
Data Mining for Motorsport Aerodynamics

Matthew Watts, Robert Carrese and Hadi Winarto

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/50298>

1. Introduction

Engineering design is the practical application of scientific theory with the goal of a real world outcome in a reasonable timeframe. Until the recent past, conducting a physical experiment was the only way an engineer could measure the merits of any reasonably complex design. With the advent and uptake of computational processes this balance has shifted and simulation is now a reasonable alternative. Although engineers have been quick to adapt to the process of digital experimentation, the methods used in analysing the data gathered have not progressed with such vigour. Statistical methods used before the digital age are still commonplace in everyday practices and this seriously limits the efficiency and final output capable. Just as the methods of testing designs have changed, the analysis processes used there after must also change.

The primary difference between simulation and experimental data sets is the amount of information present; it is now common for an engineer to spend significant time and resources selecting the most valuable data for analysis while discarding the rest. Of course this is a problem of our own creation; it would prove more prudent to use analysis tools that can handle large data sets rather than selective examination. Development of new techniques is a significant challenge, fortunately computer science has become adept at analysis of large data sets. It follows that the tools used within this field will prove useful for engineering and give rise to new design processes. There exists a wide variety of analysis tools that could be applicable. Previously the driving force behind this selection has been the requirement to arrive at a singular design that meets the user's requirements. Optimisation schemes fill this condition nicely and accordingly have seen use within various design disciplines. The relative success of this approach has polarised computational engineering design, indeed the coupling of optimisation and simulation software is now standard practice. However, a key failing of this approach is that although a honed designed is reached the path taken to this result is ultimately lost. This feature of optimisation is a big

problem for engineers; if the performance of intermediate designs is unknown, it becomes impossible to confidently adapt the final design to new conditions or features not enforced within the original optimisation scheme. Furthermore although the final design exhibits high fitness it may prove hard to evaluate why that particular configuration outperforms others without first understanding the path taken towards it. Other problematic features of optimisation codes include the limited use of user knowledge, heavy computational cost and in some cases doubt over the algorithms ability to converge to the global optimum. These latter shortcomings have been addressed, with varying success, through the application of evolutionary algorithms. Of course the primary concern of preserving and understanding the intermediate results remains. There is an emerging process that shows great promise in solving this all important problem and does so through the application of data mining techniques.

Termed design space exploration, this process identifies the relationship between the variables (design space) and the results (solution space) in both a qualitative and quantitative manner. This allows a user to not only understand the effect variables have upon the required outcome but also provides a concise method of assuring the design space encompasses the parameter bounds for optimum performance. Most importantly this provides flexibility both in the approach taken by the user and the final result. In real terms this means an engineer can be confident in the robustness of a design and even allow for quick modifications if needed.

2. Problem statement

Although the design process is a well worn path for classical fields of engineering, new computational techniques are redefining the way we approach these problems. The analysis methods once used are no longer the most efficient processes and especially when evaluating simulation data. The use of data mining techniques shows great promise in this application yet remains relatively unproven.

3. Application area

Current computational methods cannot achieve all that is experimentally capable and thus possible design disciplines are limited. Computational Fluid Dynamics (CFD) proves a pertinent choice with aerodynamic design a topical subject and relative assurance in the obtained results. To ensure engineering relevance a standardised application was considered. With a large bank of documented geometry generation methods, aerodynamic shape design and in particular aerofoil design fulfils this requirement. As the recent past has seen this field defined through the expansive use of optimisation schemes, the merit of data mining can be assessed against a competing methodology.

For the modern racing vehicle the dependency upon aerodynamic performance is paramount to overall success and in no field is this more prominent than Formula One. This

may seem an unlikely proving ground for the aerodynamic discipline, however the range of differing tracks requires a vehicle of dynamic performance and when coupled with an unrelenting development cycle this yields a challenging problem. The front wing of a modern Formula One car has an enormous effect upon the overall vehicle performance and particularly that of the aerodynamic regime. Vehicle stability and the flow fields of all downstream components hinge upon this element making it one of the most significant aerodynamic devices on the entire vehicle [1]. The flow profile of the wing is dictated by the proximity to the ground; given the low ride heights seen within this category, the operation is confined to the ground effect regime where flow characteristics can be unpredictable [1]. For this study the area of application was restricted to the outboard sections of the main plane (Figure 1); in line with the 2010 Formula One technical regulations, the centre wing portion must conform to a symmetrical profile as specified by the governing technical body [2].

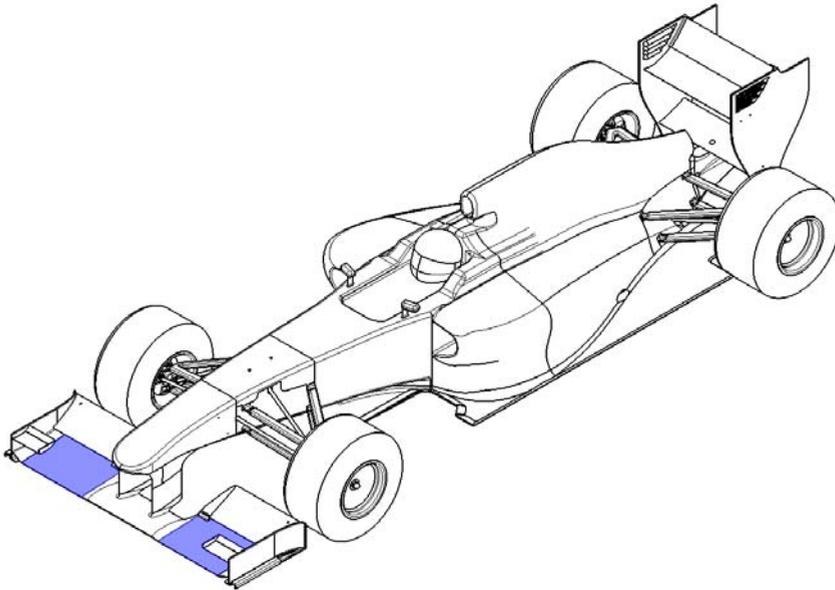


Figure 1. Generic Formula One racing car highlighting the area of application.

To enable a chordwise study, it was assumed that the wing section remains constant over the span and thus local section changes near the endplate are not considered. To simulate real flow conditions encountered by a Formula One vehicle a cornering case study was examined (Detailed in following sections).

4. Research status

Applying data mining techniques within the engineering field is not a new development, however applications to the design discipline have remain limited. Work has focused upon

production processes, fault detection, maintenance and decision support [3 - 5]. The lack of application to design and manufacturing stems from the difficulty of implementing new systems into pre-existing processes and the uncertainty in payback [6]. To overcome this, several case studies have been performed [7 & 8], although the methods used did not highlight a clear advantage over classic approaches and seemed removed for the common engineer. One data mining technique that has proven more successful and intuitive is Self-Organising Map (SOM); the ability to visualise high-dimensional data sets while preserving topological information proved the alluring characteristic [9]. Unfortunately, the initial usage of SOM did not take advantage of all available features and instead saw primary application as a visualisation tool. This limitation can be attributed to the new state of the technique and the inability to yield quantitative data [9]. As SOM is not a complete analysis tool, applications frequently coupled it with other existing methods and in particular optimisation schemes [10]. This approach has given rise to a unique hybrid design methodology where data mining, in particular SOM, is integrated in the design process as a secondary analysis technique [11]. Although this approach addresses the deficiency in analysing computational data sets, the successfulness is heavily dependant upon the user's programming skill. User's must have a fundamental understanding of the underlying mathematics and even poses the ability to create new software where off-the-shelf solutions prove unworkable. For the practical engineer this is an unwarranted difficulty that only complicates the design process further.

A simpler approach yielding the same result is to incorporate data mining as the sole analysis method. As it proves difficult to assess qualitative and quantitative information from one such process [12], it becomes necessary to couple two complimentary techniques; SOM for qualitative and Analysis Of Variance (ANOVA) for quantitative data. This approach falls into the category of design exploration which is becoming an ever expanding method in engineering design. Application of this process is still in the early stages but has seen success [12 - 14]. Aerodynamic design has been approached [12 & 13] however focus has centred upon benign flow environments and small data sets. The ability of design exploration as a methodology for practical application and not simply a high level research tool has yet to be conclusively shown. As such this chapter will focus upon the merit of design space exploration for the practical engineer through application to a particularly challenging aerodynamic problem.

5. Research model

In this section the steps taken to complete a design exploration will be covered. Although there is no fixed process to achieve this, the format implemented by [13] proves a good starting point. In this study data is gathered via a simulation technique, augmented through a surrogate model and finally data mining techniques are used to gain the required information about the design space. For the simulation process a computational flow solver is employed while SOM and ANOVA are used for data mining – a schematic of this approach is outlined within Figure 2.

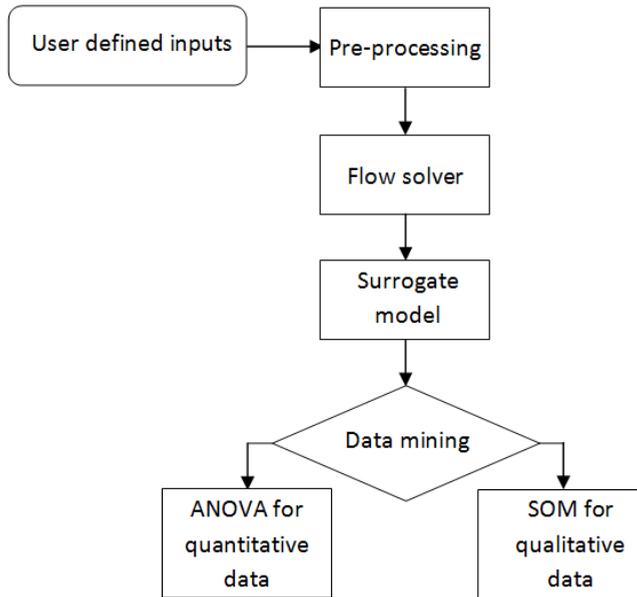


Figure 2. Design space exploration flowchart.

