Cranfield University

College of Aeronautics

Ph.D. Thesis

Academic Year 2008-2009

Michael Mifsud

Reduced-order modelling for high-speed aerial weapon aerodynamics

Supervisors: Dr. S. T. Shaw and Dr. D. G. MacManus

October 2008

This thesis is submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Cranfield University

Michael Mifsud

Reduced-order modelling for high-speed aerial weapon aerodynamics

College of Aeronautics

Ph.D. Thesis

Abstract

In this work a high-fidelity low-cost surrogate of a computational fluid dynamics analysis tool was developed. This computational tool is composed of general and physics-based approximation methods by which three dimensional high-speed aerodynamic flow-field predictions are made with high efficiency and an accuracy which is comparable with that of CFD. The tool makes use of reduced-basis methods that are suitable for both linear and non-linear problems, whereby the basis vectors are computed via the proper orthogonal decomposition (POD) of a training dataset or a set of observations.

The surrogate model was applied to two flow problems related to high-speed weapon aerodynamics. Comparisons of surrogate model predictions with high-fidelity CFD simulations suggest that POD-based reduced-order modelling together with response surface methods provide a reliable and robust approach for efficient and accurate predictions. In contrast to the many modelling efforts reported in the literature, this surrogate model provides access to information about the whole flow-field.

In an attempt to reduce the up-front cost necessary to generate the training dataset from which the surrogate model is subsequently developed, a variable-fidelity POD-based reduced-order modelling method is proposed in this work for the first time. In this model, the scalar coefficients which are obtained by projecting the solution vectors onto the basis vectors, are mapped between spaces of low and high fidelities, to achieve high-fidelity predictions with complete flow-field information. In general, this technique offers

an automatic way of fusing variable-fidelity data through interpolation and extrapolation schemes together with reduced-order modelling (ROM).

Furthermore, a study was undertaken to investigate the possibility of modelling the transonic flow over an aerofoil using a kernel POD–based reduced-order modelling method. By using this type of ROM it was noticed that the weak non-linear features of the transonic flow are accurately modelled using a small number of basis vectors. The strong non-linear features are only modelled accurately by using a large number of basis vectors.

Acknowledgements

I would like to sincerely thank my former supervisor Dr. Scott T. Shaw (now at Airbus UK) and my present supervisor Dr. David G. MacManus for their guidance and assistance while conducting this research work. I owe a great debt to these two outstanding academics and hope that in future I would have the opportunity to show them my sincere appreciation.

During my time at Cranfield University I had the good fortune to work on a number of projects which have helped to broaden my understanding of high-speed aerodynamics and CFD in general. One in particular was in close collaboration with Professor J. L. Stollery on a very interesting hypersonic flow problem. This was a unique experience for me because apart from working in an exciting field of aerodynamics, I have enjoyed the experience of learning from one of the leading experts in the field. I am sincerely grateful to Prof. Stollery for assisting and sharing with me his deep knowledge about this interesting subject.

Also, I would like to thank Trevor J. Birch of DSTL Bedford for providing some funding for this project and for some useful discussions. Moreover, I would like to thank Professor K.P.Garry Head of the Aerodynamics Group, Dr. L. Oswald Head of the High Performance Computing IT Section and Ramesh Wadher of the IT Section.

I would like to thank also all my friends and colleagues at Cranfield University who made my stay enjoyable. Finally, I would like to dedicate this work to my parents.

Nomenclature

σ^2	variance of a set of data
σ_x^2	variance of a set of data points' projections on the x-axis
σ_y^2	variance of a set of data points' projections on the y-axis
σ_i	singular value in the singular value decomposition
α_{ij}	is the scalar or projection coefficient of the j-th POD mode
Φ_{ij}	is the j-th POD basis mode
$\boldsymbol{\lambda}_i$	eigenvalue of the eigenfunction or POD mode representative of the variance captured along that eigenfunction
m	is the number of realizations or snapshots and POD modes
n	is the number of grid points in a mesh
$\mathbf{x}$	is a vector of different parameter values
$\mathbf{Y}$	is a solution vector of primitive or conservative variables
$f(x)$	computationally expensive analysis
$\hat{f}(x)$	approximation to $f(x)$
$\Psi(r)$	a typical radial basis function RBF
r	radial distance from a given centre point
w_i	weighting coefficient for a typical radial basis function $\Psi(r)$
x_i	centre data point in a typical radial basis function

Nomenclature

θ	shape parameter in a typical radial basis function
$p(x)$	a polynomial of one degree less than the radial basis function $\Psi(r)$
α	angle of incidence
δ	nose deflection angle of the nose-controlled missile configuration
D	mid-body diameter
M	Mach number
C_x	component of the aerodynamic force along the axis of the body
C_z	component of the aerodynamic force normal to the axis of the body
C_m	pitching moment acting on the body
X_{cp}	centre of pressure coordinate along the body axis
y_{hf}	response value from the high-fidelity model
y_{lf}	response value from the low-fidelity model
$\gamma_i(\mathbf{x})$	the ratio of the evaluated response values from the high-fidelity model and the low-fidelity model at the i^{th} data point
$\hat{\gamma}(\mathbf{x})$	correction response surface
$K(x, y)$	a kernel function
$\Re^n$	n - dimensional real data
$\boldsymbol{X}$	input space representation
$\boldsymbol{F}$	feature space representation
Φ	a mapping function from an input space $\boldsymbol{X}$ to an inner product feature space $\boldsymbol{F}$, also referred to as dot product space
β_k	projection coefficient onto the k^{th} non-linear basis function
$\mathbf{V}^k$	k^{th} basis vector in the feature space
P_n	projection operator

Acronyms

ANOVA	Analysis of Variance
BILU	Block Incomplete Lower-Upper
CFD	Computational Fluid Dynamics
CPU	Central Processing Unit
CSI	Cubic-spline Interpolation
CVT	Centroidal Voronoi Tessellations
CVOD	Centroidal Voronoi Orthogonal Decomposition
DD	Domain Decomposition
DoE	Design of Experiment
LHCS	Latin Hypercube Sampling
LHS	Left-hand side
IMPNS	Implicit Multi -block and -grid Parabolised Navier-Stokes
MDO	Multi-disciplinary Design and Optimisation
NOD	Non-orthogonal Decomposition
PCA	Principal Components Analysis
kPCA	kernel Principal Components Analysis
PDE	Partial Differential Equation
PNS	Parabolised Navier-Stokes
POD	Proper Orthogonal Decomposition
kPOD	kernel Proper Orthogonal Decomposition
RBF	Radial Basis Function
RHS	Right-hand side
RMS	root mean square
ROM	Reduced-Order Model
RSM	Response Surface Methods
SDE	Semi-definite Embedding
SVD	Singular Value Decomposition

Table of Contents

Chapter 1
1.0 Introduction to reduced-order modelling

1.1 Motivation

The solution of non-linear steady or unsteady aerodynamic flows by numerically solving the Navier-Stokes equations for various applications is a computationally intensive task even when implemented on modern state-of-the-art computing platforms. These tasks are very expensive with respect to data storage, data handling and processor costs. Hence, it can be a significant challenge to deal with a number of situations such as parametric studies of state solutions, multiple state solutions such as in optimization and control problems, and real-time state solutions such as in feedback control settings. In view of this difficulty, considerable attention is paid to reduce both storage and processing costs of non-linear state solutions by using reduced-order models. Generally, a reduced-order model is a mathematical model of a physical system that is derived from computational or at times even experimental data. The reduced-order model contains fewer degrees of freedom than the discretized partial differential equations and therefore it is relatively inexpensive to compute. Reduced-order models are developed and used to provide a more efficient and computationally economic way for investigating these complex problems. The use of reduced-order models is primarily motivated by the desire

to have detailed knowledge of the physics of the problem being investigated together with an efficient and reliable prediction tool.

1.2 Reduced-order modelling

In the reduced-order modelling approach, an approximation to a physical system from the results of a high-fidelity simulation model is derived. This approximate model is called a reduced-order model (ROM). Since the ROM contains fewer degrees-of-freedom[1] than the full simulation model, it is inexpensive to compute in comparison with the high-fidelity model and so this makes it appropriate for prolonged computations in multidisciplinary environments.

Generally, reduced-order modelling involves the selection of a reduced set of basis $\boldsymbol{\varphi}_j(\mathbf{x})$ from an ensemble of representative datasets $\{\mathbf{u}^k(\mathbf{x},\mathrm{t})\}_{k=1}^{N}$ of real scalar or vector fields. Once these basis functions are identified and established, an approximation of the flow-field $\mathbf{u} \in \mathrm{span}\{\boldsymbol{\varphi}_j(\mathbf{x})\}$ is made as a linear combination of the basis functions as follows:

$$\mathbf{u}(\mathbf{x},\mathrm{t}) \approx \sum_{j=1}^{M} a_j(\mathrm{t})\,\boldsymbol{\varphi}_j(\mathbf{x}) \qquad \text{Eq. (1-1)}$$

where $M < N$. In this, the coefficients a_j are determined by projecting the solution of the governing partial differential equations onto the reduced basis. The cost of such a computation is very small if the off-line computational time required for computing the reduced basis is not considered.

All reduced-order methods are in fact reduced basis methods, which in turn require the solution of high-fidelity and therefore very expensive discretized flow governing

[1] The number of degrees of freedom is the same as the number of dimensions in a physical system. The two terms are used interchangeably in this text.

equations [1]. These expensive calculations can be done off-line before a state prediction or the optimization of the design parameters or feedback control is attempted. A set of reduced basis that contains all the essential information of a set of basis of larger dimension can be extracted by various techniques, amongst which is the proper orthogonal decomposition (POD). This is an elegant and powerful data reduction technique which is often used to extract basis functions or 'mode shapes' of high-dimensional systems for subsequent use in Galerkin projections that yield reduced-order models of non-linear physical systems.

1.3 What is proper orthogonal decomposition?

The representation of eq. (1-1) is not unique. For example, if the domain of $\mathbf{x}$ is a bounded interval $\mathbf{\Omega}$, then the basis functions $\boldsymbol{\varphi}_j(\mathbf{x})$ can be chosen as a Fourier series, or Legendre polynomials, or Chebyshev polynomials, and so on. For each choice of the sequence $\boldsymbol{\varphi}_j(\mathbf{x})$ that forms a basis for the vector functions $\mathbf{u}(\mathbf{x}, t)$ which are assumed to be bounded and integrable, the sequence of functions $a_j(t)$ is different. Note that although there is no difference between the independent variables $\mathbf{x}$ and t in eq. (1-1), these may be considered as a spatial and a temporal coordinate respectively.

Considering orthonormal[2] basis functions, i.e.

$$\int_{\Omega} \boldsymbol{\varphi}_{j_1}(\mathbf{x})\, \boldsymbol{\varphi}_{j_2}(\mathbf{x})\, d\mathbf{x} = \begin{cases} 1 & \text{if} \quad j_1 = j_2 \\ 0 & \text{if} \quad j_1 \neq j_2 \end{cases}$$

then,

$$a_j(t) = \int_{\Omega} \mathbf{u}(\mathbf{x}, t)\ \boldsymbol{\varphi}_j(\mathbf{x})\, d\mathbf{x} \qquad \text{Eq. (1-2)}$$

[2] Orthonormal basis functions are orthogonal basis functions of unit magnitude ie. normalized to unit magnitude.

that is, for orthonormal basis functions, the determination of the coefficient function $a_j(\mathrm{t})$ depends only on $\boldsymbol{\varphi}_j(\mathbf{x})$ and not on the other $\boldsymbol{\varphi}$'s.

Consequently, selecting orthonormal basis functions would be very useful. Moreover, while an approximation to any desired accuracy in eq. (1-1) can always be obtained if *M* is chosen large enough (the approximation in eq. (1-1) becomes exact in the limit as $M \to \infty$), $\boldsymbol{\varphi}_j(\mathbf{x})$ may be chosen such that the approximation for each *M* is optimal in a least squares sense. This means that a sequence of orthonormal functions are found such that the first two of these basis functions give the best possible two-term approximation, the first three of these basis functions give the best possible three-term approximation and the first ten give the best possible ten-term approximation, etc. These ordered orthonormal basis functions are called the proper orthogonal modes for the vector function $\mathbf{u}(\mathbf{x},\mathrm{t})$. With these basis functions, the expression in eq. (1-1) is called the proper orthogonal decomposition of $\mathbf{u}(\mathbf{x},\mathrm{t})$.

1.4 Dimensionality reduction using POD

Dimensionality reduction is the underpinning of reduced-order modelling and for this reason it is presented here for a better understanding.

In numerous cases, the initial high-dimensional setting of a collection of scalar or vector fields, such as snapshots of the complete state of an aerodynamic flow, is a consequence of a poor coordinate system. Often such data affords a significantly reduced representation by considering a coordinate system or equivalently a set of basis that is close to an optimal representation of the data. The search for an optimal or at least a good basis for the data leads to the application of empirical transform techniques. In this context, the word empirical implies that the mappings are computed directly from the data of interest. Essentially, these are an extension of some other analytical methods such as the Fourier or Laplace transforms in which the new coordinate system is pre-determined and independent of the data. Since the structure of the data in the input space

and the size of the input space are problem-dependent, it is appropriate to examine mappings that are derived from the data. These empirical mappings are normally determined by imposing appropriate optimality criteria and often provide insight into the structure of the data because they are directly related to the data itself. By using dimensionality reduction techniques, the reduction of high-dimensional data that lies in a subspace of a larger input space is achieved in a methodical way. The following is a simple quantitative example showing what is meant by dimensionality reduction.

1.4.1 A simple example of dimensionality reduction

In this section a simple example of reducing the number of dimensions (or degrees-of-freedom) using POD for a set of discrete data points in two-dimensions is presented. This example gives a straightforward explanation of the POD as a dimensionality reduction method at an early stage of this text, so as to enable the reader to have a basic understanding of the technique while going through the initial chapters.

Consider a set of N discrete data points in a d-dimensional space. Suppose that N is equal to 10 and d is equal to 2. A two-dimensional dataset is considered for the simple reason that this can be very easily represented by a graphical plot. Table (1-1) shows the two-dimensional set of data, while figure (1-1) shows a pictorial representation of the same set of data. This set of data was taken from reference [2].

x	2.50	0.50	2.20	1.90	3.10	2.30	2.00	1.00	1.50	1.10
y	2.40	0.70	2.90	2.20	3.00	2.70	1.60	1.10	1.60	0.90

Table (1-1) Two-dimensional data set

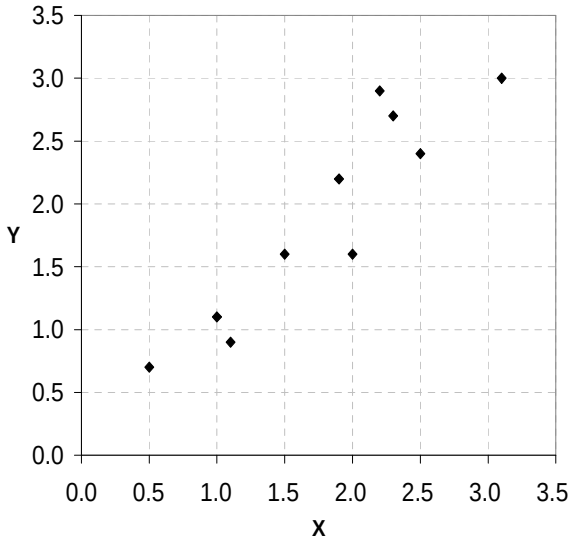


Figure (1-1) Graphical representation of a two-dimensional data set.

Generally, data points can be projected from the high-dimensional space in which they are embedded to a lower dimensional space for a better understanding of data structure. In this particular example the points can be projected at most from a two-dimensional space (d = 2) to a one-dimensional subspace (d = 1). Here two very simple projections can be made on the x and y axes, although we are interested mostly in finding vectors on which data projections are most spread out. As a measure of spread for the projected data points, the variance of the projections σ^2 can be considered. For a set of N points the mean and the variance are given by

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i \quad \text{and} \quad \sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2 \text{ respectively.}$$

Therefore, for data points projections on the x and y axes, the mean values and the variances are equal to $\bar{x}$ = 1.81, $\bar{y}$ = 1.91, σ_x^2 = 0.6166 and σ_y^2 = 0.7166. From this it is clear that the variance of the data points depends on the direction onto which the projections are made. Consequently, there must be some direction such that when the data points are projected in d dimensions it gives the largest variance. This direction vector can be determined by constructing the d x d covariance matrix and finding its eigenvalues

and their corresponding eigenvectors. The covariance matrix is determined by evaluating the terms according to:

$$C_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}[x_i - \bar{x}][y_i - \bar{y}]$$

In this example the 2 x 2 covariance matrix is worked out to be

$$\boldsymbol{C} = \begin{bmatrix} C_{xx} & C_{yx} \\ C_{xy} & C_{yy} \end{bmatrix} = \begin{bmatrix} 0.6166 & 0.6154 \\ 0.6154 & 0.7166 \end{bmatrix}$$

The positive non-diagonal elements of the covariance matrix show that the x and y variables increase together as shown in figure (1-1). Calculating the eigenvalues λ_i and eigenvectors $\mathbf{v}_i$ of the covariance matrix gives the following results,

$$\lambda_1 = 1.2840; \quad \lambda_2 = 0.0491; \quad \mathbf{v}_1 = \begin{bmatrix} -0.6779 \\ -0.7352 \end{bmatrix}; \quad \mathbf{v}_2 = \begin{bmatrix} -0.7352 \\ +0.6779 \end{bmatrix}$$

By plotting the two eigenvectors as black intermittent lines on the same plot of the centred[3] data points, it is evident that these are perpendicular to each other. See figure (1-2). Moreover, the eigenvector corresponding with the biggest eigenvalue provides a line of best fit through the data. The second eigenvector provides the other relatively much less important pattern in the data, that all the points follow the main line but are off to the side of the main line by some small quantity. So, by finding the eigenvectors of the covariance matrix, lines are extracted that characterise the data.

[3] By centred data it is implied that the mean values for x and y were subtracted from the original data points.

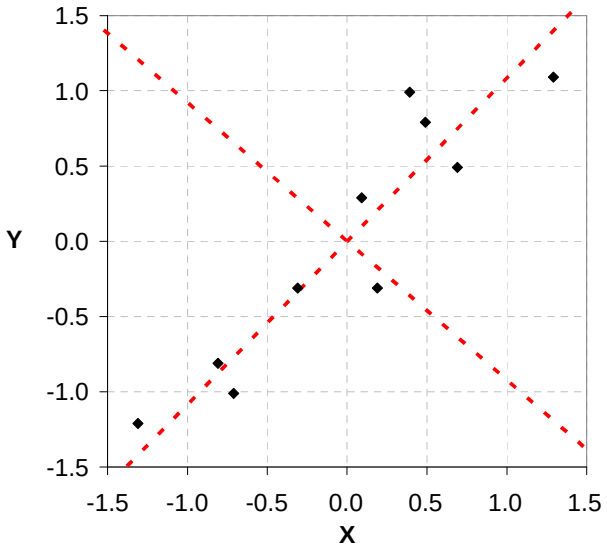


Figure (1-2) Overlaying the eigenvectors upon the centred data set of points

By multiplying the centred data with the two eigenvectors, the dataset is transformed in terms of another set of basis vectors and these are the eigenvectors. Thus, a basis transformation is made.

$$\begin{bmatrix} 0.69 & 0.49 \\ -1.31 & -1.21 \\ 0.39 & 0.99 \\ 0.09 & 0.29 \\ 1.29 & 1.09 \\ 0.49 & 0.79 \\ 0.19 & -0.31 \\ -0.81 & -0.81 \\ -0.31 & -0.31 \\ -0.71 & -1.01 \end{bmatrix} \begin{bmatrix} -0.6779 & -0.7352 \\ -0.7352 & +0.6779 \end{bmatrix} = \begin{bmatrix} -0.8280 & -0.1751 \\ +1.7776 & +0.1429 \\ -0.9922 & +0.3844 \\ -0.2742 & +0.1304 \\ -1.6758 & -0.2095 \\ -0.9129 & +0.1753 \\ +0.0991 & -0.3498 \\ +1.1446 & +0.0464 \\ +0.4380 & +0.0178 \\ +1.2238 & -0.1627 \end{bmatrix}$$

If this newly derived data is plotted with the first column along the x-axis and the second column along the y-axes for each and every data point, figure (1-3) is obtained, from which one can observe the increase in variance along the horizontal basis.

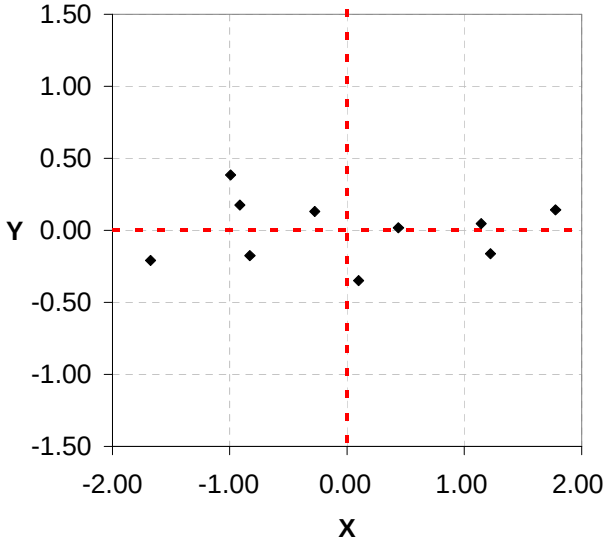


Figure (1-3) Data points expressed in terms of the newly obtained basis vectors

The following is the notion of dimensionality reduction. The magnitudes of the eigenvalues are equivalent to the variance of the projected data points along their corresponding eigenvector. Therefore, by placing the eigenvalues in descending order, this ordering sets up the eigenvectors in their order of significance. For this particular example the eigenvalue of one eigenvector is very much higher than the eigenvalue of the second one. Consequently, one can ignore the vector of lesser significance. By ignoring some eigenvectors, the original data set will be described in terms of a reduced set of dimensions. Thus, generally if the problem has n dimensions initially from which an n x n covariance matrix is formed hence obtaining n eigenvalues and eigenvectors, one can choose only the first p eigenvalues and their corresponding eigenvectors so that the final data set has only p dimensions. In this case a one dimensional dataset can be derived by multiplying the centred data with the eigenvector having the biggest variance as follows:

$$\begin{bmatrix} 0.69 & 0.49 \\ -1.31 & -1.21 \\ 0.39 & 0.99 \\ 0.09 & 0.29 \\ 1.29 & 1.09 \\ 0.49 & 0.79 \\ 0.19 & -0.31 \\ -0.81 & -0.81 \\ -0.31 & -0.31 \\ -0.71 & -1.01 \end{bmatrix} \begin{bmatrix} -0.6779 \\ -0.7352 \end{bmatrix} = \begin{bmatrix} -0.8280 \\ +1.7776 \\ -0.9922 \\ -0.2742 \\ -1.6758 \\ -0.9129 \\ +0.0991 \\ +1.1446 \\ +0.4380 \\ +1.2238 \end{bmatrix}$$

By multiplying these projections with the most significant eigenvector only, a new set of one-dimensional data points are obtained.

$$\begin{bmatrix} -0.8280 \\ +1.7776 \\ -0.9922 \\ -0.2742 \\ -1.6758 \\ -0.9129 \\ +0.0991 \\ +1.1446 \\ +0.4380 \\ +1.2238 \end{bmatrix} \begin{bmatrix} -0.6779 & -0.7352 \end{bmatrix} = \begin{bmatrix} +0.5613 & +0.6087 \\ -1.2050 & -1.3068 \\ +0.6726 & +0.7294 \\ +0.1859 & +0.2016 \\ +1.1360 & +1.2320 \\ +0.6189 & +0.6712 \\ -0.0672 & -0.0729 \\ -0.7759 & -0.8415 \\ -0.2969 & -0.3220 \\ -0.8296 & -0.8997 \end{bmatrix}$$

Plotting this derived set of data with the first column along the x-axis and the second column along the y-axis while adding the respective mean values which were subtracted initially, the result shown in figure (1-4) is obtained. Hence the original data is reconstructed using only one single eigenvector. It is evident that in the reconstruction some of the information is lost, but nevertheless an approximate set of data which is a good representation of the original data is obtained. This approximate set of data is one-dimensional as it can be represented along one basis vector only. Therefore, the original

two-dimensional set of data was reduced to a one-dimensional set of data represented by the slope of the line.

