters from the Kriging model of the inexpensive data. $\boldsymbol{R}_d(\boldsymbol{X}_e, \boldsymbol{X}_e)$ however, defines the correlation matrix between the expensive sample points using the hyperparameters obtained from the difference model. In all cases the correlation defined by equation 1 is assumed.

The construction of such a multi-fidelity model is a two step process. Firstly, an ordinary Kriging model of only the inexpensive data is constructed with the hyperparameters optimized in the usual manner, via a maximization of the likelihood function. A second ordinary Kriging model is then constructed to predict the difference between the inexpensive and expensive data. To create this a vector of differences, $\boldsymbol{d}$, is calculated where,

$$\boldsymbol{d} = \boldsymbol{y}_e - \rho \boldsymbol{y}_c(\boldsymbol{X}_e). \qquad (27)$$

The construction of this difference model is therefore identical to that of an ordinary Kriging model except that the hyperparameter, ρ, must also be optimized along with the usual θ and p parameters. The adjoint formulations presented in equations 7 and 8 can be used again to accelerate this optimization process while the derivative of the likelihood with respect to ρ can be calculated as,

$$\frac{\partial \phi}{\partial \rho} = -\sum_{i=1}^{n_e} \boldsymbol{y}_{c_i} \bar{\boldsymbol{d}}_i, \qquad (28)$$

where the adjoint of the vector of differences is given by (Toal and Keane, 2011),

$$\bar{\boldsymbol{d}} = -\boldsymbol{R}_d^{-1}\left(\boldsymbol{d} - \boldsymbol{1}\hat{\mu}_d\right)\left(\frac{1}{\hat{\sigma}_d^2}\right). \qquad (29)$$

Here σ_d^2 refers to the variance in the difference model calculated using equation 3 where $\boldsymbol{y}$, $\boldsymbol{R}$ and $\hat{\mu}$ have been replaced by $\boldsymbol{d}$, $\boldsymbol{R}_d$ and $\hat{\mu}_d$ respectively.

With the hyperparameters determined predictions of the expensive function response at an unknown point, $\boldsymbol{x}^*$ are given by,

$$y_e(\boldsymbol{x}^*) = \hat{\mu} + \boldsymbol{c}^T\boldsymbol{C}^{-1}(\boldsymbol{y} - \boldsymbol{1}\hat{\mu}), \qquad (30)$$

with the mean given by,

$$\hat{\mu} = \frac{\boldsymbol{1}^T\boldsymbol{C}^{-1}\boldsymbol{Y}}{\boldsymbol{1}^T\boldsymbol{C}^{-1}\boldsymbol{1}}, \qquad (31)$$

and where the vector of correlations is,

$$\boldsymbol{c} = \begin{bmatrix} \rho\hat{\sigma}_c^2\boldsymbol{R}_c(\boldsymbol{X}_c, \boldsymbol{x}^*) \\ \rho^2\hat{\sigma}_c^2\boldsymbol{R}_c(\boldsymbol{X}_e, \boldsymbol{x}^*) + \hat{\sigma}_d^2\boldsymbol{R}_d(\boldsymbol{X}_e, \boldsymbol{x}^*) \end{bmatrix}. \qquad (32)$$

$\boldsymbol{Y}$ is a single vector combining both inexpensive and expensive observations, i.e. $[\boldsymbol{y}_c, \boldsymbol{y}_e]$.

2.4 Multi-fidelity Multi-Output Kriging

Extending a multi-fidelity Kriging model to predict multiple outputs is a relatively simple process with the correlation matrices for the inexpensive and difference models now modified to include the correlation between outputs. In other words, $\boldsymbol{R}_c$ and $\boldsymbol{R}_d$ in the above equations are modified using the correlation matrices $\boldsymbol{K}_c$ and $\boldsymbol{K}_d$ respectively. Including two different matrices in this manner makes the model flexible enough to capture similar correlations between inexpensive data but different correlations in the difference model. The assumption of a common set of correlations for the inexpensive and difference model is perhaps not entirely sensible when, for example, additional physics is included into a simulation moving between fidelity levels. As discussed previously the construction of such a multi-output model comes at the expense of a potential compromise in the θ and p parameters but in this case the compromise impacts both the inexpensive and difference models.

As per the single fidelity multi-output Kriging model each of these additional correlation matrices is parameterized using the spherical coordinate system. As before, multiple correlation matrices can then be used for each model to capture multiple categorical variables and reduce the number hyperparameters requiring optimization.

The process of constructing such a model closely follows that of the single-output multi-fidelity model except at each stage a multi-output model is constructed. Firstly, a multi-output Kriging model is constructed from the observed inexpensive data with the hyperparameters optimized with the aid of the adjoint approach described previously. Then a second multi-output model, representing the differences between the inexpensive and expensive data, is constructed. As with the multi-output model, the scaling parameter ρ must again be determined via a likelihood optimization. Once $\bar{\boldsymbol{R}}$ has been calculated via Equation 16, the derivative of the likelihood with respect to ρ can be calculated as before using Equations 28 and 29. With efficient derivatives available for both the inexpensive and difference model, the hybridized particle swarm can again be used to optimize the hyperparameters.

Finally, it should be observed that this modelling approach assumes that the same categories are present at each level of fidelity otherwise the modification of $\boldsymbol{R}_c(\boldsymbol{X}_c, \boldsymbol{X}_e)$ to include the correlations between outputs is not possible.

3 An Illustrative Analytical Example

The, so-called Forrester function (Forrester et al, 2007), has been employed repeatedly throughout the literature to demonstrate multi-fidelity surrogate modelling techniques. Its simplicity, and one dimensional nature, make it useful for graphically illustrating the benefits of multi-fidelity surrogate modelling.

The Forrester function defines an expensive, high fidelity function,

$$f_{1e}(x) = (6x - 2)^2\sin(12x - 4), \qquad (33)$$

and an 'inexpensive', low fidelity function,

$$f_{1c}(x) = 0.5f_1(x) + 10(x - 0.5) - 5 \qquad (34)$$

where $x \in [0, 1]$. This, of course, only considers a single output so to extend the Forrester function to two outputs we add $2x$ to both $f_{1e}(x)$ and $f_{1c}(x)$ producing,

$$f_{2e}(x) = (6x - 2)^2 \sin(12x - 4) + 2x, \quad (35)$$

and

$$f_{2c}(x) = 0.5f_1(x) + 10(x - 0.5) - 5 + 2x. \quad (36)$$

This simple modification ensures that each pair of inexpensive and expensive functions is well correlated. Each of the functions is illustrated graphically in Figure 1a.

Let us first consider the benefits of multi-output Kriging compared to single output (SO) Kriging. Figure 1b illustrates the case when f_{1e} has been sampled with 11 evenly distributed points across the design space (indicated by the black circles) whereas f_{2e} has only been sampled at three locations, $x = [0.1, 0.4, 1]$. An ordinary Kriging model through only the f_{1e} data is reasonably accurate, giving an r^2 correlation of approximately 0.999. Rather unsurprisingly, an ordinary Kriging model through only the f_{2e} data is a rather poor approximation of the true function, giving an r^2 correlation of 0.188. However, when a multi-output Kriging model is constructed using both sets of data the resulting r^2 correlations are 0.999 and 0.998 for f_{1e} and f_{2e} respectively, a considerable improvement. The multi-output model's prediction of f_{2e} is able to exploit the high level of correlation between both outputs resulting in an improvement in predictive performance.

Of course, the multi-output Kriging model is effective in this instance because of the relatively high level of correlation between f_{1e} and f_{2e}. They exhibit an r^2 of approximately 0.986. One would expect a lack of correlation between the functions to result in a reduction in predictive accuracy. We can demonstrate this is indeed the case by changing our second output from f_{2e} to f_{3e} where,

$$f_{3e}(x) = f_{1e}(x + A). \quad (37)$$

As illustrated in Figure 1c, by varying A between 0 and 0.15 the level of correlation between f_{1e} and f_{3e} can be controlled. For each of the 21 values of A highlighted in Figure 1c we sample f_{3e} at $x = [0.1, 0.4, 1]$. We could use this data directly to construct a single output Kriging model or we could combine this data with 11 evenly spaced sample points of f_{1e} to construct a multi-output Kriging model. Figure 1d illustrates the variation in the predictive accuracy of the resulting single and multi-output Kriging models with respect to f_{3e} as the level of correlation between f_{1e} and f_{3e} varies. This illustrates that as the level of correlation between the outputs drops so to does the accuracy of the resulting multi-output prediction until a single output model constructed from only the f_{3e} sample points is actually superior. This is a rather crude illustration in that we consider only one test function and ignore the impact of sample density, location of sample points etc. But nevertheless, it does illustrate that the level of correlation between outputs has an important role to play in the performance of a multi-output model.

Figures 1e and 1f extend this process to multiple levels of fidelity. Figure 1e presents multi-fidelity predictions of f_{1e} while Figure 1f presents multi-fidelity predictions of f_{2e}. As illustrated in Figure 1e we have sampled f_{1e} at $x = [0.0, 0.4, 0.6, 1]$ but in addition to this, and not illustrated, we have sampled f_{1c} at 11 equally spaced locations over the design space. Figure 1f illustrates the two sample points for f_{2e} at $x = [0.1, 0.4]$ and while also not illustrated there are the four sample points for f_{2c} at $x = [0.1250, 0.375, 0.625, 0.875]$. When attempting to predict f_{1e} we therefore have a situation very similar to the classic multi-fidelity Kriging example illustrated in Forrester et al (2007) with a lot of data for f_{1c} and a smaller dataset for f_{1e}. When attempting to predict f_{2e} the amount of data available for both f_{2e} and f_{2c} has been reduced. This in analogous to an engineer attempting to reuse data from a previous multi-fidelity model of f_{1e} to inform a new multi-fidelity model of f_{2e} where less information is available, for example, at the preliminary stages of a surrogate based optimization. For ease of reference a summary of the accuracy metrics for each prediction of f_{1e} and f_{2e} are presented in Table 1. This includes the Pearson r^2 correlation, root mean square error (RMSE) and maximum absolute error (MAE).

