EngOpt 2016 - 5^{th} International Conference on Engineering Optimization
Iguassu Falls, Brazil, 19-23 June 2016.

Improving the optimisation performance of an ensemble of radial basis functions

M. Stramacchia[a,b], D.J.J. Toal[a,c] and A.J. Keane[a,d]

[a] *Faculty of Engineering and the Environment, University of Southampton, Southampton SO16 7QF, United Kingdom*
e-mail: [b] ms41g12@soton.ac.uk, [c] D.J.J.Toal@soton.ac.uk, [d] Andy.Keane@soton.ac.uk

Abstract

In this paper we investigate surrogate-based optimisation performance using two different ensemble approaches, and a novel update strategy based on the local Pearson correlation coefficient. The first ensemble, is based on a selective approach, where n_s RBFs are constructed and the most accurate RBF is selected for prediction at each iteration, while the others are ignored. The second ensemble uses a combined approach, which takes advantage of n_s different RBFs, in the hope of reducing errors in the prediction through a weighted combination of the RBFs used. The update strategy uses the local Pearson correlation coefficient as a constraint to ignore domain areas where there is disagreement between the surrogates. In total the performance of six different approaches are investigated, using five analytical test functions with 2 to 50 dimensions, and one engineering problem related to the frequency response of a satellite boom with 2 to 40 dimensions.

Keywords: Radial Basis Functions, ensemble of surrogates, Pearson correlation coefficient, bootstrap, p-value, Fisher metric.

1 Introduction and motivation

By reducing the number of objective function evaluations, surrogate-based optimisation offers an alternative to direct optimisation using both traditional and evolutionary methods, where in general hundreds or even thousands of objective function evaluations are needed. Following the seminal paper of Sacks et al. [1], intensive research on metamodeling has been carried out and success has been achieved through numerous applications and research fields, e.g. structural optimisation [2], aerodynamic optimisation [3, 4], turbomachinery optimisation [5], multiobjective optimisation [6], and multi-fidelity optimisation [7]. The most common surrogate approximation models are linear and quadratic polynomials created by performing least squares regression [8]. Besides the commonly used polynomial functions, the use of stochastic models has also been proposed (i.e. Kriging interpolation and regression) [9] to treat the deterministic computer response as a realisation of a random process. Other types of models include for example, neural networks [10], radial basis functions (RBF) [11], and support vector regression [12]. Such surrogate models, using an analytical function, $\hat{f}(\boldsymbol{x})$, for the input-output relationship, in place of the true objective function, $f(\boldsymbol{x})$, thereby reducing the computational cost associated with a design optimisation. Based on zero-order information from expensive computational models, (e.g. FEM, CFD), they construct a cheap and hopefully accurate analytical approximation.
However, in many engineering optimisation problems, the number of function evaluations is severely limited by computational cost. In this case, the lack of sufficient information describing the relationship between the response and the input variables makes it difficult for a designer to know a priori which metamodel is the best for a specific problem. One possible solution is to construct multiple metamodels based on a common training dataset [13] (i.e. a predefined design of experiment, DoE), evaluate their accuracy using a cross-validation (CV) approach based on a statistical metric such as the root-mean square error (RMSE), and then select at each iteration the one surrogate model that performs best while discarding the rest. This selective approach has some shortcomings, as it does not take complete advantage of the computational resources devoted to constructing different metamodels, and it is based on global information only gathered from the CV approach, which may be unreliable at high dimensions with sparse datasets.

In order to take full advantage of different individual surrogates, to extract as much information as possible with a predefined computational budget, researchers have recently devoted considerable attention to combining multiple metamodels in the form of an ensemble of surrogates, (i.e. combined ensemble), [14]. These studies show that the resulting ensemble of metamodels takes advantage of the prediction ability of each metamodel to enhance the global accuracy of the response prediction. However, construction of an accurate combined ensemble requires judicious selection of the weight factor [15].
The computational cost of evaluating $\hat{f}(\boldsymbol{x})$ for a given $\boldsymbol{x}$ is usually minimal, in comparison to the evaluation of the true objective function, $f(\boldsymbol{x})$. However, the problem of finding the global optimum of $\hat{f}(\boldsymbol{x})$ is not always trivial, especially in high dimensions and in the case of surrogate characterised by multimodal landscape. Furthermore, even if the surrogate prediction can be considered as the obvious search direction, sometimes this approach does not perform well due to the fact that only global information are taking into account, a misleading information in the case of an unsatisfactory dataset. Therefore, the lack of data can jeopardise the weights computation in the case of a combined ensemble, or can rank incorrectly the set of surrogates in the case of a selective ensemble, with the consequence of a misrepresentation of the true landscape. This can consequently increase the probability of selecting an update point that offers no improvement. To overcome these disadvantages, in this paper we develop an infill criteria based on surrogate prediction combined by the Pearson correlation coefficient, r. This aims to prevent updates being evaluated where there is substantial disagreement between the surrogates.
The influence of the ensemble model and the updating procedure, inside an optimisation process, is extensively studied. Five analytical test functions with various degrees of complexity, and one engineering testcase are used to investigate the performance of each ensemble strategy, with emphasise on low (i.e. less than 5 design variables), medium (i.e. less than 15 design variables) and large-scale unconstrained optimization problems (i.e. more than 15 design variables). All optimisations are based on a close-ended scenario, where the aim is to find as good a design as possible within a predetermined computational budget. Three computational budgets (denoted as $PD = 1, 2, 5$) and one DoE topology are investigated. Each ensemble performance is measured based on the solution quality after a predefined number of objective function evaluations. This performance is characterised by the mean value and 95% confidence interval of the last iteration, computed via a bootstrap approach. The performances of the algorithms are then compared using a hypothesis test approach based on the two sample Welch's t-test. This approach is used to compute the p-values of each algorithms pair comparison, and the Fisher metric is adopted to compute a vector of combined p-values used to rank the algorithms performances.

2 Experimental setup

The ensembles studied in this paper are based on a predefined set of radial basis functions (RBF), three non-tuned RBFs (i.e. Linear, Cubic, and Thin-Plate), and two tuned RBFs (i.e. Gaussian and Inverse-Multiquadric), see Table 1. Tuning parameters for Gaussian and Inverse-Multiquadric RBFs are defined based on a K-fold cross-validation approach.
A two-stage method is adopted in order to perform a surrogate-based optimization. In stage one the true objective function is first evaluated over a set of design points, then if needed the shape parameter is estimated, the RBFs are generated by fitting these points and the ensemble is constructed using a selective or a combined approach. Generally at this stage the location of the design points is only required to satisfy some space-filling criterion, as there is generally no prior knowledge of areas of interest within the design space, and here a Latin-Hypercube based on a maximin criterion is used. A total of 50 different DoEs for $d = 2, 5, 10, 15$, and only 10 for $d = 25, 40, 50$ are generated, where d is the problem dimensionality. Multiple DoEs help to negate the impact of the sampling plan in assessing the performance of each surrogate strategy and provide a meaningful average of performance. The surrogate models are constructed from sampling plans of three different point densities (PD), one, two, and five, where the point density in this study is considered as a numerical parameter used to compute the size of the computational budget, n_{cb}. The computational budget is equal to $n_{cb} = 2 \times n_{DoE}$, where n_{DoE} is the size of the initial design of experiments, with $n_{DoE} = PD \times d$. The infill points are equal to one-half the total computational budget, and only one infill point at a time is added to the dataset, using one of

Table 1: Overview of the set of RBFs used in the ensembles construction. $\varphi(\cdot)$ is the kernel of the basis function, $r = \|\boldsymbol{x} - \boldsymbol{x}^{(k)}\|_2$ is the radial basis distance between the sample point, $\boldsymbol{x}$, where we want to make a prediction, and the basis function centres, $\boldsymbol{x}^{(k)}$, while ε (if present, that is in the case of tuned-RBFs) is the shape parameter, which governs the region of influence of the kernel.

Type of basis function	$\varphi(r) \quad (r \geq 0)$
Infinitely smooth tuned RBFs	
Gaussian	$\varphi(r) = \exp(-(\varepsilon r)^2)$
Inverse-Multiquadric	$\varphi(r) = \dfrac{1}{\sqrt{(r^2 + \varepsilon^2)}}$
Piecewise smooth non-tuned RBFs	
Linear	$\varphi(r) = r$
Cubic	$\varphi(r) = r^3$
Thin-plate spline	$\varphi(r) = r^2 \ln(r)$

Table 2: Overview of the five analytical test functions used in our experimental study, where d is the problem dimension, while x_i^l and x_i^u are the lower and upper boundaries of the numerical domain.