Given the topical and preliminary nature of this research, the development of the techniques required to carry out a design exploration are just as important as the results derived from it. However as a final step a singular aerofoil will also be extracted from the optimum design space to better highlight the applications of this process.

5.1. Application

The first step of any design exploration is to determine what performance data is required and what design properties are important. These choices ultimately form the objectives (solution space) and variables (design space) respectively. This is a concurrent process that requires a firm understanding of both the engineering problem and the proposed analysis methods.

5.1.1. Objective functions

The selection of the objective functions is a highly influential step which helps determine the effectiveness and extent of the information that can be derived. An intelligent decision must be made by the user to ensure the objective functions have the greatest relevance to the case under examination. For this aerodynamic study, non-dimensional performance coefficients highlight the aptitude of the aerofoil section while ensuring data from the simulated domain can confidently be extrapolated to the real world. To understand which of these quantities prove most important, the dynamics of a Formula One vehicle must be considered.

To enable the quickest tour of a racing track, an engineer must maximise the average speed of the vehicle around the circuit, to do this the cornering and top speeds must be increased

as much as possible. Of these two profiles the cornering case proves most critical given the typical high frequency of bends within a modern racing track. To maximise the speed a car can traverse a given corner, the lateral adhesion of the vehicle must be increased. This is primarily achieved through aerodynamic means [1]. Given the relation between aerodynamic force and dynamic pressure (or velocity) is not linear, we can see that reasonable increases in lift coefficient lead to large increases in cornering speeds (Figure 3). This relationship clearly indicates why Formula One teams spend exuberant amounts of time and money in search of small aerodynamic gains. Thus lift coefficient proves a highly important parameter for this study.

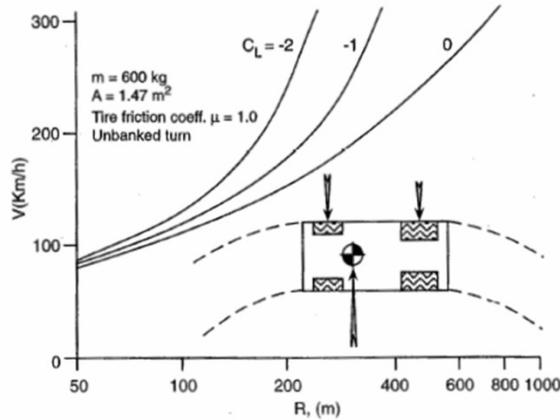


Figure 3. Relationship between cornering speed and radius with varying lift coefficient, figure courtesy of [1].

From the basic lift equation (Equation 1) it can be seen that lift coefficient is highly dependant upon flow velocity. This inherent relationship proves problematic for racing applications where a constant magnitude of downforce is required during cornering; a vehicle must slow to negotiate a typical corner and in doing so will experience a reduced magnitude of downforce. This behaviour is counter productive and it follows that an ideal vehicle would exhibit a low change in downforce for corresponding flow velocity changes.

$$C_l = \frac{2L}{\rho v^2 S} \tag{1}$$

Of course the parameters cannot be assessed in isolation. Top speed is also an important parameter for a competitive vehicle and thus drag shall be assessed. Unlike traditional aeronautic design this quantity will serve mainly as a screening parameter to ensure no undue flow effects are encountered within the optimal design area. Thus the objective functions considered within this study as (listed in decreasing importance):

- Section lift coefficient
- Section lift coefficient sensitivity to Reynolds number:
- Section drag coefficient

5.1.2. Cornering case study

Formula One vehicles have a large range of operational speeds and so a singular velocity cannot be implemented within the design phase. This situation is only augmented by the relatively low Reynolds numbers experienced by a Formula One vehicle; small changes in the local flow field lead to large changes in aerodynamic performance [15]. This is no more prevalent than during cornering where significant changes in downforce and hence grip are witnessed. Given the assessment of a large range of speeds (Reynolds numbers) is not possible within the confines of this study, a more discrete approach was considered whereby the entry/exit of a corner and the apex conditions will be assessed:

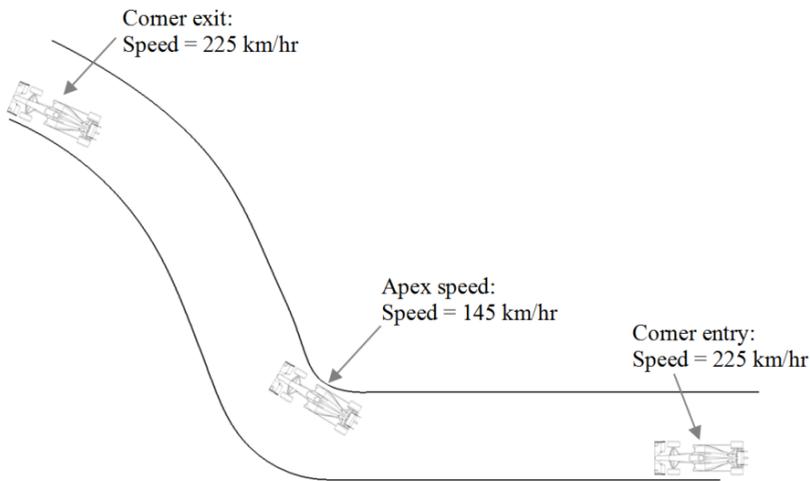


Figure 4. Cornering case study.

This combination is a typical medium speed corner that is frequently placed just before a long straight, this creates a situation whereby the speed with which the vehicle traverse the corner determines the maximum speed and thus the ability defend or overtake a competitor. For ease, entry and exit conditions were assessed jointly and thus two discrete Reynolds numbers are required. Taking sea level conditions and a characteristic length equal to a Formula One chord of 220 mm [2], the Reynolds number can be calculated:

$$\text{Re} = \frac{\rho v L^*}{\mu} \quad (2)$$

This results in a Reynolds number of 600,000 for apex conditions and 940,000 at the entry and exit of the corner.

5.1.3. Design variables

Standardised functions are the primary method of aerofoil section specification; several methods exist differing mainly in the control parameters defined. Aerofoil coordinate

functions are defined in terms of X and Z parameters, or sometimes X and Y , that are non-dimensionalised with respect to chord and thus in general terms:

$$Z = F_j(p, X) \tag{3}$$

Where p is a parameter vector of n terms and F_j is a function defined by these parameters and a switch j . In basic terms; the goal is to exercise a high level of geometric control for the smallest value of n . The most prominent of these functions being the NACA series; the spline interpolation used within these functions limits the possible shapes generated and thus are not applicable within this study. Alternatively the PARSEC method was used, here a 6th order polynomial is enforced:

$$Z_{PARSEC} = \sum_{n=1}^6 a_n(p) \cdot X^{n-1/2} \tag{4}$$

Where the coefficients a_n are prescribed from the required geometric parameters (Figure 5) and the coordinate function applied independently for the upper and lower surfaces. The ability to manipulate the PARSEC method also leads to a number of complex alterations previously unworkable; leading and trailing edge modification, introducing time variable functions (wing warping), ridge/separation-bubble additions and much more. The standard PARSEC methods contains 11 variables, however a reduction to 10 was made within this study. The parameter removed is that of trailing edge wedge thickness (ΔY_{TE}), this parameter adds significant complexity during analysis and is also not representative of Formula One style aerofoils. Furthermore as the primary goal is to force the vehicle into the ground, the geometry was inverted. Given the highly competitive nature of Formula One, the confidentiality surrounding the technologies employed is paramount. Locating reliable data upon the various aerofoil sections used proves extremely difficult. To arrive at a suitable baseline, a method termed reverse shape fitting was used.

Variable	Definition
r_{LE}	Leading edge radius
X_{up}	Upper crest location
Y_{up}	Upper crest
X_{lo}	Lower crest location
Y_{lo}	Lower crest
$Y_{xx,up}$	Curvature of upper crest
$Y_{xx,lo}$	Curvature of lower crest
Y_{TE}	Trailing edge coordinate
α_{TE}	Trailing edge direction
β_{TE}	Trailing edge wedge angle
ΔY_{TE}	Trailing edge wedge thickness

Table 1. Definition of PARSEC parameters.

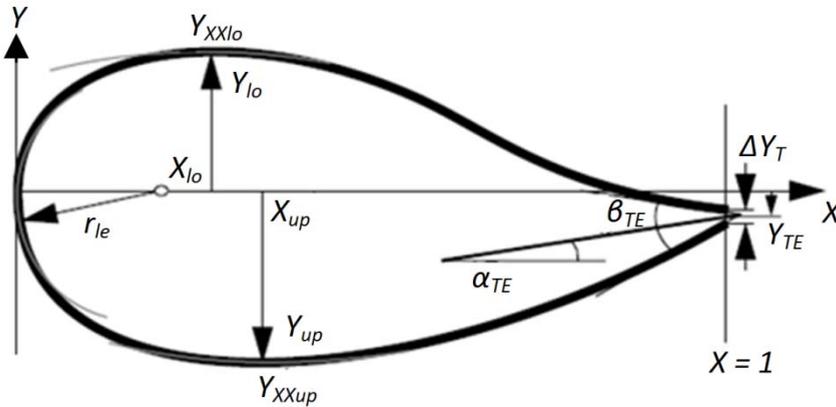


Figure 5. Modified PARSEC aerofoil geometry defined through 11 parameters [16].

In this method a known applicable aerofoil is reproduced by the PARSEC method and the associated variable values recorded. A standard deviation of 8 to 10% is applied for each value and certified through intersection tests; no connection between the upper and lower surfaces was to occur. The baseline aerofoil used here being the NASA GA(W) LS(1)-0413 due to its use upon the, now outdated, 1998 Tyrrell 026 Formula One racing car [17]. The final bounds upon the geometrical parameters are:

Variable	Range
r_{LE}	(0.0055 → 0.0215)
X_{up}	(0.2873 → 0.5343)
Y_{up}	(0.0880 → 0.1194)
X_{lo}	(0.3060 → 0.5076)
Y_{lo}	(-0.064 → -0.050)
$Y_{xx,up}$	(-1.0294 → -0.418)
$Y_{xx,lo}$	(-0.0486 → -0.8204)
Y_{TE}	(-0.3579 → 0.0231)
α_{TE}	(0.02 → 0.257)
β_{TE}	(-0.020 → 0.020)

Table 2. PARSEC parameter bounds.