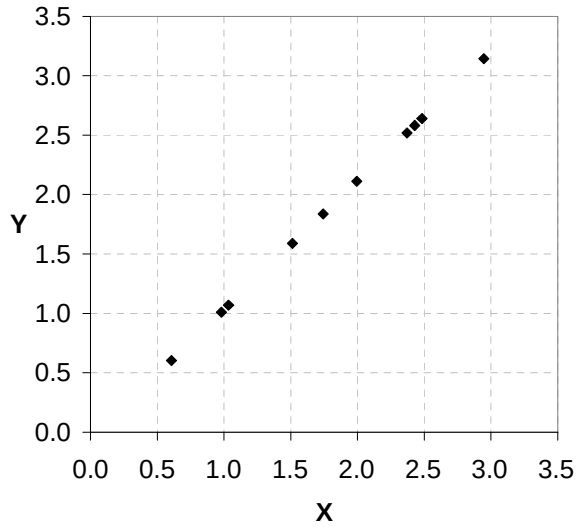


Figure (1-4) Reconstructing the original data using one single eigenvector

When this technique is applied on very large sets of data such as computational fluid dynamics (CFD) data, this technique endows us with the capability of reducing the number of dimensions of a multi-dimensional problem; thus enabling us to approximate flow solutions with low-cost with respect to both computational time and resources. Here, by resources it is implied that due to the reduction in the quantity of data attained by using this technique, a smaller memory is required for storing and handling of data. Hence, in this project we embark on developing means by which this can be realized. Such a tool would be very useful most especially in multi-disciplinary studies, where a vast amount of high-fidelity analyses are required for an accurate assignment within a reasonable time-frame. Also, this tool may be considered as a potential alternative to the semi-empirical tools presently in use for trade and trajectory studies.

1.5 Aim and objectives

The aim of this research work is to develop a computational tool based on a mathematical model by which three dimensional high-speed aerodynamic flow field predictions are made with low-cost and acceptable accuracy. This is achieved through the use of a reduced-order modelling technique, namely the proper orthogonal decomposition (POD). Ultimately, the tool can be employed in the analysis of complex aerodynamic flow problems.

In order to meet this aim, a number of objectives were identified and accomplished.

- The application of proper orthogonal decomposition as a reduced-order modelling method to a simple case of practical interest so as to understand the behaviour of this technique.

- The development of a mathematical model which integrates the proper orthogonal decomposition technique with either regression analysis or interpolation methods to predict aerodynamic flow solutions at various parameters.

- An investigation of the effects on the mathematical model when using different sampling techniques.

- The application of the developed model to realistic weapon configurations.

- The adaptation and application of the model in a variable-fidelity environment, where a combination of inexpensive low-fidelity data such as inviscid solutions with more accurate but expensive high-fidelity data such as viscous solutions is made.

- An investigation of the behaviour of a non-linear POD algorithm to model the aerodynamic flow over an aerofoil in the transonic flow regime.

1.6 Outline of thesis

This research work can be subdivided into three major parts. The first concerns the development of the computational tool based on mathematical models for making fast approximate aerodynamic predictions. The second part deals with the application of this tool to various problems related to high-speed aerodynamics. In this part the performance of the tool in terms of accuracy and efficiency is investigated. In the third part a variable-fidelity model and the POD model are combined to form a model which handles reduced variable-fidelity datasets. Some applications of this model are presented to demonstrate its effectiveness. Moreover, in this part a further investigation is conducted on the application of the very recently developed kernel POD method to highly non-linear problems in aerodynamics.

In the next chapter an extensive review of the available literature about the POD is presented. In chapter three a detailed description of the mathematical models and the key elements is made. This includes a brief review of the numerical methods employed in the flow solver used to generate the ensemble of flow realizations which form the basis for the POD technique and a review of the sampling, regression and interpolation techniques considered. In chapter four a detailed account of the POD theory and its properties is made together with a description of the different methods used in POD analysis. In chapter five the POD is presented as a reduced-order model (ROM) whereby applications of the POD as a data compression and a computational time reduction procedure are presented. In chapter six the developed model is presented as a prediction tool. Moreover, semi-empirical tools which are in common use by industry for trade and trajectory studies are discussed and a comparison is made between these tools and the computational tool developed in this work. In chapter seven, a variable-fidelity model which incorporates the POD is proposed and a comparison with another variable-fidelity model which includes POD is made. In chapter eight an attempt is made to model the transonic flow around an RAE 2822 aerofoil using kernel POD. A direct comparison is made between the reduced-order modelling effectiveness of the linear and kernel POD methods. This thesis concludes with a discussion of the main findings of this study and various suggestions for future work

Chapter 2

2.0 Literature review

In this chapter a review of the literature on the proper orthogonal decomposition is presented. The POD is a methodology for the modal decomposition of an ensemble of functions, usually this is in the form of data obtained from either experiments, measurements or numerical simulations. A modal decomposition may be useful for various reasons such as, for the identification of structures in data or for statistical analysis of sample data or as it is employed here, as a set of modes for the approximation of the numerical solution of problems with characteristics similar to the ones used in the construction of the basis modes. The latter is the subject of reduced-order modelling. This literature review includes examples of these applications together with some relevant extensions and the most recent developments.

2.1 Model reduction techniques

Antoulas *et al.* [3] have classified model reduction techniques into two categories, namely those methods based on the singular value decomposition (SVD) and those methods based on moment matching. The SVD based methods are the Hankel-norm approximation, balanced truncation, singular perturbation, cross grammian and the Karhunen-Loève expansion, while the moment matching based methods are the families

of Lanczos and Arnoldi procedures. In the first category, the Karhunen-Loève expansion turns out to be a popular method for both linear and non-linear systems. In the second category, most of the attempted applications were intended either for some specific type of non-linear system or have not been applied to general non-linear model order reduction problems in a practical way. In view of this, it is reasonable to say that for non-linear systems in aerodynamics governed by the Euler or Navier-Stokes equations, the Karhunen- Loève expansion could be an appropriate choice for model order reduction.

2.2 The Karhunen – Loève expansion

The Karhunen – Loève expansion was proposed independently by various people at different times amongst these are Karhunen [4] and Loève [5]. This method is known under a variety of names in different fields such as; Principal Component Analysis [6], Empirical Component Analysis [7], Quasiharmonic Modes [8], Proper Orthogonal Decomposition [9], Singular Value Decomposition [10] and Empirical Eigenfunction Decomposition [11]. It is important to remark that in mathematics the SVD is a powerful matrix decomposition that offers a method of solution to the principal component analysis and therefore these two are strictly not the same. However, the two methods are so closely related to one another that many times their names are used interchangeably.

From the mathematical point of view the Karhunen – Loève expansion is essentially a transformation which diagonalizes a given matrix and makes it into a canonical form. A review of the early history of the Karhunen - Loève expansion can be found in [12]. The mathematical content of the Karhunen - Loève expansion is classical and is contained in a paper by Schmidt [13]. The large number of computations required to find the eigenvectors by this technique, made this method virtually impossible to use until the middle of the last century. Radical changes came with the appearance of powerful computers and development of efficient algorithms to compute the eigenvectors. In this

work, the technique shall be referred to as the proper orthogonal decomposition (POD) which has been almost generally adopted in fluid flow analysis.

2.3 The proper orthogonal decomposition

The POD is a statistical procedure that provides a mathematical framework by which the high-energy components of a fluid flow are extracted. This may subsequently lead to a reduced model. The method provides a low-dimensional representation of a characteristic large-scale structure by decomposing it into a set of uncorrelated data-dependent linear components. The components are the eigenfunctions of a two-point correlation tensor and the expansion is optimal in several senses. For instance, the POD eigenfunctions minimize the mean square error and maximize the total energy captured in each co-ordinate direction, subject to orthogonality constraints. Also while the method is optimal with respect to second-order moments, there is no loss of higher-order moment information since the eigenfunctions form a complete basis. However, the efficiency with which higher-order information is captured is not prescribed by any optimality condition. Therefore, the POD analysis yields a set of empirical modes which describe the dominant behaviour or dynamics of a given problem. The set of empirical modes retrieved are variously called proper orthogonal modes, empirical eigenfunctions, empirical basis functions and empirical orthogonal functions.

Nowadays, the technique is used extensively in various fields such as that of detection, estimation, pattern recognition, image processing, as an efficient tool to store random processes and in system controls. The method is also used in stochastic turbulence problems [9], in which context, the associated eigenfunctions are identified with the characteristic eddies of the turbulence field. The method has also been used in steady aerodynamic analysis such as the design of inviscid aerofoils by LeGresley and Alonso [14], and parametric studies by Epureanu *et al.* [15] and Bui-Thanh *et al.* [16] and [17].

2.3.1 POD-based reduced-order modelling

Sirovich [18] introduced the method of snapshots as a way for efficiently determining the POD basis functions or modes for large problems. This technique can be used for a variety of applications, including derivation of reduced-order dynamical models for fluid dynamic applications as in Holmes *et al.* [19]. In particular, the method of snapshots has been widely applied in conjunction with computational fluid dynamic (CFD) formulations to obtain reduced-order models for unsteady aerodynamic applications as in Dowell *et al.* [20], Hall *et al.* [21], Romanowski [22] and Beran and Silva [23]. In this technique, a set of instantaneous flow solutions or snapshots is obtained using some numerical or experimental method. The POD process then computes a set of basis functions from these snapshots, which is optimal in the sense that, for any given basis size, the error between the original and reconstructed data is minimized. Reduced-order models can be derived by projecting the model onto the reduced space spanned by the POD modes. The original numerical model is usually governed by non-linear partial differential equations such as Navier-Stokes or Euler equations, which are very computationally expensive especially for a flow past a complex configuration. Moreover, the model is computationally intensive and hence unsuitable if coupling of the CFD model with another disciplinary model is required. The reduced-order model, however, is just a small system of ordinary differential equations which can be solved efficiently. In view of this, the application of POD for obtaining a model of reduced complexity has been extended to many other applications. As is common practice in dealing with non-linear partial differential equations, the linearized approximate equations are studied under some assumptions such as small disturbances of the inflow parameters as in Hall *et al.* [21]. Although these linearized models are significantly simple compared to the original non-linear models, they usually have a prohibitively high number of states and hence they are still very expensive to simulate. More recently, Kim [24] and Willcox *et al.* [25] have developed efficient POD methods for linearized systems. In those works, instead of obtaining the snapshots from a time-dependent simulation which is computationally expensive, the linearized system in time domain is first transformed to the frequency domain and then the POD snapshots are computed efficiently and accurately over the frequency range of interest. However, the major restriction of the

linearized model is the assumption of small perturbation unsteady flow and hence it is not suitable for the cases in which the non-linear effect may be important and perturbations are large. To this end, Thomas *et al.* [26] proposed a non-linear reduced order model of a CFD solver. The technique is based on a Taylor series expansion of the flow solver residual together with a Ritz type expansion using proper orthogonal decomposition modes. This method was applied to an inviscid transonic aerofoil configuration and produced better results than a linear reduced-order model when considering large changes in the flow solver input variables.

2.3.2 POD in image processing

In the context of the characterization of human faces, Kirby and Sirovich [27] have employed the POD procedure as a useful tool for the image compression of human faces. In this application, each distinguishable digital picture of a human face or snapshot is transformed into a rectangular matrix whose elements are integers in the interval [0, 255]. This corresponds to the pixel resolution. The POD procedure is then performed to obtain the POD modes. After which any image which is not a member of the ensemble, is projected onto the reduced space spanned by the dominant POD basis vectors to calculate the POD coefficients. This small number of POD coefficients is used to represent the face with a small error, instead of a big accurate matrix. The results show that roughly 100:1 compression ratio (ratio between the number of elements of a matrix representing a face and the number of POD coefficients corresponding with that face is obtained. In addition to this work, Everson and Sirovich [28] have presented a variation of the basic POD method that handles incomplete data sets. This method actually makes use of the conventional POD method for solving an unconstrained optimization problem which approximately reconstructs the full data from the incomplete one. Given a set of POD modes, an incomplete data vector can be reconstructed accurately by solving a small linear system of equations derived from the optimal conditions. It was shown that this procedure gives an unbiased estimate of the data. Moreover, if the snapshots themselves are damaged or incomplete, an iterative method can be used to derive the POD basis. The POD basis is in turn used to reconstruct the incomplete data. This method has been

successfully applied for the reconstruction of human face images, from partial data with 25% of the data missing. An interesting extension to this application was performed by Bui-Thanh [17] in external flow transonic aerodynamics.

2.3.3 Flow-field reconstruction using POD

The POD technique has been used for the reconstruction of a flow field from an incomplete data set [17]. This application is based on the procedure developed by Everson and Sirovich [28] for the reconstruction of human face images from an incomplete data set. The case considered is a NACA 0012 aerofoil at a freestream Mach number of 0.8. The POD bases were created from 51 snapshots. These were computed at uniformly spaced values of the angle of attack in the interval [-1.25^{o}, 1.25^{o}]. An incomplete flow field was generated by a computation of the flow solution at an angle of attack of 0.77^{o} and retaining only 121 surface pressure values out of a total number of 6369 pressure values throughout the whole flow field. With just this limited surface pressure data available, the complete pressure field was determined accurately with only six POD modes. Moreover, in this work the authors examined the sensitivity of the reconstruction result to both the quantity and location of this surface data. In order to select a limited number of pressure measurements, a heuristic approach as used by Cohen *et al.* [29] for unsteady flows was adopted. This approach suggests that sensors should be located in areas of high modal activity. The POD modes of a flow often exhibit sinusoidal spatial variation and sensors placed at local POD modal minima and maxima yield effective flow sensing results. This heuristic procedure was applied in that work to study the sensitivity of the reconstruction results to the amount of available data. An initial configuration of 11 measurement points corresponding to spatial optima of the first POD mode plus a few other points near the leading edge, where all POD modes were seen to vary rapidly, was chosen. Additional measurements were then considered by adding in turn the spatial optima of modes 2, 3, 4, 5 and 6 resulting in studies with 15, 21, 29, 31 and 39 sensing points respectively. From this work it was observed that the percentage error between the exact and reconstructed pressure measurements was very low even with a very small number of sensors, confirming the effectiveness of the heuristic sensor

placement algorithm. It was also noticed that subsequent reduction in the error diminishes as higher modes are considered in the sensing which is consistent with the fact that subsequent modes constitute progressively less of the total energy. The POD eigenvalues can therefore be used not only to select the number of modes but also to choose an appropriate number of sensors.

Another consideration in [17] was the effect of experimental error. To study this effect, noise was simulated by the addition of a random noise component to each computed pressure value. Maximum noise levels which were expressed as a percentage of the freestream pressure value of ± 1, 2, 5, 10 and 20 % were applied. All measurements were made with 121 measurement points and with 2, 4, 6 and 8 POD modes. It was observed that as the magnitude of the sensor noise increases, the magnitude of the reconstruction error also increases even though the prediction is very good with a relatively high noise level of 10 %. As more POD modes are used the sensitivity to noise increases. Using more POD modes in the representation allows the flow-field to be represented more accurately, but it also allows a greater degree of 'data matching' as the incomplete data formulation attempts to minimize the error between the measured and the reconstructed data. If very high levels of noise are present, the reconstruction with smaller number of modes yields more accurate results. In most fluid applications, the dominant POD modes tend to correspond to flow patterns with low spatial frequency. These modes do not have the resolution to match the high frequency components of the sensor data, causing them to naturally filter sensor noise. In practice one should use the POD eigenvalues to select the minimum possible number of POD modes that will achieve the desired resolution in nominal cases. Further, if very high levels of sensor noise are expected, consideration should be given to reducing the number of modes beyond what might be chosen under nominal circumstances.

A further investigation was conducted by creating a set of POD basis vectors from an incomplete set of snapshots. Again, the NACA 0012 aerofoil at a freestream Mach number of 0.8 was considered. A 26 snapshot ensemble was used with steady pressure solutions at angles of attack in the interval [0^o, 1.25^o]. To create the incomplete snapshot

set, 30 % of the pressure data of each snapshot was discarded randomly. The data was repaired by first repairing the missing data points in each snapshot with the average over the available data at that point. Thus a new ensemble of data was created which had no missing values. With the new ensemble, a first approximation to the POD basis was constructed. Then each snapshot in the ensemble was repaired using the first approximation of the POD basis. The repaired ensemble was then used to construct a second approximation to the POD basis. An iterative procedure was adopted and stopped after 50 iterations. From the convergence of the POD eigenvalue spectrum of the incomplete ensemble it was observed that after one iteration, the first two eigenvalues converged. After 45 iterations, only the first five eigenvalues converged; however these accounted for almost all of the flow energy.

2.3.4 POD in fluid-flow analysis

The proper orthogonal decomposition has been applied by Kirby *et al.* [30] to the analysis of an ensemble of snapshots which were obtained from sampling a supersonic shear layer large-eddy simulation. The computed optimal basis functions were used to economically characterise the sampled flow realizations. In that work it was observed that the essential flow features were captured by using 80 % of the energy while smaller-scale flow features were resolved by retaining more terms in the expansion. In the same work it is interesting to note that for this unsteady flow problem a great deal of data is necessary to provide sufficient statistics for the computations of the eigenfunctions. However this problem has been mitigated in part by exploiting the symmetry of the flow and consequently extending the data. Moreover, it has been reported that the eigenfunctions are sensitive to the scaling of the flow variables since these are in different units and have widely varying magnitudes. Consequently, three scaling factors were evaluated which make the magnitude of the fluctuating variables to be on average of order 1.

In general, the proper orthogonal decomposition technique provides a reduced order model of the field, and the subspace represented by this model is the span of the high energy modes. The goal of POD is to provide an efficient way to capture the most

information from a high-dimensional process in only a few dominant modes. The results are then synthesized and related to physical phenomena. The mathematical structure of POD analysis provides an orthogonal decomposition of this low dimensional subspace, since the modes are eigenvectors of the covariance matrix. This mathematical structure follows from the mathematical theory of Gaussian processes, not from the underlying physics of fluid flow. The orthogonal modes have no straightforward interpretation in terms of flow since they are the result of an orthogonal 'mix' of various flow mechanisms. In fact POD models of pressure field have in the past been challenged with questions of interpretation [31]. A new approach has been presented by Dunyak *et al.* [32] which seeks to capture independent but non-orthogonal flow mechanisms. This technique was termed as Non-orthogonal decomposition (NOD). In this technique the high energy subspace is spanned by a basis chosen to maintain de-correlation of coefficient time series, as in POD, while increasing the relative independence as indicated by the difference between the joint distribution and the product of the marginal distributions. Preliminary results strongly suggest that this technique effectively separates and identifies separate flow mechanisms in the area of flow vortices over a flat roof. In this work an example is brought forward in which pressure data is collected over the roof of an experimental building to capture corner vortices. Using the standard POD analysis most of the flow energy was in the first two POD modes. From this analysis an identification of the physical mechanisms for the full-scale pressure field is made. While the POD mode appears to show a mixture of phenomena; a tightly focused vortex as well as a larger-scale wake instability, the NOD modes decouple these phenomena by seeking a nearly independent decomposition. One mode captures the larger-scale wake instability without the focused vortex effect, while the other NOD mode appears to capture the vortex alone.

2.3.5 POD-based ROMs in parametric studies

While the use of POD to capture the time variation of fluid dynamic problems has been widespread, the development of reduced-order models to capture parametric variation is less commonly available in technical literature. In one of these few works,

Epureanu *et al.* [15] has used POD to develop reduced-order models for flows in turbomachinery with sampling in both time and over a range of inter-blade phase angles. Although the snapshot ensemble is computed at a single Mach number condition, the resulting reduced models were applied to flows at varying Mach numbers. Accurate results were obtained for Mach numbers close to that used in the snapshots set. In another work that considers parametric variation, Ly and Tran [33] have developed a fast computation that uses a POD basis to predict the steady-state temperature distribution of flow in a square cavity as the Rayleigh number is varied. In this case each snapshot is at a steady temperature distribution corresponding with each value of Rayleigh number. This method is a simple combination of the POD basis and an interpolation procedure, but it was shown to be effective and accurate.

The POD technique has also been applied by Bui-Thanh *et al.* [16] and [17] in a steady transonic external aerodynamic problem. In both works, all snapshots were computed by an inviscid steady-state CFD code which uses a finite volume formulation. In [16] the POD technique was coupled with a cubic-spline interpolation method in order to develop low-order models that capture the variation in parameters. The problem considered in this work is steady flow about the NACA 0012 aerofoil with varying angle of attack and Mach number. The Mach number range considered [0.75, 0.85] was divided into 20 uniform intervals and the angle of attack range [0, 1.25^{o}] was divided into 10 uniform intervals. This resulted in a total number of snapshots in the ensemble of 231. Based on this snapshot set, interpolation was used to predict the flow pressure contours at any Mach number and angle of attack within the range considered. When the pressure flow field was predicted at an angle of attack of 0.45^{o} which was not one of the snapshots and a Mach number of 0.8, with twenty-five eigenfunctions the contours of the reduced-order model matched closely with the computed one. When the pressure flow field was predicted at an angle of attack of 0.5^{o} and a Mach number of 0.812 which was not one of the snapshots, thirty eigenfunctions were required to achieve the desired level of accuracy. This indicated that the prediction is more sensitive to Mach number rather than the angle of attack. When the pressure flow field was predicted at an angle of attack of 0.45^{o} and a Mach number of 0.812, as none of these values were used to generate the

snapshots, 40 POD modes offered a satisfactory level of accuracy. These results showed that the POD method combined with interpolation allows models to be derived that accurately predict steady-state pressure fields over a range of parameter values. However, it is emphasized that in order for the interpolated result to be reliable, the properties of interest must vary smoothly with the parameters under consideration. It has been stated that the approach can be extended to the case where more than two parameters vary and which may include geometrical properties in order to apply the models in an optimisation context.

The POD combined with a response surface method was employed by Tang *et al.* [34] for the reconstruction and prediction of aerodynamic and aerothermal solutions of an X-34 configuration. In this work it has been suggested that this methodology will be used for the design and optimization of the thermal protection system of the X-34 reusable launch vehicle.

In that work, a computationally efficient, panel-method based hypersonic aero-thermoelastic software was developed and used to conduct the structural design and optimization of the thermal protection system of this reusable launch vehicle. The results obtained from this software on the windward side of the X-34 were verified with CFD results. However, discrepancies were found on the leeward side of the X-34 as the generated results lacked sufficient flow field detail due to the inadequacy of the panel-based method for representing the non-linear hypersonic effects in the high angle of attack flow regime. For this reason, they had to look upon high-level CFD methods which are not computationally efficient enough. Consequently, to predict the lee-side aerodynamics/aerothermodynamics on the X-34 with high-accuracy and efficiency, a POD/RSM based aerodynamic module was developed from the high-level CFD solutions of the X-34 configuration. In this work an extensive CFD solution database for the X-34 was created. Two parameters were considered namely the angle of attack which was varied between 0 and 30^{o} and Mach number which was varied between 2 and 10. It was not reported how many snapshots were considered. The POD technique was applied on this dataset to facilitate reduced-order modelling of steady state super- and hyper-sonic

aerothermodynamics. A response surface of the scalar coefficients versus the Mach number and the angles of attack was constructed by using RBF networks. It was reported that for all Mach numbers considered, the first three POD modes sufficed for a reasonably accurate reconstruction of the X-34 surface pressure coefficients over the entire range of angles of attack. This suggests a continuous variation of the solution in the design space. It was also reported that for the case where the Mach number is 10 and the angle of attack is 30 degrees, the POD reconstruction resulted in a slightly higher pressure near the nose than the original CFD solution. The authors failed to remark that this may have been improved by considering more modes. Alternatively by enriching the dataset over that region would have also captured the details better. Predictions were conducted with good efficiency too taking about one second on a 2.4 GHz PC for determining the surface pressure coefficients. In that work it was reported that this module proved to be not only computationally more efficient than the low-level engineering methods, but also as accurate as the high-level CFD methods, making it valid for MDO and real-time applications.