Function name	Function equation	$x_i^l \leq x_i \leq x_i^u$
Sphere	$f(\boldsymbol{x}) = \sum_{i=1}^{d} x_i^2$	$-5 \leq x_i \leq 5$
Rosenbrock	$f(\boldsymbol{x}) = \sum_{i=1}^{d-1} [100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2]$	$-1.5 \leq x_i \leq 1.5$
Trid	$f(\boldsymbol{x}) = \sum_{i=1}^{d} (x_i - 1)^2 - \sum_{i=1}^{d} x_i x_{i-1}$	$-d^2 \leq x_i \leq d^2$
Ackley	$f(\boldsymbol{x}) = 20 + \exp(1) - 20 \exp\left(0.2\sqrt{d^{-1} \sum_{i=1}^{d} x_i^2}\right) - \exp\left(d^{-1} \sum_{i=1}^{d} \cos(2\pi x_i)\right)$	$-1.5 \leq x_i \leq 1.5$
Styblinski-Tang	$f(\boldsymbol{x}) = \frac{1}{2} \sum_{i=1}^{d} (x_i^4 - 16x_i^2 + 5x_i)$	$-5 \leq x_i \leq 5$

two different infill criteria, namely the unconstrained ensemble prediction, or the constrained ensemble prediction. Once the surrogate has been obtained, the second stage is to search it for an infill point which hopefully improves the objective function. Due to their analytical nature, such surrogates can be searched cheaply using a global optimisation algorithm. In this case the Non-Dominated Sorting Genetic Algorithm II (NSGA-II) is used [16, 17], which can explore several basins of attraction simultaneously. The stopping criterion for this optimisation subroutine is $(n_{gen} + 1) \times p_{size}$ ensemble evaluations, where n_{gen} is the number of generations, and p_{size} is the population size. Note that n_{gen}, p_{size} are equal to 100 and 50 respectively, for all the simulations. These new points are then evaluated with the expensive model, and the ensemble is reconstructed based on the updated dataset. This process is then repeated until the computational budget is exhausted.
To check statistical significance between the different search strategies, a statistical resampling technique is used to compute a better representation of the average performance and the associated 95% confidence intervals. Here, bootstrapping is used considering only the best objective function values found by the final iteration of the optimisation process.

2.1 Test Problems

To carry out the numerical tests presented in this work, five analytical test functions found in the literature were used, see Table 2 and [18]. In addition an engineering testcase is also considered which is concerned with the optimization of the frequency response of a satellite boom structure, described by 40 individual Euler-Bernoulli beams connected at 20 joints. A unit force excitation is applied to node number 1, and the position of each node can be adjusted both horizontally and vertically within a 0.9×0.9 box, while the leftmost nodes are fixed. The objective is the minimization of the frequency averaged response of the beam in the range $100 - 200$ Hz and the vibration energy level is measured at the tip (i.e. rightmost node), see [19, 20] for full details. The dimensionality of all of these problems is permitted to vary from 5 to 50 dimensions, with the exception of the satellite structure which has a maximum of 40 variables. The numerical domain of the above test functions is normalised to $x_i \in [0, 1]$, before sampled.

3 Combined and Selective Ensemble based optimisation: unconstrained and constrained algorithm

The combined ensemble is a weighted average surrogate (WAS) [15], which takes advantage of n_s different RBFs, in the hope of reducing errors in the prediction through a weighted combination of the metamodels used. The predictor in this case is, $\hat{f}_{\mathrm{WAS}}(\boldsymbol{x}) = \sum_{i=1}^{n_s} w_i \hat{f}_i(\boldsymbol{x}) = \boldsymbol{w}^T \hat{\boldsymbol{f}}(\boldsymbol{x})$, where w_i is the weight associated with the ith surrogate, and $\hat{f}_i(\boldsymbol{x})$ is the predicted response by the ith surrogate. The weight vector is computed by optimising the following Lagrangian function, $\mathcal{L}(\boldsymbol{w}, \lambda) = \boldsymbol{w}^T \mathbf{C} \boldsymbol{w} - \lambda(\mathbf{1}^T \boldsymbol{w} - 1)$, which gives, $\boldsymbol{w} = \dfrac{\mathbf{C}^{-1}\mathbf{1}}{\mathbf{1}^T \mathbf{C}^{-1}\mathbf{1}}$, where $\mathbf{C}$ is the error correlation matrix, with $c_{ij} \simeq \dfrac{1}{N} \boldsymbol{e}_i^T \boldsymbol{e}_j$, where $\boldsymbol{e}_i, \boldsymbol{e}_j$ are the error vectors from CV associated with the prediction given by the surrogate i and j respectively [15]. The scalar quantity λ is the Lagrangian multiplier for the constraint $\mathbf{1}^T \boldsymbol{w} - 1$ (i.e. the weights normalisation condition). The optimal weight vector, $\boldsymbol{w}$, may include weights larger than one as well as negative weights. A practical approach to enforcing positivity is to consider only the diagonal elements of $\mathbf{C}$, because these terms are more accurately approximated than the off-diagonal terms especially in the cases of a small dataset [15]. The unconstrained version of this ensemble directly minimises $\hat{f}_{\mathrm{WAS}}(\boldsymbol{x})$, while the constrained version is based on the minimisation of $\hat{f}_{\mathrm{WAS}}(\boldsymbol{x})$ subject to the Pearson correlation coefficient constraint. This coefficient is defined as,

$$r_{s_i s_j} = \frac{\sum_{k=1}^{n_p} (s_{i_k} - \bar{s_i})(s_{j_k} - \bar{s_j})}{\sqrt{\sum_{k=1}^{n_p} (s_{i_k} - \bar{s_i})^2 \sum_{k=1}^{n_p} (s_{j_k} - \bar{s_j})^2}} \qquad (i, j = 1, 2, \ldots, n_s) \quad \text{and} \quad i \neq j, \tag{1}$$

where $(s_{i_1}, s_{i_2}, \ldots, s_{i_{n_p}})$ and $(s_{j_1}, s_{j_2}, \ldots, s_{j_{n_p}})$ are the predictions at n_p data point of RBF-i and RBF-j respectively, while $\bar{s_i}$ and $\bar{s_j}$ represent the sample mean of the n_p RBFs predictions. The n_p data point are randomly constructed and centred around each population members of the global optimiser. The correlation coefficient is then constructed for all of the available pairs of surrogate combinations, equal to $C_2^{n_s} = \binom{n_s}{2}$, resulting in a 10 vector of correlations, for each population member of the global optimiser. This is equal to $\mathbf{R}_G = \begin{bmatrix} r_{s_1 s_2} & r_{s_1 s_3} & r_{s_1 s_4} & r_{s_1 s_5} & r_{s_2 s_3} & r_{s_2 s_4} & r_{s_2 s_5} & r_{s_3 s_4} & r_{s_3 s_5} & r_{s_4 s_5} \end{bmatrix}$, where the subscript $(s_1, s_2, s_3, s_4, s_5)$ indicates the RBFs. The combined ensemble optimisation therefore becomes,

$$\begin{aligned} \underset{\boldsymbol{x} \in \mathbb{R}^d}{\text{minimize}} \quad & \hat{f}_{\mathrm{WAS}}(\boldsymbol{x}) \\ \text{subject to} \quad & \min(\mathbf{R}_G) \geq r^*, \end{aligned} \tag{2}$$

where r^* is a predefined minimum level of correlation between surrogates at a potential update point. In this study three cases are investigated, unconstrained $r^* \geq -1$, low constrained $r^* \geq 0.25$, and high constrained $r^* \geq 0.75$. For both the selective ensemble strategy, unconstrained and constrained, five RBFs are constructed, and then the most accurate RBF is selected for prediction at each iteration, while the others are ignored. The surrogate accuracy in this case is based on a generalised cross-validation approach, using the R2 factor as statistical metric. The unconstrained version of this ensemble directly minimises the RBF characterised by the largest R2 value, while the constrained version is based on the minimisation of the same RBF, but in domain region where the Pearson correlation coefficient constraint is satisfied.

