

# Black-box calibration for complex systems simulation

BY ALEXANDER I J FORRESTER

Predicting or measuring the output of complex systems is an important and challenging part of many areas of science. If multiple observations are required for parameter studies and optimization, accurate, computationally intensive predictions or expensive experiments are intractable. This paper looks at the use of Gaussian process based correlations to correct simple computer models with sparse data from physical experiments or more complex computer models. In essence, physics based computer codes and experiments are replaced by fast problem specific statistics based codes. Two aerodynamic design examples are presented. First a cheap two dimensional potential flow solver is calibrated to represent the flow over the wing of an unmanned air vehicle. The rear wing of a racing car is then optimized using rear wing simulations calibrated to include the effects of the flow over the whole car.

**Keywords:** co-Kriging; CFD; aerodynamics; optimization; surrogate model

## 1. Statistical model based simulation

This paper is concerned with the problem of determining the effects of various parameters on the output of a system. This is a problem encountered in many, perhaps all, areas of science and lies at the heart of any optimization process. Systems for which effects are to be determined may be extremely complex and depend upon many parameters, for example the human body. Scientists endeavour to capture a systems behaviour via physical and computer experiments. If experiments are quick and cheap to run, then wide ranging parameter studies or lengthy optimization processes can be performed. However, cheap experiments are often unable to capture many of the complex effects which are likely to be of interest. For example, the performance of an aerodynamic device can be simulated in a fraction of a second by ignoring viscous effects. More accurate simulations may take weeks, while physical experiments require expensive models and wind tunnel time. Parameter studies and optimization using such methods are either very expensive or cursory in nature.

Hard to obtain data can be approximated using a surrogate model which is fitted to a set of known data. Instead of relying on results from experiments at all points of interest, we can fill in the gaps between a set of observed data by fitting a surrogate model, that is a model which can be used in lieu of directly simulating or observing the system. Such models are, essentially, curve fits through known data and may take many forms, the most popular being polynomial regression [Box and Draper, 1987]. Other examples include, though are not limited to, neural networks, radial basis functions [Broomhead and Loewe, 1988], support vector regression [Smola and Schölkopf, 2004] and Kriging (a more complex radial basis function method) [Sacks et al., 1989]. Once fitted to the data a surrogate can be used in lieu of calls to computationally intensive codes or experiments to expedite design space visualization, parameter studies, optimization, etc. The time savings afforded by using a surrogate are often at the expense of accuracy. Clearly if a polynomial model is employed, we are assuming that the complex system's response can be modelled as such. More complex methods, such as Kriging, which models the response as a sum of parametric Gaussian basis functions (with parameters chosen to maximize the likelihood of the data) make fewer assumptions, but may still require large amounts of data to predict the underlying result to the required accuracy. If data is expensive and so very sparse, the surrogate may

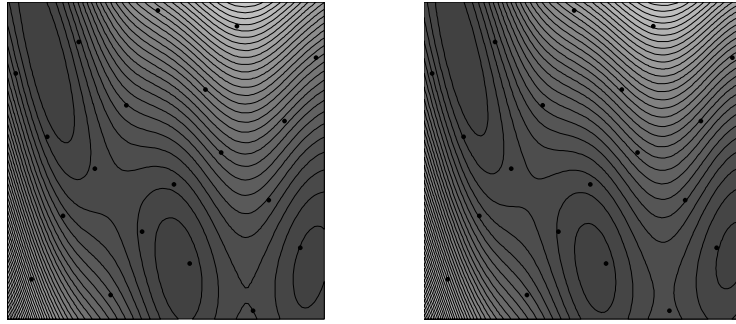


Figure 1. contours of the true Branin function (left) compared with a Kriging prediction (right) based on 20 sample points (dots) (from Forrester and Keane [2008]).

be inaccurate to the point of being of little practical use. As an example of how powerful the Kriging method can be, given sufficient data, figure 1 shows the contours of the Branin function (an analytic function often used to test optimization algorithms) and a Kriging prediction based on 20 observations. The similarity between the two contour maps is remarkable.