5.2. Formulation

Now that the required variables and outcomes have been selected, the method of analysis must be decided. For this study a four part process was used; first a finite number of aerofoils (800) were selected from the available design space, each aerofoil was then placed into a representative domain and meshed, the fluid flow was subsequently evaluated through a CFD code and the required data recorded. Finally the design and performance

data was used to train a surrogate model that provides the final data for analysis. These four steps will now be covered in further detail.

5.2.1. Sampling plan

It is intuitive that in any modelling problem the higher the number of variables under analysis the more objective function measurements required. The main concern is this relationship is not a linear one; if a single variable problem is sampled in n -locations to attain the desired accuracy, for an expanded space of k -dimensions, n^k observations will be required to achieve the same result [18]. Thus it is not feasible to assess all possible designs within the pre-allocated variable bounds and thus a limited approach must be considered. To get the most out of the available computational time, the finite simulations that can be assessed must best represent the entirety of the design space.

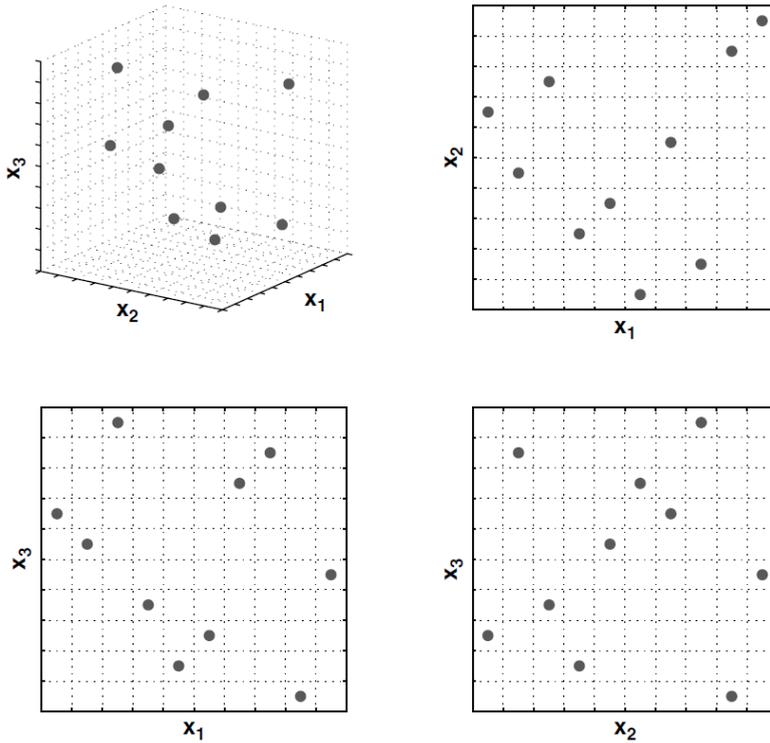


Figure 6. Latin hypercube sampling plan of a three variable problem with ten discrete points selected. The two-dimensional projections highlight that each row and column only contains one sampling point, figure courtesy of [18].

The statistical method of Latin Hypercube Sampling (LHS) was developed to fulfil this very need for a multidimensional distribution [19]. Since the inception of this technique, LHS has become a standard sampling plan implemented within statistical and engineering fields

alike. The two-dimensional form of this type of sampling can be easily understood through the Latin Square, however the multidimensional extension of this is somewhat more complex. This is performed by splitting the design space into equal sized hypercubes (bins) and placing one point in each. Making sure that from each occupied bin it is possible to exit the design space along any direction parallel with any axes without encountering any other occupied bins [18]. For a 10 dimensional case, as considered here, this is quite a complex relationship to visualise; the basic premise is easily understood for a three-dimensional space in Figure 6.

5.2.2. Pre-processing (domain and mesh generation)

Although the recent use of extensive CFD simulations has allowed the assessment of almost any flow field without aforethought, the initial steps during model creation and setup are highly important and can ultimately determine the effectiveness of the simulation. Here Gambit was implemented for pre-processing duties.

To attain real world application the simulation settings must represent actual conditions as closely as possible and thus the FIA technical regulations must be enforced; a minimum ride height of 75 mm and an average Formula One chord of 220 mm leads to a height-to-chord (h/c) ratio of 0.34 [2]. Limitations upon the ride height and angle of attack are also enforced and thus realised through fixed values. Finally to ensure a complete flow field is analysed, the nearest non-ground surface is placed 10 chord lengths from the aerofoil (Figure 7), this measure also guarantees boundary condition are not violated.

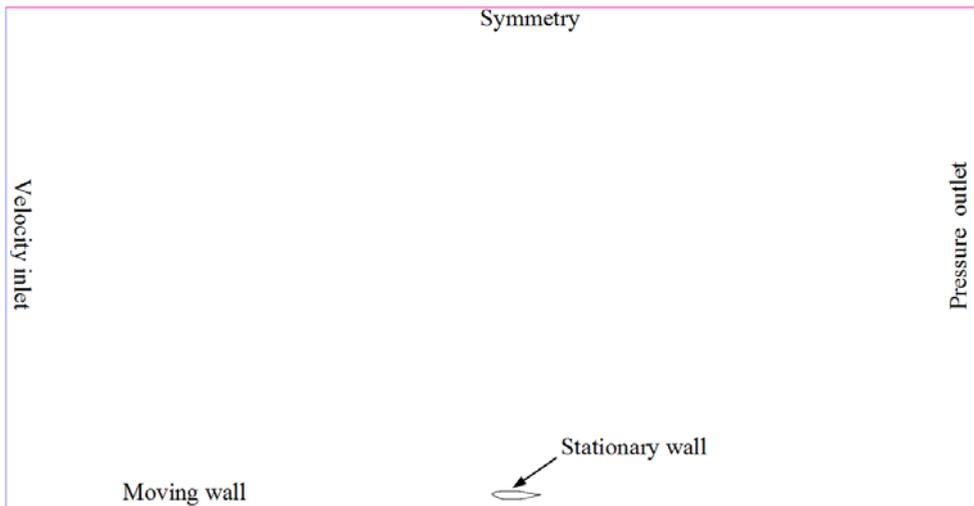


Figure 7. Computational domain and boundary conditions.

For aerofoil simulations the boundary layer mesh is a critical component within transient flow predictions, this is a direct result of the geometric dependency of near-wall functions employed

by turbulence models [20]. To ensure all close-wall flow phenomena are captured a 10 row boundary layer was extended from the surface. For the remaining domain, local flow field changes prove minimal and thus an unstructured triangular mesh is implemented (Figure 8).

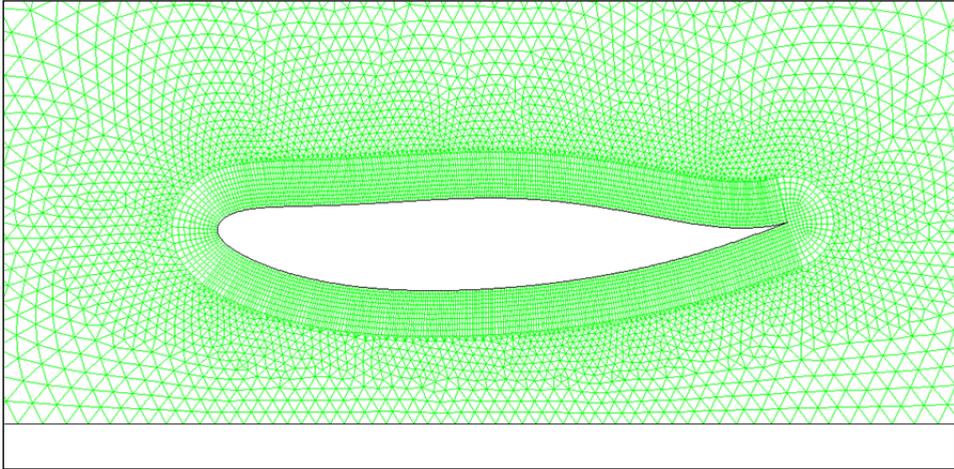


Figure 8. Characteristic aerofoil mesh.

5.2.3. Flow solver

The choice upon the solver was based upon several criteria although the simulation accuracy proves to be the overriding issue. As such the Reynolds Averaged Navier-Stokes (RANS) based commercial software package of Fluent was implemented. Although important, the mathematics behind this software package will not be discussed here in the interest of brevity. However the reader must understand that the equations governing fluid flow are currently unsolvable, modern CFD codes merely approximate the numerical solution to these equations and thus the final results are also an estimate of real world flows [20]. This simplification becomes particularly evident when dealing with the possibility of turbulent flow. To ensure confidence in the simulation results, several differing techniques can be employed within current CFD codes. These numerical techniques, termed turbulence models, are formed from real world experimental data. The aim of these models is to accurately predict the viscous fluid behaviour with as little impact to computational burden as possible. Two-equation models form the backbone of turbulence models in current usage with $k-\epsilon$ and $k-\omega$ the most widely implemented. The primary difference between these two methods being the experimental data used within the model. Naturally there have been several modifications upon the baseline models with the methods considered within this study being:

- Standard $k-\epsilon$
- RNG $k-\epsilon$
- Realisable $k-\epsilon$
- Standard $k-\omega$

5.2.4. Readers interested in the specifics of these models and CFD mathematics should consult [20] for further reading

Although several guidelines do exist for the use of these models in certain flow conditions, ground-effect modelling remains somewhat arbitrary and hence a study into the most apt model was conducted. This was achieved through a parametric approach where simulation results were compared to experimental data for a NACA0015 section at 0° angle of attack and a Reynolds number of 1.5 million [21]. As proximity to the ground greatly alters the aerodynamic performance, a range of height-to-chord ratios were simulated for each turbulence model.