4 Numerical approach

To illustrate the unconstrained and constrained ensemble ideas, we present a brief example obtained from a single design of experiment in the case of the multimodal Ackley function in two dimensions, see

Figure 1(a). Figure 1 illustrates the set of RBFs, the combined ensemble, the test function, the initial dataset and the landscape of the minimised Pearson correlation coefficient. The figure clearly shows that the initial dataset has a strong impact on all the surrogate models constructed. All of the surrogates, apart the Linear RBF, describe a misleading landscape and consequently the infill point, $\boldsymbol{x}^*$, based on the minimisation of each RBF is incorrect. At least the Linear RBF seem to describe rather well the middle part of the design space that in this case is where the global basin of attraction is located. The performance of each RBF at each iteration can be checked by calculating $f(\boldsymbol{x}^*)$, and the results for the first iteration are summarised in Table 3. As we can see, the five infill points are a poorer solution than the best solution within the initial DoE (see the best solution within the initial DoE, $f(\boldsymbol{x}^0) = 2.9018$, in Figure 1(a)). A different approach is to combine the five RBFs in an ensemble to strategically come up with a prediction that is more accurate and robust than the predictions by each individual surrogate. In this context, the objective is to create a combined ensemble that hopefully accurately captures the trends of the relationship between $\boldsymbol{x}$ and $f(\boldsymbol{x})$, and in particular, over those regions in the design space where high performance designs lie. However, the performance of such a strategy, is related to the set of surrogates used. The best set of metamodels must be based on surrogates that satisfy two conditions, 1) they must be different in terms of prediction values, $\hat{f}(\boldsymbol{x})$, and 2) they must be at the same time similar in terms of global prediction accuracy (i.e. RMSE), because in this case the chance of error cancellation would increase, [14]. In the present case, only the first condition is satisfied, as we can see in Table 3, where the five surrogates give different prediction, but only the Linear RBF is globally accurate. Obviously, the dataset plays an important role on the RBFs prediction and accuracy, and consequently, basing the construction of the ensemble only on CV information can hamper the optimisation process with an unsatisfactory dataset.

An alternative is to compare the RBFs correlation in local region around each point of the NSGA-II population, during the search of the combined ensemble. This strategy gathers some useful information in order to discard misleading infill points, and trust infill points located in regions where the various RBFs combinations are characterised by a predefined grade of correlation. This seems to help the optimisation process to leave unpromising potentially update regions dictated by some of the RBFs within the set used to build the combined ensemble, and move the choice of infill point to the constrained region of the design space identified by the Pearson correlation coefficient. This is clearly displayed in Figure 1(h) and (i) where the three infill points of the first iteration are shown. However, in this case only the strategy with a low correlation factor, (i.e. $r^* \geq 0.25$), outperform the unconstrained one, and the results are summarised in the convergence history of Figure 1(i). This behaviour opens a new question as to what is the optimal r^* value as the optimisation process progress. However, that is beyond the scope of this paper.

The selective ensemble is even more dependent on the CV result, relying solely on the R2 value at each iteration and no averaging process is considered, like in the combined ensemble. However, adding some local information also helps the optimisation, as is displayed in Figure 1(j).

In order to explain the methodology used to compare the numerical results, Figure 2 represents the complete set of results for the Ackley function in 2 dimensions when 50 different 10 point Latin-hypercube DoEs are used. This error bar plot is a graphical representation of the variability of results obtained after the bootstrap analysis. Each marker indicates the mean value estimated, while the lower and upper bars represent the 95% confidence intervals (CI) around the estimated mean. These confidence intervals can be interpreted as the range of values that would contain the true mean value 95% of the time if the optimisation were repeated on an infinite number of DoEs. The mean value, μ, is considered as a measure of the efficiency of the optimisation strategy with a lower mean value indicating a better objective function value. On the other hand, the confidence intervals can be interpreted as a measure of consistency (i.e. measured as $\theta = \theta_U - \theta_L$, where θ_U and θ_L are the upper and lower CI), therefore a low value of θ means that more confidence on the results can be placed, if and only if the efficiency is good enough, otherwise high consistency with a low efficiency could means that the optimisation get stuck in a local basin of attraction.

Table 3: Global accuracy of the RBFs (in bold are shown the best values); $\boldsymbol{w}$ is the optimal weights vector of combined ensemble, while $\hat{f}(\boldsymbol{x}^*)$ and $f(\boldsymbol{x}^*)$ are the RBFs predictions and the true objective function value based on the infill point $\boldsymbol{x}^*$ gathered from the direct minimisation of each RBF.

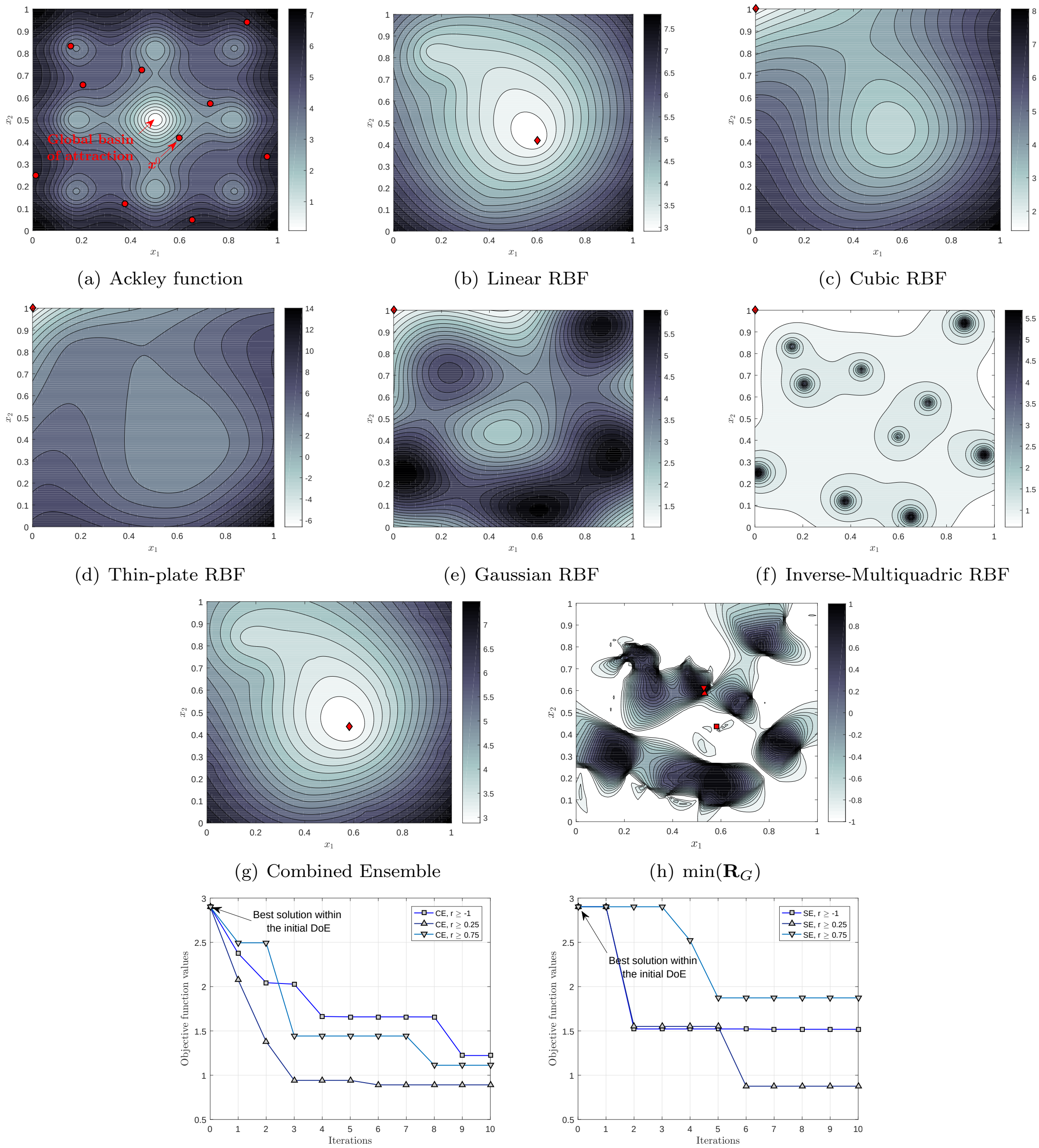
Surrogate	RMSE	MAE	R2	$\boldsymbol{w}$	$\hat{f}(\boldsymbol{x}^*)$	$f(\boldsymbol{x}^*)$
Linear	**0.7895**	**1.9668**	0.6254	0.7458	2.9144	2.9154
Cubic	1.5639	4.0642	0.0286	0.1901	1.4198	7.5340
Thin-plate	15.8758	43.4549	0.0094	0.0018	-6.6375	7.5340
Gaussian	3.5866	5.8306	0.7230	0.0361	1.0089	7.5340
Inverse-Multiquadric	4.2174	5.7001	**0.8260**	0.0261	0.6092	7.5340

4.1 Statistical comparison

Although the bootstrapped means and the associated confidence intervals provide some knowledge about the difference of the performances obtained by the different surrogate algorithms, they do not give a clear rank of the results obtained, especially in the case where the confidence intervals overlap, as in Figure 3 (a). In order to compare the surrogates, a meta-analysis approach based on a hypothesis test of the surrogates mean difference using the two sample Welch's t-test as statistical test has been performed, see [21]. The hypothesis test has been used to compute the p-value of each surrogate pair comparison (i.e. for each surrogate five comparison), resulting in a 6×6 matrix of p-values. A meta-analysis approach based on Fisher metric is then used to combine all of the individual p-values in each row into a single statistic for which a global p-value associated to each ensemble strategy can be computed, denoted as $P^F_{e_i}$, for $i = 1, 2, \ldots, 6$ (i.e. six ensemble strategies). Therefore a new vector of global p-values is defined as $\mathbf{P}^F = \begin{bmatrix} P^F_{e_1} & P^F_{e_2} & \ldots & P^F_{e_6} \end{bmatrix}$, and it is used to rank the ensemble performances. This procedure has been implemented inside the bootstrapped approach used to compute both the mean values and confidence intervals in Figure 2 (a). Following this procedure, statistical comparison between the surrogates analysed in this paper can be carried out, and the results for the Ackley function in 2 dimensions are summarised in Figure 2 (b, c), where the surrogates performance are now ranked based on the $\mathbf{P}^F$ vector (i.e. the best strategy is located on the left side of the x-axis and it is characterised by the lower $P^F_{e_i}$ value).
This analysis is used in this work to compare each optimisation strategy for all the test functions, dimensionalities and computational budget, and the numerical results are summarised in Tables 4, 5, and 6. In total 102 different optimisation problem are analysed. However, the results are related to only the last iteration, as the core of this numerical work is to study the ability of each ensemble to ultimately reach good quality designs within a close-ended optimisation scenario.