#### (a) *Co-Kriging*

Often simple physics-based simulations of systems are available, for example in §2 a simple potential flow solver will be used. We can therefore hope to obtain very accurate surrogates of the response of these codes, based on large quantities of data. We may also be able to simulate a small part of a system, which is of particular interest, at a reduced cost. For example, in §3 the rear wing of a race car is considered rather than the entire car geometry. If the error in these surrogates of simpler systems with respect to the true system response is simpler than the true system response itself we can gain accuracy by fitting a second surrogate to the error, based on a small set of expensive data and adding the two surrogates together. We then have a statistics based code of predicted behaviour plus black-box complex effects.

This way of adding a surrogate based on plentiful cheap data to a hopefully more simple surrogate based on sparse expensive data could employ any surrogate modelling method, but Kriging's multi-output variant, co-Kriging, with its parametric Gaussian basis functions is particularly attractive. This method originates in geostatistics [Krige, 1951, Cressie, 1993] where different ore grades found from core samples were correlated to give a better picture of gold distributions than possible by looking at gold in isolation. The idea being that factors which effect, say, zinc distribution are likely to effect the distribution of gold too.

More recently Hevesi et al. [1992] employed co-Kriging to predict average annual precipitation values near a potential nuclear waste disposal site using a sparse set of precipitation measurements from the region along with the more easily obtainable elevation map of the area. Altitude will naturally be correlated with precipitation, but may display a complex relationship, with features such as valleys and ridges needing to be identified, making the co-Kriging method particularly attractive for this application.

Kennedy and O'Hagan [2000] used co-Kriging to correlate finite element analyses based on a coarse mesh with those based on a fine mesh. They used a formulation which assumes the complex process is equal to the simple process, multiplied by some ratio, plus a difference process. The ratio and difference process are determined using a set of collocated data. In engineering design this method has been cast in a global optimization context and used to optimize a transonic wing design by combining an empirical drag prediction code with computational fluid dynamics (CFD) [Forrester et al., 2007].

In this paper we look at two particularly demanding calibration scenarios in the field of engineering design. Firstly in §2 we consider the calibration of simple CFD to predict full scale performance. In §3 we then consider a novel and industrially relevant case where the cheaper design simulation uses a simplified geometry, rather than the more usual simplified analysis.

The mathematics behind the co-Kriging method has been omitted for clarity. More details and the *Matlab* code used to produce the results in this paper can be found in Forrester et al. [2008].

## 2. Low-speed unmanned air vehicle wing design

While the availability of accurate turbulent flow simulations has improved in recent years, within an engineering design context, time and computer resources dictate that simulations of design alternatives may be limited to Euler or even potential flow simulations of simplified or incomplete geometries. Solutions of the Reynolds averaged Navier-Stokes (RANS) equations may also be employed on simplified geometry, though run-times preclude their use in extensive optimization processes. These, and more accurate simulations are employed for more detailed design of components. Similarly, the expense of physical experiments, e.g. wind tunnel tests, usually precludes their use in an optimization context.

In this example we will consider the lift/drag ( $L/D$ ) of the wing of a low-speed unmanned air vehicle (UAV) with a wing span of 1.4m operating at 18m/s with a Reynolds number of  $1.5 \times 10^5$  and a lift coefficient of 0.36. The wing has a constant aerofoil section throughout its span and it is the design of this aerofoil section which we are concerned with. The aerofoil is defined by six parameters which describe the shape of an upper and lower Ferguson spline (please see Sóbester and Keane [2007] for more details). We have 12 full scale wind tunnel tests at our disposal (chosen using a space filling Latin hypercube sampling plan [Morris and Mitchell, 1995]), which are unlikely to be sufficient to gain insight into the six dimensional design landscape. A Kriging prediction of  $L/D$  based on these wind tunnel data is shown in figure 2. Each sub-tile shows a clear trend, with higher  $L/D$  for lower values of  $T_A^{lower}$  (a sharper lower leading edge), but there is little variation from sub-tile to sub-tile or tile-to-tile, indicating that, although a key trend has been identified, there are insufficient data to capture trends due to all the parameter variations.

To augment the wind tunnel tests we can also use a 2D full potential code with an empirical drag correction (VGK, short for viscous Garabedian and Korn [Freestone, 1996]). Each 2D aerofoil simulation takes a few seconds and 120  $L/D$  values have been obtained. A Kriging prediction based on these data is shown in figure 3. Compared with figure 2, the trends in this plot are more complex, which we expect as there are more data. The trends are also starkly different, which we expect as the 2D potential flow model is a gross simplification of the flow over the 3D wing.