As expected the multi-fidelity (MF) Kriging model illustrated in Figure 1e produces an accurate prediction of f_{1e} resulting in a r^2 of 0.999. However, a single output multi-fidelity model constructed from only the f_{2c} and f_{2e} sample points
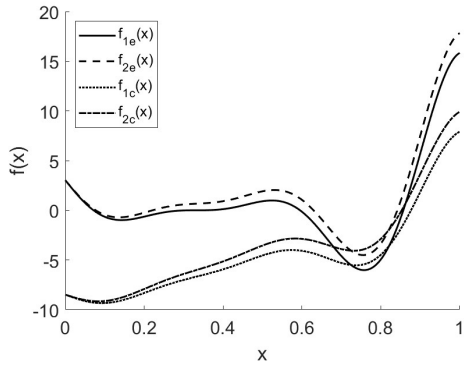
(a) True response of f_{1e}, f_{2e}, f_{1c} & f_{2c}
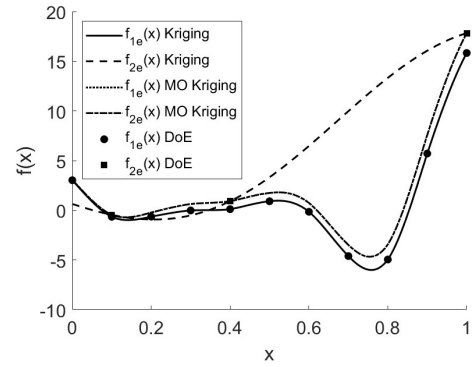
(b) Single fidelity Kriging & MO Kriging predictions
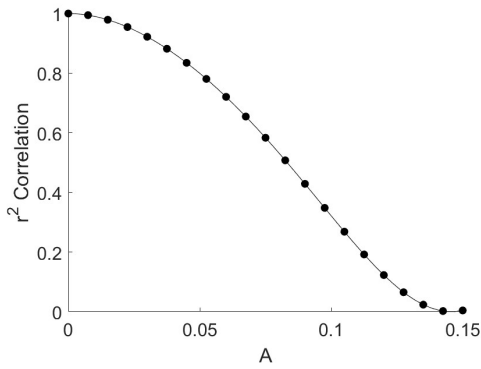
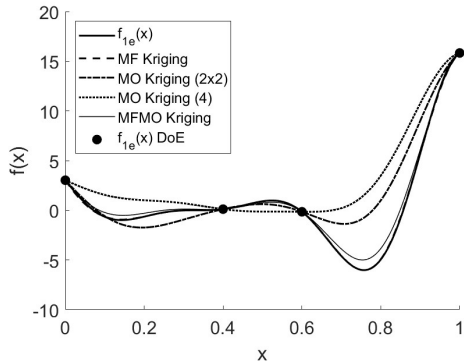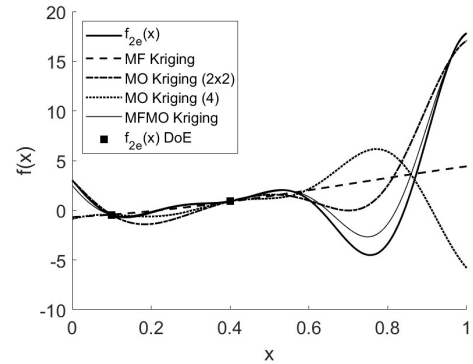(c) Correlation between f_{1e} & f_{3e} with A
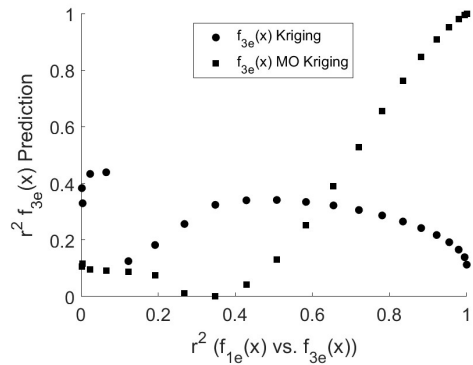
(d)

(e) Predictions of f_{1e} using MF, MO & MFMO Kriging

(f) Predictions of f_{2e} using MF, MO & MFMO Kriging

Fig. 1: Figures illustrating the the true Forrester functions (a), a comparison of a single fidelity prediction of the expensive functions using single fidelity ordinary Kriging and multi-output (MO) Kriging models (b), the correlation between f_{1e} and f_{3e} with varying A (c), predictive accuracy of f_{3e} with varying correlation level between f_{1e} and f_{3e} (d) and predictions of the expensive functions using multi-fidelity (MF) Kriging, multi-output Kriging and multi-fidelity multi-output (MFMO) Kriging modelling approaches (e) and (f).

Table 1: A summary of the accuracy of the surrogate models presented in Figures 1e and 1f.

f_{1e}	r^2	%RMSE	%MAE
MF Kriging	0.999	0.25	0.88
MFMO Kriging	0.995	2.19	5.07
MO Kriging (2×2)	0.770	11.09	27.28
MO Kriging (4)	0.636	16.40	38.68
f_{2e}	r^2	%RMSE	%MAE
MF Kriging	0.171	19.22	60.02
MFMO Kriging	0.977	3.81	8.93
MO Kriging (2×2)	0.793	10.56	26.27
MO Kriging (4)	0.513	30.38	105.83

results in a much poorer prediction with an r^2 of 0.171. In this case there is very little data upon which to build the inexpensive model resulting in a poor prediction of f_{2c} upon which to build the difference model. As can be observed graphically in Figure 1f, the resulting multi-fidelity model does little to adjust the prediction from the simple linear trend observed in the f_{2e} sample points.

Using the multi-fidelity multi-output (MFMO) Kriging model described in Section 2.4 we can combine all of our sample data from each of the four functions, f_{1e}, f_{2e}, f_{1c} and f_{2c} into a single model and use this to predict both f_{1e} and f_{2e}. Here the high level of correlation between both the inexpensive and expensive data (and therefore differences between them) results in a significant improvement in predictive performance for f_{2e} achieving a r^2 of 0.977. In this case, the multi-output model of both the inexpensive data and differences in inexpensive and expensive responses are able to exploit the high correlation between functions (as was observed in Figure 1b) to produce a more accurate model of f_{2e}. This improvement in accuracy can be seen graphically in Figure 1f.

We observed from Table 1 that the accuracy of the MFMO predictor for f_{1e} has dropped slightly, achieving a r^2 of 0.995. This is most likely a result of the compromise the hyperparameter optimization process has to make described previously, but in this case, the impact of this is negligible.

As discussed previously, an alternative method of including multiple levels of fidelity within a single model is via the multi-output Kriging model itself. In this case, the different levels of fidelity can be modelled as different outputs. For this particular example, there are two alternative approaches. The inexpensive and expensive outputs can be modelled as one categorical variable with four different outputs (MO Kriging(4)) or as two categorical variables each with two outputs (MO Kriging(2×2)). The choice between these approaches is effectively a trade-off between model flexibility and complexity. The case with four different outputs is more flexible as correlations between all combinations are explicitly parameterized but this increases the number of hyperparameters to determine and therefore the complexity of the hyperparameter optimisation. Using two separate categories (MO Kriging(2×2)) only requires two additional hyperparameters compared to the six required for MO Kriging(4). This difference in the number of parameters will become more marked as the number of output levels increases. Therefore, an important question for the current article to address is whether or not the additional flexibility offered by a more complex parameterization is worth it.

In addition to the MFMO predictions, figures 1e and 1f illustrate the predictions of both of these multi-output approaches. Once again the sample data from all four functions, f_{1e}, f_{2e}, f_{1c} and f_{2c}, is used in the construction of each model.

Focusing on the prediction of f_{2e}, we observe from Figure 1f and Table 1 that both of these approaches perform better than the single output multi-fidelity model. The MO Kriging(4) approach achieves a r^2 of 0.513 and the MO Kriging(2×2) approach achieves a r^2 of 0.793. However, these

predictions are significantly less accurate than that of the MFMO approach.

While the performance of these two strategies falls behind that of the MFMO Kriging approach it is worth noting that the nature of this analytical function makes it rather difficult to apply the simple and direct correlation that the MO model assumes between parameters. The example is constructed in such a way that the inexpensive and expensive functions are poorly correlated and that an additive term, the difference model in the MF Kriging approach, is required to correct from the inexpensive to expensive functions. Real world engineering design optimization problems, however, may have a much greater correlation between fidelities which such MO Kriging models may better exploit. This will be explored with less bias in the following engineering examples.