5 Numerical results

To compare the algorithms in a statistical manner, for each computational budget, function topology and dimensionality, a table based on mean value and confidence interval spread is constructed, and the performance in terms of μ and θ of each algorithm on each optimisation problem was ranked in the range 1-6. Then a Score-Rank in the range 6-36 is assigned as a figure of merit to each strategy, where 6 is the best value and 36 is the worst . This Score-Rank is computed by adding up the six rank values of μ or θ of each algorithm obtained on the various optimisation problems. These figures of merit are then collected in the last two columns of each table, and are used to characterise the qualitative global performance of each algorithm (a low Score Rank of both μ and θ indicates better optimisation results). Tables 4, 5, 6 and 7 clearly illustrate that when the computational budget is decreased, both of the constrained optimisation outperform the unconstrained versions. In the case of a high computational budget (i.e. $PD = 5$) all of the strategies perform in a similar way in terms of mean values, while in terms of consistency the combined ensemble seems to outperform the selective ensemble. Moreover, with a large sampling plan (i.e. $PD = 5$), the highly constrained version (i.e. $r^* \geq 0.75$) of both the ensemble consistently hamper the optimisation performance, resulting in the best approach only five times over the sixty cases analysed. Decreasing the computational budget has a strong impact on reducing the performance of the selective ensemble. In this case the combined ensemble is the best approach, in



(a) Ackley function (b) Linear RBF (c) Cubic RBF

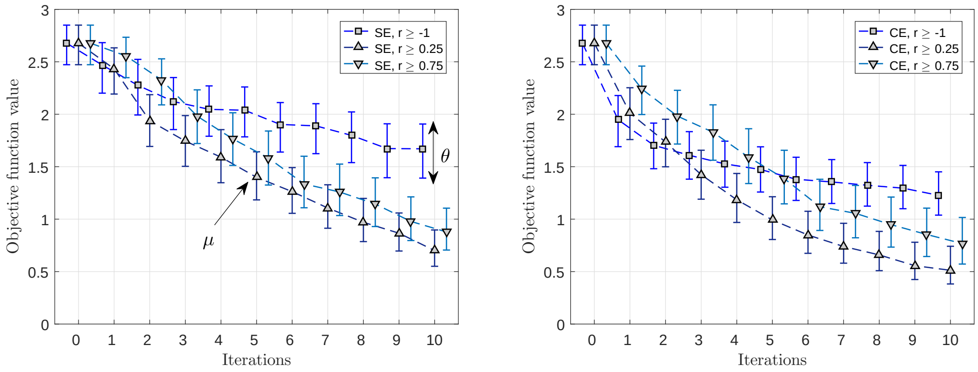
(d) Thin-plate RBF (e) Gaussian RBF (f) Inverse-Multiquadric RBF

(g) Combined Ensemble (h) $\min(\mathbf{R}_G)$

(i) Convergence history of combined ensemble (j) Convergence historye of selective ensemble

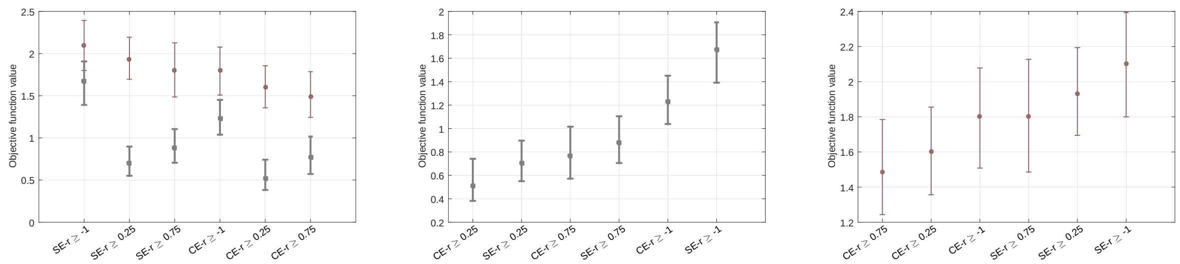
Figure 1: Contour plot of the Ackley function, set of RBFs, Combined Ensemble, minimum correlation coefficient trend, and convergence history of both unconstrained and constrained ensembles. ♦: infill point, $\boldsymbol{x}^*$, based on RBF prediction; ■, ▲, ▼, : infill point based on Combined Ensemble with $r^* \geq -1$, $r^* \geq 0.25$ and $r^* \geq 0.75$, respectively; •: 10 point Latin-Hypercube DoE. $\boldsymbol{x}^0$ indicates the best solution within the initial DoE, with $f(x^0) = 2.9018$.

regardless of the function complexity or dimensionality of the problem. Moreover, adding some local information based on the Pearson correlation coefficient helps to improve the final design, and at the same time consistently increase the rate of convergence. The r constraint is able to move the infill criteria



(a) Selective ensemble (b) Combined ensemble

Figure 2: Convergence history after the bootstrap approach of both the unconstrained and constrained ensemble strategies. The optimisation is based on 50 design of experiments each one of 10 initial points + 10 infill points. μ is the bootstrapped mean objective function value, while θ is the confidence interval spread (5000 resampling are used). The iteration number is centred on the constrained strategy with $r^* \geq 0.25$.



(a) Bootstrap results of $PD = 5$ and $PD = 2$ (b) Ranked bootstrap results, $PD = 5$ (c) Ranked bootstrap results, $PD = 2$

Figure 3: (a) error bar plot of the last iteration after the bootstrap approach, objective function: Ackley, dimensionality: 2, DoE: Latin hypercube, ■: $PD = 5$, •: $PD = 2$; (b, c) ranked bootstrap results after the meta-analysis approach.

to promising regions of the design space, even if the optimisation process starts with a small sampling plan. Interestingly, the constrained approach is less likely to get stuck in a local basin of attraction. However, an incorrect choice of the constraint level can jeopardise the optimisation, thereby a dynamic rule for the correct tuning process of the r value is needed as the optimisation progress.
In general the combined ensembles outperform the selective ensembles as Tables 7 and 8 clearly illustrate. Even if there is no single standout surrogate modelling approach which perform best on each of the 102 optimisation problems considered, Table 8 clearly shows that the combined ensemble with $r \geq 0.25$ gives the highest probability to end up with a final design in the first three rank positions (i.e. 83%), while the selective ensemble with $r \geq 0.75$ is characterised by the worst probability (i.e. 18%). These results indicate a clear advantage to creating a combined ensemble (unconstrained or constrained using the Pearson correlation coefficient) over using just the best surrogate within the same set of metamodels, as per the selective ensemble.

6 Conclusions

In the above study two different ensemble strategies have been investigated, a combined ensemble and a selective ensemble, with respect to their ability to reach as good a design solution as possible given a limited computational budget. Both the ensembles employ a novel update strategy based on the Pearson

Table 4: Results of statistical test between the ensemble strategies (low computational budget, $PD = 1$). In bold the best mean value obtained in each test function, and the best Score-Rank for μ, and θ respectively (the Score-Rank is a number in the range 6 to 36).