From Figure 9 and 10, the deficiencies that plague ground effect simulations are clear. Although reasonable agreement is seen for the behaviour of both lift and drag, magnitudes are routinely over or under predicted. It is key to note that this deficiency becomes more pronounced as the aerofoil moves towards the ground. Ground-effect flows exhibit adverse pressure gradients that increase almost exponentially as the ground is approached. This can clearly be seen in the plots where divergence is experienced below a height-to-chord ratio of 0.3. This result highlights how viscous effects dominant at very close ground proximity [21]. For very low height-to-chord values (< 0.2), the boundary layer will grow to cover the entire region between the aerofoil and the ground. Above a height of $0.4c$ the behaviour is much more agreeable notwithstanding the drag over predictions. This discrepancy arises do to two main sources; the comparison between the two-dimensional simulation domain and the real world three-dimensional experimental data and turbulence model deficiencies where small inaccuracies have major effects upon viscous drag [21]. Overall from Figure 9 and 10 it can be seen that the $k-\epsilon$ realisable model proves the best scheme to implement given the reasonable agreement for both lift and drag predictions at the prescribed h/c of 0.34. Although the $k-\omega$ model yields a slightly more accurate lift prediction, the extreme over prediction of drag is unacceptable, as is the complete model breakdown witnessed at a height below $0.2c$.

After the selection of the best turbulence model, the final flow solver settings were made. As outlined previously, two discrete Reynolds Number cases were evaluated; 600,000 and 940,000. Flow properties were set to reflect smooth operating in both cases conditions; that is a representative still air case where inlet turbulence is modelled at 0.3% and outlet settings at standard atmospheric pressure.

5.2.5. Surrogate model

The last step during formulation is the construction and use of surrogate models. The principle of a surrogate is very simple; to replace a complex simulation code with a much faster mathematical approximation. Here a surrogate model is used to augment the simulation data obtained from CFD and allow a higher fidelity within the final results. Colloquially this ‘fills in the gaps’ between simulation data thus allowing a very quick assessment of a large number of designs while maintaining accuracy. Although this method sounds easy enough, great care must be taken during execution to ensure the approximations made are done so in a high confidence while ensuring that the model remains magnitudes faster than actual simulations.

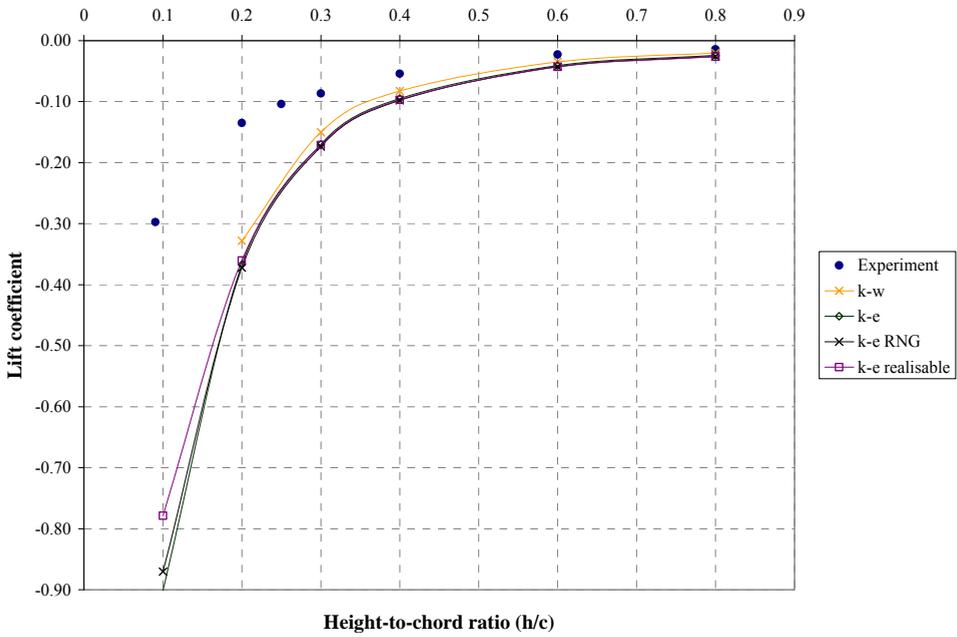


Figure 9. Lift comparison for a NACA0015 in ground effect at zero angle of attack.

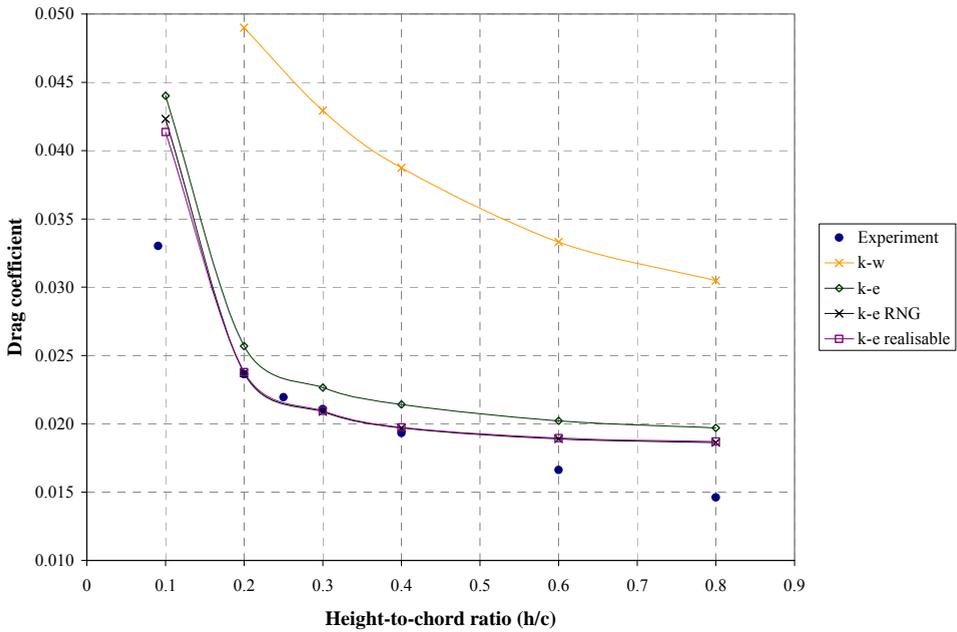


Figure 10. Drag comparison for a NACA0015 in ground effect at zero angle of attack.

The type of surrogate employed within this study is the Artificial Neural Network (ANN) that draws inspiration from the structure and/or functional aspects of biological systems. ANN is an adaptive system that is chosen due to its ability to representing complex relationships between input and output data with a high degree of accuracy. The well-known MATLAB neural network toolbox uses this model structure and provides useful user feedback including input-output mapping through the learning process. The exact architecture and mathematics of the model is not particularly relevant for this study although a brief description of how the models were used is warranted.

The first step of use is train the ANN; here the model is feed simulation data which it attempts to replicate through mathematical functions. This is an iterative process where the model learns, tests, validates and adapts this process is combined with user supervision to ensure a converged response. To guarantee confidence in the data extracted from the networks, one model was constructed per output – this leads to the development of two exclusive models; one for lift and one for drag per Reynolds number case. Although this approach demanded increased effort in construction of the extra models, the accuracy achieved was extremely high with a minimum confidence of 99.992% (Figure 11).

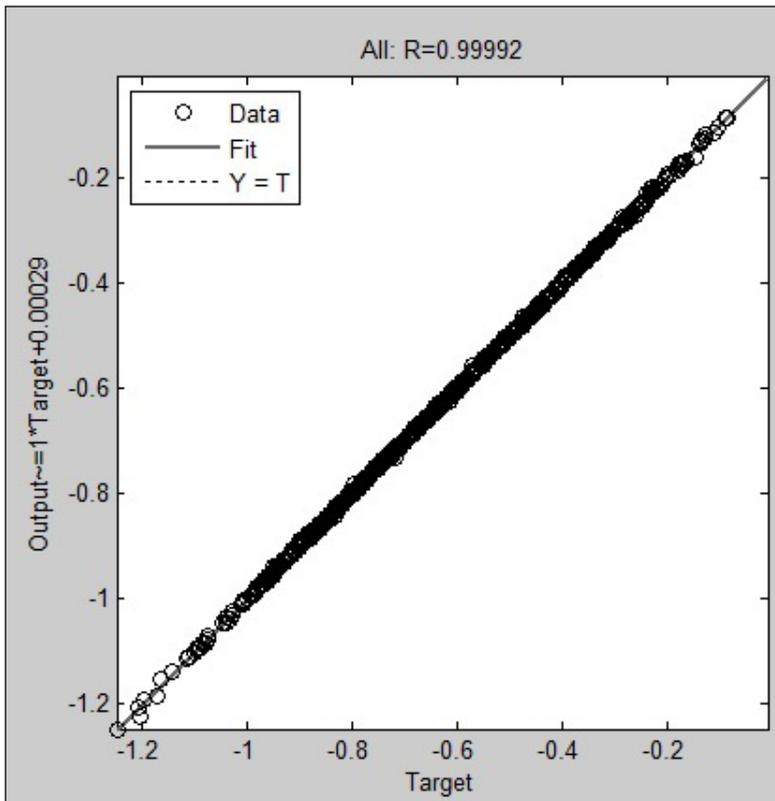


Figure 11. MATLAB NN regression plots for lift coefficient at Reynolds number of 600,000.

5.3. Analysis

The final step of this design exploration is where data mining takes focus. ANOVA and SOM techniques were applied concurrently to gain information about the design and solution space. This provides the user with the data required to complete the design process.

5.3.1. Analysis of variance

ANOVA is a data mining technique that is carried out to differentiate the contributions to the variance of the models response. This method originates from the statistical field of mathematics and reveals the influence each variable has upon the objective function in a quantitative sense [18]. Following the process outlined within [22] the relative influence of each design variable may be computed in the following manner.

The first step involves the computation of the sum of the squares; this is a measure of the variation of a subset of observations about the arithmetic mean of the entire set of observations. This represents the variation of the average observations under each design variable setting around the average of all observations. That is the relative influence for a given variable is represented by the sum of the squares of the required variable divided by the total sum of the squares for the model.