4 Multi-point Airfoil Design Optimization

The first engineering optimization case study considers the multi-point design optimization of a family of four airfoils from the NASA supercritical series, the SC(2)-0610, SC(2)-0612, SC(2)-0712 and SC(2)-0710, illustrated graphically in Figure 2a. Multi-point, in the context of airfoil optimization, refers to the optimization of an airfoil for more than one particular flight condition, or operating point, simultaneously. It is common for an airfoil optimized for one specific flight condition to actually perform worse than the baseline at off-design conditions. Considering multiple operating points within the optimization is a simple way of preventing this, see for example Szmelter (2001), Nemec et al (2004) or Painchaud-Ouellet et al (2006). Typically the multi-point optimization of an airfoil is performed in isolation, the following study therefore, not only explores the potential of a multi-fidelity, multi-output Kriging approach within an engineering optimization but the ability to improve optimization performance by considering multiple airfoil problems simultaneously. This is of practical interest if, for example, one were to optimize a series of related airfoil sections across the span of a wing in parallel.

In this case each of the four airfoils is analysed at four different design conditions, outlined in Table 2, with the overall objective function equal to the weighted summation of the drag coefficients at each of these conditions i.e.,

$$C_{d_w} = \sum_{i=1}^{4} w_i C_{d_i}, \qquad (38)$$

where $\sum_{i=1}^{4} w_i$ equals unity. The weights employed in this case study are also provided in Table 2. It should be observed that the calculation of a weighted objective function, such as that of equation 38, can be particularly sensitive to failures in any of the constituent simulations. This can reduce the amount of data available to construct an accurate surrogate model.

The airfoil geometry is modified via a free form deformation (FFD) approach where the y coordinates of ten points (five on each of the second and fourth rows of a control grid illustrated in Figure 2b) are varied by $\pm 2.5\%$ of the airfoil chord. This results in a total of 10 design variables with the baseline control grid and range of grid point movement consistent between each of the four airfoils. To simplify the problem in this particular airfoil optimization there are no additional constraints on either the pitching moment coefficient or the aerofoil thickness.

The analysis of each design is performed using the viscous Garabedian and Korn (VGK) solver (Lock and Williams, 1987) which is attractive due to its speed and accuracy (Toal and Keane, 2011). A single airfoil simulation using VGK will take approximately one second with 3-4 such analyses required to determine the drag for a fixed lift coefficient. The analysis of the weighted drag coefficient for a single design can therefore typically be completed within approximately 20 seconds. This enables the optimization to be performed a number of times to obtain a true picture of each strategy's performance.

As per the approach outlined by Toal and Keane (2011) the simulation of each airfoil at only the highest weighted operating condition, in this case condition one in Table 2, is used as a low fidelity version of the multi-point objective function. A high fidelity objective function will therefore simulate the airfoil at each of the four conditions in Table 2 but the low fidelity objective function will use only the drag coefficient from condition number one. Each high fidelity function

Table 2: Definition of the operating conditions and weightings for the calculation of the multi-point drag coefficient.

Condition	Mach No.	Required C_L	Reynolds No.	Weight
1	0.7	0.45	6.3×10^6	0.4
2	0.7	0.40	6.3×10^6	0.2
3	0.7	0.50	6.3×10^6	0.2
4	0.72	0.45	6.3×10^6	0.2
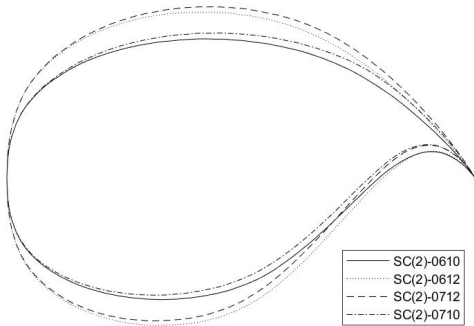
(a) Supercritical airfoil geometries
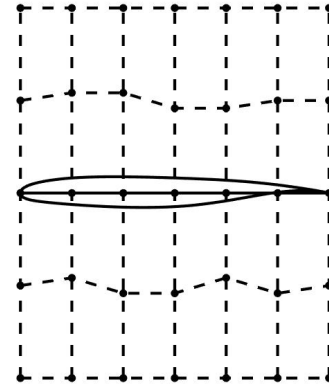
(b) Morphed SC(2)-0610 & FFD control grid

Fig. 2: Geometry for each of the four supercritical airfoils (a) and a morphed example of the SC(2)-0610 airfoil within the corresponding deformed FFD control grid.

evaluation is therefore the equivalent of four low fidelity function evaluations.

The design optimization of each airfoil assumes a total of 45 high fidelity simulations, the equivalent of 25 of which are used to define the initial sampling plan with the remaining 20 used to provide updates to the surrogate models in regions of interest. In all of the case studies presented here Matlab's genetic algorithm (GA) using a population size of 200 is run for 1000 generations to ensure the global optimum of each surrogate model's predictor is located and used as the update point. All other settings for the GA remain as the Matlab defaults. For the multi-fidelity approaches the high fidelity sampling plan contains 12 points with the effort that would have been expended calculating the remaining 13 designs instead used to make 52 low fidelity evaluations. It should be pointed out that due to

the nature of the weighted drag-coefficient calculation, the low fidelity drag coefficient will be calculated as part of the high fidelity evaluation. There will therefore be an additional 12 'free' low fidelity data points available at the 12 high fidelity locations. The multi-fidelity approach therefore comprises of 64 low and 12 high fidelity data points.

Given that the number of update cycles per airfoil optimization is identical and the initial sampling plans are equivalent in terms of simulation cost, the total simulation cost is therefore identical across all of the optimizations performed here. This is a recurring theme across each of the three engineering case studies.

A total of five different surrogate modelling strategies will be compared using each of the three engineering case studies presented within the current paper. For ease of reference, these are outlined in Table 3. A single output ordinary

Table 3: A summary of the considered surrogate modelling strategies.

Strategy Name	Description
SO Kriging	A single output ordinary Kriging model
MF Kriging	A single output multi-fidelity Kriging model
MFMO Kriging	A multi-fidelity, multi-output Kriging model
MO Kriging (x)	A multi-output Kriging model consisting of one categorical variable with x levels
MO Kriging $(x \times y)$	A multi-output Kriging model consisting of two categorical variables with x and y levels respectively

Kriging approach, referred to as SO Kriging, is used as a reference. In this approach only high fidelity data is used to construct the surrogate with independent models constructed for each categorical variable level. The multi-fidelity Kriging approach, referred to as MF Kriging, is used as the baseline multi-fidelity surrogate modelling approach. Employing both inexpensive and expensive data an independent multi-fidelity model is constructed for each categorical variable level.

The multi-output multi-fidelity Kriging approach of Lin et al (2021), referred to here as MFMO Kriging, employs both inexpensive and expensive data but constructs a single model to predict all outputs. It, therefore, combines the sampling plans across each output in its construction. As discussed previously, this approach first constructs a multi-output model of the inexpensive data and then combines this with a second multi-output difference model.

The final two approaches model the multiple levels of fidelity via categorical variables instead of via a difference model. However, the approach taken to achieve this differs between them. The strategy referred to as MO Kriging (x) models all combinations of fidelity and discrete variables within a single categorical variable and therefore via a single correlation matrix, $\boldsymbol{K}$. The x within the name of this strategy refers to the number of levels in the categorical variable. The second approach, MO Kriging $(x \times y)$ models the fidelity as a separate categorical variable to the discrete design variable. There will therefore be two correlation matrices, $\boldsymbol{k}$, one $x \times x$ in size and the second $y \times y$.

Table 4 presents the accuracy of the surrogate models of the weighted drag coefficient for each airfoil optimization problem. Accuracy in each case is computed using a 2000 point optimal

Latin hypercube sampling plan which is independent of the data used in the construction of the surrogate models. As already demonstrated by Toal and Keane (2011), a multi-fidelity approach to predicting a weighted drag coefficient is superior to a single fidelity approach. For each of the airfoil optimization problems the single output, multi-fidelity Kriging approach results in a more accurate surrogate model across all metrics than a single fidelity approach.

The multi-output Kriging approaches all result in a significant improvement in surrogate accuracy compared to the single output models. Of the three models the two multi-output approaches, where the fidelity is captured using a categorical variable, tend to result in a slightly more accurate surrogate model than the multi-fidelity multi-output approach. Using two categorical variables, one to indicate the baseline airfoil and a separate variable to indicate the fidelity tends to result in the highest r^2 correlation across the four problems but defining a single correlation matrix for all combinations (even with its relatively large number of hyperparameters) tends to result in the lowest root mean square error (RMSE) and maximum absolute error (MAE).

Table 5 presents the results for each airfoil optimization averaged over 20 different runs, each commencing from a different design of experiments. In each case, an optimal Latin hypercube has been generated using a different random number seed using Matlab's inbuilt `lhsdesign` function. No two DoEs are therefore the same and the minimum Euclidean distance between any two points over all DoE pairs is 0.197 where the design variables have been normalized between 0 and 1.

For the three multi-output surrogate modelling approaches the four airfoils are optimized in parallel i.e. for each update cycle the surrogate model

Table 4: Accuracy of single, multi-output and multi-fidelity Kriging models for each multi-point airfoil problem, metrics are averaged over 20 different sampling plans with the standard deviations shown in brackets.