2.3.6 Aerofoil inverse design optimisation using POD-based ROMs

In the field of aerofoil design, Lighthill [35] was the pioneer in aerofoil inverse design using the method of conformal mapping. This method was limited to incompressible flow problems but it was later extended by McFadden [36] to include compressible flows. By introducing the finite-difference method to evaluate the sensitivity derivatives, Hicks and Henne [37] were the first to solve a constrained aerodynamic optimization problem. Since then, gradient based methods have been used for aerodynamic design. Relatively recently, Jameson [38] has applied control theory in the context of aerofoil shape design optimization using Euler and Navier-Stokes codes. In this method, although an adjoint equation needs to be solved to obtain the gradient, the total cost for computing the gradients is independent of the number of design variables and hence it is more efficient than finite-difference methods. Significant progress has been reported on the application of this approach to the design of realistic geometries [39]. However, the use of this

adjoint-solver approach for situations involving multiple disciplines and a large number of design constraints, has been somewhat limited. This may be due to the fact that the adjoint equations, boundary conditions and gradient calculation formulae are cost function dependent and so need to be re-derived every time the cost function changes. Moreover, it is not possible to treat arbitrary forms of the cost functions, thus limiting the applicability of this procedure. Consequently, alternative techniques for aerodynamic design of wings most especially are sought, such as reduced-order methods. POD has been used in reduced-order methods for multi-disciplinary design environments where aerodynamics and structural analysis are included by Pettit and Beran [40], Thomas *et al.* [41]. Moreover, the work by LeGresley and Alonso [14] and Bui-Thanh *et al.* [17] demonstrated that the POD method could be used as a low-cost, low-order approximation for aerodynamic shape optimisation. The method proposed by Bui-Thanh *et al.* is based on the gappy POD reconstruction procedure of Everson and Sirovich [28] while that used by LeGresley and Alonso is based on the gradient approach to cost function optimisation. In both cases, conventional CFD methods were used to generate the data ensemble for the aerofoil inverse design problem. The POD procedure is then used to compute a set of optimal eigenfunctions from these snapshots. The two methods differ from each other in the way the cost function is evaluated and the way the optimal solutions are arrived at.

In particular, LeGresley and Alonso [14] have used the POD technique for both direct and inverse aerofoil design problems. In that work, a set of pressure field distributions corresponding to different aerofoil profiles were computed using an Euler solver. Different aerofoil profiles were created by perturbing the design variables of the base shape. The POD basis is then computed and used to construct a reduced-order model for Euler equations to compute new, approximate solutions for any arbitrary aerofoil at significantly lower computational costs. In this manner, both direct and inverse aerofoil design problems can be done efficiently using a gradient-based optimization procedure with the information from the reduced-order model.

Bui-Thanh *et al.* [17] applied the POD method to the problem of inverse aerofoil design. In that work, a collection of snapshots was generated by choosing a set of aerofoil

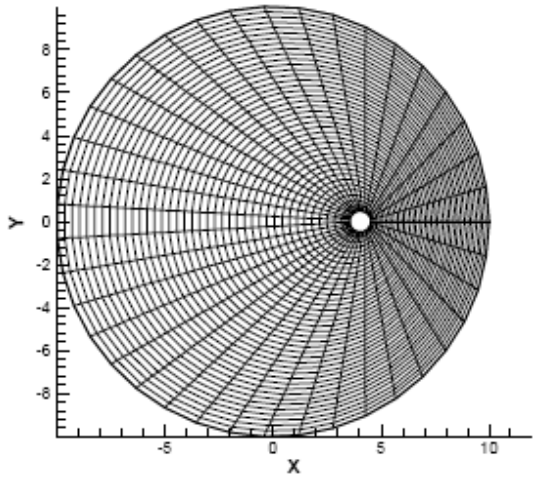
shapes and computing their corresponding surface pressure distributions. The POD technique was then used to determine the optimal aerofoil shape that produces a given target pressure distribution. An important point of this method is that rather than containing only the flow variables, the snapshots are augmented to contain also the aerofoil co-ordinates. The minimal solution of the cost function is sought with a target vector that contains the required target pressure distribution and the unknown corresponding aerofoil co-ordinates. Thus, the target vector contains both known and unknown elements and so the procedure of Everson and Sirovich [28] is used to reconstruct the missing data points.

In [42] a comparison is made between these two POD methods for aerofoil inverse design optimization. The performance of these two methods on the inverse design of various aerofoil shapes such as Korn, NACA63212, HQ2010 and GOE117 were compared and evaluated. From this it was concluded that while both methods are efficient and accurate once appropriate flow snapshots are collected, with the same data ensemble, the gradient-based POD optimization method appears to give better results for aerofoil shape design and surface pressure distribution when compared with the target.

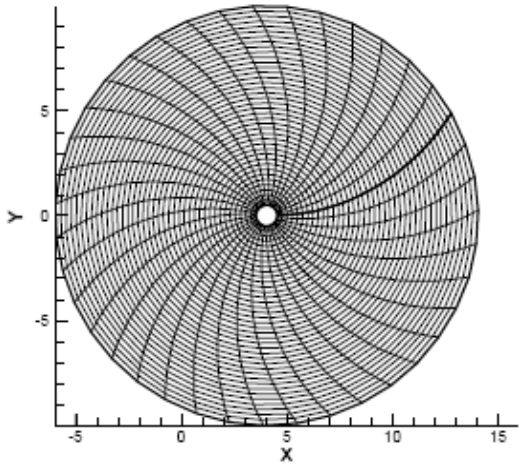
2.3.7 The effect of deforming grids on POD-based ROMs

Since POD-based reduced-order modelling is a space-index transformation from the physical to the computational domain, it does not account for changes in the grid boundaries and if possible must somehow be modified to properly reflect the changes in the physical domain. An investigation of the effect of deforming grids on POD-based ROMs was conducted by Anttonen [43], whereby various problems were considered such as potential flow about a uniformly translating 2-D cylinder in still fluid and an oscillating panel in 2-D subsonic flow. Anttonen reported that deforming grids required more modes than rigid grids for similar accuracy levels as shown in figure (2-1). Also, POD-based ROMs are less accurate on grids that differ from the grids at which these were trained. In that work, a metric was developed which correlates the accuracy of the POD-based ROM to deformations from the training grids and was used to determine

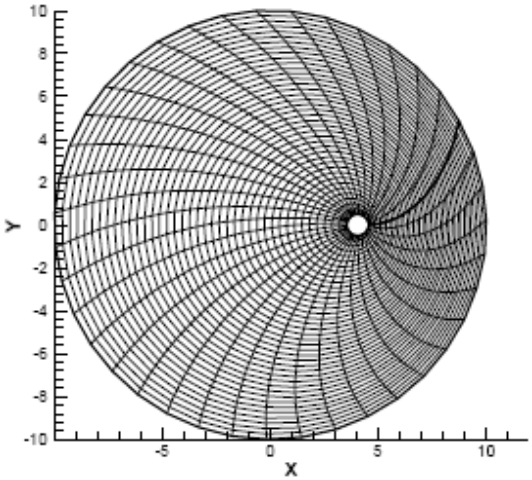
when to switch between ROMs of different grids for reduced order models of problems with deforming grids. The algorithm uses the most appropriate POD-based ROM to provide the most accurate solution.



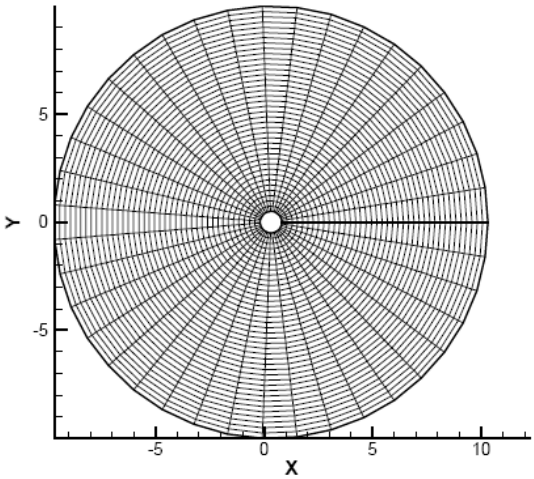
(a) Translating grid (Deforming)



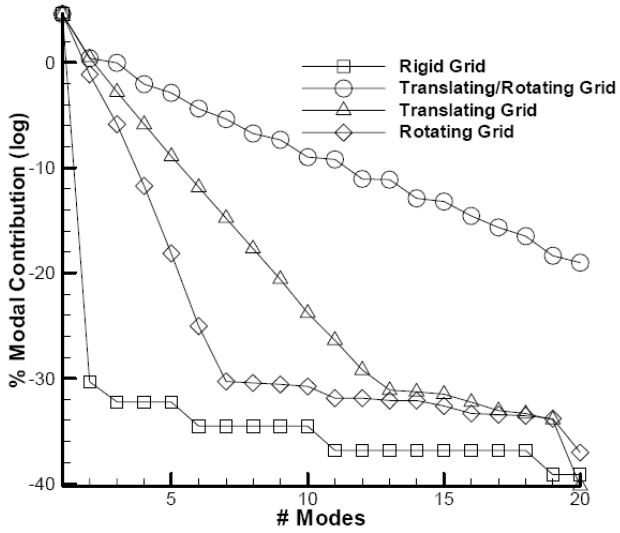
(b) Rotating grid (Deforming)



(c) Translating/Rotating Grid



(d) Rigid Grid



(e) POD-based ROM Modal Contribution

Figure (2-1) The grids considered in the analysis and the modal contribution for a potential flow about a uniformly translating 2-D cylinder in still fluid

2.3.8 POD-based ROMs for moving discontinuities in fluid flow

A significant challenge in the use of POD-based ROMs for high-speed flow applications is to capture moving shock waves as the flow parameters or boundary conditions change. Techniques that generate POD-based reduced-order models (POD/ROM) for subsonic and supersonic flows do not generate a useful ROM for a high-speed flow application with moving shocks. Excessive data and time are required for modelling this accurately. For this problem Lucia [44] proposed and used a technique to exploit POD for accurately treating moving shock waves. This technique involves the decomposition of the solution domain to isolate those regions that contain shocks and so produces internal boundaries within the flow-field between the various domains. The main idea behind this is to use the POD modes for a global approximation of the flow-field and to use a standard finite volume scheme in the region where the shocks occur. A

reduced-order model for each region is developed independently and the solution for the entire domain is formed through a linking of the boundaries of each region, using optimization based solvers to ensure a smooth solution between overlapping parts of the internal boundaries.

This technique was applied to a one-dimensional quasi-steady nozzle flow-field by Lucia [44] for demonstration. However, LeGresley and Alonso [45] applied this technique for the shape optimisation of a two-dimensional aerofoil. The results attained are good though some discrepancy could still be detected between the high-fidelity solution in dashed line and the POD/ROM with domain decomposition POD/ROM/DD as shown in figure (2-2). Though, it must be emphasized that the discrepancy is not necessarily due to a shortcoming of the POD/ROM/DD technique, but it could also be due to the changes in the physical domain.

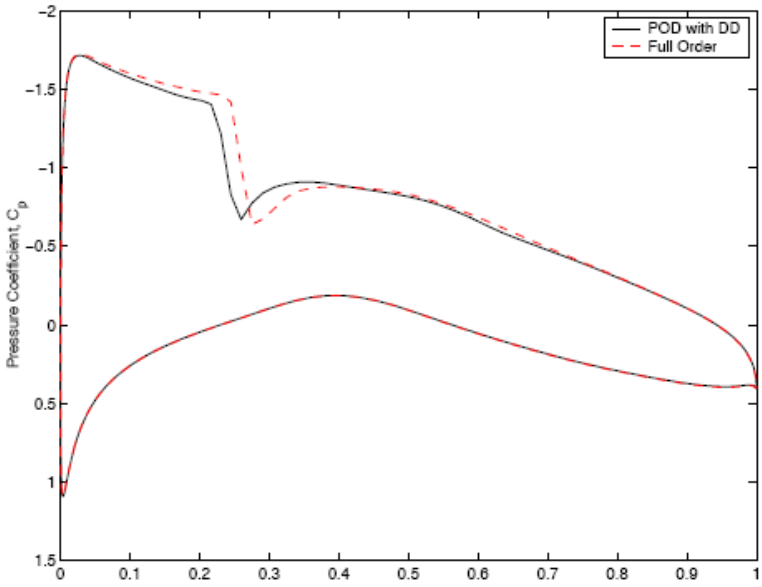


Figure (2-2) Coefficient of Pressure distribution over the RAE 2822 aerofoil using POD-based ROM with domain decomposition compared with the full order solution. Picture taken from reference [45].

It is thought that by combining POD-based ROMs with conventional spatial discretization schemes, more efficient multi-scale algorithms for the analysis of some parameter dependent PDEs can be developed. The idea of multi-scale modelling is to use two models at least – one to represent the global features of the solution and a detailed model to capture the local features. The main issues involved in designing a multi-scale scheme are (1) the numerical method employed to identify the localized regions where a full-order model should be used and (2) the approach used for handling the coupling between the two models in order to arrive at consistent values for the flow-field variables in the interface region. Thus, while it is recognised that this domain-decomposition technique is reasonably good, it does offer some challenging issues.

2.3.9 POD-based ROMs for aeroelastic analysis of a complete F-16 aircraft configuration

In [46] and [47], Lieu *et al.* applied POD-based ROMs for an aeroelastic analysis of a complete F-16 aircraft configuration with clean wings for varying Mach number and low angles of attack. In that work it has been shown that the POD method produces accurate ROMs for the aeroelastic analysis of a complete aircraft configuration at a fixed flight condition. However, changes in the Mach number or the angle of attack often require the reconstruction of the ROM in order to maintain accuracy. Consequently, all this defeats the most important and the very much sought-after computational efficiency. In that work it was shown that straightforward approaches for ROM adaptation lead to inaccurate POD bases in the transonic flight regime. Thus, a new ROM adaptation scheme is proposed and evaluated for varying Mach number and angle of attack. This scheme interpolates the subspace angles between two POD subspaces and then generates a new POD basis through an orthogonal transformation based on the interpolated subspace angles. This computational methodology is applied to a complete F-16 configuration in various airstreams. The predicted aeroelastic frequencies and damping ratio coefficients were compared with counterparts obtained from full-order non-linear aeroelastic simulations and flight test data. Good correlations are reported in the transonic

flow regime. It is reported that this technique has a significant potential for accurate, real-time, aeroelastic predictions.

2.3.10 POD combined with centroidal Voronoi tessellations

The computational cost of the POD methodology may become exorbitant for problems with a very large number of snapshots. For such problems, Du *et al.* [48] proposed a method by which the dominant basis functions are computed using centroidal Voronoi tessellations (CVT). This approach has the advantage that it does not involve the solution of an eigenvalue problem and can also be efficiently parallelized. Moreover, Du and Gunzburger [49] presented an approach for model reduction that hybridizes centroidal Voronoi tessellations with POD. The idea behind this technique is that the POD have been used to define reduced bases for low-dimensional approximations of complex systems, while the CVT have been used in a variety of data compression and clustering settings. Thus, it is expected that this proposed technique which is referred to as the centroidal Voronoi orthogonal decomposition CVOD inherits favourable characteristics from both its parents and so introduces the concept of clustering into the decomposition. By imposing various clusters within some domain, each sub-CVOD basis for a specific cluster can be used to capture the dynamics of that cluster. Another interesting feature of CVOD is that it avoids the over-crowding of the reduced basis into a few dominant modes which is a possible drawback of POD.

Chapter 3

3.0 The mathematical model

In this chapter a detailed description of the numerical methods employed in setting up the mathematical model is presented. We start with a description of the CFD tool utilised in this work followed by an in-depth account of the various numerical methods adopted to develop the low-cost surrogate of the CFD tool. These include techniques for sampling the parametric space using a design-of-experiment approach, a reduced-order model for the appropriate representation of the dominant characteristics of the fluid-flow structure and generalized linear models for the construction of response surfaces. The latter includes both regression and interpolation techniques.

3.1 The CFD analysis tools

One of the objectives of this work was to develop a high-fidelity low-cost surrogate of a computational fluid dynamics analysis tool, namely the IMPNS (Implicit Multiple in blocks and grids Parabolised Navier-Stokes) flow solver [51] [52]. This is a parabolised Navier-Stokes solver which has been developed at Cranfield University. The IMPNS software has been developed to provide a practical flow solver for problems in high-speed external weapon aerodynamics. Most of the flow computations in this research thesis were conducted using the IMPNS, even though the transonic flow computations over the RAE 2822 aerofoil in chapter 8 were conducted using the commercial software Fluent [53].

3.1.1 The IMPNS flow solver

IMPNS provides algorithms for the solution of the Euler, Thin-layer or Parabolized Navier Stokes equations together with a range of turbulence closures that includes the algebraic model of Baldwin and Lomax [54] enhanced with modifications due to either Degani and Schiff [55] or the curvature method of Qin and Jayatunga [56] and variants of the one-equation model of Spalart and Allmaras [57]. The turbulence closure is coupled with the mean flow equations in a segregated fashion.

The governing equations are formulated for a finite control volume and solved using an implicit space marching procedure. For flows in which there is no upstream influence a single sweep is employed starting at the nose of the configuration and proceeding in the stream-wise direction. The approach has been extended to allow for flows with upstream influence, for example blunt body flows and flows exhibiting axial separation. In this case a multi-sweep procedure in which the solver marches backwards and forwards is employed to capture the elliptic characteristics of the governing equations. A combination of single sweep and multi-sweep strategies can be used to solve for flows that contain embedded regions of flow where upstream influence is important.

IMPNS provides a number of schemes that can be employed in the spatial discretization, see for example Qin and Ludlow [58]. In the present calculations the spatial discretization is performed using an approximate Riemann solver based on the work of Osher and Solomon [59] together with a central difference based scheme for the viscous fluxes.

An implicit system of equations arises at each marching plane following the spatial discretization. This system is solved using a relaxation approach in which an additional pseudo-time derivative is added to the steady governing equations. The implicit system is then solved by marching to the steady state in pseudo-time. Convergence of the pseudo time relaxation is accelerated through the use of a combination of an implicit Newton-Krylov method [60] and full multi-grid [61].

IMPNS uses standard structured multi-block grids. To provide geometric flexibility and to reduce computational expense non-matching block faces are permitted in the stream-wise direction. This permits changes of grid topology in the stream-wise direction, allowing the grid to conform to the geometric characteristics of the configuration being studied. IMPNS has been used extensively to study the aerodynamics of high-speed weapon configurations with remarkable success, further details of its development and application can be found in references [62], [63], [64], [65], [66], [67], [68], [69], [70].

3.1.2 Mathematical formulation of the IMPNS

The governing equations are the Favre-averaged Navier-Stokes equations describing the conservation of mass, momentum and energy for the steady flow of a compressible ideal Newtonian fluid. This system of equations can be written in conservative integral form as,

$$\int_S \mathbf{F}_n dS = \mathbf{0} \qquad \text{Eq.(3-1)}$$

in which $\mathbf{F}_n$ is the flux through the surface of an arbitrary control volume bounded by the surface S [51] [71]. The flux may be conveniently split into convective, $\mathbf{F}_n^{\text{inv}}$, and diffusive, $\mathbf{F}_n^{\text{vis}}$, terms so that,

$$\mathbf{F}_n = \mathbf{F}_n^{\text{inv}} - \frac{1}{\text{Re}}\mathbf{F}_n^{\text{vis}} \qquad \text{Eq.(3-2)}$$

The convective flux $\mathbf{F}_n^{\text{inv}}$ may be obtained from,

$$\mathbf{F}_n^{\text{inv}} = (\hat{\mathbf{n}}^{\text{T}}\mathbf{u})\boldsymbol{\Phi} + p(0, \hat{\mathbf{n}}^{\text{T}}, 0)^{\text{T}} \qquad \text{Eq.(3-3)}$$

where $\hat{\mathbf{n}}$ is the unit vector normal to the surface of the control volume and

$$\mathbf{\Phi} = \mathbf{Q} + p(0,0,0,0,1)^{\mathrm{T}} \qquad \text{Eq.(3-4)}$$

and $\mathbf{Q} = \left(\rho, \rho\mathbf{u}^{\mathrm{T}}, \rho E\right)^{\mathrm{T}}$ is the vector of conservative variables.

The diffusive flux may be calculated from,

$$\mathbf{F}_n^{\mathrm{vis}} = \hat{\mathbf{n}}^{\mathbf{T}} \left[\boldsymbol{\tau} \begin{pmatrix} \mathbf{0} \\ \hat{\mathbf{i}} \\ \hat{\mathbf{j}} \\ \hat{\mathbf{k}} \\ \mathbf{u} \end{pmatrix} + k \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \nabla T \end{pmatrix} \right] \qquad \text{Eq.(3-5)}$$

where the components of the shear-stress are

$$\tau_{ij} = \left(\mu + \mu_T\right)\left(\partial_i u_j + \partial_j u_i - \frac{2}{3}\delta_{ij} \nabla T \mathbf{u} \right) \quad i, j \in \{1,2,3\} \qquad \text{Eq.(3-6)}$$

This system of equations is closed by assuming that air is an ideal Newtonian fluid for which Stokes' hypothesis is valid. The coefficients of thermal conductivity and molecular viscosity are modelled using Sutherland's law.

For a structured grid containing hexahedral cells a discrete form of Eq.(3-1) can be obtained,

$$[\mathbf{F}_\xi s_\xi]_{i-½,j,k}^{i+½,j,k} + [\mathbf{F}_\eta s_\eta]_{i,j-½,k}^{i,j+½,k} + [\mathbf{F}_\zeta s_\zeta]_{i,j,k-½}^{i,j,k+½} = \mathbf{0} \qquad \text{Eq.(3-7)}$$

where $\mathbf{F}_\xi, \mathbf{F}_\eta$ and $\mathbf{F}_\zeta$ are the fluxes through cell faces aligned with the streamwise and cross-flow directions respectively. By neglecting the viscous flux in the streamwise direction a reduced form of the Navier-Stokes equations, the Parabolized Navier-Stokes equations (PNS) [51], can be obtained,

$$[\mathbf{F}_{\xi}^{\text{inv}} s_{\xi}]_{i-½,j,k}^{i+½,j,k} + [\mathbf{F}_{\eta} s_{\eta}]_{i,j-½,k}^{i,j+½,k} + [\mathbf{F}_{\zeta} s_{\zeta}]_{i,j,k-½}^{i,j,k+½} = \mathbf{0} \qquad \text{Eq.(3-8)}$$

In Eq.(3-8) the cross-flow fluxes remain unaltered, except for neglected streamwise contributions to derivatives in the viscous fluxes. Provided a suitable discretisation of the streamwise flux is used to maintain stability, Eq.(3-8) may be solved using an implicit *multiple-sweep* space-marching methodology, where flow information is propagated in both directions by marching in both the positive and negative streamwise direction.

If the flow outside of the boundary layer is supersonic in the streamwise direction and there is no streamwise separation present, an approximation due to Vigneron *et al*. [72] enables the PNS equations to be solved by a well-defined (stable) *single-sweep* space-marching procedure in the streamwise direction. Vigneron obtained such a procedure by modifying the pressure gradient in the subsonic portion of the boundary layer to suppress the elliptic character of Eq. (3-8); this can be done by replacing the streamwise convective term $\mathbf{F}_{\xi}^{\text{inv}}$ by,

$$\mathbf{F}_{\xi}^{\text{vig}}(\mathbf{Q}) = \mathbf{F}_{\xi}^{\text{inv}} - [1-\omega(\mathbf{m},\hat{\xi})]\,p(0,\hat{\xi},0)^{\text{T}} \qquad \text{Eq.(3-9)}$$

where $\hat{\xi}$ is a unit vector in the ξ direction and

$$\omega(\mathbf{m},\hat{\boldsymbol{\xi}}) = \text{Min}\left\{1, \frac{0.9\gamma(\mathbf{m}^{\text{T}}\hat{\xi})^2}{1+(\gamma-1)(\mathbf{m}^{\text{T}}\hat{\xi})^2}\right\}, \mathbf{m} = \mathbf{u}/\text{a} \qquad \text{Eq.(3-10)}$$

3.1.2.1 Numerical Method

In order to obtain an efficient solution of the flow governing Eq.(3-9) without stability restriction on the maximum size of the space step a pseudo time term is introduced,

$$\frac{\partial \mathbf{Q}}{\partial \tau} V\Big|_{i,j,k} + [\mathbf{F}_{\xi}^{\text{inv}} s_{\xi}]_{i-½,j,k}^{i+½,j,k} + [\mathbf{F}_{\eta} s_{\eta}]_{i,j-½,k}^{i,j+½,k} + [\mathbf{F}_{\zeta} s_{\zeta}]_{i,j,k-½}^{i,j,k+½} = \mathbf{0} \qquad \text{Eq.(3-11)}$$

As $\tau \to \infty$ the pseudo-time term vanishes and the governing equations are recovered.

3.1.2.2 Spatial discretization

The streamwise convective flux is evaluated using a simple first order upwind scheme for single sweep calculations. For multi-sweep calculations in which flow disturbances are permitted to propagate in both upstream and downstream directions the flux vector splitting proposed by Steger and Warming [73] is employed. A high-resolution finite volume scheme based on Osher and Solomon's [59] upwind flux difference splitting is employed for the spatial discretisation of the convective flux terms in the cross flow direction. To enhance the spatial resolution in the cross-flow plane, a nominally third-order accurate slope limited MUSCL interpolation of the primitive variables is employed [74].