If N is the total number of design observations and \bar{T} is the arithmetic mean of the entire set, the total sum of the squares (SS_T) is given by:

$$SS_T = \sum_{i=1}^N (y_i - \bar{T})^2 \quad (5)$$

And the sum of the squares for each design variable (SS_γ) is:

$$SS_\gamma = \sum_{i=1}^n n_{\gamma i} (\bar{y}_{\gamma i} - \bar{T})^2 \quad (6)$$

Where n is the total number of possible settings for the design variable γ , $n_{\gamma i}$ is the number of observations for which design variable γ is at setting i , and $\bar{y}_{\gamma i}$ is the arithmetic mean of all observations for which design variable γ is at setting i . With this information the relative percentage contribution or relative influence (RI) can be easily calculated:

$$RI = \left(\frac{SS_\gamma}{SS_T} \right) \times 100 \quad (7)$$

The easiest way to execute this methodology is to vary one parameter across the range while holding all others constant, in this manner a one-way analysis of variance is achieved where the individual effects of each variable can be assessed in terms of the global variance. This approach also facilitates the implementation of the surrogate models in a much easier fashion than a varying all parameters simultaneously. As a one-way analysis of variance

does not detail in the interaction between variables, this feature will be addressed through SOM techniques.

5.3.2. Self-Organising Map

The mathematics of a SOM system are highly developed and not particularly relevant to the goals of this study, thus direct discussion here is not merited. In this study, the commercial software Viscovery SOMine® 5 is implemented for the purposes of data mining. This software utilises a hexagonal grid of an order matching the neighbourhood data, the features can thus be directly read from the data distribution map [13].

A self-organising map is another form of artificial neural network that uses unsupervised learning, nonlinear projection algorithm to produce a low-dimensionalised representation of a higher dimension input space [13]. Here a low-dimensional array of neurons (nodes) is modified to represent the input vector through weight adjustment. The closer the two patterns are in the original space, the closer the low-dimensional system response to input data. In this manner SOM acts to reduce the dimension of the input data while preserving the topological features and thus qualitative information can be obtained. As with any neural network there are two main phases of operation; learning and mapping. During the learning phase, the program must be 'trained' using input examples that represent the expected data during mapping. This process can be simplified to 5 primary steps:

1. Randomise the map's nodal weight vectors
2. Select an input vector
3. Traverse each node within the map
 - 3.1. Use Euclidean (ordinary) distance formula to find similarity between input vector and the map's nodal weight vector
 - 3.2. Track the node that produces the smallest distance (this node is the best matching unit - BMU)
4. Update the nodes in the neighbourhood of BMU by pulling them closer to the input vector
5. Advanced to the next iteration and repeat from 2

Firstly the weight of the neurons is initialised to either a random small value or a representative average value of the largest input vectors (1). The network is then fed a large number of training vectors that correspond to data that is expected to be experienced (2). This is an iterative approach with the number of trials dependant upon the accuracy of the initial weightings. To increase the learning speed, a competitive system is incorporated that computes the magnitude of the difference between the input vector and the weight vector for every neuron (3) and then reassigns the weight function for the neuron with the smallest difference (4). This process is then repeated for each input vector a number of times to better match the training data (5). A visualisation of this process can be seen in Figure 4.15 where the blue bead represents the training data distribution and the red circle is the current training data. During mapping there will be a singular neuron whose weight vector is the closest to the input vector and thus is taken as the solution.

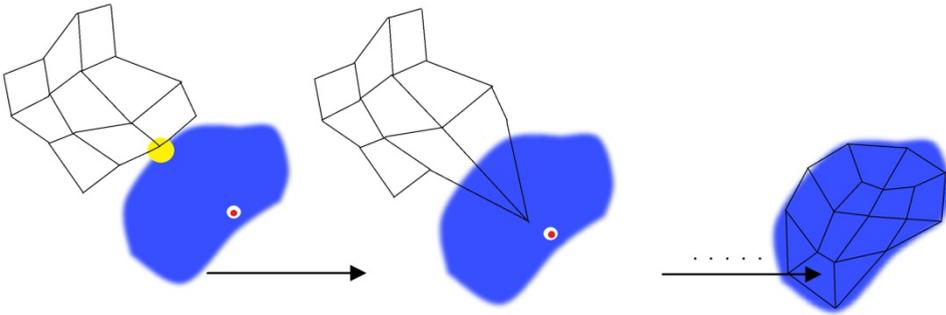


Figure 12. SOM training scheme.

Once the maps have been formed, data upon the required parameters can be read directly from all dimensions. SOM clusters data based upon neuron weights with areas of similar characteristics grouped together, in this way SOM forms semantic maps that allows the user to understand complex relationships with ease.

6. Results

Although nine permutations of the objective functions were analysed only the most pertinent are covered forthwith. For the remaining data, a direct correlation to the forthcoming conclusions was seen. The objective functions under direct analysis, in list of decreasing importance, are:

- Maximise section absolute lift coefficient at apex conditions ($Re = 600,000$)
- Minimise lift coefficient sensitivity to Reynolds number (0 represents an optimum while negative values highlight a downforce decrease with decreasing speed)
- Minimise section drag coefficient at corner entry/exit ($Re = 940,000$)

ANOVA was performed first with the ensuing results carried through to the SOM analysis making certain a full comprehension of the design space was achieved. Finally to display the application and power of this approach a candidate aerofoil was extracted from the design space highlighting a local optimised design that best meets the above objective functions.

6.1. ANOVA

Results from Analysis of Variance are most easily understood in terms of global variance and visualised through a pie chart. Here the most striking feature revealed is the magnitude of influence attributed to the parameters controlling the ground facing geometry and particularly that of Y_{TE} , α_{TE} and Y_{UP} . It can also be seen that X_{up} plays a significant role in lift coefficient sensitivity and somewhat in drag generation. Although the magnitude of influence changes between the objective functions, the parameters involved remain constant. This behaviour clearly highlights the unique nature of ground effect flows and the need for a differing design methodology than that of free stream aerodynamics.

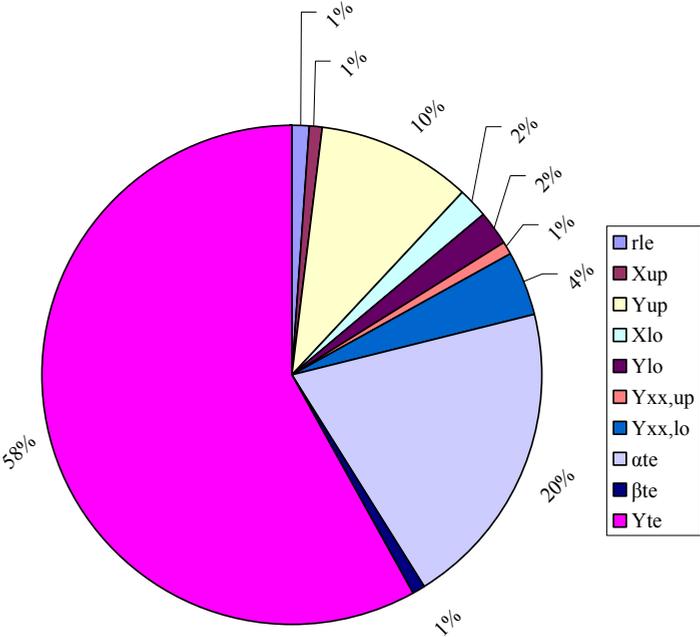


Figure 13. Influence for lift coefficient at apex.

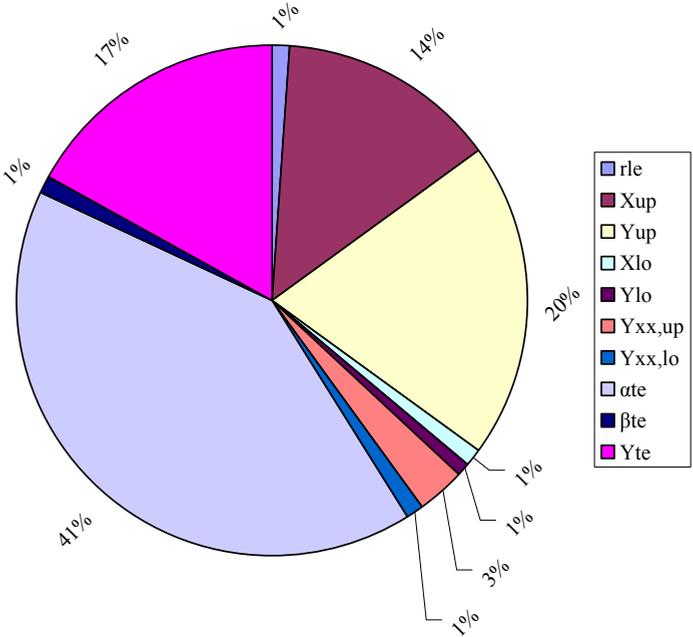


Figure 14. Influence for lift coefficient sensitivity.

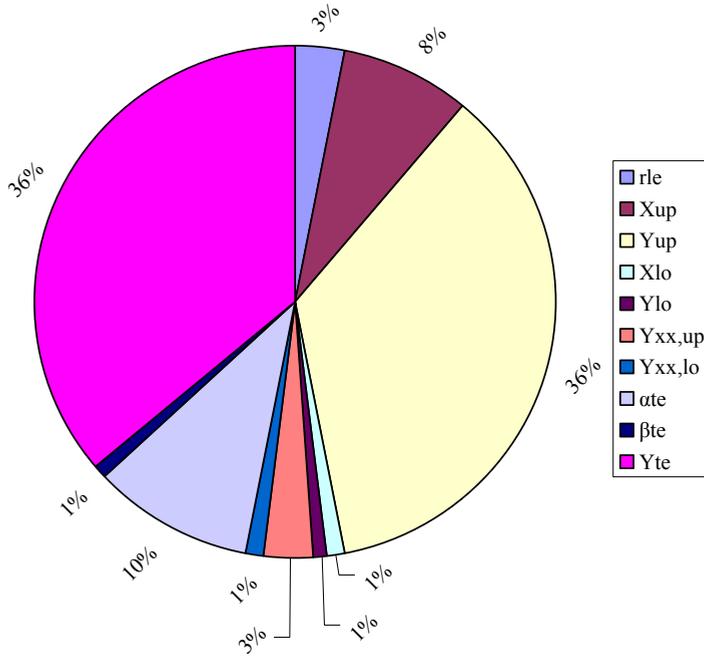


Figure 15. Influence for drag coefficient at corner entry/exit.