SC(2)-0610	r^2	%RMSE	%MAE
SO Kriging	0.304 (0.168)	14.72 (2.93)	65.91 (13.17)
MF Kriging	0.682 (0.190)	8.98 (2.62)	46.55 (11.40)
MFMO Kriging	**0.761** (0.197)	9.27 (5.03)	**41.40** (13.07)
MO Kriging (4×2)	0.759 (0.188)	8.39 (2.42)	44.49 (17.62)
MO Kriging (8)	0.747 (0.196)	**7.65** (2.87)	42.62 (19.19)
SC(2)-0612	r^2	%RMSE	%MAE
SO Kriging	0.433 (0.252)	10.96 (3.08)	65.16 (7.11)
MF Kriging	0.717 (0.182)	7.65 (2.25)	49.94 (9.61)
MFMO Kriging	0.800 (0.195)	7.84 (4.47)	**52.77** (8.68)
MO Kriging (4×2)	**0.805** (0.180)	7.13 (1.86)	54.28 (13.10)
MO Kriging (8)	0.784 (0.184)	**6.30** (2.17)	53.32 (12.75)
SC(2)-0712	r^2	%RMSE	%MAE
SO Kriging	0.676 (0.259)	9.90 (4.29)	41.36 (15.52)
MF Kriging	0.860 (0.194)	6.21 (2.99)	32.22 (10.08)
MFMO Kriging	0.865 (0.208)	8.32 (6.72)	29.13 (12.83)
MO Kriging (4×2)	**0.889** (0.206)	7.16 (2.45)	29.04 (9.58)
MO Kriging (8)	0.876 (0.209)	**5.37** (3.03)	**24.00** (10.28)
SC(2)-0710	r^2	%RMSE	%MAE
SO Kriging	0.654 (0.219)	10.27 (3.99)	50.74 (9.96)
MF Kriging	0.814 (0.196)	7.28 (3.35)	44.89 (10.94)
MFMO Kriging	0.852 (0.205)	8.23 (5.78)	39.10 (10.41)
MO Kriging (4×2)	**0.868** (0.201)	6.72 (2.61)	42.07 (11.77)
MO Kriging (8)	0.850 (0.211)	**6.07** (3.27)	**34.67** (11.36)

Table 5: Average percentage reduction in weighted drag coefficient for each multi-point airfoil optimization problem taken over 20 independent optimization runs. Standard deviation in reduction presented in brackets.

	SC(2)-0610	SC(2)-0612	SC(2)-0712	SC(2)-0710
SO Kriging	2.62 (0.266)	3.12 (0.416)	3.78 (0.212)	3.59 (0.367)
MF Kriging	2.59 (0.225)	3.17 (0.320)	3.66 (0.232)	3.61 (0.244)
MFMO Kriging	**2.64** (0.183)	3.29 (0.201)	3.79 (0.231)	3.64 (0.121)
MO Kriging (8)	2.64 (0.251)	**3.35** (0.257)	**3.86** (0.216)	**3.71** (0.129)
MO Kriging (4×2)	2.60 (0.295)	3.31 (0.340)	3.85 (0.242)	3.65 (0.200)

is used to find the best design for each airfoil. This results in a total of four simulations being performed at each update but the total simulation effort is identical to that of four independent optimizations of each airfoil.

For each airfoil optimization the multi-output surrogate modelling strategies produce the greatest reduction in weighted drag coefficient with the MO Kriging approach using a single categorical
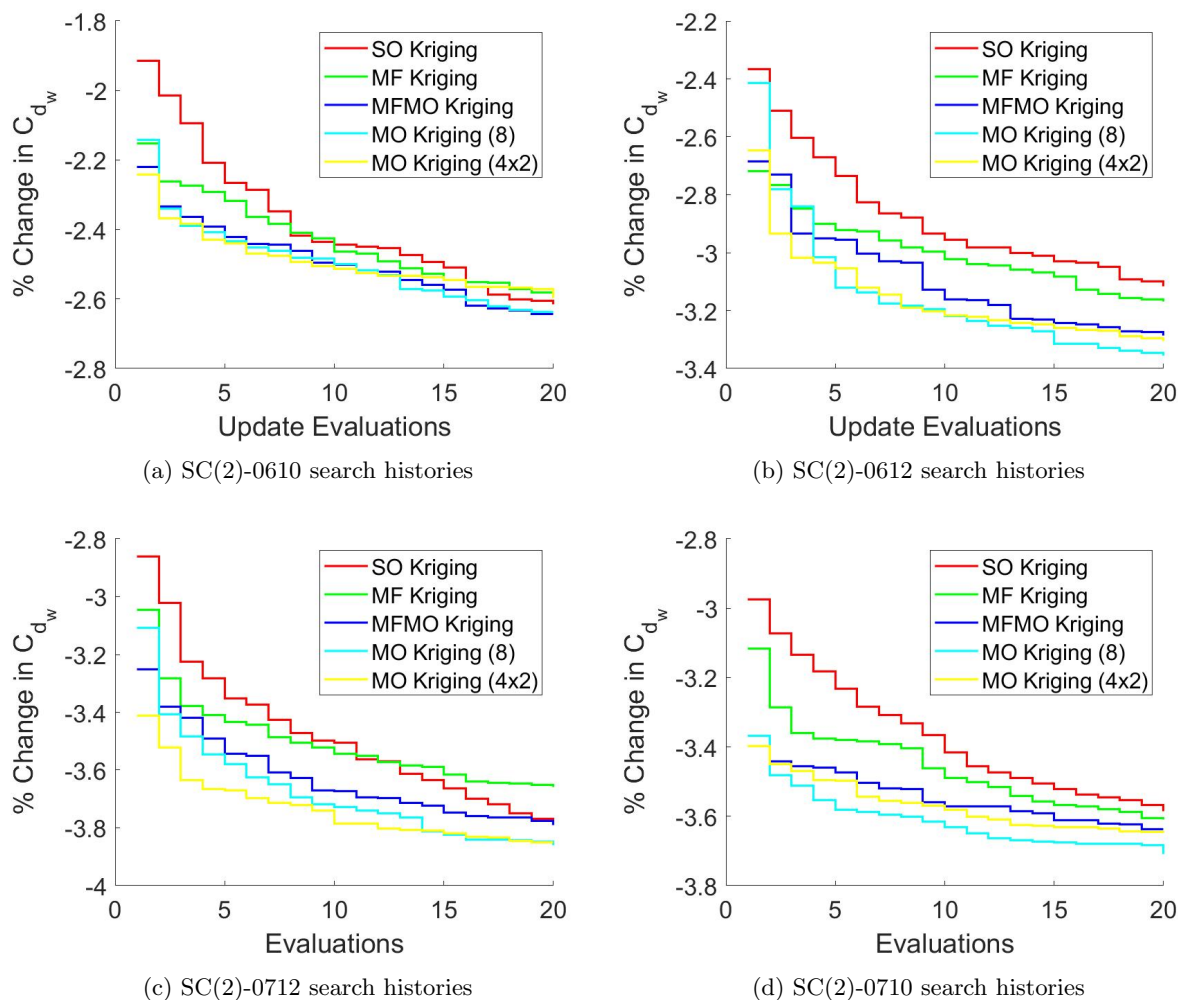
(a) SC(2)-0610 search histories

(b) SC(2)-0612 search histories

(c) SC(2)-0712 search histories

(d) SC(2)-0710 search histories

Fig. 3: Average search histories for the 10 variable airfoil optimization using single and multi-output surrogate modelling approaches for the SC(2)-0610 (a), SC(2)-0612 (b), SC(2)-0712 (c) and SC(2)-0710 (d) airfoil optimizations respectively

variable performing the best on three of the four problems.

Figure 3 presents the optimization histories for each of the airfoil problems averaged over the 20 runs. To simplify the plots only the weighted drag coefficients of the update evaluations are plotted here. These plots reinforce the performance advantage of a multi-output approach over a single output approach in all cases. The superiority of a pure multi-output approach compared to a multi-fidelity multi-output approach in terms of both rate of convergence and final design quality is demonstrated clearly in Figures 3b, 3c and 3d.

5 Vibrating Truss Optimization

The second engineering optimization problem considers the design of a vibrating 2D truss similar to that found in a satellite structure. Taken from Nair and Keane (1999), the geometry of the 2D truss illustrated in Figure 4 is modified in order to minimize the band-averaged vibration attenuation at the tip of the truss. The truss is modelled as a set of 42 Euler-Bernoulli beams each defined by two finite elements. The leftmost nodes are

assumed to be encastre and a unit force excitation is applied to node one (highlighted with the diamond in Figure 4) over a 100-200Hz frequency range. The x and y coordinates of the remaining numbered nodes in Figure 4 are permitted to vary within a 0.9×0.9 box centred on the node, resulting in a total of 40 design variables. In addition to the geometric parameters a categorical variable has been introduced, the choice of material, which alters the properties of all of the beams within the model. In this case, a total of three different materials are considered steel, aluminium and titanium. Apart from the design variable bounds this optimization problem is unconstrained.

This particular engineering problem (excluding the consideration of material choice) has been used a number of times within the literature to test the performance of optimization algorithms. The complexity of the resulting design space (Toal and Keane, 2012) coupled with the relatively low simulation cost provides a realistic, relatively high dimensional, structural design problem for which enough cases can be run to gain confidence in the resulting performance comparisons.

The calculation of the band-averaged vibration attenuation is performed in the range 100-200Hz with a step size of 1Hz. In order to create a low fidelity version of this objective function the step size has been increased to 4Hz resulting in a simulation approximately one quarter the cost of the original high fidelity simulation.

Table 6 presents the resulting surrogate model accuracy when the five modelling approaches are considered. In this case a 200 point sampling plan of only the expensive function is used for the single output Kriging model while a sampling plan consisting of 100 expensive and 400 inexpensive evaluations is used for each multi-fidelity case. A separate sampling plan has been defined for each material choice with the multi-output models combining the sampling plans across all three materials. Due to a combination of sampling plan size and problem dimensionality, the results are averaged over only 10 different cases.