We hope to calibrate this grossly simplified model using the wind tunnel data and co-Kriging. If the Kriging model of the VGK data has captured at least some important effects, calibrating this model will prove more accurate than working with the wind tunnel data alone. The resulting co-Kriging model, with complex trends afforded by the density of cheap data, but accuracy in these trends afforded by the wind tunnel data, is shown in figure 4. Note how the key sub-tile trend from figure 2 has been retained (though slightly modified), but there are now more complex variations across the other variables. The leave-one-out cross-validation error<sup>†</sup> of this model is a 21% improvement compared with the Kriging model based on wind tunnel data alone and confirms the benefits of utilising the cheap data via co-Kriging.

Seemingly ignoring the underlying physics of a system, as we have in figure 4, could be considered a little controversial. However, although the physics is not being modelled directly, it is taken into account

<sup>†</sup> Validation using an independent set of data is desirable, but rarely practical when conducting physical or very expensive computer experiments. Leave-one-out cross-validation, where a model is assessed by its ability to back-predict successively omitted points, is though an indicator of generalization properties.

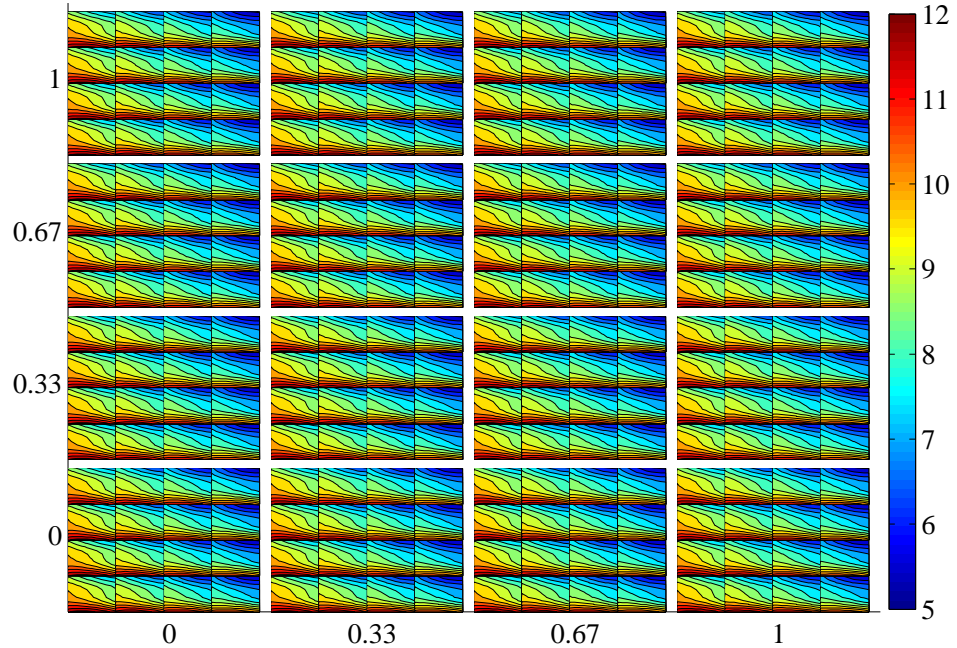


Figure 2. Kriging prediction of  $L/D$  based on data from 12 wind tunnel tests. Variation across all six dimensions is shown via nested axes. The tensions in the splines at the leading edge ( $T_A^{upper}$  and  $T_A^{lower}$ , which control the bluntness of the leading edge) vary on the  $x,y$  axis of each sub-tile. The tensions in the splines at the trailing edge ( $T_B^{upper}$  and  $T_B^{lower}$ , which determine the shape of the middle section of the airfoil) vary in each tile. The camber and boattail angles ( $\alpha_c$  and  $\alpha_b$ , which determine the camber and trailing edge bluntness) vary from tile-to-tile. All variable ranges are normalised between zero and one.

via the black-box calibration process. In essence figure 4 shows the results of a new, quick-running CFD code, specific to this problem, which takes into account all physical effects. The accuracy of the code is limited by the quantity and quality of underlying data (inevitably there are experimental errors in figure 4), though the Gaussian process based method employed does provide us with error estimates, which will come into play in §3 in the context of global optimization.