Algorithms	**Sphere**		**Trid**		**Rosenbrock**		**Ackley**		**Styblinski-Tang**		**Truss**		**Score-Rank**	
	μ	θ	μ	θ	μ	θ	μ	θ	μ	θ	μ	θ	μ	θ
Dimension = 5														
SE, $r \geq -1$	1.96E+01	5.74E+00	3.91E+02	1.17E+02	1.82E+02	4.36E+01	3.00E+00	6.67E-01	-1.04E+02	1.32E+01	**-7.78E+00**	9.22E-01	32	31
SE, $r \geq 0.25$	**5.72E+00**	2.44E+00	1.69E+02	5.75E+01	6.38E+01	3.49E+01	**2.49E+00**	5.43E-01	**-1.09E+02**	9.92E+00	-6.67E+00	9.90E-01	**21**	**19**
SE, $r \geq 0.75$	7.36E+00	2.86E+00	**1.43E+02**	5.62E+01	**5.23E+01**	2.03E+01	2.66E+00	5.50E-01	-1.05E+02	1.14E+01	-6.58E+00	1.02E+00	25	24
CE, $r \geq -1$	6.40E+00	3.03E+00	2.30E+02	8.41E+01	1.02E+02	4.39E+01	2.14E+00	4.07E-01	-1.05E+02	1.23E+01	**-8.09E+00**	8.22E-01	21	23
CE, $r \geq 0.25$	**3.43E+00**	1.14E+00	**8.03E+01**	4.70E+01	**2.98E+01**	7.68E+00	**2.06E+00**	4.46E-01	**-1.09E+02**	1.04E+01	-6.93E+00	9.79E-01	**9**	**10**
CE, $r \geq 0.75$	5.38E+00	2.80E+00	1.14E+02	5.29E+01	3.54E+01	1.95E+01	2.35E+00	5.12E-01	-1.05E+02	1.23E+01	-6.75E+00	9.95E-01	18	19
Dimension = 10														
SE, $r \geq -1$	**3.68E+00**	6.45E+00	9.33E+03	3.22E+03	2.69E+02	1.07E+02	**1.57E+00**	6.07E-01	**-2.07E+02**	1.75E+01	**-1.11E+01**	1.12E+00	**22**	31
SE, $r \geq 0.25$	9.04E+00	4.77E+00	**5.35E+03**	2.26E+03	1.21E+02	4.73E+01	2.54E+00	5.05E-01	-2.00E+02	2.00E+01	-9.74E+00	1.00E+00	29	**25**
SE, $r \geq 0.75$	1.31E+01	6.40E+00	5.48E+03	1.97E+03	**1.16E+02**	4.40E+01	2.50E+00	5.33E-01	-2.04E+02	1.94E+01	-1.00E+01	1.04E+00	27	**25**
CE, $r \geq -1$	1.39E+01	2.84E+00	4.81E+03	1.16E+03	1.54E+02	4.32E+01	1.56E+00	2.64E-01	-2.04E+02	2.01E+01	**-1.13E+01**	1.05E+00	21	20
CE, $r \geq 0.25$	**3.52E+00**	1.19E+00	**2.99E+03**	9.86E+02	1.01E+02	2.96E+01	**1.52E+00**	3.89E-01	**-2.10E+02**	1.90E+01	-1.01E+01	9.40E-01	**10**	**10**
CE, $r \geq 0.75$	4.28E+00	1.44E+00	3.34E+03	1.08E+03	**9.91E+01**	2.67E+01	1.73E+00	4.03E-01	-2.10E+02	2.15E+01	-9.65E+00	9.11E-01	17	15
Dimension = 15														
SE, $r \geq -1$	**1.54E+00**	6.86E-01	4.89E+04	1.94E+04	3.88E+02	2.10E+02	**1.29E+00**	5.30E-01	-2.87E+02	2.65E+01	**-1.26E+01**	1.09E+00	**25**	26
SE, $r \geq 0.25$	8.34E+00	6.11E+00	**2.95E+04**	7.68E+03	2.41E+02	1.47E+02	2.35E+00	5.79E-01	**-2.90E+02**	2.48E+01	-1.19E+01	1.15E+00	28	30
SE, $r \geq 0.75$	1.22E+01	7.87E+00	3.72E+04	1.34E+04	**2.13E+02**	1.10E+02	2.37E+00	5.48E-01	-2.88E+02	2.40E+01	-1.22E+01	1.04E+00	31	**25**
CE, $r \geq -1$	1.12E+01	3.74E+00	2.58E+04	6.08E+03	1.82E+02	4.44E+01	1.23E+00	2.31E-01	**-3.06E+02**	2.34E+01	**-1.31E+01**	9.68E-01	16	**13**
CE, $r \geq 0.25$	**4.16E+00**	1.36E+00	**2.06E+04**	4.48E+03	**1.47E+02**	3.69E+01	**1.04E+00**	2.86E-01	-3.01E+02	2.41E+01	-1.24E+01	1.14E+00	**10**	16
CE, $r \geq 0.75$	4.56E+00	1.92E+00	2.10E+04	4.97E+03	1.49E+02	2.94E+01	1.12E+00	2.72E-01	-2.98E+02	2.72E+01	-1.23E+01	1.04E+00	16	16
Dimension = 25														
SE, $r \geq -1$	1.46E+01	7.05E+01	5.08E+05	7.42E+05	1.09E+03	6.09E+02	**1.08E+00**	1.77E+00	**-4.87E+02**	9.14E+01	-2.23E+01	4.12E+00	**27**	33
SE, $r \geq 0.25$	**7.31E+00**	3.74E+00	6.11E+05	8.31E+05	5.71E+02	4.66E+02	2.67E+00	6.26E-01	-4.30E+02	7.16E+01	-2.25E+01	4.74E+00	28	27
SE, $r \geq 0.75$	3.24E+01	1.45E+01	**4.54E+05**	3.46E+05	**5.04E+02**	3.87E+02	2.53E+00	6.57E-01	-4.12E+02	5.72E+01	**-2.26E+01**	3.51E+00	29	**21**
CE, $r \geq -1$	1.21E+01	6.26E+00	**1.51E+05**	5.35E+04	2.03E+02	7.81E+01	8.28E-01	2.25E-01	**-4.57E+02**	7.20E+01	**-2.28E+01**	4.70E+00	**12**	17
CE, $r \geq 0.25$	9.24E+00	3.86E+00	2.28E+05	1.08E+05	2.05E+02	1.07E+02	9.98E-01	5.85E-01	-4.50E+02	6.31E+01	-2.27E+01	3.56E+00	18	17
CE, $r \geq 0.75$	**8.92E+00**	3.44E+00	2.19E+05	9.58E+04	**1.86E+02**	8.49E+01	**7.49E-01**	4.22E-01	-4.38E+02	3.35E+01	-2.27E+01	3.62E+00	**12**	**11**
Dimension = 40/50														
SE, $r \geq -1$	**6.98E+00**	8.79E-01	**5.72E+06**	1.39E+06	2.21E+03	8.99E+02	2.21E+00	1.70E+00	-7.13E+02	1.48E+02	**-3.01E+01**	6.75E+00	**19**	**23**
SE, $r \geq 0.25$	2.38E+01	1.92E+01	7.14E+06	6.39E+06	1.22E+03	9.30E+02	2.30E+00	1.19E+00	**-7.42E+02**	1.00E+02	-2.82E+01	6.86E+00	25	28
SE, $r \geq 0.75$	6.39E+01	3.46E+01	7.23E+06	6.20E+06	**1.21E+03**	9.37E+02	**1.82E+00**	3.84E-01	-7.13E+02	1.11E+02	-2.77E+01	6.55E+00	29	27
CE, $r \geq -1$	3.94E+01	1.28E+01	7.27E+06	4.43E+06	4.71E+02	1.57E+02	1.30E+00	2.35E-01	**-8.16E+02**	8.49E+01	**-2.90E+01**	3.54E+00	19	**14**
CE, $r \geq 0.25$	**2.25E+01**	2.73E+00	**6.08E+06**	1.25E+06	**3.56E+02**	8.20E+01	**1.05E+00**	1.54E-01	-7.94E+02	9.73E+01	-2.89E+01	8.32E+00	**11**	**14**
CE, $r \geq 0.75$	2.51E+01	1.98E+01	8.79E+06	1.15E+07	4.64E+02	5.14E+02	1.17E+00	5.20E-01	-7.87E+02	8.27E+01	-2.62E+01	3.32E+00	23	20

Table 5: Results of statistical test between the ensemble strategies (medium computational budget, $PD = 2$). In bold the best mean value obtained in each test function, and the best Score-Rank for μ, and θ respectively (the Score-Rank is a number in the range 6 to 36).