For the single output modelling strategies the multi-fidelity approach results in a more accurate surrogate model across all metrics for each of the three materials. Of the multi-output modelling approaches both the multi-fidelity and two-category multi-output approach result in surrogate models with an improved accuracy over the multi-fidelity single output approach for each of the three materials. The single category multi-output approach, however, appears to struggle in each case producing surrogate models which perform worse than the baseline single output Kriging model. This result is rather curious as one would expect the additional data and the relatively high level of correlation between the results of each material to result in a performance improvement. However, the number of additional hyperparameters in this formulation compared to the two-category model is significant. In terms of overall surrogate model prediction accuracy the multi-fidelity multi-output Kriging model is the clear winner across all categories.

The optimization strategy adopted here employs a total of 30 update iterations where at each point the predicted optimum for each of the three materials is evaluated using the expensive FE simulation. For the single output surrogates each optimization is performed independently and there is no transfer of information between the material problems. For the multi-output models, all three of the new data points calculated at each update cycle are appended to the overall data set. The prediction of the design space at each iteration is therefore impacted by the updates across all three materials from the preceding iterations. Given the expense of constructing the surrogate models at each iteration, the results are restricted to being averaged over five independent optimizations.

Table 7 presents the average results of these optimizations for each of the five surrogate modelling approaches while Figure 5 presents the average search histories for each update iteration.

The optimization of the steel truss shows a moderate performance improvement when adopting a single output multi-fidelity approach. In this case, the multi-fidelity approach offers a more rapid initial rate of performance improvement and a noticeably better final design. The performance advantage for the other materials is not quite as clear-cut. For aluminium, the rate of convergence is relatively similar as the optimization progresses with the single fidelity approach producing a slightly better final design. For titanium, again the rate of convergence is similar but this time the multi-fidelity approach offers slightly more design improvement.
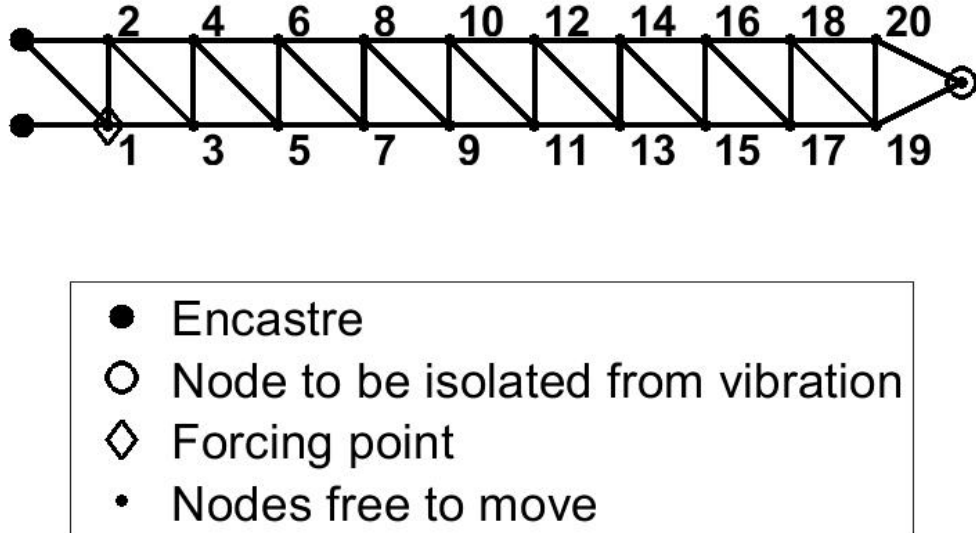
Fig. 4: An illustration of the baseline truss geometry recreated from Toal and Keane (2012).

All three multi-output approaches, where the data from all three material optimizations has been combined, offer a considerable improvement in optimization performance, resulting in similar final designs. However, on two out of the three material optimizations the two pure multi-output approaches, where the simulation fidelity is represented within a categorical variable rather than via a difference model, perform better. The exception is for steel where the MFMO Kriging strategy produces the best design on average.

The optimization histories presented in Figure 5 illustrate that over the initial few updates the additional predictive accuracy offered by both the MFMO Kriging and MO Kriging (3×2) strategies leads to a more rapid improvement in design performance than for the MO Kriging (6) strategy. However, after approximately 10 updates this strategy has reached or surpassed the level of improvement achieved by the other two. This suggests an improvement in the predictive accuracy of this particular model as the optimization has progressed and indeed by the end of the optimization, the surrogate model produced using the MO Kriging (6) strategy has a r^2 of 0.467, 0.478 and 0.499 when predicting, respectively, steel, aluminium and titanium. This is much closer to the predictive accuracy of the MFMO Kriging approach, for example, which achieves r^2 of 0.506, 0.522 and 0.538 for the same materials.

6 Combustor Optimization

The final engineering design problem concerns the optimization of a gas turbine combustor. Based on the problem outlined by Toal et al (2021), the porting arrangement of the combustor illustrated in Figure 6 is optimized in order to minimize the

Table 6: Accuracy of single and multi-output co-Kriging models for the 40 variable truss problem, metrics are averaged over 10 different sampling plans with standard deviations in brackets.

Steel	r^2	%RMSE	%MAE
SO Kriging	0.331 (0.044)	10.47 (0.632)	52.02 (4.946)
MF Kriging	0.428 (0.029)	9.59 (0.269)	47.11 (3.023)
MFMO Kriging	**0.502** (0.014)	**8.70** (0.136)	**42.65** (1.702)
MO Kriging (3×2)	0.460 (0.065)	9.01 (0.511)	45.17 (2.234)
MO Kriging (6)	0.226 (0.199)	11.11 (1.887)	55.32 (9.067)
Aluminium	r^2	%RMSE	%MAE
SO Kriging	0.314 (0.063)	10.85 (0.730)	52.85 (5.774)
MF Kriging	0.452 (0.023)	9.56 (0.296)	46.46 (3.000)
MFMO Kriging	**0.523** (0.013)	**8.63** (0.074)	**44.34** (1.567)
MO Kriging (3×2)	0.477 (0.073)	8.99 (0.588)	46.08 (3.956)
MO Kriging (6)	0.233 (0.201)	11.18 (1.877)	52.62 (6.335)
Titanium	r^2	%RMSE	%MAE
SO Kriging	0.381 (0.069)	10.07 (0.676)	49.55 (4.677)
MF Kriging	0.444 (0.034)	9.60 (0.306)	49.05 (5.444)
MFMO Kriging	**0.534** (0.012)	**8.58** (0.115)	**42.92** (2.556)
MO Kriging (3×2)	0.504 (0.063)	8.83 (0.541)	44.32 (2.862)
MO Kriging (6)	0.277 (0.187)	10.71 (1.508)	52.63 (6.754)

Table 7: Optimization results for each material from a 40 variable design optimization employing single and multi-output surrogate modelling strategies averaged over five optimizations with standard deviations presented in brackets.
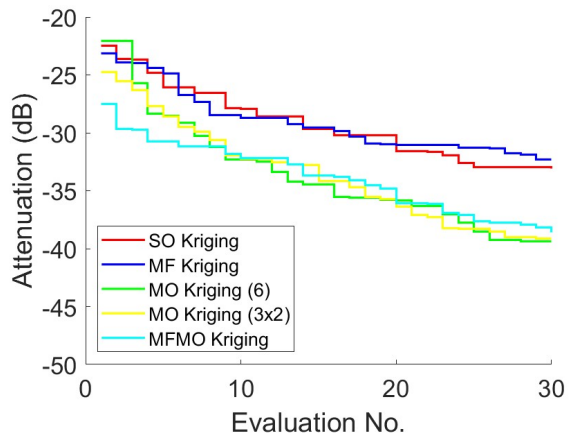
	Steel	Aluminium	Titanium
SO Kriging	-38.39 (4.50)	-33.07 (3.88)	-34.71 (3.69)
MF Kriging	-40.15 (5.57)	-32.32 (4.47)	-36.73 (2.63)
MFMO Kriging	**−46.01** (4.39)	-38.60 (3.59)	-43.06 (3.05)
MO Kriging (3×2)	-44.76 (3.36)	-39.11 (3.67)	-43.40 (3.56)
MO Kriging (6)	-45.46 (3.77)	**−39.40** (3.00)	**−44.94** (3.09)

production of NOx. In this case there are a total of eight design variables. Each of the four rows of ports are permitted to move axially with the radius of the ports in each row also free to vary. In addition to these eight variables two different port topologies are considered. One in which there are no chutes present on any of the ports, as illustrated in Figure 6a, and one in which there are only chutes present on the initial row of ports on both the inner and outer walls (see Figure 6b). Other than the variable bounds there are no additional constraints on this particular optimization problem.