### 3. Race car wing design

In this study we consider the design of the rear wing of an open cockpit hill-climb race car – similar in appearance to a Formula One car–. We are interested in the effects of varying the geometry of the rear wing by changing the 12 design parameters shown in figure 5. The geometry of the car, along with boxes depicting areas of mesh refinement around the turbulent regions of the wheels, wings and wake for a 6,000,000 cell mesh is shown in figure 6. RANS simulations, using the one equation Spalart-Allmaras model in *Fluent*, based on this mesh take between 12 and 15 hours on an eight processor 16GB Linux node. Cluster scheduling dictates that only two nodes can be used at any one time. Constraints such as these on computing resources are representative of those experienced throughout academia and industry.

We hope to expedite the search of this twelve dimensional design space by augmenting full-car

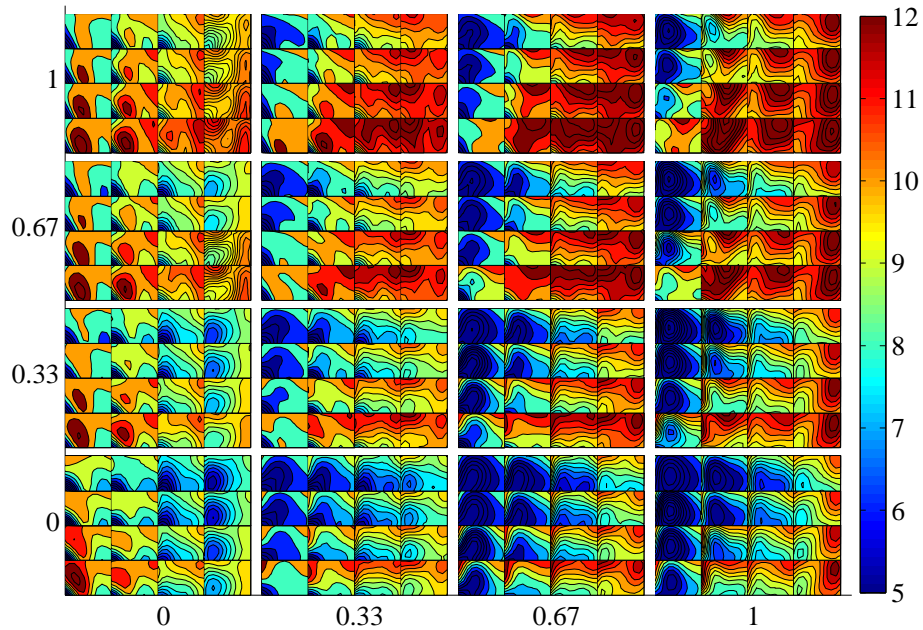


Figure 3. Kriging prediction of  $L/D$  based on data from 120 VGK simulations (figure 2 caption explains plot).

simulations with rear-wing-only simulations. The rear wing geometry and mesh refinement zones for a 1,500,000 mesh are shown in figure 6. Spalart-Allmaras simulations of this geometry, using inflow conditions taken from the baseline full-car simulation take approximately 1.5 hours on the computer resources used for the full-car simulations.

Naturally the simulated flow over such a simplified geometry (figure 7) is quite different from that which includes the rest of the car (figure 8). Nevertheless, by exploiting the fact that the difference between the effects of the two flows is simpler than the effects of the full-car flow itself, useful time savings can be had by employing the co-Kriging methodology.

Our first study of this problem compares a Kriging model built with  $L/D$  data from 20 full-car simulations with a co-Kriging model built using 10 full-car simulations and 120 rear-wing-only simulations. This represents a similar level of computing resource (CPU hours) for each method. The Kriging prediction of  $L/D$  based on 20 full-car simulations is shown in figure 9. For visualization purposes only four variables are shown. These are those which the co-Kriging model indicates have the greatest discrepancy between the cheap and expensive data (this is determined by examining the widths of the difference model basis functions). Based on such sparse data, the trends are very simple indeed, with only one variable leading to a significant variation in  $L/D$ .