The viscous fluxes are evaluated using a second-order finite volume scheme. Flow gradients required in the evaluation of the stress tensor and heat flux are obtained using Gauss' theorem and auxiliary cells constructed around the face for which the data is required.

3.1.2.3 Pseudo-time discretization

Following spatial discretization a semi-discrete system of ordinary differential equations is obtained,

$$\frac{\partial \mathbf{Q}}{\partial \tau} V \Big|_{i,j,k} + R_{i,j,k} = \mathbf{0} \qquad \text{Eq.(3-12)}$$

Eq.(3-12) is solved using an efficient implicit method that employs a BILU pre-conditioned Krylov sub-space method [60]. The multi-grid Full Approximation Scheme (FAS) is used to further accelerate convergence to the steady state [61].

3.2 Approximation concepts and surrogates

The computational cost associated with the use of high-fidelity simulation models is extremely high, most especially when multi-disciplinary studies are conducted. Although advances in computing hardware and software are constantly being made and have reduced the costs significantly over the last few decades, the fidelity with which engineers desire to model engineering systems has also increased considerably. Consequently, general approximation techniques were developed and are constantly being developed for improving the efficiency of processes and it has now become popular to employ computationally cheap approximation models instead of exact calculations in design and analysis.

In this section a few computational approaches are described for building approximation models of high-fidelity analysis codes. Here, the high-fidelity analysis model is represented by the relationship $y = f(\mathbf{x})$ where $\mathbf{x} \in \Re^r$ is the vector of inputs to the simulation code and y is a scalar output. The objective of the approximation is to construct a model $\hat{y} = \hat{f}(\mathbf{x},\boldsymbol{\alpha})$ that is computationally cheaper to evaluate than the high-fidelity code. In this model $\boldsymbol{\alpha}$ is a vector of undetermined parameters, which is estimated either by employing what is called a black-box surrogate approach or a physics-based surrogate approach.

In the black-box approach the aim is to construct a surrogate without using any domain-specific knowledge of the analysis code, that is, the code is considered to be a computational module that cannot be modified. The black-box approach involves running the analysis code at a number of preselected inputs to generate a set of input-output data. In this text this data is interchangeably called either the training dataset or observational data. From this data, a surrogate model is trained to learn the input – output mapping by minimizing some appropriate loss function. This approach to surrogate construction has a number of practical advantages, although it has some disadvantages too.

On the other hand, physics-based approaches exploit to some extent the governing equations solved by the analysis code either in the continuous or discrete form and therefore make use of domain-specific knowledge. It is to be remarked that in this text, the words surrogates and meta-models are used interchangeably and refer to approximation models.

In this work, an approximation model which couples both the black-box and the physics-based approaches is considered.

3.2.1 Interpolation versus regression techniques

Generally surrogate models are developed such that it is allowed to either interpolate or else regress the data. The following is a quantitative description of what is meant by regression and interpolation of data.

Given a set of data points $\mathbf{x}^{(i)} \in \Re^r$ where $i = 1,.....,m$ and the corresponding function values $f(\mathbf{x}^{(i)})$, a global approximation function $\hat{f}(\mathbf{x}^{(i)})$ can be obtained by

(i) Linear regression

Considering a second order polynomial as a model such as,

$$\hat{f}(\mathbf{x}^{(i)}) = c_0 + \sum_{i=1}^{r} c_i x_i + \sum_{i=1}^{r} c_{ii} x_i^2 + \sum_{i=1}^{r} \sum_{\substack{j=1 \\ j \neq i}}^{r} c_{ij} x_i x_j \qquad \text{Eq. (3-13)}$$

where the *c*'s are unknown coefficients determined by the least-squares approach. This response surface equation (3-13) can then be used to determine the approximate function values at any arbitrary parameter values. This technique is generally inadequate in providing a globally accurate representation because of its smoothing effect.

(ii) Interpolation

An interpolating model satisfies the following condition

$$\hat{f}(\mathbf{x}^{(i)}) = f(\mathbf{x}^{(i)}), \quad i = 1,.....,m \qquad \text{Eq. (3-14)}$$

which shows that the function and the approximation are equal at all data points.

In physical experiments, the surrogate models are normally not allowed to interpolate the data due to the presence of random errors and so regression techniques are used. In computational experiments, for example, when the analysis CFD solver is run to incomplete convergence, it is sensible not to use an interpolating surrogate model. Moreover, numerical errors arising in similar problems are not directly related with random noise and so the use of regression techniques in similar situations is far from ideal. This is because it is statistically debatable to treat numerical errors as random quantities that are independent and identically distributed. It therefore appears that there is a need for a more accurate framework for surrogate modelling in the presence of numerical errors. Consequently, it is important that computational experiments for subsequent use in surrogate modelling are appropriately converged. In addition, one should be cautious in interpreting the results of model significance based on least-square residuals and t-statistics etc., since these do not have any statistical meaning for deterministic experiments. It therefore appears that interpolation techniques are a more appropriate form for surrogate modelling when training datasets are obtained by running a deterministic computer model. This does not mean that for a given problem an interpolation model will necessarily work better than a regression model. In fact what is more important in practice is the generalization performance of the surrogate model.

3.2.2 Generalized linear models

Two very common methods for constructing surrogate models are polynomial response surface models and radial basis functions. Both are examples of generalized linear models and can be expressed in the following form

$$\hat{y}(x) = \sum_{i=1}^{m} \alpha_i \Phi_i(x) \qquad \text{Eq. (3-15)}$$

where Φ_i i = 1, m denote a set of fixed basis functions and α_i are unknown coefficients in the approximation. For information, generalized non-linear models use basis functions such as $\Phi = \Phi(x, \theta)$ where θ is a set of hyper-parameters.

3.2.2.1 Response Surface Methods: Linear regression

The response surface methodology (RSM) consists basically of (i) the application of design of experiments techniques (ii) regression analysis and (iii) analysis of variance techniques for data interpretation from physical experiments [75] [76]. The origin of RSM dates back to the early twentieth century when scientists in the field of agriculture required and developed tools for modelling data from physical experiments [77]. Since these are subject to measurement errors, least-square regression techniques were employed under the assumption that the measurement errors for each data point are independent and identically distributed. In particular RSM is useful in situations where there are a few observations because the physical experiment is both very expensive and time consuming to perform. The accuracy of the response surface can be quantified by examining the statistics of the least-squares residuals.

3.2.2.1.1 RSM and Physical Experiments

In RSM, the relationship between the observations and the independent variables is defined as

$$y = f(\mathbf{x}^{(i)}) \qquad \text{Eq. (3-16)}$$

where y is the observed response, $\mathbf{x}^{(i)}$ is the vector of n_v independent variables defined as

$$\mathbf{x} = [x_1, x_2, ..., x_{n_v}], \qquad \text{Eq. (3-17)}$$

and $f(\mathbf{x}^{(i)})$ is the unknown function. The empirical model of the unknown function found via the application of RSM is defined as

$$\hat{y} = \hat{f}(\mathbf{x}^{(i)}) \qquad \text{Eq. (3-18)}$$

where $\hat{f}(\mathbf{x}^{(i)})$ typically is a first or second order polynomial in **x**. Note that the random error (uncertainty) present in stochastic experimental data is implicit in both $f(\mathbf{x}^{(i)})$ and $\hat{f}(\mathbf{x}^{(i)})$.

RSM employs the statistical techniques of regression analysis and analysis of variance ANOVA to determine $\hat{f}(\mathbf{x}^{(i)})$ through a systematic decomposition of the variability in the observed response values. The empirical model is then estimated by assigning portions of the variability to either the effect of an independent variable or to a random error.

3.2.2.1.2 RSM and Computational Experiments

Computer simulations are nowadays very common in science and engineering. In many instances these simulations require enormous amounts of computer time and memory when performed. Moreover, a single simulation can necessitate weeks or months of effort to set-up and debug. As a consequence, it is attractive to consider applying RSM methods, originally developed for physical experiments, to computer simulations as well. It should be noted that in this research RSM is applied to the results obtained from deterministic computer simulations, where simulations performed with identical starting values and initial conditions produce identical results.

3.2.2.1.3 Polynomial Models for RSM

Normally in RSM applications, either linear or quadratic polynomials are used to model the observed response values. RSM becomes very expensive when higher-order polynomials are considered for experiments involving several variables. Giunta [78] remarks that quadratic polynomials provide the best compromise between modelling accuracy and computational cost of all models investigated in his work. In addition cubic and higher-order polynomials may contain one or more inflection points and so in gradient-based numerical optimization schemes, the optimizer may converge to an inflection point rather than to an optimum.

If n_s analyses are conducted and $p = 1, ..., n_s$, then a quadratic response surface model has the form

$$y^{(p)} = c_o + \sum_{1 \le j \le n_v} c_j x_j^{(p)} + \sum_{1 \le j \le k \le n_v} c_{(n_v - 1 + j + k)} x_j^{(p)} x_k^{(p)}, \qquad \text{Eq. (3-19)}$$

where $y^{(p)}$ is the response; $x_j^{(p)}$ and $x_k^{(p)}$ are the n_v design variables or parameters; and $c_o, c_j, c_{(n_v-1+j+k)}$ are the unknown polynomial coefficients. There are $n_t = (n_v+1)(n_v+2)/2$ coefficients in the quadratic polynomial. This polynomial model may be written in matrix notation as

$$y^{(p)} = \mathbf{c}^T \bar{\mathbf{x}}^{(p)}, \qquad \text{Eq. (3-20)}$$

where **c** is the vector of length n_t of unknown coefficients to be estimated,

$$\mathbf{c} = [c_0, c_1, ..., c_{n_t-1}], \qquad \text{Eq. (3-21)}$$

and $\bar{\mathbf{x}}^{(p)}$ is the vector of length n_t corresponding to the form of the $x_j^{(p)}$ and $x_k^{(p)}$ terms in the polynomial model Eq. (3-19). For the p th observation this is

$$\bar{\mathbf{x}}^{(p)} = \left[1, x_1^{(p)}, x_2^{(p)}, ..., x_1^{(p)} x_2^{(p)}, ..., (x_{n_v}^{(p)})^2\right] \qquad \text{Eq. (3-22)}$$

Note that there is a difference between the p^{th} vector of independent variables, $\mathbf{x}^{(p)}$ and the p^{th} vector of independent variables mapped into the form of the polynomial model, $\bar{\mathbf{x}}^{(p)}$.

Estimating the unknown coefficients requires n_s analyses, where $n_s \geq n_t$. Under such conditions, the estimation problem may be formulated in matrix notation as

$$\mathbf{y} \approx \mathbf{X}\mathbf{c} \qquad \text{Eq. (3-23)}$$

where **y** is the vector of n_s observed response values,

$$y = [y^{(1)}, y^{(2)}, ..., y^{(n_s)}], \qquad \text{Eq. (3-24)}$$

and **X** is the matrix formed by the p row vectors $\bar{\mathbf{x}}^{(p)}$ which is assumed to have rank n_t. Thus, **X** may be expressed as

$$\mathbf{X} = \begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)} & \ldots & (x_{n_v}^{(1)})^2 \\ . & . & . & \ldots & . \\ . & . & . & \ldots & . \\ . & . & . & \ldots & . \\ 1 & x_1^{(n_s)} & x_2^{(n_s)} & \ldots & (x_{n_v}^{(n_s)})^2 \end{bmatrix}$$

The unique least squares solution to Eq. (3-23) is

$$\hat{\mathbf{c}} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y} \qquad \text{Eq. (3-25)}$$

where $\left(\mathbf{X}^T\mathbf{X}\right)^{-1}$ exists if the rows of **X** are linearly independent. When $\hat{\mathbf{c}}$ is substituted for **c** into Eq. (3-20), values of the response may be predicted at any location **x** by mapping **x** into $\bar{\mathbf{x}}^{(p)}$. In matrix notation this corresponds to

$$\hat{\mathbf{y}} = \hat{\mathbf{c}}^T\bar{\mathbf{x}}^{(p)} \qquad \text{Eq. (3-26)}$$

If $n_s > n_t$, the system of equations is over-determined. Thus, the predicted response values (from the polynomial model) at the original sample locations may differ from the observed response values at the sampled locations.

In practice, the matrix **X** may be poorly conditioned, which may cause numerical instabilities to arise when the normal equations are directly solved to compute the RSM coefficients. Hence, it is preferable to employ the singular value or QR decomposition scheme to estimate the coefficients. Further to ensure robust estimates for the coefficients, the number of training points are chosen to be greater than the number of unknown coefficients in the model.

3.2.2.1.4 ANOVA and regression analysis

Apart from estimating the coefficients in the quadratic polynomial model, ANOVA and regression analysis also yield a measure of uncertainty in the coefficients. This uncertainty estimation is achieved by the t-statistic. This is defined in Myers and Montgomery [76] as

$$t = \frac{c_{j-1}}{\sqrt{\hat{\sigma}^2 (\mathbf{X}^T \mathbf{X})^{-1}_{jj}}} \qquad \text{Eq. (3-27)}$$

where $\hat{\sigma}^2$ is the estimate of the variance in the observed response data and $j = 1,..., n_t$. Note that the reciprocal of the t-statistic is an estimate of the standard derivation of each coefficient as a fraction of its value. Accordingly, coefficients with low values for the t-statistic are not accurately estimated. It is normally the choice of the user to select the minimum allowable t-statistic. The choice typically depends on the number of observed response values used to create the response surface model.

A number of applications of RSM in design optimization can be found in the literature, see for example the dissertations by Balabanov [79], Giunta [78], Venter [80], Vavalle [81] and the references therein.

This technique is generally inadequate in providing a globally accurate representation because of its smoothing effect and so is of limited utility when modelling complex input-output relationship. Hence polynomial models are useful only when it is required to approximate an input-output relationship over a small region of the parametric space. However, the drawbacks associated with parametric models can be alleviated to some extent via non-parametric models that employ more flexible basis functions. The next section deals with this type of function.

3.2.2.2 Radial Basis Functions Approximations

A radial basis function (RBF) is a real-valued function whose value depends on the Euclidean distance from some point called a centre [82] [83]. Radial basis functions are typically used to build up function approximation of the form

$$\hat{f}(x) = p(x) + \sum_{i=1}^{N} w_i \Psi\left(\|x - x_i\|\right) \qquad \text{Eq. (3-28)}$$

where the approximating function $\hat{f}(x)$ is represented as a sum of N radial basis functions Ψ, each associated with a different centre x_i and weighted by an appropriate coefficient w_i. $p(x)$ is a polynomial of one degree less than the RBF and is included to ensure a unique solution for the weight vector. Thus, an RBF is a weighted sum of translations of a radially symmetric basis function augmented by a polynomial term. The basis function in this context is a real function of a positive real value r which is the distance or radius from the origin.

Typical radial basis functions are

i. Gaussian $\Psi(r) = e^{-\frac{r^2}{\theta}}$

ii. multi-quadric $\Psi(r) = \sqrt{1 + \frac{r^2}{\theta}}$

iii. inverse multi-quadric $\Psi(r) = \left(1 + \frac{r^2}{\theta}\right)^{-1/2}$

iv. poly-harmonics such as the tri-harmonic $\Psi(r) = r^3$ and

v. thin-plate splines $\Psi(r) = r^k \ln r \quad k \in [2,4,.....]$

The constant θ in (i), (ii) and (iii) is called the shape parameter. [84] [85]

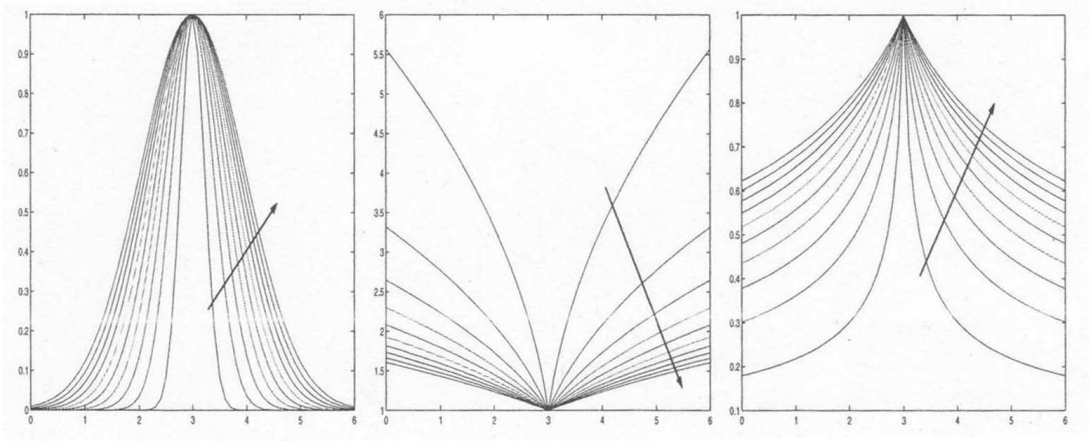


Figure (3-1) Shapes of Gaussian, multi-quadric and inverse multi-quadric RBFs as these vary with the shape parameter. The arrow shows the direction in which θ increases. Pictures taken from reference [93]

RBFs provide an effective way for interpolating scattered data as the associated system of linear equations is guaranteed to be invertible under very mild conditions on the locations of the data points. For example, the thin-plate spline only requires that the points are not co-linear while the Gaussian and the multi-quadric place no restrictions on the locations of the points. In particular RBFs do not require that the data lie on any sort of regular grid. Many researchers remark that RBFs are suitable for problems with up to a few thousand training data points [86].

The main attraction of RBF-type networks is that the model parameters may be determined by using linear methods. Another major attraction of such expansion functions is their flexibility, in fact these may be chosen to be either local such as the Gaussian and the inverse multi-quadric or global such as the multi-quadric and poly-harmonics. They may or may not incorporate shape parameters which may be tuned to reflect the nature of the data.

The type of RBF to use for a given set of data is one of the active areas of research. A basic question is whether to choose a local or a global function and how to set the parameters. The parameter values may be determined through solving a non-linear optimization problem or simply fixed uniformly. In particular, the Gaussian RBF which is a local basis function has a specific domain of influence which is located at the RBF

centre and with a size proportional to the radius r. Outside of this domain which is also called as the receptive field, the Gaussian RBF does not contribute significantly to the expansion. The locality of the Gaussian RBF leads to the useful feature that each weight is associated with a region of the input space and is to a large extent unaffected by data outside its receptive field.

In contrast with the local Gaussian RBF, it is possible to employ the global cubic RBF. This RBF is attractive in that the global functions tend to require less data for training. In addition the lack of a radius parameter simplifies the fitting process.

The RBF interpolant $\hat{f}(x)$ is defined by the coefficients of the polynomial $p(x)$ and the weights w_i . Since this produces an under-determined system, the orthogonality condition

$$\sum_{j=1}^{N} w_j p(x_j) = 0 \qquad \text{Eq.(3-29)}$$

is further imposed on the coefficients $\boldsymbol{w} = (w_1, ..., w_N)$.

Let $\Lambda = (p_1, ..., p_l)$ be a basis for the polynomial and let $\boldsymbol{c} = (c_1, ..., c_l)$ be the coefficients that give $p(x)$ in terms of this basis. Then Eq. (3-28) and Eq. (3-29) may be written in matrix form as

$$\begin{pmatrix} \boldsymbol{\Psi} & \Lambda \\ \Lambda^{\mathrm{T}} & 0 \end{pmatrix} \begin{pmatrix} w \\ c \end{pmatrix} = \begin{pmatrix} f \\ 0 \end{pmatrix} \qquad \text{Eq.(3-30)}$$

where $\boldsymbol{\Psi} = \begin{pmatrix} \psi(0) & \psi(\|x_2 - x_1\|) & .. & \psi(\|x_N - x_1\|) \\ \psi(\|x_1 - x_2\|) & \psi(0) & .. & \psi(\|x_N - x_2\|) \\ : & : & : & : \\ \psi(\|x_1 - x_N\|) & \psi(\|x_2 - x_N\|) & .. & \psi(0) \end{pmatrix}$

and $\Lambda_{i,j} = p_j(x_i), \quad i = 1,.....N,\ j = 1,.....l.$

Solving Eq. (3-30) determines $\boldsymbol{c}$ and $\boldsymbol{w}$, hence $\hat{f}(x)$. There are two different approaches for solving Eq. (3-30), each with its own advantages. Direct methods such as the SVD and the QR factorizations have a particular appeal. The SVD is most especially useful if the left-hand side matrix of Eq. (3-30) is ill- conditioned, that is, if the ratio of its largest to its smallest singular value is large, say 10^6 - 10^{10}. On the other hand, the QR factorization is attractive if training data or centres will be added or subtracted since it has inexpensive update procedures. Iterative methods such as the steepest-descent algorithm and the conjugate-gradient algorithm for solving Eq. (3-30) determine a sequence of improving estimates for the solution.

Analysis of the approximation characteristics of a broad class of radial basis functions can be found in the literature, see for example Liao *et al.* [87] and the references listed therein. The results in that work suggest that by using suitable radial basis functions in a generalized linear model, it is possible to approximate any function to an arbitrary degree of accuracy.

Franke [88] compared a variety of RBF interpolation techniques of two-dimensional problems for a scattered set of data. The comparison of the algorithms was made according to various criteria such as accuracy, visual aspect, sensitivity to parameters, execution time, storage requirements and ease of implementation. It was shown that multi-quadrics give best results compared with the other techniques. This observation has then led to the widespread application of multi-quadrics for modelling complex two- and three dimensional surfaces in various scientific areas [89]. Essentially, multi-quadrics are continuously differentiable, integrable and are capable of representing with very high accuracy functions with steep gradients. Monotonicity and convexity are some observed properties as a result of such high accuracy. Kansa [90] proposed a modified multi-quadric scheme which is an excellent method for both accurate interpolation and partial derivative estimate.

Michelli [91] showed that the Gram matrix is guaranteed to be non-singular only for a class of RBFs, when the set of input vectors in the training dataset are distinct. Specifically, non-singularity can be guaranteed only if positive-definite kernels are used such as the Gaussian and the inverse multi-quadric. If a positive-definite RBF is used, then the Gram matrix is symmetric positive definite. Such basis functions are also commonly referred to as Mercer kernels in learning theory. Nevertheless, for high values of θ, that is as the basis functions become flatter, the $\mathbf{\Psi}$ matrix can become ill-conditioned. A monotonic relationship between the condition number of $\mathbf{\Psi}$ and the shape parameter for multi-quadrics was proved by Wang [92]. In practice, it is often observed that high accuracy is achieved only at the verge of numerical stability.

In this work, a relatively small number of training points were considered and so the computational time to evaluate the weights of the RBF was small. However, when dealing with a few thousand training points, the computational cost and memory requirements become considerable. The construction of RBF approximations requires $O(n^2)$ memory and $O(n^3)$ operations. Such costs are increased when there is also the need to establish suitable values of θ for a given dataset. Consequently some other approaches for constructing RBF approximations such as the greedy approach [86] [93], which significantly reduces the computational cost and memory requirements for large datasets have been proposed.

Krishnamurthy [94] investigated and compared the ability of RBFs, Kriging and least squares methods for function and function's derivative estimation by considering some numerical examples. Krishnamurthy concluded that Kriging yield the best results followed closely by RBFs, but the free parameter in Kriging was optimised while the free parameter in Gaussians and multi-quadric RBFs was maintained constant at one. The optimisation process of the free parameter in Kriging was expensive.

To the best of the author's knowledge RBFs were not extensively used in aerodynamics related work, but one particular example will be discussed in detail in chapter seven where variable-fidelity modelling is considered.

3.2.2.2.1 Pre-conditioning techniques

The radial basis function interpolation matrix $\mathbf{\Psi}$ suffers many times from ill-conditioning when the discrete set of data points is a large set and when the radial basis function is for example the multi-quadric function, its reciprocal or the thin-plate spline. In the case of the multi-quadric function, the conditioning is bad for any value of the parameter, but the situation gets worse for larger parameters. In this section, a possibility of how the matrix's condition number may be improved before beginning the computation of the interpolation coefficients is mentioned. This is the standard approach when large linear systems of whatever origin become numerically intractable due to large condition numbers and ensuing serious numerical problems with rounding errors. However, there is significant influence of the form of ψ on the preconditioning technique we choose.

Assuming that $\mathbf{\Psi}$ is positive definite, as is the case when ψ is the reciprocal multi-quadric function. In principle this can be solved using a standard conjugate gradient method or even a direct method such as a Cholesky factorisation[4]. The convergence speed of conjugate gradients, however, depends on the condition number of the matrix of the linear system which is being solved. Convergence under these conditions may therefore be slow or may fail completely due to rounding errors.