As per the approach outlined by Toal et al (2021), manipulation of the geometry is achieved via a parameterization implemented within Siemens NX. The Prometheus combustor design system (Zhang et al, 2014) is used to automatically extract an appropriate CFD fluid volume and define an appropriate meshing strategy. In this case the CFD mesh is generated using ICEM CFD with the simulation then performed using the proprietary Rolls-Royce CFD package, PRECISE-UNS (Anand et al, 2013). The K-ϵ realizable turbulence model is used with each simulation running for a total of 5000 iterations. Fuel
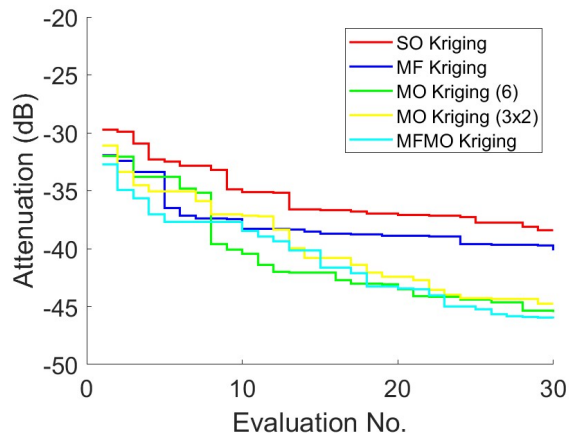
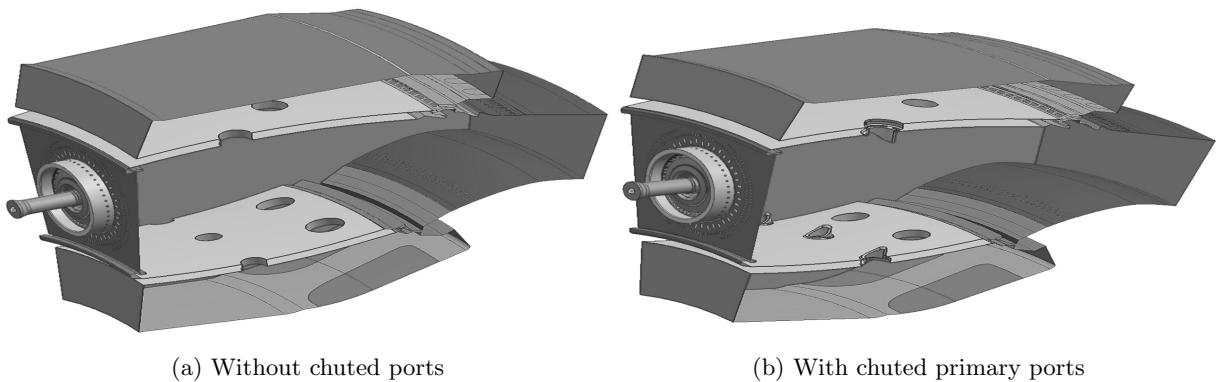(a) Steel search histories

(b) Aluminium search histories
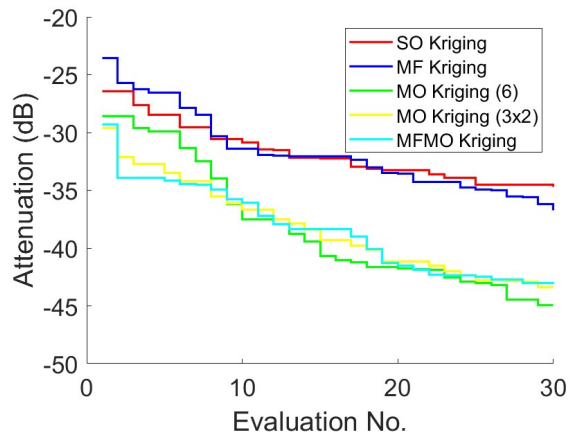
(c) Titanium search histories

Fig. 5: Average search histories for the 40 variable truss optimization problem using single and multi-output surrogate modelling approaches for steel (a), aluminium (b) and titanium (c).

(a) Without chuted ports

(b) With chuted primary ports

Fig. 6: Gas turbine combustor geometry without chuted ports and within only primary ports chuted.

is introduced into the model after 500 iterations with ignition occurring after 600. Combustion is modelled via the flamelet-generated manifold technique with a constant set of velocity profiles provided at the annuli and swirler inlets. Simulation conditions are representative of a 100% throttle condition in a large civil airliner gas turbine engine.

In order to alter the fidelity of the simulation the density of the underlying mesh is varied using the ICEM CFD scripts. A high fidelity simulation consists of approximately 17.0M elements while a low fidelity equivalent consists of approximately 3.5M elements. This results in a 5:1 cost ratio i.e. for each high fidelity simulation it is possible to perform five low fidelity simulations (including both meshing and simulation effort). The change in simulation fidelity naturally has an impact on the accuracy of the simulation results. Across the design space the low fidelity prediction of NOx, for example, is reasonably well correlated with the high fidelity prediction giving an r^2 of 0.924 but with a RMSE of approximately 7.2%. While providing a reasonable prediction of trend, the precise level of NOx production, something which is required for certification reasons, is not something which the low fidelity model can be relied upon to provide. This is, of course, what we are aiming to correct using our multi-fidelity surrogate modelling approaches.

Each high fidelity CFD simulation is naturally significantly more expensive than the airfoil and truss simulations used in the previous engineering test problems. Running enough simulations to gain appropriate statistics upon which to judge the performance of each of the surrogate modelling strategies therefore becomes difficult. Instead, and as per Toal et al (2021), a single large sampling plan is defined and used to construct a relatively accurate surrogate model of NOx for each fidelity and port topology. This surrogate model is then used as the basis for the subsequent surrogate based design optimizations. In this case a sampling plan consisting of 560 CFD simulations is performed for each simulation fidelity and port topology, the results from which are used to create a Kriging based surrogate model using the CM02 toolbox. The sampling plan used here is an optimal Latin hypercube generated using Matlab's `lhsdesign` function. The accuracy of each of these models is calculated via leave-one-out cross validation and presented in Table 8.

Table 9 presents the prediction accuracy of each of the five surrogate modelling strategies for both combustor port topologies. For the single output Kriging model a 20 point sampling plan is defined for each port topology. For the multi-fidelity modelling approaches the sampling plan consists of 10 expensive simulations and 50 inexpensive simulations. The sampling plan simulation effort is therefore identical between the single and multi-fidelity approaches. The results presented in Table 9 have been averaged over 50 different sampling plans where all 'simulations' are evaluated using the surrogate models presented in Table 8. Prediction accuracy is determined using a separate 8000 point optimal Latin hypercube sampling plan.

As expected the single output multi-fidelity approach offers a more accurate prediction than the single output, single fidelity approach. The multi-output approaches, combining the sampling plans from both port topologies, however, offer the best-performing surrogate models across all metrics. In this particular engineering example, the multi-output surrogate modelling approach where two separate categorical variables is employed (MO Kriging (2×2)) performs noticeably better than all of the other strategies. This strategy is closely followed by the multi-output Kriging strategy employing a single combined categorical variable (MO Kriging (4)).

Table 10 and Figure 7 present the optimization results for each of the five surrogate modelling strategies. Each optimization commences from either a 20 point sampling plan of the expensive function or 10 and 50 point sampling plans of the expensive and inexpensive functions respectively. A total of 20 update cycles are performed with an optimal design for each port topology evaluated each time. Figure 7 presents the search history for each optimization averaged over 50 independent optimization runs while Table 10 presents the average reduction in NOx over the 50 optimizations. It should be noted that the percentage reduction in NOx is relative to each port topology's baseline performance.

In the optimization of both port styles the multi-output approaches tend to result in a better performing final design. The performance of

Table 8: Accuracy of the surrogate models of raw CFD data used within the subsequent design optimization and surrogate accuracy studies.
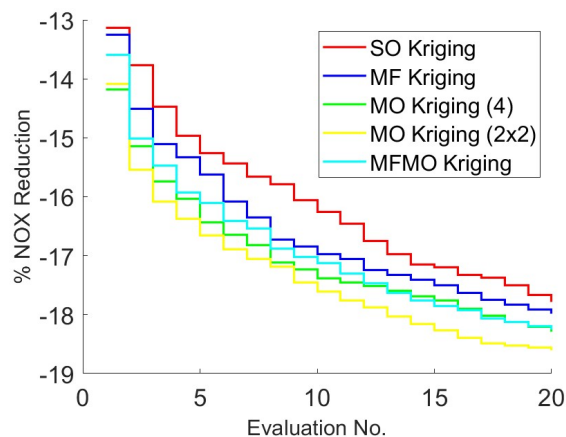
Port Topology	Simulation Fidelity	r^2	%RMSE	%MAE
Without Chutes	High Fidelity	0.8171	7.36	31.79
Without Chutes	Low Fidelity	0.837	6.39	45.49
Primary Chutes Only	High Fidelity	0.892	6.98	43.02
Primary Chutes Only	Low Fidelity	0.876	7.79	25.37

Table 9: Accuracy of single, multi-output and multi-fidelity Kriging models for each of the 8 variable combustor problems, metrics are averaged over 50 different sampling plans with standard deviations presented in brackets.

Without Chutes	r^2	%RMSE	%MAE
SO Kriging	0.448 (0.154)	20.81 (3.30)	87.22 (15.22)
MF Kriging	0.627 (0.078)	16.89 (2.05)	79.66 (12.12)
MFMO Kriging	0.626 (0.087)	17.17 (2.50)	79.97 (13.31)
MO Kriging (2×2)	**0.688** (0.058)	**15.27** (1.58)	**78.26** (11.62)
MO Kriging (4)	0.626 (0.116)	16.69 (2.48)	82.04 (11.88)
With Primary Chutes	r^2	%RMSE	%MAE
SO Kriging	0.591 (0.129)	18.50 (3.08)	72.03 (10.11)
MF Kriging	0.624 (0.107)	18.66 (2.88)	72.58 (9.93)
MFMO Kriging	0.670 (0.083)	16.86 (2.41)	67.34 (8.92)
MO Kriging (2×2)	**0.734** (0.078)	**14.99** (2.02)	60.97 (9.75)
MO Kriging (4)	0.703 (0.114)	15.33 (2.75)	**59.47** (8.27)
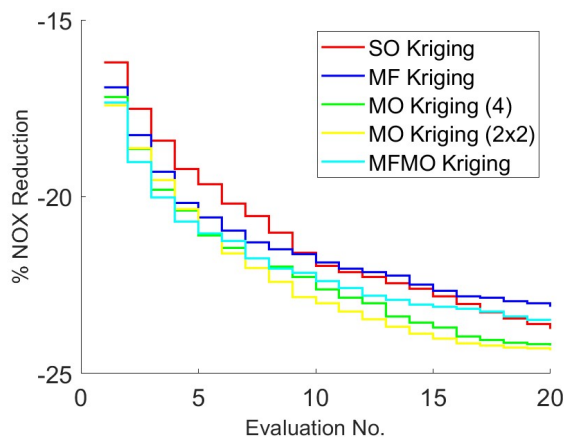
(a) Search histories without chutes   (b) Search histories with chutes

Fig. 7: Search histories for the 8 variable combustor optimization problem, without chutes (a) and with primary chutes present (b), averaged over 50 different optimizations.