Figure 10 shows a Kriging prediction based on the 120 rear-wing-only simulations. As in the UAV example, the higher density of data allows more complex trends to be derived. However, here the trends show some similarity to the full-car model in figure 9.

The co-Kriging prediction in figure 11 can be seen (visually) to combine aspects of both figure 9 and 10. Now there is a clear variation in  $L/D$  for all variables. This model will be used as a basis for the search for promising designs in a surrogate model based optimization process.

In this 12 dimensional design space the true global optimum is likely to remain elusive. It is perhaps

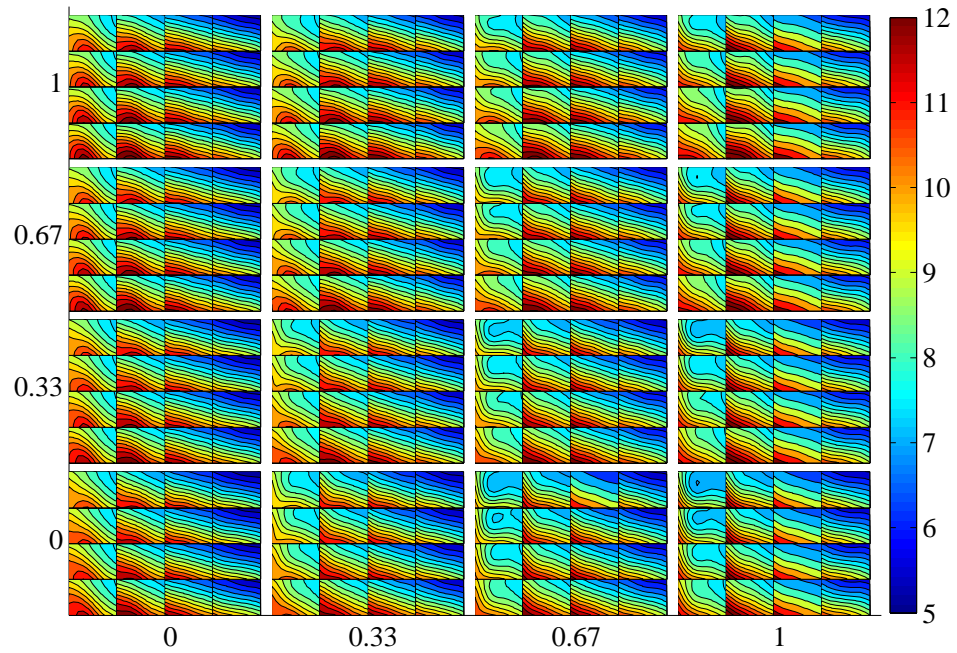


Figure 4. co-Kriging prediction of  $L/D$  based on data from 120 VGK simulations and 12 wind tunnel tests (figure 2 caption explains plot).

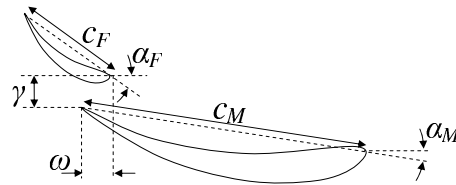


Figure 5. Rear wing design variables. The six parameters shown are varied for the upper and lower wing (figure 6 shows the location of the upper and lower wings), giving 12 parameters.

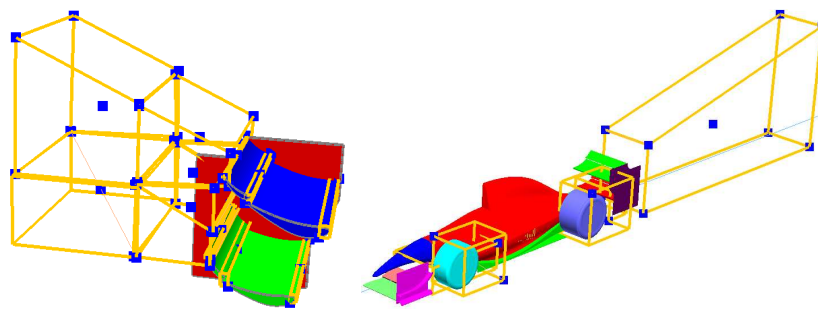


Figure 6. Hill-climb race car rear wing geometry (left) and half-car geometry (right, which is used for the “full-car” simulations). Yellow boxes depict areas of mesh refinement. Meshing is performed using *HARPOON*.