Algorithms	**Sphere**		**Trid**		**Rosenbrock**		**Ackley**		**Styblinski-Tang**		**Truss**		**Score-Rank**	
	μ	θ	μ	θ	μ	θ	μ	θ	μ	θ	μ	θ	μ	θ
Dimension = 2														
SE, $r \geq -1$	2.10E+00	5.94E-01	-1.20E+00	5.95E-01	1.25E+01	9.88E+00	2.10E+00	5.94E-01	-5.04E+01	9.24E+00	**-6.32E+00**	5.88E-01	25	25
SE, $r \geq 0.25$	1.93E+00	4.99E-01	**-1.70E+00**	2.40E-01	**2.95E+00**	2.31E+00	1.93E+00	4.99E-01	-5.24E+01	7.91E+00	-4.64E+00	9.27E-01	**16**	**13**
SE, $r \geq 0.75$	**1.80E+00**	6.42E-01	-1.14E+00	7.45E-01	6.71E+00	1.07E+01	**1.80E+00**	6.42E-01	**-5.25E+01**	7.41E+00	-4.32E+00	8.92E-01	19	23
CE, $r \geq -1$	1.80E+00	5.70E-01	-9.86E-01	6.67E-01	6.74E+00	5.71E+00	1.80E+00	5.70E-01	-5.04E+01	8.38E+00	**-6.09E+00**	7.60E-01	27	24
CE, $r \geq 0.25$	1.60E+00	4.99E-01	**-1.22E+00**	5.15E-01	**2.75E+00**	2.39E+00	1.60E+00	4.99E-01	**-5.22E+01**	7.57E+00	-4.69E+00	9.61E-01	**16**	**18**
CE, $r \geq 0.75$	**1.48E+00**	5.41E-01	-1.12E+00	7.84E-01	5.19E+00	9.12E+00	**1.48E+00**	5.41E-01	-5.21E+01	7.57E+00	-4.26E+00	8.97E-01	23	23
Dimension = 5														
SE, $r \geq -1$	1.13E+00	7.44E-01	8.47E+01	6.51E+01	5.35E+01	2.29E+01	1.92E+00	5.14E-01	-1.25E+02	1.18E+01	**-8.38E+00**	8.52E-01	26	28
SE, $r \geq 0.25$	**1.09E+00**	8.91E-01	**2.56E+01**	2.49E+01	3.15E+01	1.09E+01	**1.71E+00**	5.29E-01	**-1.30E+02**	1.04E+01	-6.88E+00	8.63E-01	**14**	**22**
SE, $r \geq 0.75$	2.01E+00	1.05E+00	5.70E+01	3.72E+01	**2.66E+01**	1.05E+01	2.10E+00	4.65E-01	-1.24E+02	1.10E+01	-6.63E+00	8.47E-01	31	23
CE, $r \geq -1$	5.39E+00	1.70E+00	4.41E+01	2.75E+01	3.21E+01	1.27E+01	1.76E+00	2.94E-01	-1.28E+02	1.26E+01	**-8.50E+00**	7.63E-01	22	22
CE, $r \geq 0.25$	**1.23E+00**	8.03E-01	**1.44E+01**	1.99E+01	**2.28E+01**	6.41E+00	**1.30E+00**	3.62E-01	**-1.29E+02**	1.13E+01	-7.08E+00	8.97E-01	**11**	17
CE, $r \geq 0.75$	1.75E+00	7.38E-01	4.27E+01	2.99E+01	2.53E+01	7.34E+00	1.86E+00	4.43E-01	-1.28E+02	1.07E+01	-6.79E+00	7.76E-01	22	**14**
Dimension = 10														
SE, $r \geq -1$	6.36E-01	1.19E+00	3.03E+03	1.62E+03	6.35E+01	4.71E+01	1.22E+00	4.30E-01	**-2.43E+02**	1.89E+01	**-1.29E+01**	9.56E-01	23	34
SE, $r \geq 0.25$	**1.90E-01**	1.75E-01	**7.12E+02**	5.76E+02	**4.77E+01**	1.90E+01	**1.14E+00**	2.48E-01	-2.34E+02	1.59E+01	-1.25E+01	8.70E-01	**17**	19
SE, $r \geq 0.75$	7.84E-01	5.41E-01	8.04E+02	4.11E+02	5.17E+01	1.64E+01	1.73E+00	4.01E-01	-2.23E+02	1.22E+01	-1.21E+01	8.64E-01	29	**17**
CE, $r \geq -1$	3.67E+00	1.81E+00	9.16E+02	4.33E+02	5.02E+01	1.33E+01	1.03E+00	1.98E-01	**-2.49E+02**	1.87E+01	**-1.31E+01**	6.97E-01	18	19
CE, $r \geq 0.25$	**6.96E-01**	3.77E-01	**5.10E+02**	2.13E+02	4.52E+01	1.02E+01	**8.63E-01**	2.04E-01	-2.38E+02	1.78E+01	-1.25E+01	8.96E-01	**13**	**14**
CE, $r \geq 0.75$	1.14E+00	5.63E-01	9.93E+02	5.89E+02	**4.43E+01**	1.06E+01	1.14E+00	2.79E-01	-2.28E+02	1.45E+01	-1.16E+01	1.01E+00	26	23
Dimension = 15														
SE, $r \geq -1$	**3.99E-02**	3.49E-02	3.86E+04	2.21E+04	8.26E+01	3.06E+01	1.55E+00	8.25E-01	**-3.27E+02**	2.70E+01	-1.52E+01	1.05E+00	29	29
SE, $r \geq 0.25$	1.04E-01	7.11E-02	8.99E+03	5.77E+03	**7.67E+01**	5.53E+01	**1.05E+00**	3.76E-01	-3.24E+02	2.18E+01	**-1.56E+01**	7.93E-01	**23**	**20**
SE, $r \geq 0.75$	7.72E-01	4.44E-01	**7.66E+03**	3.78E+03	7.88E+01	7.49E+01	1.23E+00	4.47E-01	-3.08E+02	1.75E+01	-1.53E+01	9.10E-01	29	25
CE, $r \geq -1$	1.81E+00	7.61E-01	5.02E+03	1.88E+03	**6.70E+01**	1.14E+01	**5.68E-01**	1.30E-01	**-3.61E+02**	2.47E+01	-1.56E+01	8.65E-01	13	20
CE, $r \geq 0.25$	**5.58E-01**	2.85E-01	**4.06E+03**	1.43E+03	7.11E+01	1.64E+01	7.16E-01	1.37E-01	-3.53E+02	2.37E+01	**-1.60E+01**	7.82E-01	**11**	**14**
CE, $r \geq 0.75$	9.54E-01	5.35E-01	6.31E+03	2.32E+03	7.42E+01	1.61E+01	7.96E-01	1.22E-01	-3.29E+02	2.38E+01	-1.53E+01	8.29E-01	21	18
Dimension = 25														
SE, $r \geq -1$	**8.32E-02**	4.33E-02	3.57E+05	1.24E+04	**9.81E+01**	3.12E+01	1.25E+00	1.98E+00	**-5.81E+02**	1.23E+02	-2.43E+01	4.29E+00	21	**18**
SE, $r \geq 0.25$	2.92E-01	1.43E-01	**2.67E+05**	9.11E+04	1.77E+02	1.01E+02	**6.89E-01**	2.68E-01	-5.33E+02	7.72E+01	**-2.67E+01**	3.64E+00	**18**	21
SE, $r \geq 0.75$	1.39E+00	5.69E-01	2.72E+05	1.19E+05	3.03E+02	3.85E+02	9.41E-01	2.91E-01	-4.77E+02	6.20E+01	-2.64E+01	5.52E+00	30	28
CE, $r \geq -1$	**7.67E-01**	6.84E-01	**3.83E+04**	4.03E+04	**1.03E+02**	3.50E+01	**2.66E-01**	1.10E-01	-5.26E+02	1.03E+02	-2.56E+01	4.97E+00	16	21
CE, $r \geq 0.25$	1.20E+00	1.31E+00	4.10E+04	2.17E+04	1.07E+02	4.63E+01	7.01E-01	1.75E-01	**-5.40E+02**	6.71E+01	**-2.69E+01**	4.66E+00	**15**	**19**
CE, $r \geq 0.75$	3.18E+00	1.95E+00	8.17E+04	3.83E+04	1.50E+02	1.18E+02	7.30E-01	2.01E-01	-4.88E+02	5.63E+01	-2.60E+01	3.15E+00	26	**19**
Dimension = 40/50														
SE, $r \geq -1$	**4.15E+00**	5.53E-01	5.61E+06	9.44E+05	2.02E+03	5.91E+02	**9.81E-01**	8.94E-02	**-8.38E+02**	6.35E+01	-3.41E+01	6.62E+00	**21**	**19**
SE, $r \geq 0.25$	5.31E+00	1.06E+00	5.25E+06	7.62E+05	1.06E+03	8.01E+02	1.52E+00	1.08E+00	-7.96E+02	1.14E+02	**-3.43E+01**	5.24E+00	26	27
SE, $r \geq 0.75$	4.70E+01	2.91E+01	**4.52E+06**	1.41E+06	**9.36E+02**	7.31E+02	1.94E+00	7.30E-01	-7.64E+02	6.43E+01	-3.08E+01	5.42E+00	32	27
CE, $r \geq -1$	**4.21E+00**	8.89E-01	2.19E+06	4.77E+05	2.88E+02	6.07E+01	**9.81E-01**	8.18E-02	**-9.31E+02**	1.00E+02	-3.39E+01	9.95E+00	**12**	20
CE, $r \geq 0.25$	4.36E+00	7.60E-01	**2.02E+06**	4.42E+05	**2.59E+02**	4.40E+01	1.11E+00	8.20E-02	-8.91E+02	9.36E+01	**-3.50E+01**	7.25E+00	**12**	**16**
CE, $r \geq 0.75$	9.42E+00	1.41E+00	3.21E+06	5.51E+05	3.38E+02	9.05E+01	1.07E+00	6.09E-02	-8.01E+02	5.33E+01	-3.14E+01	6.78E+00	23	17

Table 6: Results of statistical test between the ensemble strategies (high computational budget, $PD = 5$). In bold the best mean value obtained in each test function, and the best Score-Rank for μ, and θ respectively (the Score-Rank is a number in the range 6 to 36).