Table 10: Optimization results (percentage reduction in NOx) for each combustor topology from the 8 variable combustor optimization problem employing single and multi-output surrogate modelling strategies (results averaged over 50 optimizations with standard deviations in brackets).

	Without Chutes	With Chutes
SO Kriging	-23.73 (0.014)	-17.78 (0.010)
MF Kriging	-23.12 (0.015)	-17.98 (0.010)
MFMO Kriging	-23.51 (0.015)	-18.26 (0.008)
MO Kriging (2×2)	**−24.35** (0.003)	**−18.60** (0.005)
MO Kriging (4)	-24.21 (0.005)	-18.29 (0.005)

the single output approaches appears to be somewhat problem dependent. For the case where no chutes are present the single output Kriging model eventually results in a better design than the multi-fidelity Kriging model even though the better accuracy of the multi-fidelity model appears to produce an initial faster rate of convergence. Interestingly, in the case of the multi-fidelity multi-output Kriging model, performance is initially on a par with that of the other multi-output models but decays after 6-7 update cycles and eventually performs slightly worse than the single output Kriging model. This coupled with the poorer than expected performance of the multi-fidelity model suggests a poor correlation between fidelity levels for this particular problem which may limit the performance of a traditional multi-fidelity approach. The multi-output approaches, where the fidelity level is included as part of the categorical variable(s) are able to better tune the correlation between the fidelity levels to take this into account thereby outperforming the other approaches.

For the second chute topology, where chutes are only present on the primary ports, the story is slightly different. The multi-output approaches are the best performing strategies. The single output multi-fidelity Kriging approach marginally performs better than the single output Kriging approach suggesting that, in this case, the presence of the inexpensive data can be exploited. The multi-output Kriging approach where port topology and fidelity are treated as separate categorical variables, however, performs noticeably better than the other four strategies.

7 Conclusions

The current article has investigated the performance of three different multi-output, multi-fidelity Kriging surrogate modelling strategies with respect to both their predictive and optimization performance on three very different engineering design problems. The mathematics and modelling philosophy behind each approach has been explained in full and an adjoint of the likelihood function for both single and multi-fidelity, multi-output Kriging, has been presented for the first time.

Overall the results on each of the engineering case studies demonstrate the viability of a multi-fidelity, multi-output approach to engineering design optimization when faced with either categorical variables or topological decisions. The multi-point airfoil design optimization demonstrated that, through a multi-output approach, multiple airfoils from the same family could be optimized more effectively in parallel rather than in series. The truss attenuation optimization case study demonstrated that multi-fidelity, multi-output modelling, could be employed to capture the impact of a material choice and accelerate the design optimization of a structural optimization problem. Finally, the combustor optimization problem demonstrated that a multi-output, multi-fidelity approach could accelerate the design optimization of a gas turbine combustion system when presented with a topological decision, in this case the presence or absence of chuted ports.

Of the three multi-output, multi-fidelity approaches compared, the results across all three case studies demonstrated that a pure multi-output modelling approach, where the fidelity is considered as a level within a categorical variable,

is a viable alternative to a more traditional multi-fidelity approach employing a difference model. In a number of cases, both surrogate model accuracy and optimization performance were better when using a multi-output predictor approach. There is, however, an important caveat to place on this general conclusion in that the multi-output modelling approach generally requires a high level of correlation between both low and high fidelity responses. If when moving between output fidelities there is a significant enough shift in the physics modelled to result in a reduction in correlation, a more traditional approach, employing a difference model, may, in fact, be more appropriate. The three case studies presented within this article, with perhaps the exception of the combustor case, exhibit relatively high levels of correlation between fidelities which the multi-output approach can exploit. It should, of course, be noted that even a difference model approach can in some circumstances fail if the differences between fidelity levels cannot be accurately modelled given the potentially limited amount of data available (Toal (2015)).

Of the two purely multi-output modelling approaches investigated, overall optimization performance appears to be better when including all combinations of the simulation outputs within a single categorical variable. For the airfoil case study this amounted to a total of eight levels within the categorical variable; two levels of simulation fidelity for each of the four airfoils. While this offers the greatest flexibility in terms of modelling correlations between data sets it also results in a significantly higher number of hyperparameters to optimize. In this case, it may be that the relatively large designs of experiments (DoEs) used in these case studies simplify the hyperparameter optimization and a combination of a smaller DoE size and more levels per category would cause a degradation in performance.

Not addressed within this paper is the potential for switching between single output and multi-output models as an optimization progresses. This could potentially offer an improvement in overall optimization performance by, for example, exploiting multiple categories at the beginning of an optimization when data is sparse and then the more precise hyperparameters of a single output model when there is more data. An evaluation of each model's accuracy via leave-one-out cross-validation would enable the optimizer to use the most accurate model at every stage.

As demonstrated using the simple 1D Forrester function in Section 3, the multi-output Kriging formulation requires the outputs to be well correlated in order to offer a better prediction than separate single output Kriging models. However, it is feasible that the level of correlation between outputs could vary across the design space. While this aspect is not actively addressed within the current article the literature does suggest potential mechanisms in which this could be accounted for. Snoek et al (2014) suggest, for example, the use of a beta distribution to warp the spaces so that they are better correlated which results in an improvement in the predictive accuracy of a multi-output Kriging model. Warping the design space in such a manner may also help to prevent the compromise in the hyperparameters that can occur as the amount of sample data increases.

Declarations

Conflict of Interest Statement

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Replication of Results

The results presented within the paper have been created by a number of propitiatory Rolls-Royce software tools. The raw results are therefore sensitive and cannot be published.

References

Alizadeh R, Allen JK, Mistree F (2020) Managing computational complexity using surrogate models: a critical review. Research in Engineering Design 31(3):275–298. https://doi.org/10.1007/s00163-020-00336-7

Allmendinger R, Emmerich MTM, Hakanen J, et al (2017) Surrogate-assisted multicriteria

optimization: Complexities, prospective solutions, and business case. Journal of Multi-Criteria Decision Analysis 24(1-2):5–24. https://doi.org/10.1002/mcda.1605

Anand MS, Eggels R, Staufer M, et al (2013) An advanced unstructured-grid finite-volume design system for gas turbine combustion analysis. In: ASME 2013 Gas Turbine India Conference. American Society of Mechanical Engineers, https://doi.org/10.1115/gtindia2013-3537

Black N, Najafi AR (2022) Learning finite element convergence with the multi-fidelity graph neural network. Computer Methods in Applied Mechanics and Engineering 397:115,120. https://doi.org/10.1016/j.cma.2022.115120

Bonilla EV, Chai K, Williams C (2007) Multi-task gaussian process prediction. In: Platt J, Koller D, Singer Y, et al (eds) Advances in Neural Information Processing Systems, vol 20. Curran Associates, Inc.

Brooks C, Forrester A, Keane A, et al (2011) Multi-fidelity design optimisation of a transonic compressor rotor. In: 9th European Turbomachinery Conference, Istanbul, Turkey, 21st-25th March

Chatterjee T, Chakraborty S, Chowdhury R (2017) A critical review of surrogate assisted robust design optimization. Archives of Computational Methods in Engineering 26(1):245–274. https://doi.org/10.1007/s11831-017-9240-5

Deb K, Roy PC, Hussein R (2020) Surrogate modeling approaches for multiobjective optimization: Methods, taxonomy, and results. Mathematical and Computational Applications 26(1):5. https://doi.org/10.3390/mca26010005

Eberhart R, Kennedy J (1995) A new optimizer using particle swarm theory. In: 6th International Symposium on Micro Machine and Human Science, pp 39–43

Eddine DS, Smail K, Mahfoudh C, et al (2022) Robust optimization of the aerodynamic design of a helicopter rotor blade based on multi-fidelity meta-models. Proceedings of the Institution of Mechanical Engineers, Part E: Journal of Process Mechanical Engineering p 095440892211115. https://doi.org/10.1177/09544089221111593

Forrester A, Keane A, Bressloff N (2006) Design and analysis of "noisy" computer experiments. AIAA Journal 44(10):2331–2339

Forrester A, Sóbester A, Keane A (2007) Multi-fidelity optimization via surrogate modelling. Proceedings of the Royal Society A 463(2088):3251–3269. https://doi.org/10.1098/rspa.2007.1900

Garud SS, Karimi IA, Kraft M (2017) Design of computer experiments: A review. Computers & Chemical Engineering 106:71–95. https://doi.org/10.1016/j.compchemeng.2017.05.010