In order to improve the condition number of the positive definite matrix $\mathbf{\Psi}$, the following can be solved

$$\mathbf{P}\,\mathbf{\Psi}\,\mathbf{P}\,\boldsymbol{w} = \mathbf{P}\mathbf{f} \qquad \text{Eq. (3-31)}$$

where $\mathbf{P}$ is a preconditioning matrix, non-singular and usually symmetric- if it is non symmetric, the left–multiplication in Eq. (3-31) would have to be by $\mathbf{P}^{\mathbf{T}}$. $\mathbf{P}$ is chosen such that it is not too expensive to compute the matrix product on the left-hand side and such that the product is positive definite ie. a banded matrix. If $\mathbf{P}$ is symmetric and $\mathbf{P}^2 = \mathbf{C}$, then

[4] A symmetric positive-definite matrix can be decomposed into a lower triangular matrix and the transpose of the lower triangular matrix.

the matrix product $\mathbf{C}\boldsymbol{\Psi}$ should ideally have a spectrum on the positive real half-axis consisting of a small number of clusters, one of them near one, because the conjugate gradient algorithm can deal particularly well with such situations and because $\mathbf{P}^2\boldsymbol{\Psi} \approx \mathbf{I}$ is equivalent to $\mathbf{P}\boldsymbol{\Psi}\mathbf{P} = \mathbf{I}$. Of course, the choice of $\mathbf{C} = \boldsymbol{\Psi}^{-1}$ would be optimal, $\mathbf{C}\boldsymbol{\Psi}$ being the identity matrix. If $\mathbf{P}$ is chosen suitably, then the condition number of the preconditioned matrix is usually also small. There are always several choices of $\mathbf{P}$ possible and having made one choice, the desired coefficient vector is $\boldsymbol{\lambda} = \mathbf{P}\boldsymbol{w}$ which is evaluated at the end.

This subject is vast but the interested reader may consult reference [83] or [95] for further details.

3.2.2.2.2 Parameter optimization in multi-quadric response surface approximation

In general the performance of an RBF to approximate a response surface depends to a great extent on the shape parameter. Hence, some form of shape parameter tuning is important and consequently techniques for optimizing the shape parameter are constantly sought. An efficient method of computing the optimal shift parameter based on the leave-one-out cross validation technique is proposed by Wang [92] and was implemented in this work. A detailed exposition of this method is presented in Appendix A.

3.2.2.2.3 Hermite interpolation using RBFs

If the sensitivity of some response function can be computed efficiently, it would make sense to construct surrogate models with sensitivity data as these can be more accurate than those built using function values only. In generalized linear models, the output sensitivities can be incorporated into the surrogate model using the idea of Hermite interpolation [96]. A description of this interpolation method is presented in Appendix B.

3.2.3 Physics-based approximations

The approximation methods described previously are general in scope and being so can be applied to various simulation codes since no assumptions are made to the governing equations being solved. In this section a consideration is made to physics-based approximation concepts, which in general require a deeper understanding of the governing equations and the numerical methods employed for their solution. These physics-based approximation methods are another attempt to circumvent the curse of dimensionality associated with certain surrogate modelling by exploiting domain-specific knowledge. One example of physics-based approximation methods are reduced basis methods of linear and non-linear systems. Keane and Nair [93] claim that a notable advantage of most physics-based approaches is that they do not suffer from the curse of dimensionality and can hence be readily applied to problems with greater numbers of design variables.

3.2.3.1 Reduced basis methods

By using reduced basis methods it is possible to build models with fewer unknowns than the original high-fidelity model. To illustrate reduced basis methods consider the discrete mathematical model of a physical system written in the form

$$\mathbf{R}(\mathbf{w}, \mathbf{x}) = 0 \qquad \text{Eq. (3-32)}$$

where $\mathbf{w} \in R^n$ denotes the discretized vector of field variables and $\mathbf{x} \in R^p$ denotes the vector of design variables. Note that **w** is an implicit function of the design variables. The above form of discrete equations is normally obtained by spatial and temporal discretization of PDEs governing the system response and so it is a representative of a wide class of problems encountered in structural analysis, fluid mechanics, heat transfer, etc.

The main idea used in reduced basis methods is to approximate the field variable vector $\mathbf{w}$ as

$$\hat{\mathbf{w}} = c_1\varphi_1 + c_2\varphi_2 + ... + c_m\varphi_m = \mathbf{\Phi}\,\mathbf{c} \qquad \text{Eq. (3-33)}$$

where $\mathbf{\Phi} = \{\varphi_1, \varphi_2, ..., \varphi_m\} \in R^{nxm}$ denotes a matrix of known basis vectors and $\mathbf{c} = \{c_1, c_2, ..., c_m\}^T \in R^m$ is a vector of undetermined coefficients.

An assumption is made in reduced basis methods that the response vector lies somewhere in the subspace spanned by a set of basis vectors. This allows the original problem with *n* unknowns to be recast into a problem with *m* unknowns and so if $m << n$ then it is possible to approximate **w** very efficiently**.** Generally, the undetermined coefficients are estimated so that the governing equations are satisfied in some way for example by minimizing the integral of an appropriate error norm evaluated over the parametric space thus ensuring that the reduced basis approximation is global in nature. Moreover, the model parameters are evaluated using the governing equations, hence explains the reason why reduced basis methods belong to the class of physics-based approximation methods.

Thus, the main elements of reduced basis methods are the choice of the basis vectors and the numerical scheme employed to compute the vector of undetermined coefficients. Both the accuracy and the convergence rate depend on this choice. The following sections outline the possible choices associated with these elements.

3.2.3.1.1 Choice of basis vectors

A good set of basis vectors is a set of vectors which is easy to compute as well as guaranteed to be linearly independent. For the general form of discrete governing

equations specified in Eq. (3-32) there are three commonly used options for the basis vectors. Considering the case when there is only one parameter, say x.

The Lagrange subspace is formed by using state solutions of the discrete governing equations which correspond to various different values of the parameters say $x^{(1)}$, $x^{(2)}$,...., $x^{(m)}$, that is,

$$\mathbf{\Phi} = \text{span}\left\{\mathbf{w}\left(x^{(1)}\right), \mathbf{w}\left(x^{(2)}\right), ..., \mathbf{w}\left(x^{(m)}\right)\right\} \in R^{nxm} \qquad \text{Eq. (3-34)}$$

These solutions are obtained by standard techniques such as finite element or finite volume methods.

The Hermite subspace is formed by combining the Lagrange subspace together with the first-order derivatives of the field variable vector **w** at $x^{(1)}$, $x^{(2)}$,....,$x^{(m)}$, that is,

$$\mathbf{\Phi} = \text{span}\left\{\mathbf{w}\left(x^{(1)}\right), \frac{\partial \mathbf{w}\left(x^{(1)}\right)}{\partial x}, ..., \mathbf{w}\left(x^{(m)}\right), \frac{\partial \mathbf{w}\left(x^{(m)}\right)}{\partial x}\right\} \in R^{nx2m} \qquad \text{Eq. (3-35)}$$

The state and sensitivity approximations are obtained through standard techniques such as finite element or finite volume methods.

The Taylor subspace is written in terms of the Hermite subspace together with the higher derivatives of the field variable vector **w** (sensitivities and higher-order sensitivities) at $x^{(1)}$,

$$\mathbf{\Phi} = \text{span}\left\{\mathbf{w}\left(x^{(1)}\right), \frac{\partial \mathbf{w}\left(x^{(1)}\right)}{\partial x}, \frac{\partial^2 \mathbf{w}\left(x^{(1)}\right)}{\partial x^2}, ..., \frac{\partial^{m-1} \mathbf{w}\left(x^{(1)}\right)}{\partial x^{m-1}}, ...\right\} \in R^{nxm} \qquad \text{Eq. (3-36)}$$

The state and derivative approximations are obtained through expensive techniques such as finite element or finite volume methods. The Taylor basis may be somewhat complicated to program due to the complexity of the partial differential equations that determine the higher-order sensitivities. In addition, the number of higher-order derivatives grows very rapidly with the number of design parameters, e.g., if one has 10

design parameters, there are 55 different second derivative sensitivities. Thus, the dimension of the Taylor reduced basis grows quickly with the number of parameters and the number of derivatives used.

In general, the Lagrange subspace is the most attractive from a practical viewpoint since the basis vectors spanning this subspace can be easily computed. A small point of concern with all three subspaces is the possibility that some of the basis vectors could be linearly dependent but this can be overcome for example by employing the SVD scheme. Reduced basis methods using the subspaces aforementioned were applied to solve a wide range of problems including fluid flow and structural mechanics for example, Balmes [97], Ito and Ravindran [98] and Prud'homme *et al.* [99].

3.2.3.1.2 Schemes for computing the coefficients

Once a set of linearly independent basis vectors is developed, a scheme for computing the vector of unknown coefficients is required. The main objective of this scheme is that an estimation of the coefficients is made such that the discrete governing equations are satisfied in some sense, in contrast with the general approximation concepts mentioned earlier where the coefficients in the surrogate model are computed without regard to the governing equations. Two schemes in particular are the Bubnov-Galerkin scheme and the Petrov-Galerkin scheme.

In the Bubnov-Galerkin scheme, $\mathbf{c}$ is computed such that the residual error vector obtained by substituting $\hat{\mathbf{w}} = \mathbf{\Phi c}$ into the governing equations is orthogonal to the approximating subspace. For the discrete governing equations considered earlier, this implies that the following condition is imposed $\boldsymbol{\varphi}_i^T \, \mathbf{R}(\mathbf{\Phi c}, \mathbf{x}) = 0$, $i = 1,2,.....m.$

In the Petrov-Galerkin scheme $\mathbf{c}$ is computed by directly minimizing the L_2 norm of the residual error, that is, $\left\| \mathbf{R}\left(\mathbf{\Phi c}, \mathbf{x}\right) \right\|_2$.

It must be pointed out that when the governing equations are non-linear, then the residual error vector is a non-linear function of the vector **c** and therefore ideally non-linear least-squares minimization techniques must be employed to estimate the unknown coefficients. However, if the governing equations are linear, then linear regression techniques can be used to solve for the vector **c**.

3.2.3.2 Reduced basis methods for non-linear problems

It is possible to apply linear reduced basis methods to non-linear problems provided a suitably rich set of basis vectors is available. This is apparent from section 3.2.3.1 where a general discrete governing equation is considered to introduce the idea of reduced basis approximations. Moreover, most numerical algorithms for analyzing non-linear systems are based on the solution of a sequence of linear sub-problems and so makes it possible to employ a reduced basis method developed for linear systems to solve a series of linear sub-problems encountered in non-linear problems. In this thesis, an approach that uses the proper orthogonal decomposition method to derive the basis vectors is considered. This methodology has already been introduced in chapter 1 and a further detailed mathematical description is presented in the next chapter.

3.2.3.2.1 The reduced-order modelling approach

In the reduced-order modelling approach an approximation to the physical system from some limited results of the high-fidelity simulation model is sought. This approximate model is called a reduced-order model (ROM). A reduced-order model is an inexpensive mathematical model of a physical system that is derived from experimental or numerical simulation data. The aim of the ROM is to reproduce the behaviour of the original system of PDEs over a range of input parameters as accurately as possible, but at a much smaller cost than the cost for solving the PDEs. The reduced-order model contains many fewer degrees of freedom than the full simulation and is therefore inexpensive to compute.

Reduced-order models can be constructed using basis vectors generated via POD. Assuming that the reduced basis approximation is applied to the discrete form of the governing equations and substituting the approximation into the discrete governing equations $\mathbf{R}(\mathbf{w},\mathbf{x})=0$, the residual can be expressed in terms of the undetermined coefficients in the reduced basis approximation as $\mathbf{R}(\mathbf{\Phi c},\mathbf{x})$. Subsequently, the vector of coefficients can be computed by minimizing $\|\mathbf{R}(\mathbf{\Phi c},\mathbf{x})\|$ using a non-linear least squares minimizing technique.

3.2.3.2.1.1 Error Estimation

The usefulness of a reduced-order model is measured by its ability to reproduce the behaviour of a high-fidelity simulation model at a much lower cost. Therefore one possible definition of the error of a reduced-order model is the difference between the solution of the reduced-order model and the solution that would be obtained by running a high-fidelity simulation for the same input parameters, initial conditions and boundary conditions. This error estimation was adopted in this work.

3.3 Design-of-Experiment Theory

In the description of the surrogate modelling techniques covered so far, it has been assumed that a set of observational data was generated by some technique and is already available. In the development of any surrogate model, the location of sample points within the parametric space has an important influence on both the cost of constructing the model and on the accuracy of model predictions. In this section, the problem of generating training data that leads to approximation models that generalize well is examined.

Design-of-Experiment theory (DoE) [100] is a branch of statistics which provides the statistician with methods for selecting the independent variable values for a limited number of experiments, while achieving a good representation over the parametric space of interest. The various experimental design methods create certain combinations of analyses in which the independent variables are prescribed at specific values or levels. The results of these planned experiments are then used to study and investigate the response and sensitivity of some dependent quantity to the independent variables.

Classical DoE techniques include full- and fractional-factorial designs. A common feature of these techniques is that the sample points are placed at the extremes of the parameter space to alleviate the effects of noise and so these are appropriate if the data is contaminated by noise, which make it necessary to employ regression techniques to filter noise. In contrast to physical experiments, observations made using computer experiments are not subject to random errors and so to extract the maximum information about the input-output relationship, the sample points are chosen to fill the design space in an optimal sense.

In this work we have investigated the performance of two different sampling strategies; the full factorial and the Latin Hyper-Cube design-of-experiments. The following sections deal with these two sampling methods.

3.3.1 The Full-Factorial Experimental Design

Prior to creating an experimental design, the allowable range of each of the n_v variables is defined by lower and upper bounds. The allowable range is then discretized at equally-spaced intervals, thus forming different levels. The region enclosed by the lower and upper bounds on the variables is termed the design or parametric space, the vertices of which determine an n_v - dimensional cube. If each of the variables is specified at only the lower and upper bounds (two levels), the experimental design is called a 2^{n_v} full factorial. Similarly, a 3^{n_v} full factorial design is created by specifying the lower bound, mid-point and upper bound (three levels) for each of the n_v variables.

The full factorial design of experiment uniformly samples the design parameters across the whole parameter space of interest. This technique while easy to implement is expensive requiring $\prod_{i=1}^{n} m_i + 1$ samples where m_i is the number of intervals used to resolve the individual parameters and *n* is the number of parameters. This number can become excessively large for even a modest number of design parameters.

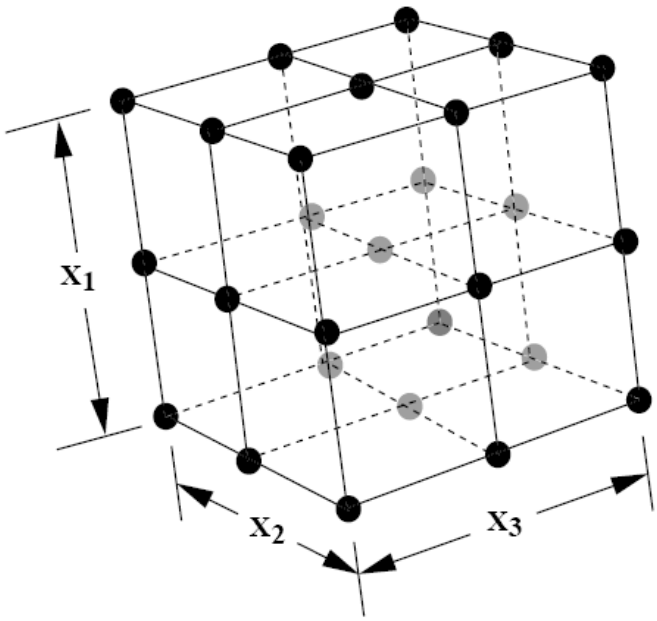


Figure (3-2) A 3^3 full factorial experimental design resulting in 27 points

3.3.2 Monte Carlo Techniques

Monte Carlo techniques are probably the most straightforward of all DoE methods, in which the main purpose is to use a random number generator to sample the parametric space. When the parametric space is regular, Monte Carlo techniques are relatively easy to implement but the implementation turns out to be difficult for irregular parametric spaces. Moreover, the implementation becomes more involved with an assumed joint probability distribution of the input vector. The major drawback of the Monte Carlo technique is that the points generated may not fill the parametric space uniformly and so large areas of the space may be left unexplored. In some other variants of the Monte Carlo

technique such as the stratified sampling methods, the sampling points are generated more uniformly by dividing the parametric space into bins of equal probability and at least one point is placed within a bin. Thus, if each parameter has two intervals, the total number of bins is 2^n. This method suffers from the same disadvantage as full-factorial designs.

3.3.3 The Latin Hyper-Cube Sampling

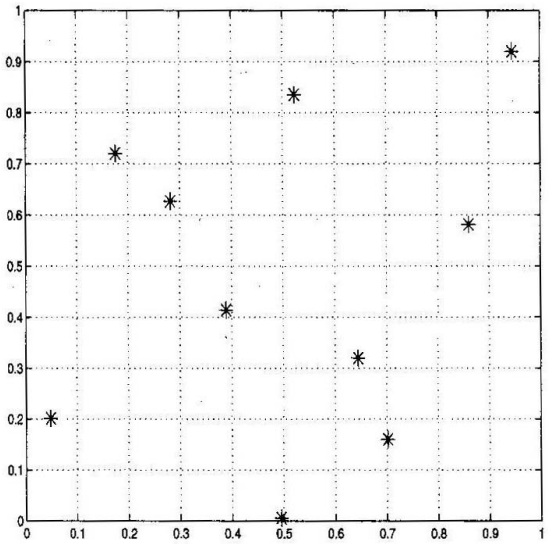
The Latin Hyper-Cube sampling (LHCS) technique was first proposed by McKay et al. [101] as an alternative to Monte Carlo techniques for the design of computer experiments. In LHCS, each parameter range is divided into m intervals or bins of equal probability. This leads to a total of m^n bins in the whole space. Subsequently, m samples are generated such that for each parameter, when a one-dimensional projection is taken, there will be only one sample in each bin. The LHCS algorithm produces samples as follows:

$$x_j^{(i)} = \frac{\pi_j^{(i)} + \kappa_j^{(i)}}{m}, \qquad \forall \quad 1 \le j \le n,\ 1 \le i \le m \qquad \text{Eq. (3-37)}$$

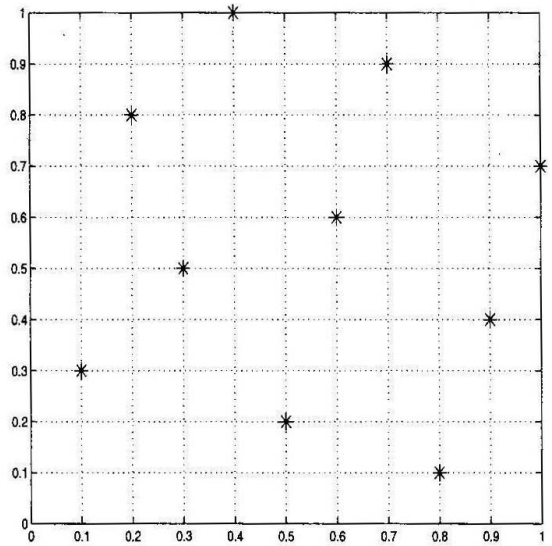
where m is the number of samples, $\kappa \in [\ 0,1\ ]$ is a random number and π is an independent random number permutation. The subscript denotes the parameter number and the superscript in brackets denotes the sample number. From each parameter, one of the points on the interval is selected randomly and the response is evaluated. This is done until all points are used up. This method is useful because there is no correlation between parameters and the samples are chosen randomly. However, the space-filling characteristics produced using the standard LHCS are not guaranteed to be optimal. In an effort to provide an optimal LHCS design, the algorithm is modified using the approach of Audze and Englais [102] . In this approach a m^n grid is first generated and then the sample points are placed so that no two points lie along the same grid line and the metric,

$$\sum_{i=1}^{n}\sum_{j=i+1}^{n}\frac{1}{d_{ij}^{2}} \qquad \text{Eq. (3-38)}$$

is minimized. In Eq. (3-38) d_{ij} is the Euclidean distance between points i and j. In order to ensure that each point generated is placed in the centre of its bin, a lattice sampling technique is adopted in which the value of κ is set equal to 0.5.



(a) Distribution of points generated using the standard LHS method



(b) Distribution of points generated using the optimized LHS method

Figure (3-3) Comparison of the space-filling capability between the standard LHS algorithm and the optimized LHS algorithm.
Pictures taken from reference [91]

Chapter 4

4.0 The proper orthogonal decomposition

In this chapter a review of the mathematical theory of the proper orthogonal decomposition is presented. As it has already been stated earlier, the proper orthogonal decomposition is a mathematical procedure that provides a basis for the modal decomposition of an ensemble of multi-dimensional data. This data may be either experimental or numerical solutions of a system of partial differential equations, which allow a reduction in the order of the system under consideration. A modal decomposition may be required for various reasons such as for the identification of structures in the data, or for the statistical analysis of sample data, or as it is employed in this work, as a set of basis modes for approximating the solution of numerical problems with characteristics similar to the ones used in the construction of the basis modes. The mathematical development of POD for fluid flow applications in particular is described in some detail in [19] and [103]. The essentials of this development and the properties of POD, in particular those most important to reduced-order modelling, are presented in this chapter. Here, the POD is also described for steady-state problems in terms of the singular value decomposition, which is a more straightforward method.

Note that although the POD methodology is almost exclusively applied to non-linear problems, it is important that one recognizes that it is a linear procedure, and the nested sequence of subspaces are linear spaces, even if the data that generates it, is non-linear.

4.1 The POD theory

Consider an ensemble $\{\mathbf{u}^k(\mathbf{x})\}$ of real vector fields on the domain $\mathbf{x} \in \Omega$. Here, it is assumed that the ensemble consists of a set of instantaneous snapshots of a numerical simulation solution field. In seeking good representations of members of $\{\mathbf{u}^k(\mathbf{x})\}$, it is required to project each **u** onto candidate basis functions and so it is assumed that the **u** 's belong to an inner product space: the linear, infinite-dimensional Hilbert space $H(\Omega)$ of square integrable functions with inner product

$$(\mathbf{f},\mathbf{g}) = \int_{\Omega} f(x)g(x)\,dx \qquad \text{Eq. (4-1)}$$

The POD basis is a set of functions $\{\boldsymbol{\varphi}_j(\mathbf{x})\}$ that is the best linear set of basis for describing the ensemble of observations. Since the basis is linear, a flow-field $\mathbf{u} \in span\{\boldsymbol{\varphi}_j\}$ can be represented as a linear combination of the POD modes,

$$\mathbf{u}(\mathbf{x},\mathrm{t}) = \sum_j a_j(\mathrm{t})\boldsymbol{\varphi}_j(\mathbf{x}) \qquad \text{Eq. (4-2)}$$

The POD modes, or empirical eigenfunctions, are defined by requiring that the averaged projection of the ensemble $\mathbf{u}^k$ onto $\boldsymbol{\varphi}$ is a maximum:

$$\max_{\boldsymbol{\varphi}\in H(\Omega)} \frac{\langle (\mathbf{u},\boldsymbol{\varphi})^2 \rangle}{\|\boldsymbol{\varphi}\|^2}, \qquad \text{Eq. (4-3)}$$

where $\|.\|$ is the norm generated by the inner product. The averaging operator $\langle . \rangle$ used in Eq. (4-3) could be an ensemble average over many numerical or experimental realizations, or it could be a time-average taken from different samples of a single experiment. The main assumption regarding the averaging operator is that it commutes with the inner product. This assumption is shown to hold for the scalar case defined on the Hilbert space L^2 under certain conditions on **u**.

The constrained optimization problem Eq. (4-3) with constraint $\|\boldsymbol{\varphi}\| = 1$ reduces to the eigenvalue problem

$$\mathbf{R}\boldsymbol{\varphi} = \lambda\boldsymbol{\varphi} \qquad \text{Eq. (4-4)}$$

where

$$\mathbf{R}\boldsymbol{\varphi} \equiv \left\langle \mathbf{u}^k (\mathbf{u}^k, \boldsymbol{\varphi}) \right\rangle \qquad \text{Eq. (4-5)}$$

The operator **R** is self-adjoint and non-negative definite; if we further assume that **R** is compact, then there exists a countable set of non-negative eigenvalues λ_i, with associated eigenfunctions $\boldsymbol{\varphi}_i$. The eigenfunctions, appropriately normalized, form an orthonormal subspace of *H*, i.e. $(\boldsymbol{\varphi}_i, \boldsymbol{\varphi}_j) = \delta_{ij}$. The notions of compactness of operators and spaces, as well as the theory of self-adjoint operators, come from the mathematical discipline of functional analysis; see, e.g., [104]. For more details on the compactness of **R** and the required assumptions, refer to section 3.8.2 of [19].