Algorithms	Sphere		Trid		Rosenbrock		Ackley		Styblinski-Tang		Truss		Score-Rank	
	μ	θ	μ	θ	μ	θ	μ	θ	μ	θ	μ	θ	μ	θ
Dimension = 2														
SE, $r \geq -1$	**6.00E-07**	5.01E-07	**-2.00E+00**	2.21E-06	**8.03E-01**	3.83E-01	1.67E+00	5.15E-01	**-6.82E+01**	4.29E+00	**-6.77E+00**	0.00E+00	**14**	**17**
SE, $r \geq 0.25$	4.22E-03	1.16E-02	-1.90E+00	2.42E-01	8.53E-01	5.48E-01	**7.00E-01**	3.46E-01	**-6.82E+01**	4.34E+00	-6.21E+00	6.42E-01	18	23
SE, $r \geq 0.75$	1.33E-02	2.35E-02	-1.78E+00	2.77E-01	1.46E+00	6.47E-01	8.78E-01	3.99E-01	-6.71E+01	4.05E+00	-6.25E+00	6.09E-01	26	22
CE, $r \geq -1$	**7.06E-06**	1.13E-05	**-2.00E+00**	1.31E-05	**8.36E-01**	3.53E-01	1.23E+00	4.13E-01	**-6.66E+01**	4.04E+00	**-6.77E+00**	0.00E+00	**15**	**11**
CE, $r \geq 0.25$	2.49E-03	2.97E-03	-1.89E+00	2.32E-01	1.25E+00	1.21E+00	**5.12E-01**	3.59E-01	-6.49E+01	4.24E+00	-6.31E+00	6.28E-01	21	21
CE, $r \geq 0.75$	1.66E-02	6.11E-02	-1.77E+00	2.84E-01	1.82E+00	1.10E+00	7.67E-01	4.44E-01	-6.50E+01	4.28E+00	-6.19E+00	6.53E-01	32	32
Dimension = 5														
SE, $r \geq -1$	**2.44E-04**	2.65E-04	**-3.00E+01**	2.34E-02	1.04E+01	5.63E+00	1.08E+00	4.31E-01	-1.56E+02	1.16E+01	-1.01E+01	5.16E-01	23	25
SE, $r \geq 0.25$	3.83E-04	3.10E-04	**-3.00E+01**	2.65E-02	**6.72E+00**	2.30E+00	**5.49E-01**	1.72E-01	**-1.59E+02**	1.03E+01	-1.02E+01	5.95E-01	**15**	**21**
SE, $r \geq 0.75$	3.91E-03	3.66E-03	-2.61E+01	3.24E+00	6.93E+00	1.83E+00	6.09E-01	1.95E-01	-1.53E+02	1.19E+01	**-1.05E+01**	3.91E-01	24	23
CE, $r \geq -1$	**3.07E-04**	2.89E-04	**-3.00E+01**	3.98E-02	7.98E+00	2.94E+00	7.22E-01	3.39E-01	**-1.62E+02**	1.02E+01	-9.95E+00	4.50E-01	22	23
CE, $r \geq 0.25$	4.09E-04	2.94E-04	**-3.00E+01**	2.18E-02	**7.43E+00**	2.74E+00	**3.30E-01**	1.19E-01	-1.59E+02	9.69E+00	-1.04E+01	3.60E-01	**17**	**12**
CE, $r \geq 0.75$	2.67E-03	1.98E-03	-2.48E+01	3.37E+00	8.47E+00	2.60E+00	4.52E-01	2.05E-01	-1.47E+02	8.26E+00	**-1.05E+01**	4.04E-01	25	22
Dimension = 10														
SE, $r \geq -1$	4.31E-03	2.15E-03	1.74E+01	2.17E+02	2.41E+01	7.61E+00	1.47E+00	7.16E-01	-3.04E+02	1.93E+01	-1.40E+01	8.13E-01	29	30
SE, $r \geq 0.25$	**1.92E-03**	1.08E-03	**-5.85E+01**	1.28E+02	2.05E+01	7.27E+00	**5.31E-01**	1.86E-01	**-3.05E+02**	1.99E+01	-1.45E+01	7.65E-01	**19**	25
SE, $r \geq 0.75$	8.57E-03	5.01E-03	-1.13E+02	3.74E+01	**2.00E+01**	4.89E+00	5.56E-01	1.62E-01	-2.77E+02	1.96E+01	**-1.50E+01**	4.90E-01	22	**19**
CE, $r \geq -1$	8.67E-03	3.23E-03	**-5.92E+01**	1.31E+02	**2.05E+01**	4.42E+00	**1.87E-01**	2.86E-01	**-3.21E+02**	1.46E+01	-1.46E+01	5.59E-01	17	19
CE, $r \geq 0.25$	**3.35E-03**	8.36E-03	-1.35E+02	3.65E+01	2.36E+01	4.93E+00	3.00E-01	9.39E-02	-3.02E+02	1.55E+01	-1.51E+01	5.39E-01	**15**	**16**
CE, $r \geq 0.75$	6.61E-03	3.74E-03	-4.87E+01	4.71E+01	2.54E+01	5.35E+00	3.17E-01	9.88E-02	-2.78E+02	1.79E+01	**-1.55E+01**	4.23E-01	24	17
Dimension = 15														
SE, $r \geq -1$	**1.20E-03**	3.61E-04	9.15E+02	1.42E+03	**3.09E+01**	8.40E+00	**1.00E-01**	4.49E-02	-4.80E+02	2.48E+01	-1.75E+01	7.38E-01	**18**	**18**
SE, $r \geq 0.25$	3.17E-03	1.31E-03	9.96E+02	1.50E+03	**3.09E+01**	1.20E+01	1.21E-01	5.01E-02	**-4.83E+02**	2.35E+01	**-1.77E+01**	8.22E-01	19	27
SE, $r \geq 0.75$	1.34E-02	1.27E-02	**1.26E+02**	5.59E+02	4.13E+01	5.22E+01	3.69E-01	1.08E-01	-4.07E+02	2.58E+01	**-1.77E+01**	7.68E-01	25	31
CE, $r \geq -1$	**2.10E-03**	5.28E-04	2.07E+02	2.12E+02	**3.70E+01**	7.72E+00	**2.09E-02**	8.31E-03	**-4.91E+02**	1.99E+01	-1.74E+01	8.61E-01	**16**	**13**
CE, $r \geq 0.25$	2.92E-03	1.25E-03	**1.03E+02**	2.46E+02	4.35E+01	8.24E+00	2.10E-01	5.56E-02	-4.63E+02	2.10E+01	-1.76E+01	8.18E-01	21	19
CE, $r \geq 0.75$	1.40E-02	4.00E-03	2.36E+02	1.41E+02	4.99E+01	8.99E+00	2.14E-01	3.13E-02	-4.09E+02	2.44E+01	**-1.79E+01**	7.55E-01	27	18
Dimension = 25														
SE, $r \geq -1$	**3.01E-02**	8.72E-03	5.71E+04	6.50E+04	5.70E+01	6.99E+01	**7.88E-02**	2.92E-02	**-7.55E+02**	1.38E+02	-2.93E+01	4.17E+00	**19**	27
SE, $r \geq 0.25$	3.49E-02	8.70E-03	7.40E+04	1.41E+05	4.67E+01	1.76E+01	9.87E-02	2.80E-02	-7.29E+02	6.55E+01	-2.94E+01	2.64E+00	22	**20**
SE, $r \geq 0.75$	9.38E-02	3.76E-02	**4.81E+04**	1.23E+05	**4.43E+01**	7.71E+00	4.48E-01	2.34E-01	-5.63E+02	5.41E+01	**-2.97E+01**	2.96E+00	25	23
CE, $r \geq -1$	**2.33E-02**	6.88E-03	**5.52E+03**	6.62E+03	**6.08E+01**	1.03E+01	**5.47E-02**	1.49E-02	**-8.00E+02**	5.89E+01	-2.84E+01	2.45E+00	**14**	**11**
CE, $r \geq 0.25$	3.95E-02	1.39E-02	6.28E+03	5.89E+03	7.44E+01	1.39E+01	1.99E-01	1.12E-01	-7.40E+02	9.18E+01	**-3.02E+01**	5.32E+00	19	25
CE, $r \geq 0.75$	1.10E-01	3.15E-02	6.53E+03	2.36E+03	8.50E+01	2.27E+01	3.08E-01	6.77E-02	-5.66E+02	5.77E+01	-3.01E+01	2.83E+00	27	20
Dimension = 40/50														
SE, $r \geq -1$	3.47E+00	4.21E-01	**2.96E+06**	1.15E+06	1.81E+02	2.75E+01	**7.55E-01**	1.03E-01	-1.03E+03	2.12E+02	-4.12E+01	8.25E+00	19	22
SE, $r \geq 0.25$	**3.43E+00**	5.78E-01	3.44E+06	7.33E+05	**1.70E+02**	1.38E+01	8.17E-01	1.13E-01	**-1.14E+03**	1.56E+02	**-4.34E+01**	6.92E+00	**17**	**18**
SE, $r \geq 0.75$	3.85E+00	7.57E-01	3.96E+06	1.55E+06	1.97E+02	2.62E+01	1.62E+00	1.04E+00	-9.20E+02	1.58E+02	-4.01E+01	9.18E+00	32	31
CE, $r \geq -1$	**3.22E+00**	2.66E-01	1.26E+06	1.81E+05	2.06E+02	2.84E+01	**8.02E-01**	1.40E-01	**-1.19E+03**	9.49E+01	**-4.59E+01**	9.14E+00	**13**	20
CE, $r \geq 0.25$	3.41E+00	4.45E-01	**1.14E+06**	2.21E+05	**1.91E+02**	2.60E+01	8.22E-01	9.56E-02	-1.17E+03	1.77E+02	-4.21E+01	7.35E+00	15	**16**
CE, $r \geq 0.75$	3.92E+00	4.54E-01	1.18E+06	1.55E+05	3.00E+02	8.09E+01	1.03E+00	1.12E-01	-9.46E+02	7.06E+01	-3.82E+01	9.06E+00	30	19