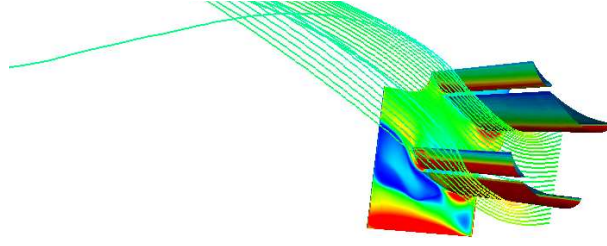


Figure 7. Simulated flow over a baseline rear wing in isolation. Streamlines are coloured by velocity magnitude and the filled contours on the wing represent skin friction coefficient.

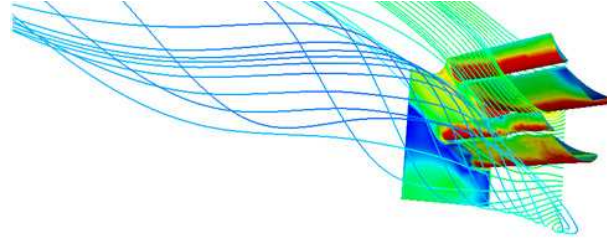


Figure 8. Simulated flow over a baseline rear wing behind the full-car.

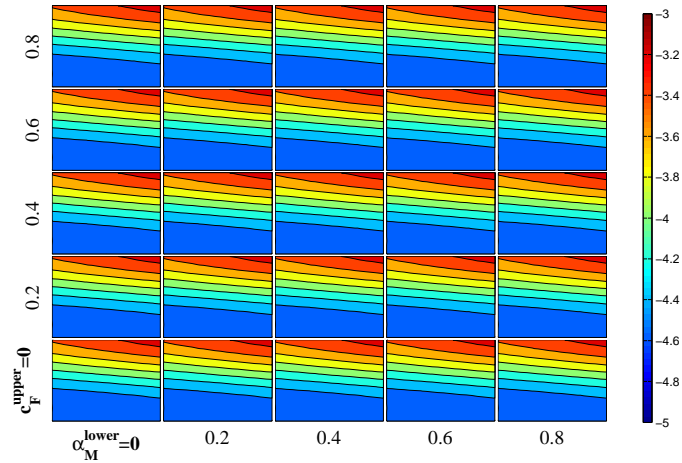


Figure 9. Kriging prediction of  $L/D$  based on 20 full-car simulations. Variation across four dimensions is shown via nested axes:  $\alpha_M^{upper}$  and  $\alpha_F^{upper}$  vary in each tile while  $\alpha_M^{lower}$  and  $c_F^{upper}$  vary from tile-to-tile. All variable ranges are normalised between zero and one.

worth taking a moment to consider the magnitude of the problem we face. Given one design variable we might need at least three  $L/D$  evaluations to gain some insight into how this variable affects  $L/D$ . To extend this sampling density to 12 variables, we would require  $3^{12} = 531,441$   $L/D$  evaluations. We only have 120 rear wing and 10 full-car simulations and so we can hope only to achieve a significant level of design improvement. A leave-one-out cross-validation correlation coefficient of  $r^2 = 0.90$  is encouraging though.

To decide upon a new promising design we maximize the expectation of improving on the best design so far [Schonlau, 1997]. The expected improvement can be calculated because a Kriging (and co-Kriging)

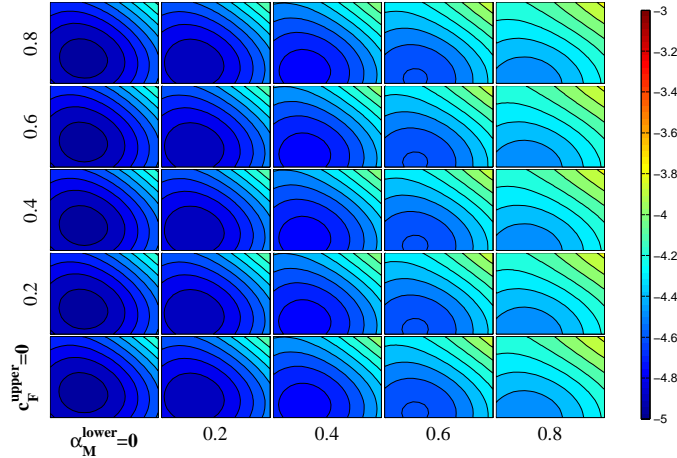


Figure 10. Kriging prediction of  $L/D$  based on 120 rear wing simulations (figure 9 caption explains plot).