The POD modes are the eigenfunctions $\boldsymbol{\varphi}_i$ associated with non-zero λ_i. Taking the inner product of Eq.(4-4) with $\boldsymbol{\varphi}$, it is straightforward to show that $\left\langle (\mathbf{u}^k, \boldsymbol{\varphi}_i)^2 \right\rangle = \lambda_i$. In other words, the magnitude of the eigenvalue is equivalent to the average energy of the projection of the ensemble onto the associated eigenfunction, where the square of the inner product is interpreted as an energy measure. It is important to remark that in incompressible fluid mechanics with velocity measurements, this energy is related to the fluid's kinetic energy. However, in structural dynamics problems with, say, displacement and/or velocity measurements, there is generally no direct correspondence between this energy and either the system's kinetic energy or its potential energy or any combination thereof. Therefore, thinking of the eigenvalues as 'energies' in a general mechanical context is incorrect in principle and may be misleading [105]. In signal processing, this energy is not a physical energy. The POD modes may be ordered according to the magnitude of their eigenvalue, with $\lambda_1 / \boldsymbol{\varphi}_1$ equal to the eigenvalue/eigenfunction pair with the largest eigenvalue, λ_n equal to the smallest non-zero eigenvalue, and

$\lambda_1 > \lambda_2 > ... > \lambda_n > ... > \lambda_n$. In building reduced order models one is interested in truncating the POD basis and retaining only the most energetic modes. It can be shown that the sequence of truncated POD bases form an optimal set, in the sense that a POD basis comprised of *K* modes describes more energy (on average) of the ensemble than any other linear basis of the same dimension *K*. This compression of the ensemble energy into a minimum number of modes makes the POD basis attractive for reduced order modelling.

The span of the POD basis is not complete in $H(\Omega)$, but it is complete in the sense that, on average, any snapshot used to construct it can be represented, i.e. $\left\langle \left\| \mathbf{u}^k - \sum_j (\mathbf{u}^k, \boldsymbol{\varphi}_j)\boldsymbol{\varphi}_j \right\| \right\rangle = 0$. Conversely, each POD mode can be reconstructed as a linear combination of the observations used to construct the basis. Thus,

$$\boldsymbol{\varphi}(\mathbf{x}) = \sum_{k=1}^{m} a_k \mathbf{u}^k(\mathbf{x}) \qquad \text{Eq. (4-6)}$$

Eq. (4-6) is a mathematical expression of the intuitive argument that the POD basis contains only information on the kinematics of the flow-field that were already encoded in the observations. Further down the road in the reduced order modelling process, the dynamical features of a model will depend critically on the observations used to construct the reduced basis. A consequence of Eq. (4-6) is that the POD eigenfunctions share any closed linear property shared by all the ensemble members $\mathbf{u}^k$. Examples of such properties are the divergence-free property of incompressible flow and satisfaction of linear boundary conditions such as the no-slip surface condition.

In practice, the $\mathbf{u}^k$ are vectors of state variables at discrete grid point locations, each containing a single solution from the numerical simulation. They will have length *nl*, where *n* is the total number of grid points and *l* is the number of dependent variables describing the flow state. Thus, the discretized version of Eq. (4-4) will be an eigenvalue

problem of order *nl*. For $n >> m$, where m is the number of flow-field snapshots used, this procedure is costly and, it turns out, inefficient.

Sirovich [106] showed how the eigenvalue problem Eq.(4-4) can be reduced to order m, resulting in a much more efficient procedure for $n >> m$. Assume that the averaging operator $\langle . \rangle$ is a time average over a finite number of samples. Substituting the modal decomposition Eq. (4-2) into Eq. (4-4) to obtain

$$\frac{1}{m}\sum_{i=1}^{m}\mathbf{u}^i\left(\mathbf{u}^i,\sum_{k=1}^{m}a_k(t)\mathbf{u}^k\right)=\lambda\sum_{k=1}^{m}a_k\mathbf{u}^k \qquad \text{Eq. (4-7)}$$

Using the property (x +y,z)=(x,z)+(y,z),

$$\frac{1}{m}\sum_{i=1}^{m}\mathbf{u}^i\left(\mathbf{u}^i,\sum_{k=1}^{m}a_k\mathbf{u}^k\right)=\frac{1}{m}\sum_{i=1}^{m}\mathbf{u}^i\sum_{k=1}^{m}a_k\left(\mathbf{u}^i,\mathbf{u}^k\right) \qquad \text{Eq. (4-8)}$$

$$=\sum_{i=1}^{m}\left[\sum_{k=1}^{m}\frac{1}{m}\left(\mathbf{u}^i,\mathbf{u}^k\right)a_k\right]\mathbf{u}^i=\lambda\sum_{k=1}^{m}a_k\mathbf{u}^k \qquad \text{Eq. (4-9)}$$

A sufficient condition for the solution of Eq. (4-4) is then

$$\sum_{k=1}^{m}\frac{1}{m}\left(\mathbf{u}^i,\mathbf{u}^k\right)a_k=\lambda a_i;\quad i=1,.....,m \qquad \text{Eq. (4-10)}$$

Equation (4-10) is one row of a new eigenvalue problem with row index *i* and column index *k.* Once the eigenvectors for Eq. (4-10) are computed, the POD modes are computed using Eq. (4-6). This is the so-called ‘method of snapshots’ for computing a POD basis.

4.1.1 The POD analysis for snapshots with a changing physical domain

When the physical domain between snapshots is different, that is there exists a geometrical change between snapshots, the modes derived from the snapshots will no longer remain at fixed places within the computational domain and consequently an error is introduced in this modelling technique being a space-index transformation. The inner product is computed as

$$\left(\mathbf{u}^i,\mathbf{u}^k\right)=\int_{\Omega}\mathbf{u}^i(\mathbf{x})\,\mathbf{u}^k(\mathbf{x})d\Omega \qquad \text{Eq. (4-11)}$$

For simplicity, here we consider a steady-state problem ie. $\mathbf{u}=\mathbf{u}(\mathbf{x})$ only.

The evaluation of the inner product in Eq. (4-11) is a problem when the domain of the *i* and *k* snapshots are not the same. One approach to this problem is to use a common domain for every snapshot and apply transpiration boundary conditions to account for the changes in the boundary [107]. However, in this work the method as suggested and adopted by LeGresley and Alonso [14] has been considered. In this case, the structured physical mesh is transformed into the computational domain with indices denoted by ξ,η,ζ. The state variables are considered to be constant in each cell and so the discrete equivalent of Eq. (4-11) is as follows

$$\left(\mathbf{u}^i,\mathbf{u}^k\right)=\sum_{\xi,\eta,\zeta}\mathbf{u}^i(\xi,\eta,\zeta)\mathbf{u}^k(\xi,\eta,\zeta)\;\mathbf{V}(\xi,\eta,\zeta) \qquad \text{Eq. (4-12)}$$

where $\mathbf{V}(\xi,\eta,\zeta)$ is the ensemble average of the ξ,η,ζ -th cell volume as calculated in the physical domain. In case that the physical domain is the same for all snapshots, Eq. (4-12) is then equivalent to the continuous definition given in Eq. (4-11) . Once this is completed, the flow solution variables can be expanded in the form of

$$\mathbf{u}(\mathrm{x},\mathrm{y},\mathrm{z})=\sum_j a_j\boldsymbol{\varphi}_j(\mathrm{x},\mathrm{y},\mathrm{z}) \qquad \text{Eq. (4-13)}$$

4.2 The POD in terms of the SVD

In this section, the POD is described for steady-state problems in terms of the singular value decomposition (SVD). This approach is normally preferred as it is more straightforward and easy to be implemented.

Considering an ensemble of data $\mathbf{A}\left(\mathbf{Y(x)}\right)$ where $\mathbf{A} \in \mathrm{R}^{mxn}$ is obtained from the solutions $\mathbf{Y(x)}$ of a high-fidelity model at various design points $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, ..., \mathbf{x}^{(m)}$ where $\mathbf{Y} \in \mathrm{R}^{n}$ represents the solution vector of primitive or conservative field variables and $\mathbf{x} \in \mathrm{R}^{r}$ represents the vector of different design variables or parameters. From these the ensemble of data is formed as follows,

$$\mathbf{A}\left(\mathbf{Y(x)}\right) = \begin{pmatrix} \mathbf{Y}\left(\mathbf{x}^{(1)}\right) \\ \mathbf{Y}\left(\mathbf{x}^{(2)}\right) \\ . \\ . \\ \mathbf{Y}\left(\mathbf{x}^{(m)}\right) \end{pmatrix} \qquad \text{Eq. (4-14)}$$

where $\mathbf{Y}\left(\mathbf{x}^{(i)}\right) = \left\{\mathbf{Y}_1\left(\mathbf{x}^{(i)}\right), \mathbf{Y}_2\left(\mathbf{x}^{(i)}\right), ..., \mathbf{Y}_n\left(\mathbf{x}^{(i)}\right)\right\}$, n is the number of grid points over which the computational calculation is performed and m is the number of realizations or parameters combination.

The sub-space spanned by these solutions can be used to approximate $\mathbf{Y}$ by representing it in terms of orthogonal basis functions or vectors $\mathbf{\Phi}$ which span the parameter space of interest.

$$\mathbf{\Phi} = \mathbf{span}\ \left\{\mathbf{Y}\left(\mathbf{x}^{(1)}\right), \mathbf{Y}\left(\mathbf{x}^{(2)}\right), \mathbf{Y}\left(\mathbf{x}^{(3)}\right), ..., \mathbf{Y}\left(\mathbf{x}^{(m)}\right)\right\} \qquad \text{Eq. (4-15)}$$

In POD each realization or solution vector is referred to as a snapshot.
The objective of the POD is to obtain a set of m optimal basis vectors which span the subspace formed by m snapshots. In this context the snapshots are obtained for different

values of the parameters, so that the solution at a particular parameter can be reconstructed as

$$\mathbf{Y} = \langle \mathbf{Y} \rangle + \sum_{i=1}^{m} \boldsymbol{\alpha}_i \boldsymbol{\Phi}_i \qquad \text{Eq. (4-16)}$$

where $\langle \mathbf{Y} \rangle = \frac{1}{m} \sum_{i=1}^{m} \mathbf{Y}\left(\mathbf{x}^{(i)}\right)$ is the arithmetic mean of the vector of field variables at each grid point across the whole set of snapshots, $\boldsymbol{\Phi}_i$, i = 1, 2, ……,m denotes the set of basis vectors at a particular grid point and $\boldsymbol{\alpha}_i$, i = 1, 2,……,m are scalar coefficients to be determined. Hence a snapshot can be reconstructed by adding up the contribution from each basis vector in turn at each grid point.

Defining a set of modified snapshots obtained by subtracting $\langle \mathbf{Y} \rangle$ from $\mathbf{Y(x}^{(i)}\mathbf{)}$,

$$\tilde{\mathbf{Y}}^{(i)} = \mathbf{Y(x}^{(i)}\mathbf{)} - \langle \mathbf{Y} \rangle, \quad i = 1,2,.....,m \qquad \text{Eq. (4-17)}$$

where m is total number of snapshots.

Let $\mathbf{A} \in \mathrm{R}^{mxn}$ denote the matrix whose rows are the modified snapshots. Each snapshot is constructed by placing in order the solution at each grid point for the whole grid. This order can be determined arbitrarily, but is subject to the constraint that it must be consistent throughout the whole set of snapshots.

$$\mathbf{A} = \begin{pmatrix} \tilde{Y}_1^{(1)} & . & . & . & \tilde{Y}_n^{(1)} \\ \tilde{Y}_1^{(2)} & . & . & . & \tilde{Y}_n^{(2)} \\ \tilde{Y}_1^{(3)} & . & . & . & \tilde{Y}_n^{(3)} \\ . & . & . & . & . \\ \tilde{Y}_1^{(m)} & . & . & . & \tilde{Y}_n^{(m)} \end{pmatrix} \qquad \text{Eq. (4-18)}$$

The SVD of $\mathbf{A}$ can be written as

$$\mathbf{A} = \mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^{\mathrm{T}} \qquad \text{Eq. (4-19)}$$

where $\mathbf{U} \in \mathrm{R}^{mxm}$ and $\mathbf{V} \in \mathrm{R}^{nxn}$ are orthogonal matrices. These matrices are the left and right singular vectors respectively. $\mathbf{\Sigma} \in \mathrm{R}^{mxn}$ is a diagonal matrix whose diagonal elements consist of $q = \min(m,n)$ non-negative numbers σ_i arranged in decreasing order, that is,

$$\sigma_1 \geq \sigma_2 \geq \sigma_3 \geq .... \geq \sigma_q$$

σ_i are referred to as the singular values of $\mathbf{A}$, hence the name singular value decomposition. In expanded matrix form the SVD of $\mathbf{A}$ can be expressed as follows,

$$\begin{pmatrix} \tilde{Y}_1^{(1)} & \cdots & \tilde{Y}_n^{(1)} \\ \tilde{Y}_1^{(2)} & \cdots & \tilde{Y}_n^{(2)} \\ \tilde{Y}_1^{(3)} & \cdots & \tilde{Y}_n^{(3)} \\ \vdots & \cdots & \vdots \\ \tilde{Y}_1^{(m)} & \cdots & \tilde{Y}_n^{(m)} \end{pmatrix} = \begin{pmatrix} u_{11} & \cdots & u_{1m} \\ u_{21} & \cdots & u_{2m} \\ u_{31} & \cdots & u_{3m} \\ \vdots & \cdots & \vdots \\ u_{m1} & \cdots & u_{mm} \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_3 & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & 0 & \sigma_m & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} v_{11} & \cdots & v_{1n} \\ v_{21} & \cdots & v_{2n} \\ v_{31} & \cdots & v_{3n} \\ \vdots & \cdots & \vdots \\ \vdots & \cdots & \vdots \\ \vdots & \cdots & \vdots \\ v_{n1} & \cdots & v_{nn} \end{pmatrix}$$

This is the full form of the SVD. Since $\mathbf{\Sigma}$ is a diagonal *m* x *n* matrix, then the above matrix equation can be written in reduced form as follows if we assume that $m < n$,

$$\begin{pmatrix} \tilde{Y}_1^{(1)} & \cdots & \tilde{Y}_n^{(1)} \\ \tilde{Y}_1^{(2)} & \cdots & \tilde{Y}_n^{(2)} \\ \tilde{Y}_1^{(3)} & \cdots & \tilde{Y}_n^{(3)} \\ \vdots & \cdots & \vdots \\ \tilde{Y}_1^{(m)} & \cdots & \tilde{Y}_n^{(m)} \end{pmatrix} = \begin{pmatrix} u_{11} & \cdots & u_{1m} \\ u_{21} & \cdots & u_{2m} \\ u_{31} & \cdots & u_{3m} \\ \vdots & \cdots & \vdots \\ u_{m1} & \cdots & u_{mm} \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & 0 & \cdots & 0 \\ 0 & 0 & \sigma_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_m \end{pmatrix} \begin{pmatrix} v_{11} & \cdots & v_{1n} \\ v_{21} & \cdots & v_{2n} \\ v_{31} & \cdots & v_{3n} \\ \vdots & \cdots & \vdots \\ v_{m1} & \cdots & v_{mn} \end{pmatrix}$$

where the matrices $\mathbf{\Sigma}$ and $\mathbf{V}^{\mathbf{T}}$ are reduced in size. The columns of $\mathbf{V}$ and hence the rows of $\mathbf{V}^{\mathbf{T}}$ are the proper orthogonal modes of the system. Hence the set of basis vectors $\mathbf{\Phi}_i = \mathbf{V}^{\mathbf{T}}(i,:) \in \mathrm{R}^n$. These basis vectors are of unit magnitude and orthogonal, hence orthonormal.

Writing the product of $\mathbf{U}$ and $\mathbf{\Sigma}$ as a matrix $[\alpha_{ij}]$

$$\begin{pmatrix} \tilde{Y}_1^{(1)} & \cdots & \tilde{Y}_n^{(1)} \\ \tilde{Y}_1^{(2)} & \cdots & \tilde{Y}_n^{(2)} \\ \tilde{Y}_1^{(3)} & \cdots & \tilde{Y}_n^{(3)} \\ \vdots & \cdots & \vdots \\ \tilde{Y}_1^{(m)} & \cdots & \tilde{Y}_n^{(m)} \end{pmatrix} = \begin{pmatrix} \alpha_{11} & \cdots & \alpha_{1m} \\ \alpha_{21} & \cdots & \alpha_{2m} \\ \alpha_{31} & \cdots & \alpha_{3m} \\ \vdots & \cdots & \vdots \\ \alpha_{m1} & \cdots & \alpha_{mm} \end{pmatrix} \begin{pmatrix} v_{11} & \cdots & v_{1n} \\ v_{21} & \cdots & v_{2n} \\ v_{31} & \cdots & v_{3n} \\ \vdots & \cdots & \vdots \\ v_{m1} & \cdots & v_{mn} \end{pmatrix} \qquad \text{Eq. (4-20)}$$

which has the form of $\tilde{Y}_k^{(i)} = \sum_{j=1}^{m} \alpha_{ij} v_{jk}$. Thus, the scalar coefficients are obtained directly from the multiplication of $\mathbf{U}$ and $\mathbf{\Sigma}$. From the SVD of $\mathbf{A}$ it can be noticed that since $\mathbf{V}$ is an orthogonal matrix, the transpose of $\mathbf{V}$ is equal to its inverse ie. $\mathbf{V}^{\mathbf{T}} = \mathbf{V}^{-1}$, therefore

$$\mathbf{AV} = \mathbf{U\Sigma} \qquad \text{Eq. (4-21)}$$

that is,

$$\alpha_{ij} = \tilde{\mathbf{Y}}^{(i)}.\boldsymbol{\Phi}_{j} \qquad \text{Eq. (4-22)}$$

where $\boldsymbol{\Phi}_{j} = \mathbf{V}(:, j) \in \mathrm{R}^{n}$. The scalar coefficients α_{ij} are also referred to as projection coefficients because these are obtained by projecting the solution vector onto the basis vectors.

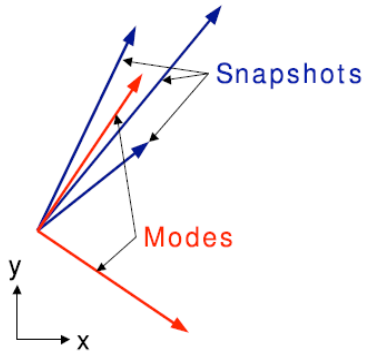


Figure (4-1) Vector representation of the solution vectors and the POD modes

A complete reconstruction of the snapshots can be obtained from

$$\mathbf{Y} = \langle \mathbf{Y} \rangle + \tilde{\mathbf{Y}} \qquad \text{Eq. (4-23)}$$

Now, $\mathbf{Y}$ may represent a vector of scalar functions such as the primitive or conservative variables and therefore the method described can be applied to each variable in turn to form a distinct basis for each variable. However, an improvement in the ability of the basis to represent the system may be achieved by considering not only how the individual variables vary from one snapshot to another but also how variables change relative to one another. Hence $\mathbf{Y}$ is developed from state variable vectors consisting of all the primitive or conservative variables [14]. In this case, the POD modes are sensitive to the scaling of the flow variables as these are in different units and have significantly

varying magnitudes. Consequently, appropriate scaling factors are necessary for each fluctuating flow variable which makes their magnitude of the same order [30].

If a problem is represented by a suitable number of snapshots from which a suitably rich set of basis vectors is available, the singular values become small rapidly and a small number of basis vectors are adequate to reconstruct and approximate the snapshots. In this way, POD provides an efficient means of capturing the dominant features of a multi-degree of freedom system and representing it to the desired precision by using the relevant set of modes, thus reducing the order of the system. In other words, the reduced-order model is derived by projecting the CFD model onto a reduced space spanned by only some of the proper orthogonal modes or POD eigenfunctions.

Assuming that *p* modes which correspond to the largest *p* singular values are dominant, then the energy E or variance in the data captured by the first *p* modes can be computed as

$$\mathrm{E}(p) = \frac{\sum_{i=1}^{i=p} \sigma_i^2}{\sum_{i=1}^{i=m} \sigma_i^2} \qquad \text{Eq. (4-24)}$$

If this energy is high say over 99% of the total energy, then *p* modes are adequate to capture the principal features and approximately reconstruct the dataset. Thus, a reduced subspace is formed which is only spanned by *p* modes.

Instead of the SVD approach, the proper orthogonal modes can also be computed by solving for the largest *p* eigenvalues and corresponding eigenvectors of the matrix $\mathbf{K} = \mathbf{A}^{\mathbf{T}}\mathbf{A} \in \mathbf{R}^{nxn}$ where $\mathbf{A}^{\mathbf{T}}\mathbf{A} = \mathbf{V}\boldsymbol{\Sigma}^{\mathbf{2}}\mathbf{V}^{\mathbf{T}}$. Hence it follows that $\mathbf{V}$ is the matrix of eigenvectors of $\mathbf{K}$ and σ_i^2, i = 1, 2,……., m are its eigenvalues. For cases where m<<n, instead of using the SVD procedure, it is computationally more efficient to use the 'method of snapshots' proposed by Sirovich [106]. In this approach an m x m eigenvalue problem $\mathbf{A}\mathbf{A}^{\mathbf{T}}\mathbf{U} = \boldsymbol{\Sigma}^{\mathbf{2}}\mathbf{U}$ is formed, from which $\mathbf{U}$ is computed. Hence, pre-multiplying

by $\mathbf{U}^{\mathrm{T}}$ Eq. (4-19) we get $\mathbf{U}^{\mathrm{T}}\mathbf{A} = \mathbf{\Sigma}\mathbf{V}^{\mathrm{T}}$. Therefore, the first m rows of $\mathbf{U}^{\mathrm{T}}\mathbf{A}$ normalized to unit magnitude represent the proper orthogonal modes.

Using this model reduction technique a low dimensional system is produced that has the same essential characteristics as the original system but with far less storage requirements and a much lower evaluation time. In the next chapter the proper orthogonal decomposition is applied as a reduced-order modelling method to a simple case of practical interest so as to understand its behaviour and subtlety.

4.2.1 Pseudo-Continuous Representation

The use of reduced-order models based upon the proper orthogonal decomposition for prediction requires the transformation of the projection coefficients, α_i from the discrete sample space for which they have been computed to a continuous space. If α_i varies as a smooth function with the change in parameters then a meta-model may be used to determine the POD projection coefficients at intermediate parametric values not included in the original data ensemble. The predicted solution vector $\mathbf{Y}\left(\mathbf{x}^{(\beta)}\right)$ for any variable β within the parametric space is given by,

$$\mathbf{Y}\left(\mathbf{x}^{(\beta)}\right) = \left\langle \mathbf{Y} \right\rangle + \sum_{i=1}^{p'} \alpha_i^{\beta} \mathbf{\Phi}_i \qquad \text{Eq. (4-25)}$$

where p' is normally greater than p and the weighting coefficients α_i^{β} are found using the meta-model.

A variety of meta-modeling techniques suitable for this purpose are described in the literature of which the response surface methodology (RSM) has found general acceptance. The RSM is a statistical tool originally developed for experimental design and subsequently adapted to approximate computational simulations. The RSM is founded on the assumption that the data can be described by a set of simple basis

functions, such as second-order polynomials, that are fitted to the data ensemble using a least-squares regression technique.

While regression techniques work well for experimental data, where noise due to random errors is smoothed out from the data, they are less appropriate when dealing with the results from deterministic numerical simulations or when working with complex data sets. Of particular concern in the context of the present work are two problems; firstly that the response surface constructed using regression analysis may not exactly fit the sample data from which it has been constructed and secondly that the method smoothes local variations in the data.

To investigate and understand these problems further parametric/non-parametric meta-models are also explored, in particular spline interpolation methods and radial basis functions. In the latter case multi-quadric, polyharmonic and Gaussian kernels were considered. All these models produce an interpolative fit through all of the sample points and provide improved representations of data sets that have localized minima and maxima. This is dealt with in detail in chapter 6.