Table 7: Total Score-Rank considering the impact of the computational budget. The best and worst Score-Rank are 36 and 216 for $PD = 5, 2$ respectively, while for $PD = 1$ they are 30 and 180.

	Computational budget		
	PD = 5	PD = 2	PD = 1
SE, $r \geq -1$	120	145	125
SE, $r \geq 0.25$	110	114	131
SE, $r \geq 0.75$	154	170	141
CE, $r \geq -1$	**99**	108	89
CE, $r \geq 0.25$	108	**78**	**58**
CE, $r \geq 0.75$	165	141	86

Table 8: Ensembles total Score-Rank at each ranking position. In bold the surrogate appearing with greatest frequency (the maximum achievable frequency is 102 equal to the total number of optimisation problems analysed).

	Rank Position					
	1st	2nd	3rd	4th	5th	6th
SE, $r \geq -1$	19	17	10	8	13	**35**
SE, $r \geq 0.25$	11	18	19	**27**	19	8
SE, $r \geq 0.75$	4	5	10	22	**33**	28
CE, $r \geq -1$	29	22	14	14	13	10
CE, $r \geq 0.25$	**30**	**24**	**31**	13	3	1
CE, $r \geq 0.75$	9	16	18	18	21	20

correlation coefficient, r, using two different level of constraint, $r^* \geq 0.25$ and $r^* \geq 0.75$. In total six surrogate-based optimisation strategies have been compared and investigated, considering as a design of experiments a Latin-hypercube approach, six test function, six dimensionalities and three computational budgets. Statistical tests have been used to estimate the confidence intervals and to test a statistical hypothesis for the difference of surrogate best design. In particular a bootstrap resampling technique coupled with a meta-analysis approach based on combined p-values computed using the Fisher approach has been used to rank the surrogate strategies.

The results presented show that in general the constrained version of both the ensembles outperform consistently the unconstrained versions, especially when the computational budget is small. These results underline the importance to incorporate into a search strategy a measure of the local agreement between the surrogates available. In conclusion, when encountering a black-box optimisation problem for the first time without any prior knowledge of the nature of the landscape, the application of the combined ensemble presented here, offers a reasonable level of optimisation performance whilst negating the impact of selecting a poor or inappropriate surrogate model within the set available, like in the case of selective ensemble. Moreover, in most of the case analysed, adding some local information to the global accuracy gathered from a CV approach, can improve significantly both the performance and the rate of convergence. However, an incorrect choice of the level of the correlation constraint can hamper the optimisation process, thereby a tuning approach able to evaluate the correct r constraint, as the optimisation progress, seems to be a reasonable future study.

7 Acknowledgements

We thank Rolls-Royce plc. for the financial support received as a part of the University Technology Center in Computational Engineering. The encouragement and advice of Dr I. Voutchkov of the University of Southampton, is also greatly appreciated.

References

[1] J. Sacks, W. Welch, T. Mitchell, and H. Wynn, "Design and analysis of computer experiments," *Statistical Science*, pp. 433–435, 1989.

[2] B. Glaz, T. Goel, L. Liu, P. Friedmann, and R. Haftka, "Multiple-surrogate approach to helicopter rotor blade vibration reduction," *AIAA Journal*, vol. 47, no. 1, pp. 271–282, 2009.

[3] A. Keane, "Wing optimization using design of experiment, response surface, and data fusion methods," *Journal of Aircraft*, vol. 40, no. 4, pp. 741–750, 2003.

[4] S. Leary, A. Bhaskar, and A. Keane, "Global approximation and optimization using adjoint computational fluid dynamics codes," *AIAA Journal*, vol. 42, no. 3, pp. 631–41, 2004.

[5] D. Toal, A. Keane, D. Benito, J. Dixon, J. Yang, M. Price, T. Robinson, A. Remouchamps, and N. Kill, "Multi-fidelity multidisciplinary whole engine thermo-mechanical design optimization," *Journal of Propulsion and Power*, vol. 30, no. 6, pp. 1654–1666, 2014.

[6] A. Keane, "Statistical improvement criteria for use in multiobjective design optimization," *AIAA Journal*, vol. 44, no. 4, pp. 879–891, 2006.

[7] A. Forrester, A. Sobester, and A. Keane, "Multi-fidelity optimization via surrogate modelling," *Proceedings of the Royal Society A*, vol. 463, no. 2088, pp. 3251–3269, 2007.

[8] R. Myers, D. Montgomery, and C. Anderson-Cook, *Response Surface Methodology: Process and Product Optimization Using Designed Experiments.* Wiley, 2011.

[9] N. Cressie, *Statistics for spatial data.* J. Wiley, 1993.

[10] C. Bishop, *Pattern Recognition and Machine Learning.* Springer, 2006.

[11] G. Fasshauer, *Meshfree Approximation Methods with MATLAB.* World Scientific, 2007.

[12] B. Schlkopf and A. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond.* MIT Press, 2002.

[13] E. Acar and M. Rais-Rohani, "Ensemble of metamodels with optimized weight factors," *Structural and Multidisciplinary Optimization*, vol. 37, no. 3, pp. 279–294, 2009.

[14] F. A. C. Viana, R. T. Haftka, and L. T. Watson, "Efficient global optimization algorithm assisted by multiple surrogate techniques," *Journal of Global Optimization*, vol. 56, no. 2, pp. 669–689, 2013.

[15] F. A. C. Viana, R. T. Haftka, and V. Steffen, "Multiple surrogates: how cross-validation errors can help us to obtain the best predictor," *Structural and Multidisciplinary Optimization*, vol. 39, no. 4, pp. 439–457, 2009.

[16] K. Deb, *Multi-Objective Optimization Using Evolutionary Algorithms.* Wiley, 2001.

[17] I. Voutchkov, A. Keane, and R. Fox, "Robust structural design of a simplified jet engine model, using multiobjective optimization," 2006.

[18] J. Momin and Y. Xin-She, "A literature survey of benchmark functions for global optimization problems," *CoRR*, vol. abs/1308.4008, 2013.

[19] A. Keane and P. Nair, *Computational Approaches for Aerospace Design: The Pursuit of Excellence.* Wiley, 2005.

[20] D. Toal and A. Keane, "Non-stationary kriging for design optimization," *Engineering Optimization*, vol. 44, no. 6, pp. 741–765, 2012.

[21] D. Montgomery and G. Runger, *Applied Statistics and Probability for Engineers, Student Solutions Manual.* John Wiley & Sons, 2010.