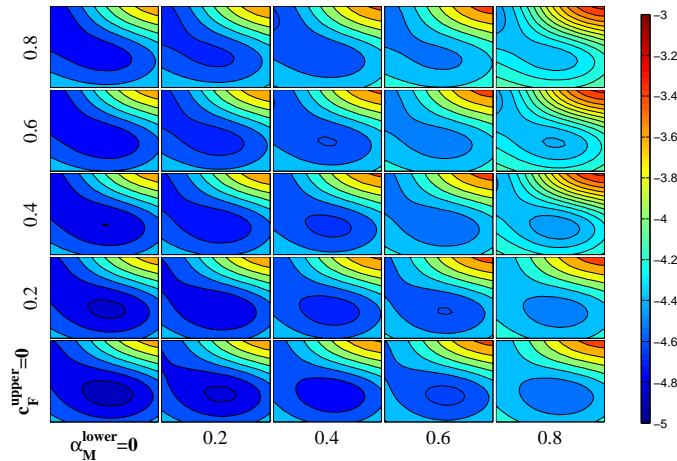


Figure 11. co-Kriging prediction of  $L/D$  based on 10 full-car and 120 rear wing simulations (figure 9 caption explains plot).

prediction represents the mean of a Gaussian process, that is, for any combination of design variables, it is the most probable value at that point, but other values do have probabilities. By integrating the probabilities below the best value so far we can then calculate the expected improvement that may be obtained at that point. Jones [2001] gives an excellent description of this method. This expected improvement, although a difficult, multi-modal function to search, can be computed in a fraction of a second and so its maximum can be found using an extensive genetic algorithm followed by sequential quadratic programming.

Once the 12 variable values which correspond to the design with the maximum expected improvement in  $L/D$  have been found, a new full-car and rear wing simulation are performed and the  $L/D$  values are appended to the data set, i.e. there are now data for 121 rear wing and 11 full-car simulations. The co-Kriging model is then re-built, the maximum expected improvement found, new simulations performed, and the process repeated until a convergence criterion is met, which is often simply that time allocated for the design optimization is exhausted.



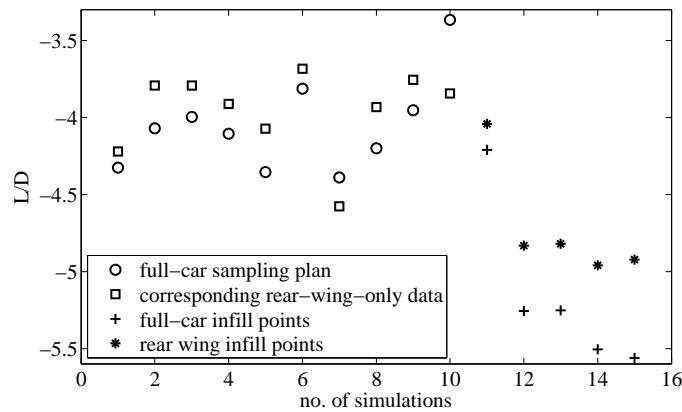


Figure 12.  $L/D$  values for the 10 full-car simulations and five infill points and their rear-wing-only simulation counterparts (it is the full-car values we are minimizing).

Here, five so-called *infill points* have been added to the original data set and their corresponding  $L/D$  values are shown in figure 12. A significant improvement over the designs in the original data set has been achieved. The rear-wing-only counterparts to the full-car simulations are also shown. Note how the  $L/D$  values follow a similar trend, with the rear-wing-only tending to have slightly higher (worse)  $L/D$ . Exceptions to this are points seven and 10, where the trend is reversed. Note also that the difference between the simulations becomes wider as the optimization progresses and that the optimum predicted by the full-car simulations is not the best point according to the rear-wing-only simulations. Clearly the black-box calibration process has captured the various unknown trends in order that design improvements can be made.