Chapter 5

5.0 POD-based reduced order modelling

In this chapter the model described in the preceding one is used to exhibit the utility of POD-based ROMs for high-speed aerodynamic problems. This is demonstrated by considering an inviscid steady flow about an axi-symmetric flare stabilized projectile similar to the one studied by Schmidt *et al.* [108] and Plostins *et al.* [109]. This case was selected as it is one for which the industry standard semi-empirical modelling tools perform poorly, most especially in determining the axial force and the pitch damping coefficients [113] and therefore poses a reasonable challenge to the proposed approach, particularly as a prediction tool. Moreover, the semi-empirical tools output only those properties defined in the code. In addition this methodology offers an effective and valuable data management system for both data storage and handling. This is demonstrated by an illustration.

5.1 Inviscid flow over an axi-symmetric projectile

To demonstrate the effectiveness of POD-based ROMs, an inviscid steady flow over an axi-symmetric flare stabilized projectile with a parametric variation of the inflow Mach number (flow condition) and the flare base radius (geometry) was considered. It is well-known that the forces acting on a cone-cylinder-flare body at hypersonic speeds vary linearly provided no flow separation develops ahead of the flare due to for example a large flare angle, in which case the forces start to deviate from a linear relationship. In

this work, the flare angle was maintained low enough so as to create an attached shock at the flare. Moreover, an angle of incidence of 0º was considered resulting in an axi-symmetric flow over the projectile.

5.1.1 Generation of the dataset of observations

A Lagrangian sub-space was developed by generating snapshot sets for two varying parameters using the IMPNS software. Snapshot sets consist of state solutions corresponding to several sets of parameter values that emerge from the problem specification and must contain sufficient information to accurately represent the characteristics of the solutions of the full-order equations. For this problem, the Mach number range [4.0, 6.0] was divided into 20 uniform intervals while the non-dimensionalized flare base radius r/D range [0.75, 1.25] was divided into 10 uniform intervals. The full factorial design-of-experiment (Section 3.3.1) was employed to generate the sample points resulting in an ensemble containing a total of 231 snapshots. This was used as yet no systematic and effective procedure is available in generating snapshot sets. However, this problem is investigated in the next chapter.

5.1.2 Geometry and the computational grid

Figure (5-1) shows the geometry of the projectile with the maximum and minimum flare angles considered, while Figure (5-2) shows an example of the grid on which the axi-symmetric inviscid computations were accomplished.

Nose tip included angle = 18.9º **Flare included angle = 28.1º**

(a) r/D = 1.25

Nose tip included angle = 18.9º **Flare included angle = 9.5º**

(b) r/D = 0.75

Figure (5-1) Geometry of the axi-symmetric projectile

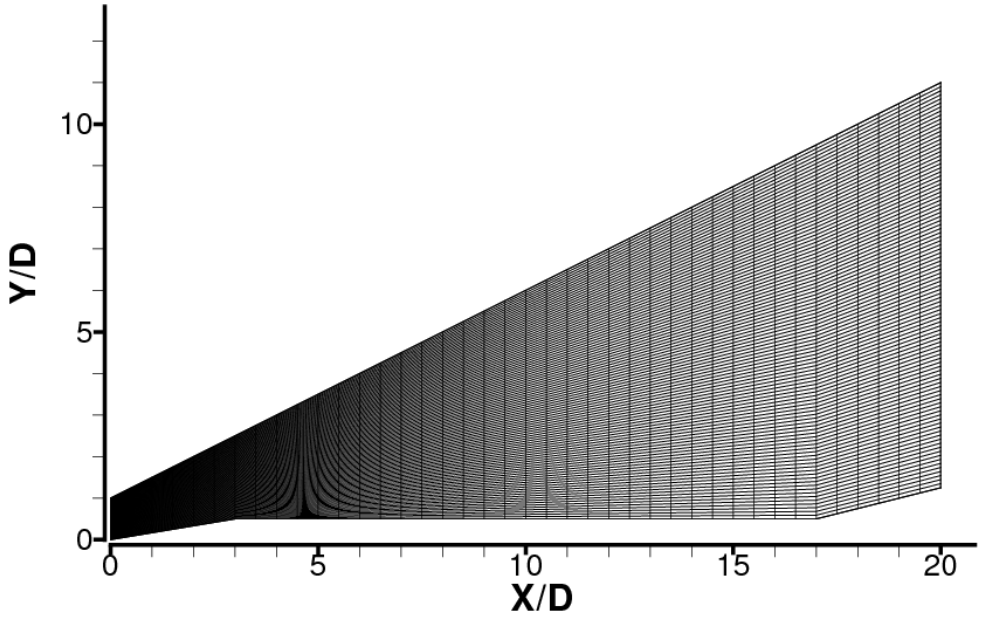
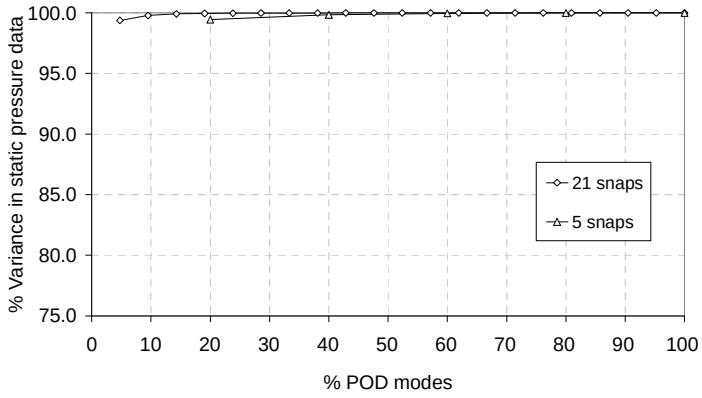


Figure (5-2) Two-dimensional grid with 41 x 101 grid points used for the Euler computations at r/D = 1.25

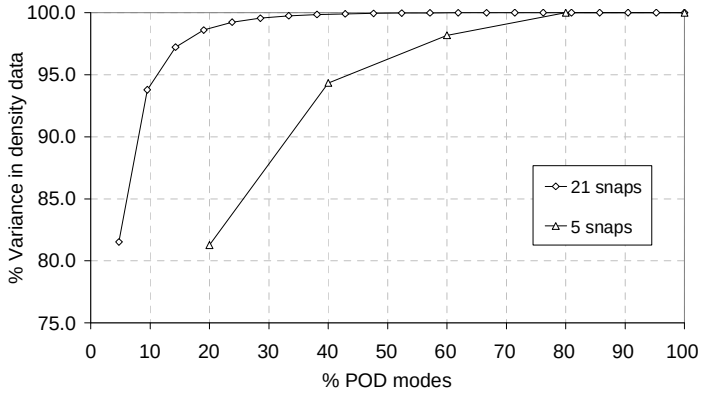
5.1.3 Investigation of the problem behaviour

In order to investigate the linear/non-linear behaviour of this particular problem, a comparison of the variance as captured by the POD modes for each of the primitive variables and using datasets with a different number of snapshots was conducted. POD analysis was applied on two subsets of the generated dataset of observations with one varying parameter, that is, the Mach number. A constant flare angle corresponding to a flare base radius r/D of 0.75 was considered. One set of data had 21 snapshots with the Mach number varying within the range [4.0, 6.0] with 20 uniform intervals, while the other set of data had 5 snapshots with 4 uniform intervals within the same Mach number range. The POD was applied on centered data. By centered data it is meant that the average value of the primitive variables over the series of snapshots is subtracted from each computed raw value.

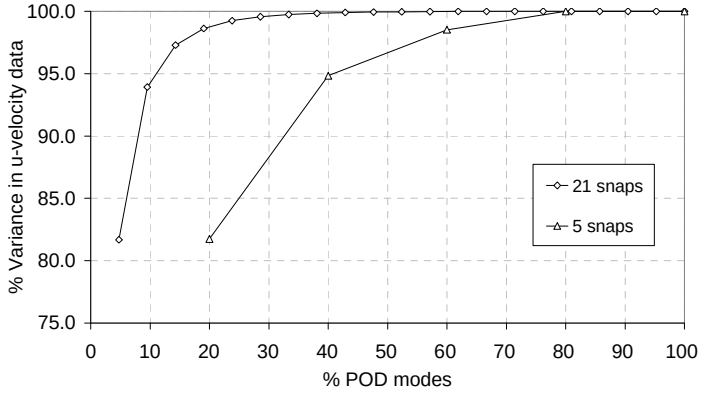
From these analyses, plots were made of the percentage energy or variance in data captured versus the percentage number of POD modes utilised for each of the primitive variables. Now, it is expected that if the primitive variables vary linearly across the set of snapshots, then the variance in data captured from the two analyses should be the same and hence superimpose. Figure (5-3)(a) reveals that indeed the static pressure varies linearly across the set of snapshots since the percentage variance in the data as captured by the percentage number of POD modes for both analyses superimpose, but there exists some considerable difference in the percentage variance for the density, u-velocity and v-velocity. This implies that the latter primitive variables vary non-linearly across the set of snapshots. Thus, from this investigation it can be concluded that this problem is non-linear from a POD analysis point of view.


% Variance in static pressure data
100.0
95.0
90.0
85.0
80.0
75.0
0
10
20
30
40
50
60
70
80
90
100
% POD modes
21 snaps
5 snaps

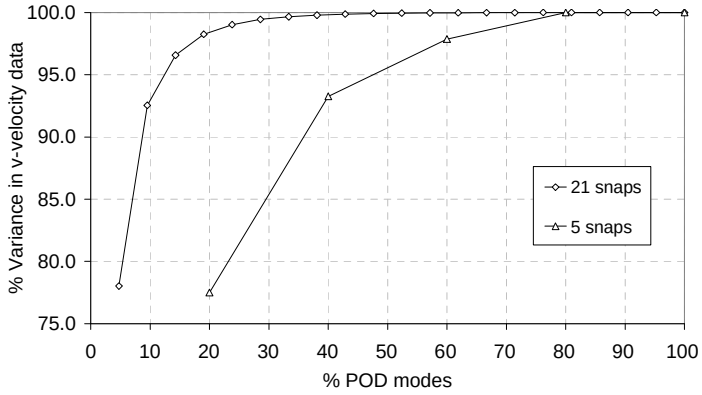
(a)


% Variance in density data
100.0
95.0
90.0
85.0
80.0
75.0
0
10
20
30
40
50
60
70
80
90
100
% POD modes
21 snaps
5 snaps

(b)


% Variance in u-velocity data
100.0
95.0
90.0
85.0
80.0
75.0
0
10
20
30
40
50
60
70
80
90
100
% POD modes
21 snaps
5 snaps

(c)



(d)

Figure (5-3) – Percentage variance captured versus the percentage number of POD modes for each of the primitive variables for different snapshot sets

From these analyses it was observed that for the 21 snapshot dataset, 47.6 % of the POD modes were required to capture over 99.9 % of the variance in the data for all the variables, while for the 5 snapshots dataset, close to 80 % of the POD modes are required to capture the same amount of variance in the data. These values were mostly determined by the cross-flow velocity. It is expected that the introduction of the second parameter would affect these results, causing a further increase in the percentage number of POD modes to capture the 99.9 % of the variance in the data.

5.1.4 Data reconstruction

In this case the whole dataset of observations was considered. The POD-based ROM method was applied to the resulting dataset and a model containing 231 orthogonal basis vectors or POD modes was obtained. This model was used to reconstruct the solution at a Mach number of 4.0 and a flare base radius of r/D = 0.75. The results are pictorially represented in Figures (5-4) to (5-6). Figure (5-4) shows the variation of the energy defined in Eq. (4-21) captured by the POD modes against the number of POD modes for each of the primitive variables. It is evident that most of the energy that is 99.9 % is captured within the first 16 POD modes. This indicates that the dataset of observations is rich enough to capture the flow details. Moreover the number of numerically significant singular values gives an estimate of the effective dimension of the Lagrangian sub-space or snapshots set.

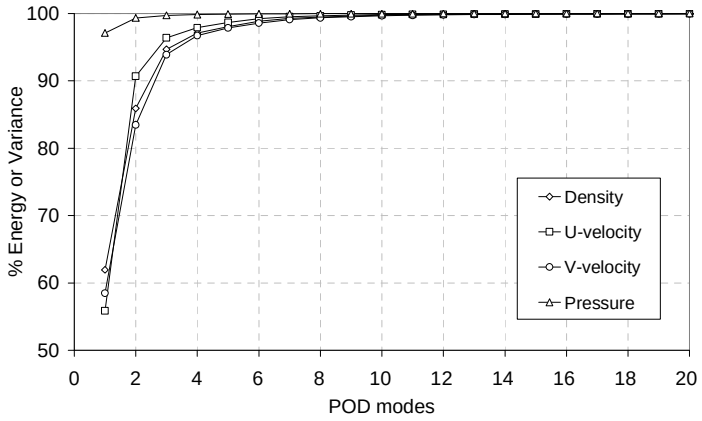


Figure (5-4) Percentage Energy captured versus the number of POD modes for each of the primitive variables

While the energy captured by a given set of basis vectors provides some understanding of the relative accuracy of the model, the actual error remains unquantified. This quantity can be determined by comparing the original sample data with that generated by the reduced-order model. In this work an assessment is made by determining the percentage variation of the root mean square value of the global error in the whole flow field as it varies with the number of POD modes. This was calculated according to the following;

$$Global\ \ Error_{RMS} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(\frac{var_{ROM}^{i} - var_{Hi-Fi}^{i}}{var_{Hi-Fi}^{i}}\right)^{2}} \qquad \text{Eq. (5-1)}$$

where *m* is the number of grid points and *var* represent the primitive variable.

For this case the percentage root mean square global error in the whole flow field reduces to approximately 0.1% when using only 60 of the 231 available POD modes. The biggest error was registered by the cross flow v-velocity. Refer to Figure (5-5).

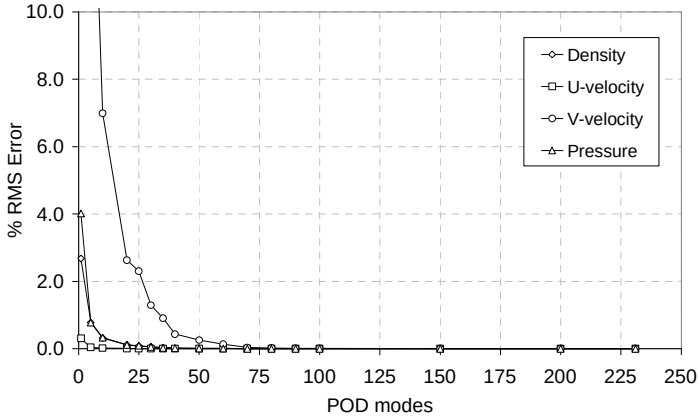
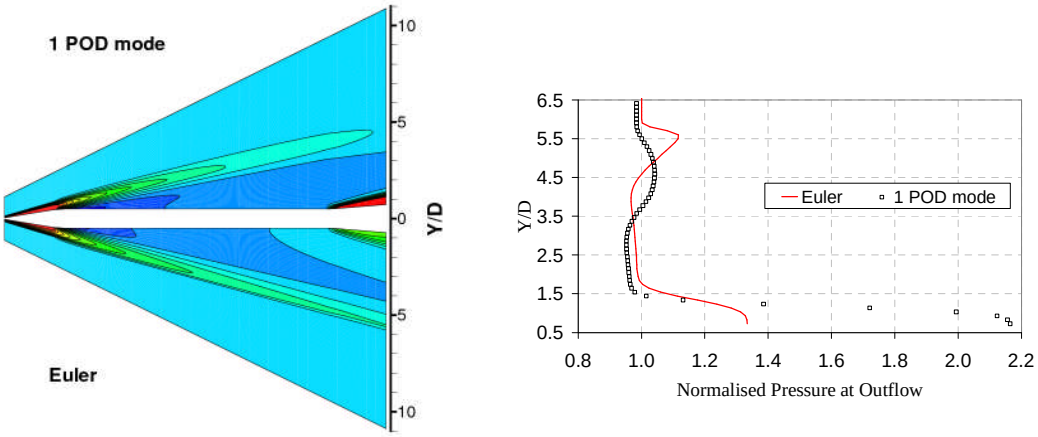


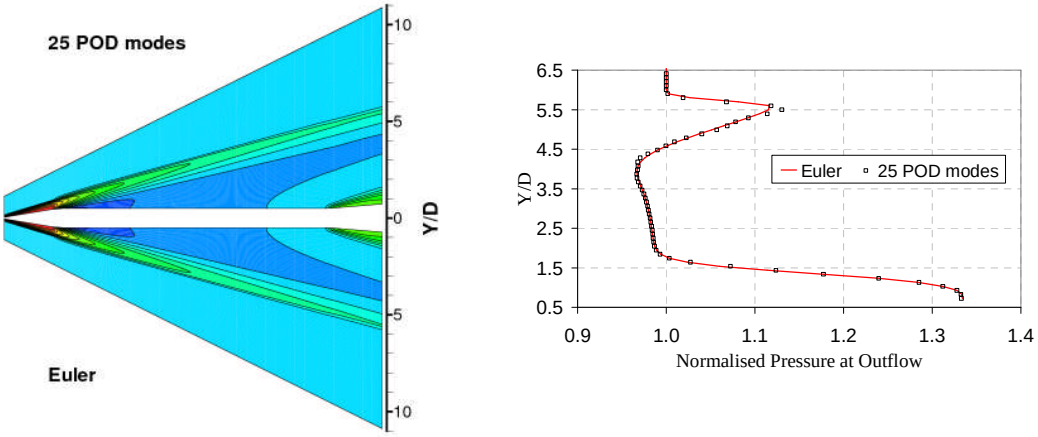
Figure (5-5) R.M.S. Error versus the number of POD modes for each of the primitive variables

Figure (5-6) illustrates this more clearly by comparing static pressure contours obtained from the model using 1, 25 and 35 POD modes (upper-part) with those of the original Euler computation (lower-part). A comparison of the normalized static pressure distribution at the outflow is also included. Note that the static pressure is normalized using the free-stream pressure. Using a single POD mode, the main flow features, the conical shock at the nose and the compression over the flare, are reasonably represented though there is considerable disagreement in the outflow pressure distribution. For one single POD mode, the RMS global error in the whole static pressure flow field is 4.0 %. With 25 POD modes significant improvements are obtained. The conical shock is now essentially identical to that of the Euler computation, while the detail of the compression is much better represented than with just one mode. In this case, only slight discrepancies are noticeable in the static pressure distribution at the outflow. In fact the RMS global error within the whole pressure field is only 0.09 %. Further improvements are obtained with 35 POD modes, with the entire flow field now essentially identical to that obtained by solution of the Euler equations. This is also evident from the outflow pressure distribution. The RMS global error within the whole static pressure field is 0.03 %.

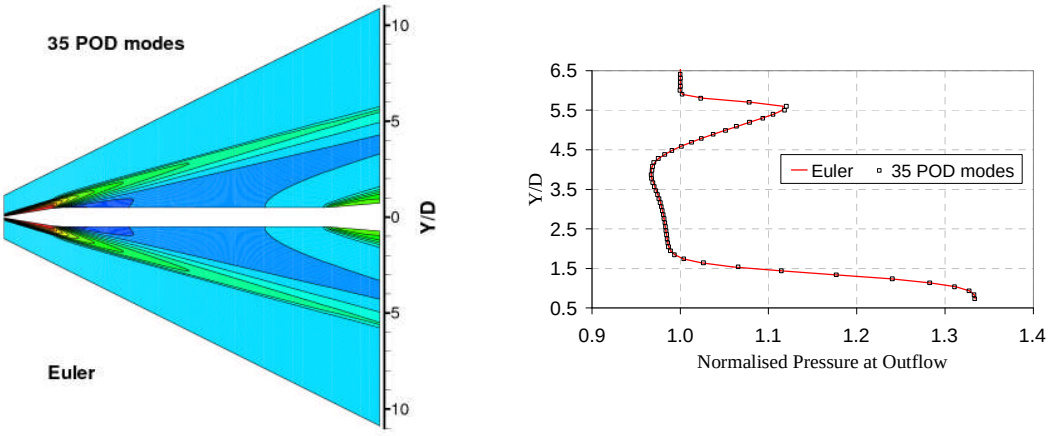
Thus, in this section it has been shown that the complete static pressure field has been reconstructed to a very high level of accuracy using only 16 % of the POD modes or basis vectors derived from the set of snapshots. Consequently, the pressure field is now represented or modelled by a reduced set of basis vectors.



(a) 1 POD mode



(b) 25 POD modes



(c) 35 POD modes

Figure (5-6) Comparison of pressure contours and outflow pressures at M = 4.0 and r/D = 0.75 using the POD/ROM (upper-half) and Euler computation (lower-half).

5.1.5 Data de-noising

For a set of noisy data such as some physical response from an experiment, the same method can also be applied to de-noise it by retaining the essential information while neglecting the least energetic POD modes, which may be considered noise. Note that this applies to data with random errors which may be considered as a Gaussian noise. Unfortunately, this does not apply to incompletely converged computer simulations in which the error is normally biased.

5.2 POD as a data and time management system

5.2.1 Data compression

A useful application of this technique is that of data compression. This is achieved by keeping the most informative POD modes whilst neglecting the rest without much loss of detail. Figure (5-7) shows pictorially this potential by making a direct comparison between the percentage RMS global error in the cross-flow field and the percentage effective data in use versus the number of POD modes. With the consideration of only ten POD modes and hence only 4.4 % of the total amount of data, the global RMS error in the cross-flow velocity field is 7.0 %. By considering one hundred POD modes and therefore using 43.4 % of the total data, the RMS global error in the cross-flow velocity field goes down to 0.01 %. This shows that by using 43.4 % of the data only, the solutions can be reconstructed with very high accuracy. Thus, this methodology offers an effective and valuable data management system for both data storage and handling by compressing the data.

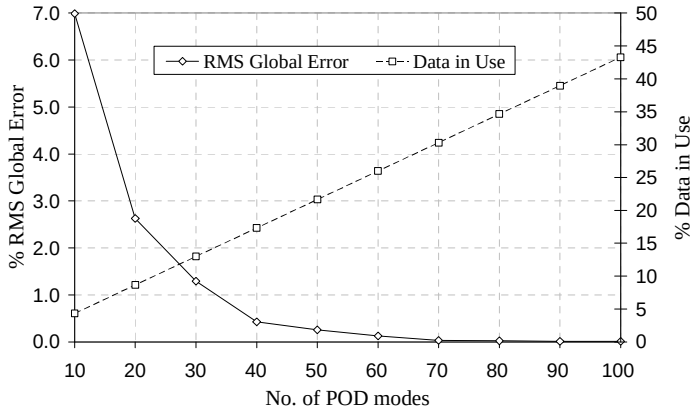


Figure (5-7) Data compression capability of the POD/ROM

5.2.2 Computational Efficiency

Apart from providing a means for data compression, the POD-based ROM methodology offers a significant on-line computational time reduction. Examples of this POD-based ROM feature will be presented and discussed in the following chapter, wherein model predictions are considered and therefore this aspect comes out more clearly.

5.3 Concluding remarks

In this chapter the utility of POD-based ROMs is demonstrated by considering an inviscid steady flow about an axi-symmetric flare stabilized projectile. It was shown that the complete solutions for the snapshots which formed part of the original dataset of observations were reconstructed by using a few POD modes. In addition it was illustrated that this methodology offers an effective and valuable data management system for both data storage and handling.

In this work all the POD analysis was conducted on the primitive variables after the data was centred. In the literature one finds that some do refer to the centralization of data but some others do not. While this has some important application in the interpretation of the data, it is important to remark that the eigenvalues of the Gram matrix or equivalently the square of the singular values extracted from the singular value decomposition represent the variance in the data if and only if the data is centred.

During POD analysis, some difference has been noted in the RMS errors between performing the decomposition on centred and raw data. The difference is greatest along the first few POD modes but becomes negligible as the number of POD modes increases.

Chapter 6

6.0 The surrogate model predictions

In the previous chapter it has been established that the POD method can be used as an effective reduced-order modelling technique. This is so because it provides highly accurate reconstructions of the snapshots that are used in the ensemble of observations, with a small number of POD modes. In this chapter the attention is now directed towards making fast and approximate predictions at parameter combinations not considered when generating the dataset of observations, using the same POD method coupled with either an interpolation or a regression technique. For model predictions, the scalar coefficients obtained by projecting the solution vectors onto the basis vectors are used to generate an interpolated or regressed response surface. In doing so, the projection coefficients are transformed from a discrete space into a pseudo-continuous one, thus creating a foundation for computing the flow field at any parameter value encompassed within the sampled space.