Giles M (2008) Collected matrix derivative results for forward and reverse mode algorithmic differentiation. Lecture Notes in Computational Science and Engineering 64:35–44. https://doi.org/10.1007/978-3-540-68942-3-4

Guo Q, Hang J, Wang S, et al (2020) Design optimization of variable stiffness composites by using multi-fidelity surrogate models. Structural and Multidisciplinary Optimization 63(1):439–461. https://doi.org/10.1007/s00158-020-02684-3

Haftka RT, Villanueva D, Chaudhuri A (2016) Parallel surrogate-assisted global optimization with expensive functions – a survey. Structural and Multidisciplinary Optimization 54(1):3–13. https://doi.org/10.1007/s00158-016-1432-3

Jones D, Schonlau M, Welch W (1998) Efficient global optimization of expensive black-box functions. Journal of Global Optimization 13(4):455–492. https://doi.org/10.1023/A:1008306431147

Kennedy M, O'Hagan A (2000) Predicting the output from a complex computer code when fast approximations are available. Biometrika 87(1):1–13. https://doi.org/10.1093/biomet/87.1.1

Kleijnen JP, Mehdad E (2014) Multivariate versus univariate kriging metamodels for multiresponse simulation models. European Journal of Operational Research 236(2):573–582. https://doi.org/10.1016/j.ejor.2014.02.001

Li M, Liu Z, Huang L, et al (2022) A new multi-fidelity surrogate modelling method for engineering design based on neural network and transfer learning. Engineering Computations 39(6):2209–2230. https://doi.org/10.1108/EC-06-2021-0353

Lin Q, Hu J, Zhou Q, et al (2021) Multi-output gaussian process prediction for computationally expensive problems with multiple levels of fidelity. Knowledge-Based Systems 227:107,151. https://doi.org/10.1016/j.knosys.2021.107151

Liu H, Ong YS, Cai J (2017) A survey of adaptive sampling for global metamodeling in support of simulation-based complex engineering design. Structural and Multidisciplinary Optimization 57(1):393–416. https://doi.org/10.1007/s00158-017-1739-8

Liu H, Cai J, Ong YS (2018a) Remarks on multioutput gaussian process regression. Knowledge-Based Systems 144:102–121. https://doi.org/10.1016/j.knosys.2017.12.034

Liu H, Ong YS, Cai J, et al (2018b) Cope with diverse data structures in multi-fidelity modeling: A gaussian process method. Engineering Applications of Artificial Intelligence 67:211–225. https://doi.org/10.1016/j.engappai.2017.10.008

Liu X, Zhu Q, Lu H (2014) Modeling multiresponse surfaces for airfoil design with multiple-output-gaussian-process regression. Journal of Aircraft 51(3):740–747. https://doi.org/10.2514/1.c032465

Liu X, Zhao W, Wan D (2022) Multi-fidelity co-kriging surrogate model for ship hull form optimization. Ocean Engineering 243:110,239. https://doi.org/10.1016/j.oceaneng.2021.110239

Lock R, Williams B (1987) Viscous-inviscid interactions in external aerodynamics. Progress in Aerospace Sciences 24(2):51–171. https://doi.org/10.1016/0376-0421(87)90003-0

Nair P, Keane A (1999) Design optimization of space structures with nonperiodic geometries for vibration suppresion. In: Proceedings of the 40th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference, St. Louis, MO

Nasti A, Voutchkov II, Toal DJJ, et al (2022) Multi-fidelity simulation for secondary air system seal design in aero engines. In: Proceedings of ASME Turbo Expo 2022 Turbomachinery Technical Conference and Exposition

Nemec M, Zingg D, Pulliam T (2004) Multipoint and multi-objective aerodynamic shape optimization. AIAA Journal 42(6):1057–1065. https://doi.org/10.2514/1.10415

Painchaud-Ouellet S, Tribes C, Trépanier J, et al (2006) Airfoil shaped optimization using a nonuniform rational B-spline parameterization under thickness constraint. AIAA Journal 44(10):2170–2178. https://doi.org/10.2514/1.15117

Peherstorfer B, Willcox K, Gunzburger M (2018) Survey of multifidelity methods in uncertainty propagation, inference, and optimization. SIAM Review 60(3):550–591. https://doi.org/10.1137/16M1082469

Pelamatti J, Brevault L, Balesdent M, et al (2018) Efficient global optimization of constrained mixed variable problems. Journal of Global Optimization 73(3):583–613. https://doi.org/10.1007/s10898-018-0715-1

Phiboon T, Khankwa K, Petcharat N, et al (2021) Experiment and computation multifidelity multi-objective airfoil design optimization of fixed-wing UAV. Journal of Mechanical Science and Technology 35(9):4065–4072. https://doi.org/10.1007/s12206-021-0818-3

Quan L, Jiexiang H, Lili Z, et al (2022) Gradient-enhanced multi-output gaussian process model for simulation-based engineering design. AIAA JOURNAL 60(1). https://doi.org/10.2514/1.J060728

Serani A, Pellegrini R, Wackers J, et al (2019) Adaptive multi-fidelity sampling for CFD-based optimisation via radial basis function metamodels. International Journal of Computational Fluid Dynamics 33(6-7):237–255. https://doi.org/10.1080/10618562.2019.1683164

Shi M, Lv L, Sun W, et al (2020) A multi-fidelity surrogate model based on support vector regression. Structural and Multidisciplinary Optimization 61(6):2363–2375. https://doi.org/10.1007/s00158-020-02522-6

Snoek J, Swersky K, Zemel RS, et al (2014) Input warping for bayesian optimization of non-stationary functions. In: Proceedings of The 31st International Conference on Machine Learning

Stork J, Eiben AE, Bartz-Beielstein T (2020) A new taxonomy of global optimization algorithms. Natural Computing https://doi.org/10.1007/s11047-020-09820-4

Szmelter J (2001) Multipoint aerodynamic wing optimization in viscous flow. Journal of Aircraft 38(5):860–867. https://doi.org/10.2514/2.2845

Tabatabaei M, Hakanen J, Hartikainen M, et al (2015) A survey on handling computationally expensive multiobjective optimization problems using surrogates: non-nature inspired methods. Structural and Multidisciplinary Optimization 52(1):1–25. https://doi.org/10.1007/s00158-015-1226-z

Toal D (2015) Some considerations regarding the use of multi-fidelity kriging in the construction of surrogate models. Structural and Multidisciplinary Optimization https://doi.org/10.1007/s00158-014-1209-5

Toal D, Keane A (2011) Efficient multi-point aerodynamic design optimization via co-kriging. Journal of Aircraft 48(5):1685–1695. https://doi.org/10.2514/1.C031342

Toal D, Keane A (2012) Non-stationary kriging for design optimization. Engineering Optimization 44(6):741–765. https://doi.org/10.1080/0305215X.2011.607816

Toal D, Bressloff N, Keane A, et al (2011) The development of a hybridized particle swarm for kriging hyperparameter tuning. Engineering Optimization https://doi.org/10.1080/0305215X.2010.508524

Toal DJJ, Zhang X, Keane AJ, et al (2021) The potential of a multifidelity approach to gas turbine combustor design optimization. Journal of Engineering for Gas Turbines and Power 143(5). https://doi.org/https://doi.org/10.1115/1.4048654

Tran A, Tran M, Wang Y (2019) Constrained mixed-integer gaussian mixture bayesian optimization and its applications in designing fractal and auxetic metamaterials. Structural and Multidisciplinary Optimization 59(6):2131–2154. https://doi.org/10.1007/s00158-018-2182-1

Wang B, Chen T (2015) Gaussian process regression with multiple response variables. Chemometrics and Intelligent Laboratory Systems 142:159–165. https://doi.org/10.1016/j.chemolab.2015.01.016

Wang L, Tao S, Zhu P, et al (2020) Data-driven topology optimization with multiclass microstructures using latent variable gaussian process. Journal of Mechanical Design 143(3). https://doi.org/10.1115/1.4048628

Yong HK, Wang L, Toal DJJ, et al (2019) Multi-fidelity kriging-assisted structural optimization of whole engine models employing medial meshes. Structural and Multidisciplinary Optimization 60(3):1209–1226. https://doi.org/10.1007/s00158-019-02242-6

Zhan D, Xing H (2020) Expected improvement for expensive optimization: a review. Journal of Global Optimization 78(3):507–544. https://doi.org/10.1007/s10898-020-00923-x

Zhang C, Liu L, Wang H, et al (2022) SCGAN: stacking-based generative adversarial networks for multi-fidelity surrogate modeling. Structural and Multidisciplinary Optimization 65(6). https://doi.org/10.1007/s00158-022-03255-4

Zhang R, Zhao X (2020) Inverse method of centrifugal pump blade based on gaussian process regression. Mathematical Problems in Engineering 2020:1–10. https://doi.org/10.1155/2020/4605625

Zhang X, Toal D, Bressloff N, et al (2014) Prometheus: a geometry-centric optimisation system for combustor design. In: ASME Turbo Expo 2014: Turbine Technical Conference and Exposition, Dusseldorf, DE

Zhang X, Toal DJJ, Keane AJ, et al (2015) Isothermal combustor prediffuser and fuel injector feed arm design optimization using the prometheus design system. Journal of Engineering for Gas Turbines and Power 138(6). https://doi.org/10.1115/1.4031711

Zhou Q, Qian PZG, Zhou S (2011) A simple approach to emulation for computer models with qualitative and quantitative factors. Technometrics 53(3):266–273. https://doi.org/10.1198/TECH.2011.10025