#### 4. Conclusions

Two applications of black box calibration for complex system simulation have been presented. The first shows the calibration of a simple 2D CFD code to predict full scale 3D experimental data. Far from being ignored, that which cannot be analysed and understood within permitted time and cost constraints is accounted for using a statistical model. Here effects due to, among others, 3D and turbulent flow and surface roughness have been embedded in the co-Kriging model. While efficient computer models which directly simulate complex systems are always desirable, this potential for constructing quick problem specific codes which may incorporate effects that are not understood is clearly attractive.

The second application tackled the case of component design – the rear wing of a race car – where the rest of a complex geometry – i.e. the car – may influence the result. Here the complex effects include the upstream effect of changes in the wing geometry and the effect of the wake of the car on the flow over the wing. For this example, the calibrated model is successfully used in a surrogate model based optimization processes. Despite the large difference between the flow over a rear wing in isolation and the whole car, the co-Kriging model yields design improvements with very few additional simulations.

Although this paper has examined aerodynamic applications of co-Kriging, as the methodology sees systems as black-boxes, it is naturally applicable to other areas of science. For example, the author is currently working on bio-mechanical models where musculoskeletal simulations using quick inverse dynamics are calibrated with experimental data. We have considered a single system output ( $L/D$ ), however, correlating and optimizing multiple quantities is also possible and further research in this area would be welcomed.

## 5. Acknowledgements

The author gratefully acknowledges the work of Alistair Ward and Nishant Taluja in computing data for the example applications.

## References

- E. P. Box and N. R. Draper. *Empirical Model Building and Response Surfaces*. John Wiley & Sons, 1987.
- D. S. Broomhead and D. Loewe. Multivariate functional interpolation and adaptive networks. *Complex Systems*, 2:321–355, 1988.
- N. A. C. Cressie. *Statistics for Spatial Data*. Probability and mathematical statistics. Wiley, revised edition, 1993.
- A. I. J. Forrester and A. J. Keane. Recent advances in surrogate-based optimization. *Progress in Aerospace Sciences*, 45(1-3):50–79, 2008.
- A. I. J. Forrester, A. Sóbester, and A. J. Keane. Multi-fidelity optimization via surrogate modelling. *Proc. R. Soc. A*, 463(2088):3251–3269, 2007.
- A. I. J. Forrester, A. Sóbester, and A. J. Keane. *Engineering Design via Surrogate Modelling: A Practical Guide*. John Wiley & Sons, Chichester, 2008.
- M. M. Freestone. VGK method for two-dimensional aerofoil sections. Technical Report 96028, Engineering Sciences Data Unit, 1996.
- J. Hevesi, A. Flint, and J. Istok. Precipitation estimation in mountainous terrain using multivariate geostatistics. part II: Isohyetal maps. *Journal of Applied Meteorology*, 31:677–688, 1992.
- D. R. Jones. A taxonomy of global optimization methods based on response surfaces. *Journal of Global Optimization*, 21:345–383, 2001.
- M. C. Kennedy and A. O’Hagan. Predicting the output from complex computer code when fast approximations are available. *Biometrika*, 87(1):1–13, 2000.
- D. G. Krige. A statistical approach to some basic mine valuation problems on the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Engineering Society of South Africa*, 52(6):119–139, December 1951.
- M. D. Morris and T. J. Mitchell. Exploratory designs for computational experiments. *Journal of Statistical Planning and Inference*, 43:381–402, 1995.
- J. Sacks, W. J. Welch, T. J. Mitchell, and H. Wynn. Design and analysis of computer experiments. *Statistical Science*, 4(4):409–423, 1989.
- M. Schonlau. *Computer Experiments and Global Optimization*. PhD thesis, University of Waterloo, Waterloo, Ontario, Canada, 1997.
- A. J. Smola and B. Schölkopf. A tutorial on support vector regression. *Statistics and Computing*, 14: 199–222, 2004.
- A. Sóbester and A. J. Keane. Airframe design via cubic splines Ferguson’s curves revisited. *In-fotech@Aerospace, Sonoma Wine County, CA*, 2007.