The behaviour witnessed through Figure 13 – 15 can be attributed to the inclusion of a closed surface that creates an area of high velocity and hence low static pressure between the aerofoil and ground; it is this area of suction that works to pull the section downwards (Figure 16).

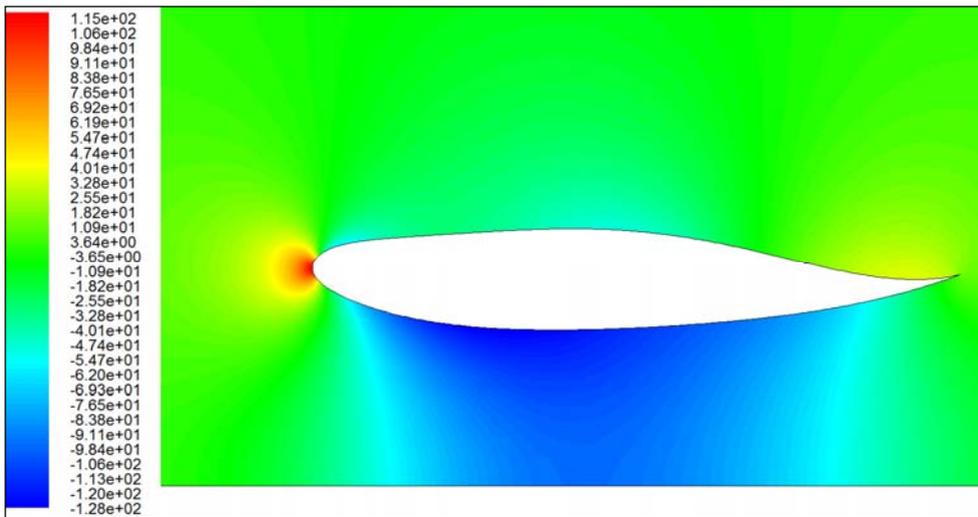


Figure 16. Aerofoil in ground effect highlighting the venturi effect; figure coloured by static pressure (Pa).

This phenomenon is primarily controlled by the geometric properties of the ground facing side that combine to create a quasi convergent-divergent duct. Termed the venturi effect, this flow phenomenon is a highly prevalent feature of ground effect flows that can prove either a hindrance or help depending upon the application. For racing vehicles this suction proves highly beneficial due to the augmentation of downforce and a subsequent Lift-to-Drag ratio increase. Although this phenomena has been exploited for decades [1], until now it has not been properly quantified. The results from the ANOVA plots clearly indicate the dominance of this effect; the four most influential parameters that account for 80 to 90% of the variance define the aerofoil geometry of the ground facing side. From these results the most important PARSEC parameters have been identified, although an understanding of modal interaction and a measurement of the design space validity remains inaccessible. For this purpose SOM was employed with the results from ANOVA were carried through this analysis.

6.2. SOM

The large amounts of information presented through SOM may be analysed from several different standpoints. The user's familiarity upon gathering meaningful conclusions from this data is critical to the overall effectiveness of this approach. Comparing all maps simultaneously proves a laborious and an inefficient task even for a small number of inputs. It proves far easier to analyse and draw conclusions from the objective function maps first and carry these results over to the design variable maps.

For objective function maps, the it can be seen that all objective functions are met to a high standard in the highlighted region. This area of high fitness is located within the far left of the design maps (Figure 17 – 19) and indicates an optimum performance is achieved within the current design space. This proves the validity of the process used to locate the design space (detailed within the design variable section) and that PARSEC bounds are large enough to contain an optimal solution.

To determine the area of high fitness, a trade-off approach was considered. The most important objective functions are met first and so on through the three maps; lift coefficient at apex conditions proves the most critical condition followed closely by sensitivity. Drag coefficient at corner entry/exit is less important but must be optimised where possible. Thus the area of most favourable design exhibits a lift coefficient between -0.62 and -1.15 at apex conditions (Figure 17) while the loss in negative lift during corner entry and exit is limited to some 5% (Figure 18). Finally at the onset of braking (the average top speed condition), the magnitude of drag is minimised to no more than 50% of the global maxima (Figure 19).

Shifting focus to the design variables, Figure 20 – 22 shows the most significant PARSEC parameters identified through SOM. These maps highlight the relationship between high fitness and Y_{TE} , α_{TE} and β_{TE} . This is a telling result; these parameters control the ground facing geometry of the aerofoil and the former two were identified as highly important through ANOVA. Furthermore it can be seen that these afore mentioned parameters all require values within the lower range to achieve high fitness. In terms of the effect upon section geometry these requirements have a definite effect:

- Reducing Y_{TE} leads to the trailing edge coordinate moving away from the ground – considering the convergent-divergent duct that is formed by the aerofoil/ground interaction this condition leads to an overall increase in the area downstream of the ‘throat’ aiding pressure recovery and aerofoil performance.
- Reducing α_{TE} corresponds to a more aggressive taper at the tail of the aerofoil – similar to the condition for Y_{TE} this works to open the divergent section of the duct with similar performance changes pressure recovering and thus increasing the aerofoil performance.
- Reducing β_{TE} results in a thinner trailing edge and thus an increased **total** area between the ground and the aerofoil – this works in combination with the above conditions to increase aerodynamic performance.

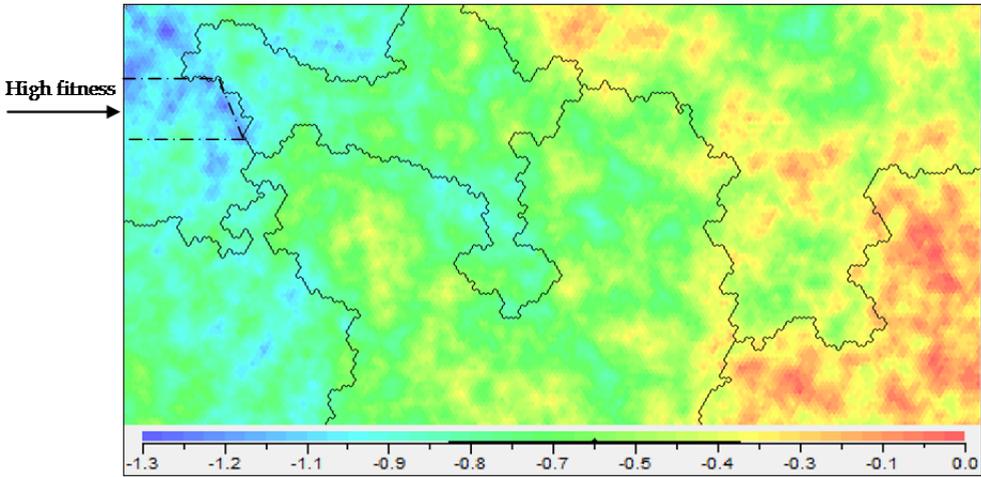


Figure 17. SOM coloured by lift coefficient at apex conditions.

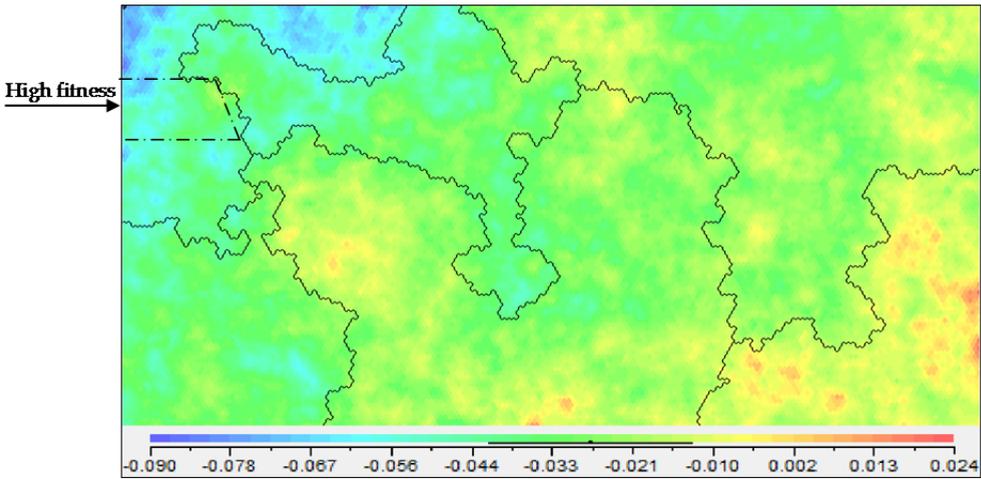


Figure 18. SOM coloured by lift sensitivity to Reynolds number.

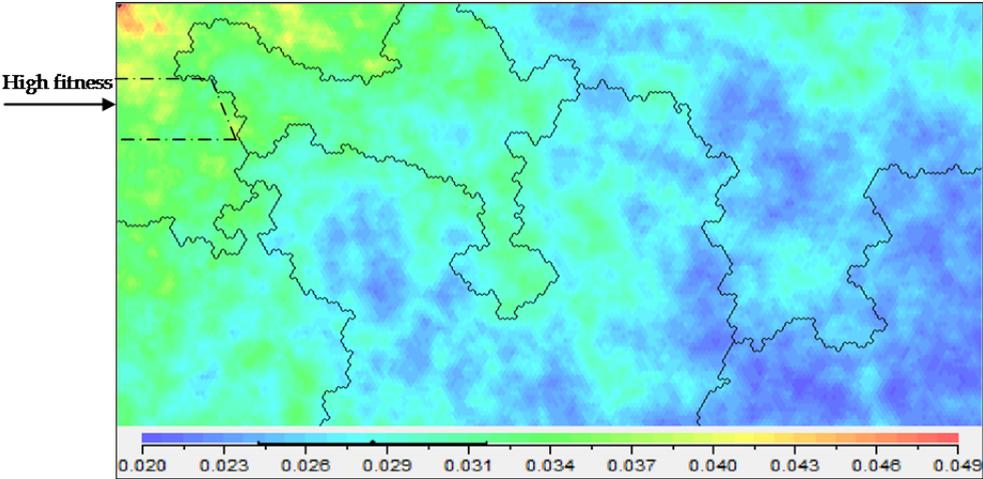


Figure 19. SOM coloured by drag coefficient at corner entry/exit.

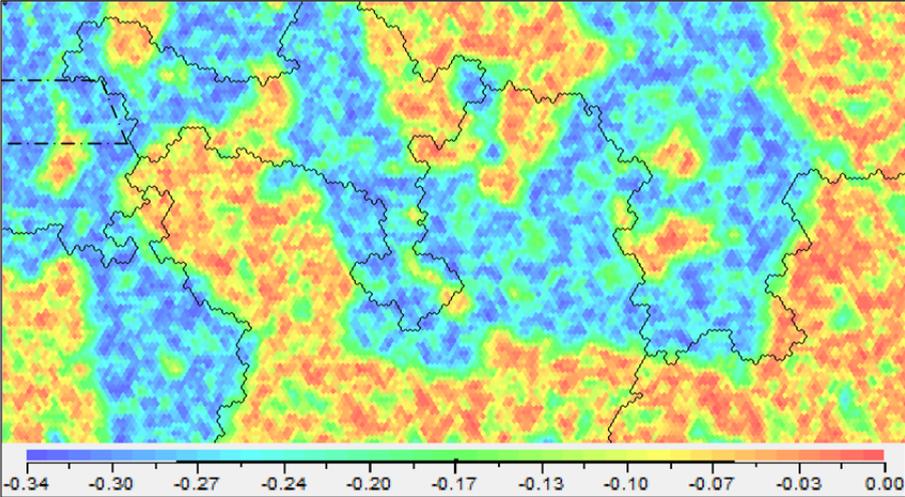


Figure 20. SOM coloured by $\alpha\epsilon$

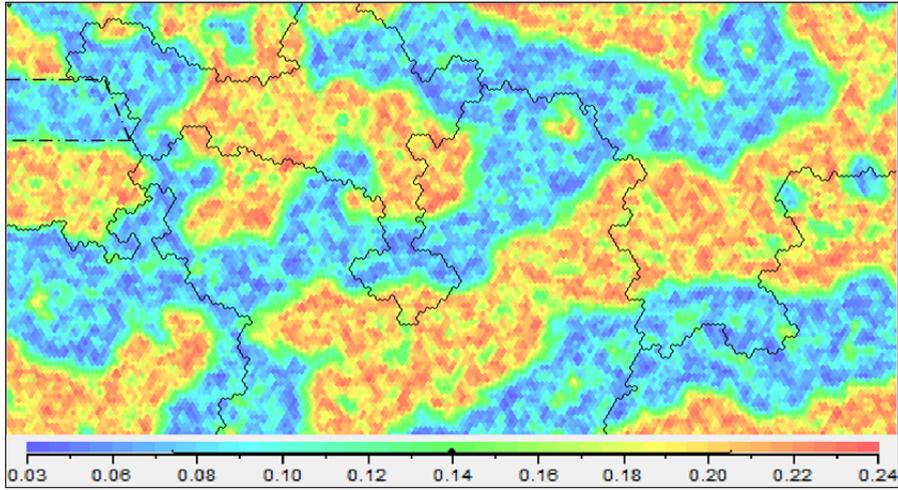


Figure 21. SOM coloured by β_{TE}

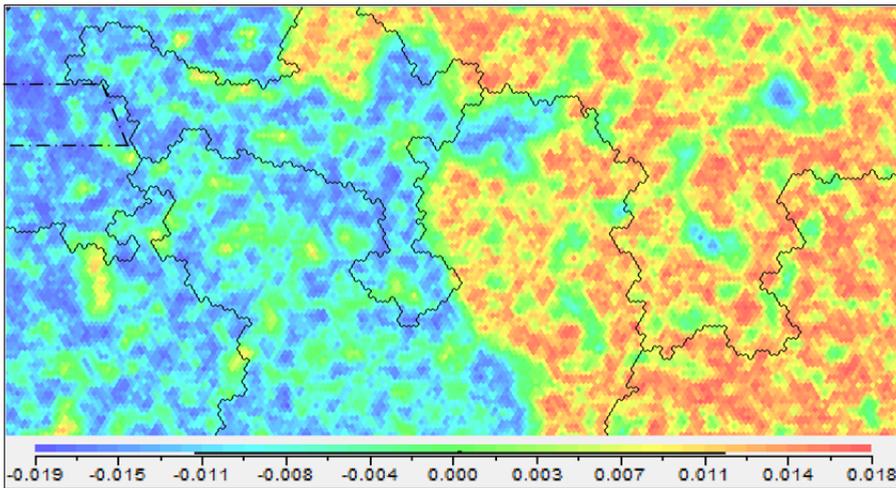


Figure 22. SOM coloured by Y_{TE}

From the variable maps the best range of the afore mentioned parameters can also be derived. From the prescribed relationships all three important parameters can be reduced to between 50 and 75 percent of the original domain:

Design variable and condition		Bounds for optimal operation
α_{TE}	Decrease	$[-0.3579 \rightarrow -0.094]$
β_{TE}	Decrease	$[0.000 \rightarrow -0.020]$
Y_{TE}	Decrease	$[-0.02 \rightarrow 0.000]$

Table 3. Variable condition and bounds for best operation identified through SOM.

For the remaining variable SOM maps, a more randomised pattern is seen that highlights modal interaction and thus no discrete relationship can be formed in confidence.

6.3. Summary

Through the previous section, the complimentary nature of the ANOVA and SOM results has been highlighted however the final results are most easily comprehended through the summarised tables below:

Design variable	Objective Function influence (%)		
	Lift coefficient at apex	Lift sensitivity to Reynolds number	Drag coefficient at entry/exit
Y_{TE}	58	17	36
α_{TE}	20	41	10
Y_{UP}	11	20	36
X_{UP}	2	14	8

Table 4. Variables of highest importance and relative influence as identified through ANOVA.

Design variable and condition		Lift coefficient at apex	Lift sensitivity to Reynolds number	Drag coefficient at entry/exit	Bounds for optimal operation
α_{TE}	Decrease	Increase	Decrease	Small increase	[-0.3579 → -0.094]
β_{TE}	Decrease	Increase	Decrease	Small increase	[0.000 → -0.020]
Y_{TE}	Decrease	Increase	Decrease	Small increase	[-0.02 → 0.000]

Table 5. Effects of variables upon objective functions including bounds for optimum performance as identified through SOM.

With this summary the design exploration is complete; the next step is entirely user defined and dependant upon the individual requirements.

6.4. Candidate extraction

As the final step of this study, the area of high fitness previously identified was extracted from the design space and the optimum aerofoil located. The tools provided within SOMine® allow for user determined extraction based on a number of criteria, here a simple graphical approach was used that saw the discrete selection of aerofoils that conform to the selection within (Figure 23).

To avoid the construction of a three-dimensional Pareto front, a weighting function (Equation 8) was constructed that summated the absolute values of the three objective functions. The design with the highest total to this function was thus the local optimal.

$$Aptitude = 0.5C_L + 0.3\Delta C_L + 0.2C_D \tag{8}$$

The selection of the above coefficients was made at the author’s discretion and could be tailored to suit individual requirements such as maximum negative lift, minimum drag and alike. The data extracted through SOMine® is of the same format as that input by the user; here ten PARSEC parameters and nine performance parameters are provided for each aerofoils. To aid quick assessment, this data was imported into an Excel spreadsheet where the above calculation was performed and a simple minimum/maximum search used to determine the best performing candidate aerofoil. The aptitude of the final design was compared to the front wing of the championship winning Ferrari F-2000 Formula One (Table 6).

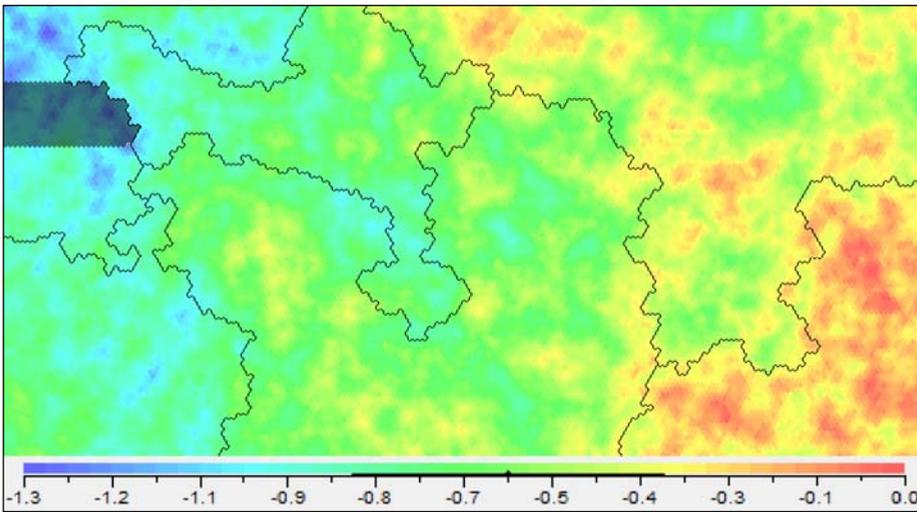


Figure 23. SOM map highlighting candidate selection area common to all maps (Coloured by lift coefficient at apex).

Aerofoil	Lift coefficient at apex	Lift sensitivity to Reynolds number	Drag coefficient at entry/exit
Ferrari F-2000 [23]	1.0	N/A	0.034
Candidate	1.043	0.0350	0.0315

Table 6. Optimum aerofoil performance.

Table 6 highlights the exceptional performance of the candidate aerofoil across the objective functions and the respective increase over a real Formula One front wing. When simulating this section, the basis of this high performance can be clearly related to the conclusions drawn from the ANOVA and SOM analysis, Figure 24 shows that:

- The area under the aerofoil is maximised to allow the largest mass flow rate while the ground facing geometry is smoothed to form the optimal duct shape; this corresponds to a low value of Y_{TE} , α_{TE} and Y_{UP} .

- The ground facing and free-stream surfaces of the aerofoil require differing approaches; the ground facing side requires a highly streamlined semi-symmetrical design that best replicates the venturi effect while the upper side exhibits great changes camber and a loaded trailing edge.

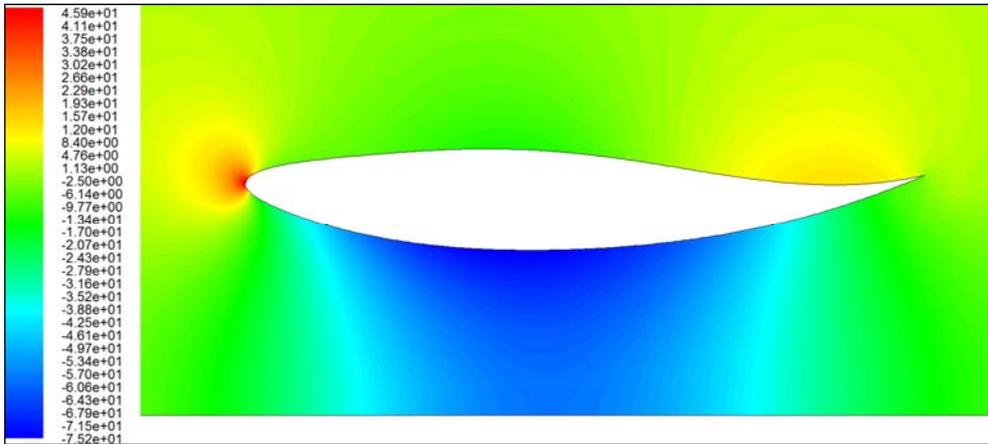


Figure 24. Static pressure (Pa) contours of the local optimal aerofoil ($Re = 940,000$).

It can be seen when the optimal conditions found through SOM and ANOVA are enforced a highly streamlined aerofoil with a ground facing side that replicates a convergent-divergent duct results. This result demonstrates not only the aptitude of the final design but also further verifies the design space as encompassing the parameter bounds required to achieve an optimised design.

It must also be noted that given the data upon the whole design space has already been computed, the selection of another specific aerofoil based upon varying conditions can be made within seconds. This is a highly advantageous outcome given the ability to assess any combination within the design space takes only a fraction of the time when compared to an optimisation scheme.

7. Further research

Given the topical and original nature of design exploration, there exists several areas where the work carried out within this chapter can be expanded. For the racing industry an additional study that extends the current results to a higher dimensional space would prove useful; representation of a multi-element wing and three-dimensional geometry will contribute to the real world applications. Although the magnitude of variables for this type of approach may preclude a full analysis, removal of the non-influential parameters could assist this cause without undue effect upon the objective functions. More generally, the development of this design technique would benefit from applications to other engineering disciplines particularly where competing methodologies are in use.

8. Conclusion

In this chapter a multi-objective design exploration for a Formula One front wing element was carried out. Insight was obtained into the most influential and important parameters and their range for optimal performance. Three discrete objective functions were considered that proved imperative to a Formula One vehicle while the representative speed range was achieved by varying the Reynolds number during simulation. A methodology for design space exploration was also presented and verified; to ensure accuracy a RANS based flow solver was implemented and coupled with ANN surrogate models that augmented the obtained results to achieve the required fidelity. ANOVA and SOM data mining techniques were then applied yielding quantitative and qualitative information upon the design space.

From the expansive data obtained a number of key results were determined;

- The range of the PARSEC bounds were proven to encompass the conditions for optimal performance and thus the aptitude of the shape fitting method used to determine a suitable baseline was proven.
- A discrete method of determining the area of high fitness for a multi-objective study was presented.
- The complimentary nature of ANOVA and SOM techniques enabled the most influential and important parameters to be determined, the range of best operation was also identified which represented a 25 to 50% reduction in the original variable bounds.
- After extraction of the high fitness aerofoils a weighting function determined the ideal aerofoil within this area. Comparison to an actual wing also highlighted this candidate performed better than its real world counterpart.

In total, a new design technique that removes the ambiguity of classical techniques was covered, the aptitude and flexibility of design space exploration both within the conceptual and ongoing design phases was proven. The successful outcome of this study has also made ground toward the wider acceptance of this methodology and it is hoped this work will proved inspiration to others within the engineering community to embrace and expand design space exploration.

Author details

Matthew Watts, Robert Carrese and Hadi Winarto
*School of Aerospace, Mechanical and Manufacturing Engineering,
RMIT University, Melbourne Australia*

Acknowledgement

Professor Simon Watkins, SAMME, RMIT University: For his assistance and feedback during writing of this chapter.

Nomenclature

ΔC_l	Lift Coefficient sensitivity
ANOVA	Analysis Of Variance
c	Aerofoil chord length
C_D	Drag coefficient
CFD	Computational Fluid Dynamics
C_l	Lift Coefficient
D	Aerodynamic drag
FIA	Federation Internationale de l'Automobile
h	Height above the ground
L	Aerodynamic lift
L^*	Characteristic length
NACA	National Advisory Committee for Aeronautics
Pa	Pascals
RANS	Reynolds-Averaged Navier Stokes
Re	Reynolds number
S	Reference area
SOM	Self-Organising Map
v	Velocity
μ	Dynamic viscosity
ρ	Density

9. References

- [1] Katz, K., (1995) "Race Car Aerodynamics – Designing for speed", Bentley Publishers, United Kingdom.
- [2] Federation Internationale de l'Automobile [FIA] (2010), Formula One Technical Regulations, 8 Place de la Concorde, France, Available: <http://www.fia.com/en-GB/sport/regulations/Pages/InternationalSportingCode.aspx> Accessed 2011 Oct 10.
- [3] Kusiak, A., (2006) "Data Mining: Manufacturing and Service Applications", International Journal of Production Research, Taylor and Francis, Abingdon, Oxfordshire, UK, 44(18-19), pp 4175-4191.
- [4] Harding, J. A. et al (2006), "Data Mining in Manufacturing; A Review", Transactions of the ASME Journal of Manufacturing Science and Engineering, ASME, New York, USA, 128(4), pp. 969-976.
- [5] Feng, J. C. and Kusiak, A., "Data Mining Applications in Engineering Design, Manufacturing and Logistics", International Journal of Production Research, Taylor and Francis, Abingdon, UK, 14(4), pp. 2689-2694.
- [6] Wang, K., (2007) "Review on Application of Data Mining in Product Design and Manufacturing", Fourth International Conference on Fuzzy Systems and Knowledge Discovery.

- [7] Kusiak, A. and Tseng, T. L., (2000) "Data Mining in Engineering Design: A Case Study" Proceedings of the 2000 IEEE International Conference on Robotics & Automation, San Francisco, CA, 2000
- [8] Rokach, L. and Maimon, O., (2006) "Data Mining for Improving the quality of Manufacturing: A Feature Set Decomposition Approach", Journal of Intelligent Manufacturing, Kluwer Academic Publishers, Dordrecht, Netherlands, 17(3), pp. 285-299.
- [9] Kohonen, T. et al, (1996) "Engineering Applications of the Self-Organizing Map" Proceeding of the IEEE, Vol. 84, No. 10.
- [10] Khurana, M., (2008) "Application of an Hybrid Optimization Approach in the Design of Long Endurance Airfoils", 26th International Congress of the Aeronautical Sciences, Alaska, USA.
- [11] Khurana, M. and Winarto, H., (2010) "Development and Validation of an Efficient Direct Numerical Optimisation Approach for Aerofoil Shape Design", The Aeronautical Journal, Vol. 114, No. 1159.
- [12] Chiba, K., Oyama, A., (2007) Obayashi, S. and Nakahashi, K., "Multidisciplinary Design Optimisation and Data Mining for Transonic Regional-Jet Wing", Journal of Aircraft, Vol. 44, No. 4.
- [13] Kanazaki, M., Tanaka, K., Jeong, S. and Yamamoto, K., (2007) "Multi-Objective Aerodynamic Exploration of Elements' Settings for High-Lift Airfoil Using Kriging Model", Journal of Aircraft, Vol 44. No. 3.
- [14] Shimoyama, K., Lim, J., Jeong, S., Obayashi, S. and Koishi, M., (2009) "Practical Implementation of Robust Design Assisted by Response Surface Approximation and Visual Data-Mining", Journal of Mechanical Design, Vol. 131, No. 6.
- [15] Anderson, J., (2007) "Fundamentals of Aerodynamics", Forth Edition, McGraw-Hill, New York, United States.
- [16] Sobieczky, H. (1998) "Parametric airfoils and wings," *Notes on Numerical Fluid Mechanics*, Vol. 68, pp 71-85.
- [17] Kuya, Y. et al., (2009) "Flow Physics of a Race Car Wing with Vortex Generators in Ground Effect", Journal of Fluids Engineering, Vol 131, no. 1.
- [18] Forrester, A. (2008) "Engineering Design via Surrogate Modelling: A Practical Approach," A John Wiley and Sons Publishing, United Kingdom.
- [19] McKay, M., & Beckman, R., (1979) "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output form a Computer Code", *Technometrics*, Vol. 21 No. 2.
- [20] Tu, J., et al, (2008) "Computational Fluid Dynamics: A Practical Approach", Butterworth-Heinemann, Burlington, United States.
- [21] Ranzenbach, R. and Barlow, K., (1996) "Cambered Airfoil In Ground-Effect – An Experimental and Computational Study", SAE paper 960909.
- [22] Jilla, C., (1999) "A Multiobjective, Multidisciplinary Design Optimization Methodology of the Conceptual Design of Distributed Satellite Systems", Massachusetts Institute of Technology, United States.
- [23] Wright, P., (2004) "Ferrari Formula 1: Under the Skin of the Championship-Winning F1 – 2000", David Bull Publishing, Phoenix, United States