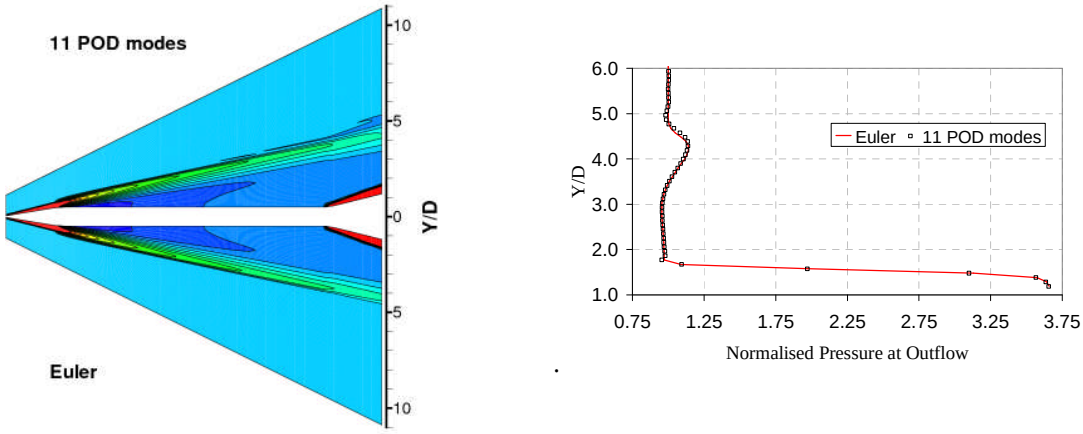
In this chapter, the meta-model is applied to two flow problems related to high-speed weapon aerodynamics; inviscid flow around the flare stabilized hypersonic projectile considered earlier and supersonic turbulent flow around a fin stabilized projectile with drooping nose control. Comparisons of model predictions with high-fidelity CFD simulations are presented. In the end of this chapter a comparison is made between the accuracy and efficiency of the meta-modelling tool with the industry's standard semi-empirical modelling tools.

6.1 Predicting the inviscid axi-symmetric flow over the flare stabilized projectile

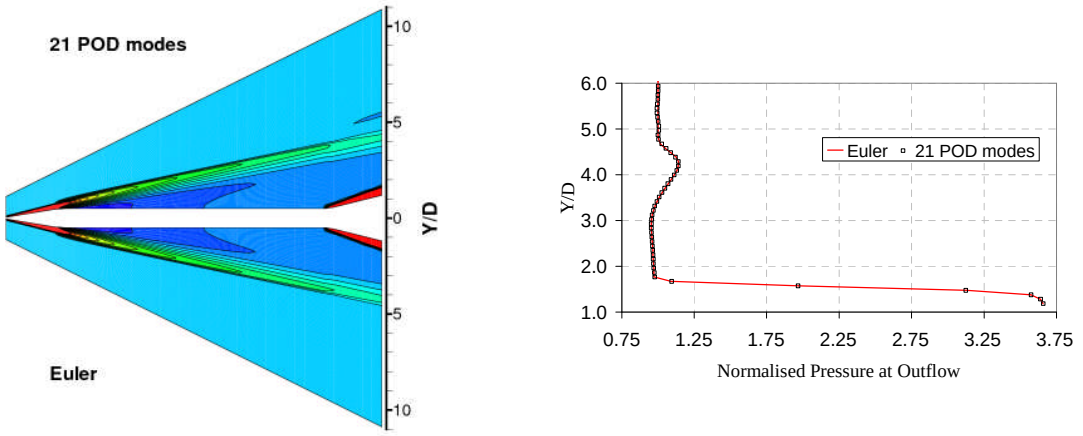
In this section the full dataset of observations used in the preceding chapter for snapshots reconstruction was used again and therefore parametric variation of both flow condition (Mach number) and geometry (flare base radius) were considered. The Mach number range [4.0, 6.0] was divided into 20 uniform intervals while the non-dimensionalized flare base radius r/D range [0.75, 1.25] was divided into 10 uniform intervals. The full factorial design-of-experiment was employed to generate the sample points resulting in an ensemble containing a total of 231 snapshots. Again, an axi-symmetric flow about the projectile is considered since the angle of incidence was maintained at 0^{o}.

6.1.1 One-dimensional interpolation using cubic-splines

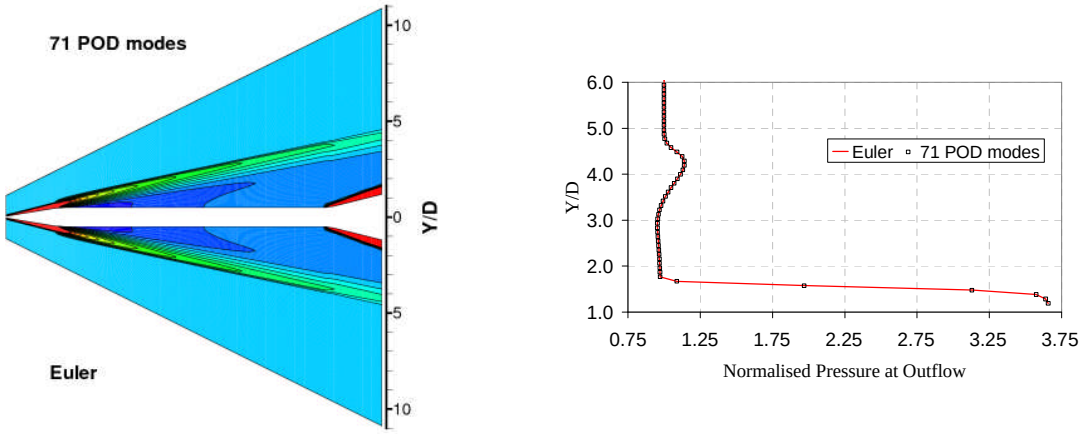
Comparisons of the model predicted static pressure fields with high-fidelity (CFD) solutions for a Mach number of 5.37 and a non-dimensionalized base radius r/D of 1.25 are presented in figure (6-1). The interpolation in this example is accomplished by using cubic-splines along one parameter which is the Mach number as data exists at this base radius. For this case, the predicted solution's behaviour with increasing number of POD modes is similar to that of the previous one, with as few as 11 POD modes a reasonable representation of the main flow features can be obtained, while increasing the number of POD modes leads to successive improvements in the detail of the captured pressure variation. A comparison of the outflow pressure normalized by the free-stream value between that predicted by the meta-model and the high-fidelity solution is also included. For this case, just 71 of the 231 available POD modes resulted in a global RMS error in the pressure field of 0.012 %. This compares with a typical global RMS error in the pressure field of 0.0025 % when the model is used to re-construct members of the ensemble from which it has been deduced.



(a) 11 POD modes



(b) 21 POD modes



(c) 71 POD modes

Figure (6-1) Comparison of pressure contours and outflow pressure at M = 5.37 and r/D = 1.25 obtained using the reduced-order model (upper) and Euler computation (lower)

6.1.2 Two-dimensional interpolation using cubic-splines

Predictions carried out by interpolating along both parameters show similar behaviour. For a Mach number of 4.87 and a base flare radius r/D of 1.03, the use of 71 POD modes resulted in a global RMS error between the meta-model predicted and CFD pressure fields of 0.016 %. As expected the global RMS error for a two-dimensional interpolation is greater than that for a one-dimensional interpolation at a fixed number of POD modes due to the errors associated with each interpolation.

6.1.3 Relationship between the number of samples and model accuracy

The influence of the number of samples in the initial ensemble on the model predictions was investigated. Proper orthogonal decompositions employing 21, 11 and 5 samples at equal intervals to discretize the Mach number range were considered. The same numbers of samples as in the previous case were considered along the geometric parameter, thus resulting in 231, 121 and 55 snapshots. In an effort to isolate the effects of sample size, it was ensured that the predictions obtained for individual ensembles were independent of the number of modes in the reduced model. This was typically obtained using just 40% of the available POD modes. A strong relationship between the sample size and error was found. For the smallest ensemble the global RMS error in the pressure field was found to be 0.68% while for 121 and 231 snapshots the error reduced to 0.15% and 0.016% respectively.

6.2 Meta-model predictions for three-dimensional inviscid flow over the flare stabilized projectile

6.2.1 Full-factorial design-of-experiment

In an effort to understand more clearly the relationship between the predictions of the meta-model and the method used to obtain a pseudo-continuous representation of the scalar coefficients a second study was undertaken. For this case the parameters of interest were extended to include incidence. The Mach number range [4.0, 6.0] was divided into 4 uniform intervals instead of the 20 used in the previous axi-symmetric study, the incidence range [0.0º, 4.0º] was also divided into 4 uniform intervals and the non-dimensionalized flare base radius range [r/D=0.8, r/D=1.2] was divided using a further 4 intervals, instead of the 8 intervals used previously. The observations in this case were sampled according to a full factorial design-of-experiment requiring a total of 125 evaluations of the IMPNS analysis code.

Predictions were made at two points in the parameter space, a point towards the centre of the space (α=2.2˚, M=5.3, r/D=1.05) and a point close to the boundary of the sampled space (α=0.7˚, M=4.2, r/D=0.85). In both cases, none of the parameters of interest correspond to values used to generate the ensemble of snapshots. The predictions obtained for four different modelling approaches (linear regression, linear spline interpolation, cubic spline interpolation and a polynomial augmented multi-quadric radial basis function) are summarized in Table (6-1). In this table, forces and moments are presented that have been obtained by integrating the predicted surface pressures. The reduced-order model used 40 of the 125 available POD modes, increasing this number had no significant effect on the predicted values.

		α=2.2° M=5.3 r/D=1.05			α=0.7° M=4.2 r/D=0.85		
		Meta-Model	CFD	% Error	Meta-Model	CFD	% Error
Linear Regression	**Cx**	0.3518	0.3536	0.51	0.1670	0.1658	0.72
	Cz	0.2591	0.2580	0.43	0.0566	0.0570	0.70
	Cm	-2.7795	-2.7709	0.31	-0.4714	-0.4557	3.45
	Xcp	10.7257	10.7414	0.15	8.3290	7.9985	4.13
Linear Spline Interpolation	**Cx**	0.3573	0.3536	1.05	0.1694	0.1658	2.17
	Cz	0.2594	0.2580	0.54	0.0575	0.0570	0.88
	Cm	-2.7897	-2.7709	0.67	-0.4658	-0.4557	2.22
	Xcp	10.7559	10.7414	0.13	8.1002	7.9985	1.27
Cubic Spline Interpolation	**Cx**	0.3526	0.3536	0.28	0.1682	0.1658	1.45
	Cz	0.2575	0.2580	0.19	0.0572	0.0570	0.35
	Cm	-2.7644	-2.7709	0.23	-0.4616	-0.4557	1.29
	Xcp	10.7345	10.7414	0.06	8.0715	7.9985	0.91
Polynomial Augmented Multi-quadric RBF	**Cx**	0.3537	0.3536	0.03	0.1660	0.1658	0.12
	Cz	0.2579	0.2580	0.04	0.0570	0.0570	0.00
	Cm	-2.7702	-2.7709	0.03	-0.4570	-0.4557	0.29
	Xcp	10.7402	10.7414	0.01	8.0197	7.9985	0.27

Table (6-1) Comparison of model predictions with CFD simulations with various modelling strategies

The results suggest that of the methods considered, the response surface constructed using a polynomial augmented multi-quadric radial basis function is most accurate. Surprisingly the linear regression, which employs quadratic polynomials, performs no better than the linear spline interpolation. This is attributed to the global nature of the regression analysis used to obtain the response surface which appears to excessively smooth the projection coefficients. As the interpolation schemes pass through the supporting data exactly they provide a much better representation of local variations in the data.

To confirm this hypothesis the effects of data locality on the predictions obtained from the linear regression based model were investigated. In this case the parameters' range was reduced in size as shown in Table (6-2). Each parameters' range was discretized using two equal intervals resulting in a total ensemble of 27 snapshots. By successively localizing the data, improved predictions were obtained. This behaviour is attributed to improved representation of the local parametric variations by the response surface rather than improvements in the reduced-order model.

Parameters' Range		α = 2.2° M = 5.3 r/D = 1.05		
		Meta-Model	CFD	% Error
α = [2°, 4°] M = [5, 6] r/D = [0.9, 1.1]	Cx	0.3542	0.3536	0.17
	Cz	0.2556	0.2580	0.93
	Cm	-2.7266	-2.7709	1.60
	Xcp	10.6681	10.7414	0.68
α = [2°, 4°] M = [5, 6] r/D = [1.0, 1.1]	Cx	0.3540	0.3536	0.11
	Cz	0.2563	0.2580	0.66
	Cm	-2.7395	-2.7709	1.13
	Xcp	10.6901	10.7414	0.48
α = [1°, 3°] M = [5, 6] r/D = [1.0, 1.1]	Cx	0.3540	0.3536	0.11
	Cz	0.2575	0.2580	0.19
	Cm	-2.7611	-2.7709	0.35
	Xcp	10.7230	10.7414	0.17

Table (6-2) Effect of data localization on model predictions

6.2.2 Latin Hyper-Cube Sampling

The use of parameter sampling techniques based upon the full factorial design-of-experiment can be prohibitively expensive for problems involving large numbers of parameters. To overcome this problem sampling techniques based upon Latin Hyper-Cube sampling were explored. In this approach a given number of samples are distributed in an optimal manner within the parameter space as described earlier. Refer to Vavalle [81] for a more detailed discussion. Table (6-3) presents results obtained using the Latin Hyper-Cube sampling. For this example the parameter space considered was that with an angle of attack varying between [1°, 3°], a Mach number between [5.0, 6.0] and a base flare radius r/D between 0.9 and 1.1. The parameter space was populated using only 27 snapshots. Despite the limited number of snapshots considered, the achieved accuracy is better than that obtained for the examples presented earlier which used many more or the same number of snapshots. This is evident from both response surface construction methods utilized in this section. In particular, the prediction obtained by using the polynomial augmented multi-quadric RBF is highly accurate.

		α = 2.2 M = 5.3 r/D = 1.05		
		Model	**CFD**	**% Error**
Linear Regression	**Cx**	0.3538	0.3536	0.06
	Cz	0.2577	0.2580	0.12
	Cm	-2.7647	-2.7709	0.22
	Xcp	10.7281	10.7414	0.12
Polynomial Augmented Multi-quadric RBF	**Cx**	0.3536	0.3536	0.00
	Cz	0.2581	0.2580	0.04
	Cm	-2.7722	-2.7709	0.05
	Xcp	10.7404	10.7414	0.01

Table (6-3) Comparison of model predictions with CFD computations using the Latin Hyper-Cube Sampling technique

6.2.3 Predicting the individual integrated forces by a response surface method

For this particular example a meta-model of the individual integrated forces, pitching moment and centre of pressure was also developed by generating response surfaces for each property using a linear regression method employing a second-order model. This model was used to predict the forces, moment, and centre of pressure at the previous point. A comparison with computed data is presented in Table (6-4).

		α = 2.2 M = 5.3 r/D = 1.05		
		Predicted	**CFD**	**% Error**
Linear Regression	**Cx**	0.3536	0.3536	0.00
	Cz	0.2577	0.2580	0.12
	Cm	-2.7645	-2.7709	0.23
	Xcp	10.7282	10.7414	0.12

Table (6-4) Comparison of a RSM based meta-model for the forces and moments with CFD computations

From tables (6-3) and (6-4), it is evident that the two regression methods provide results that are almost identical. While the evaluation of forces and moments from a meta-model of the individual integrated forces and moment data is less expensive than from the POD model based on the full CFD output, the resulting information is limited to

that in the model. In the case of the POD model any result that can be deduced from the CFD data can now be also deduced from the POD model. This represents a significant advantage in many practical design situations.

6.3 Fin stabilized projectile with nose control

The second problem considered in the current study relates to supersonic turbulent flow around a realistic weapon configuration which was investigated experimentally by Landers *et al.* [110] and computationally by Meunier [111] and Shoesmith *et al.* [112]. The configuration consists of a 0.7 power series nose followed by a cylindrical body with 8 stabilizing fins. The nose is free to rotate in the pitch plane, providing a means of control. Figure (6-2) presents a visualization of the geometry and a typical grid for a nose deflection of 8º. Two parameters, the nose deflection and the flow angle of attack were varied. The nose deflection was allowed to vary in the range [0º, 8º] while the flow angle of attack was varied in the range [0º, 6º]. In this case a full factorial design-of-experiment was used to sample the parameters space since the grids which were originally generated for those nose deflections considered in the physical experiment for the validation of the computational results were used. All of the computations were performed with an inflow Mach number of 3.0, a free-stream static temperature of 110 ºK and a Reynolds number of 9.5 million per foot. All these were computed in accordance with the experimental conditions of Landers *et al.* [110].

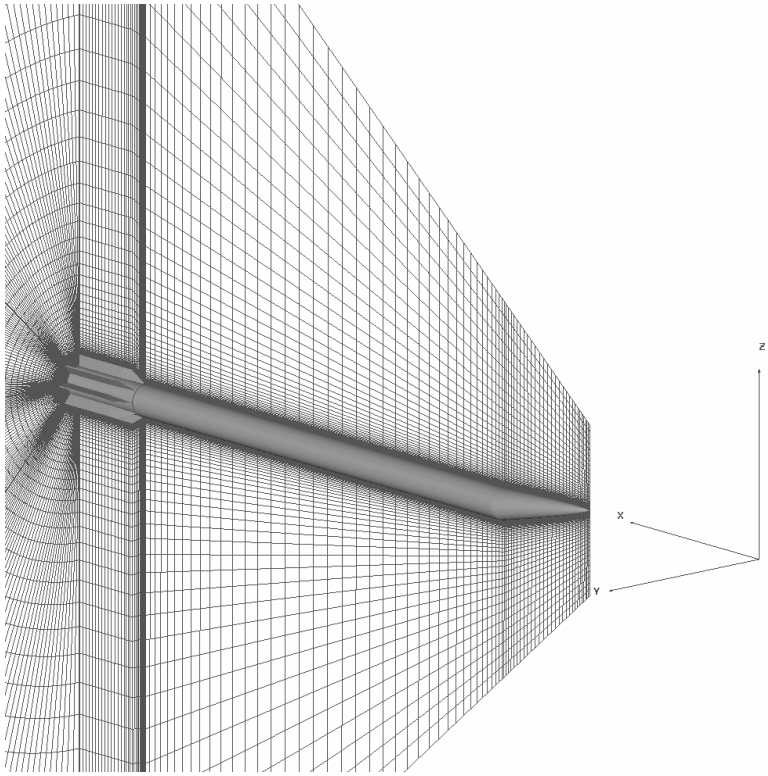


Figure (6-2) Nose controlled weapon configuration with a nose deflection of 8º and the computational grid with approx. 1 million grid points.

6.3.1 Meta-model predictions

For the first prediction a subset of the available data was employed. The nose deflection was varied in the range [2º, 8º] while angle of attack was varied between [0º, 6º], uniform intervals of 2º were used for both parameters. A proper orthogonal decomposition was performed on the ensemble of computational experiments. In this example all sixteen of the available POD modes were utilized and so this problem can be considered as a basis transformation rather than a basis reduction. Pseudo-continuous models of the projection coefficients were obtained using cubic-spline interpolation and a tri-harmonic radial basis function. These two modelling methods were used as (i) they are non-parametric and (ii) they offered good generalization ability. Predictions are compared with high-fidelity simulations in Table (6-5) for a nose deflection of 5º at an angle of attack of 1º, 3º and 5º. The comparisons are considered to be acceptable given the small number of modes available within the proper orthogonal decomposition.

		δ = 5º								
		α = 1º			α = 3º			α = 5º		
		Model	CFD	% Error	Model	CFD	% Error	Model	CFD	% Error
CSI	Cx	0.1864	0.1868	0.21	0.1904	0.1934	1.55	0.2034	0.2044	0.49
	Cz	0.1944	0.1916	1.46	0.5040	0.5048	0.16	0.9262	0.9191	0.77
	Cm	-1.497	-1.491	0.40	-5.755	-5.717	0.66	-11.320	-11.261	0.53
	Xcp	7.701	7.779	1.01	11.419	11.324	0.84	12.222	12.252	0.25
Tri-har. RBF	Cx	0.1853	0.1868	0.80	0.1860	0.1934	3.83	0.2029	0.2044	0.73
	Cz	0.1921	0.1916	0.26	0.5035	0.5048	0.26	0.9189	0.9191	0.02
	Cm	-1.451	-1.491	2.68	-5.752	-5.717	0.61	-11.243	-11.261	0.16
	Xcp	7.553	7.779	2.92	11.424	11.324	0.89	12.235	12.252	0.14

Table (6-5) Comparison of model predictions with CFD simulations for δ = 5º and α = 1º, α = 3º, α = 5º

As seen before, improving the locality of the data results in slight improved predictions, although in this case it is thought that, the errors arise principally as a consequence of grid deformation. This can be seen in Table (6-6) in which predictions obtained from an ensemble with the nose deflection varying between [4º, 6º] in intervals of 1º and an angle of attack varying between [0º, 6º] in intervals of 2º are compared with CFD simulations. In addition to the previously mentioned two interpolating techniques, a Gaussian kernel was considered in this as well. A comparison between the results obtained by the cubic splines and the Gaussian RBF shows that indeed the cubic splines provide a good generalization ability in this problem.

		δ = 5º								
		α = 1º			α = 3º			α = 5º		
		Model	CFD	% Error	Model	CFD	% Error	Model	CFD	% Error
CSI	Cx	0.1881	0.1868	0.70	0.1918	0.1934	0.83	0.2051	0.2044	0.34
	Cz	0.1941	0.1916	1.30	0.5041	0.5048	0.14	0.9278	0.9191	0.95
	Cm	-1.486	-1.491	0.30	-5.747	-5.717	0.54	-11.324	-11.261	0.56
	Xcp	7.660	7.779	1.54	11.401	11.324	0.68	12.205	12.252	0.38
Tri-harmonic RBF	Cx	0.1906	0.1868	2.03	0.1894	0.1934	2.07	0.2087	0.2044	2.10
	Cz	0.1961	0.1916	2.35	0.5033	0.5048	0.30	0.9236	0.9191	0.49
	Cm	-1.497	-1.491	0.38	-5.739	-5.717	0.38	-11.286	-11.261	0.22
	Xcp	7.630	7.779	1.92	11.403	11.324	0.70	12.2185	12.2517	0.27
Gaussian RBF	Cx	0.1898	0.1868	1.61	0.1911	0.1934	1.19	0.2058	0.2044	0.68
	Cz	0.1923	0.1916	0.37	0.5043	0.5048	0.10	0.9212	0.9191	0.23
	Cm	-1.431	-1.491	4.00	-5.750	-5.717	0.59	-11.282	-11.261	0.19
	Xcp	7.441	7.780	4.37	11.402	11.324	0.68	12.247	12.252	0.04

Table (6-6) Comparison of model predictions with CFD simulations due to improved data locality, δ = 5º and α = 1º, α = 3º, α = 5º

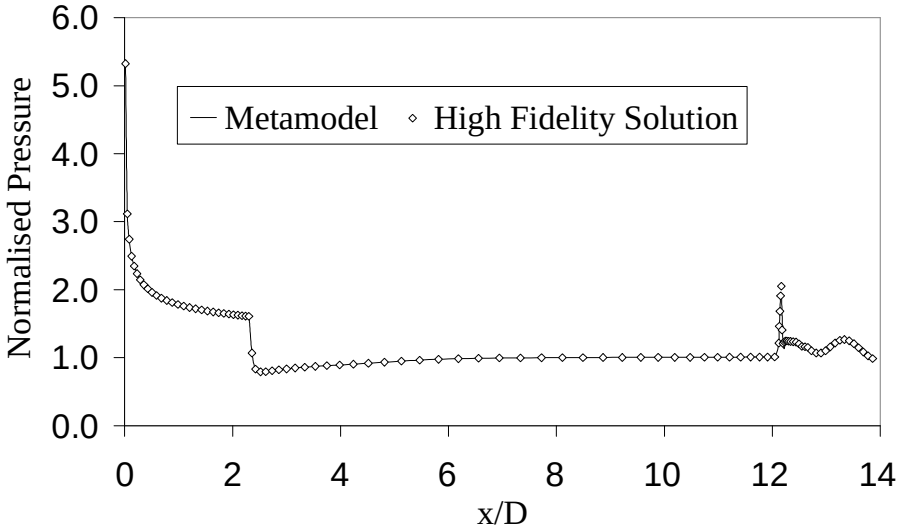
Improving the resolution of incidence α from 2º to 1º intervals provides further improvements in the predicted data, Table (6-7). Tables (6-5), (6-6) and (6-7) show that the cubic-spline interpolation and the tri-harmonic RBF are of comparable accuracy as expected since the tri-harmonic radial basis function is in fact a cubic-spline. These results suggest that by using suitable kernels in a generalized linear model it is possible to approximate any function to an acceptable degree of accuracy [87].

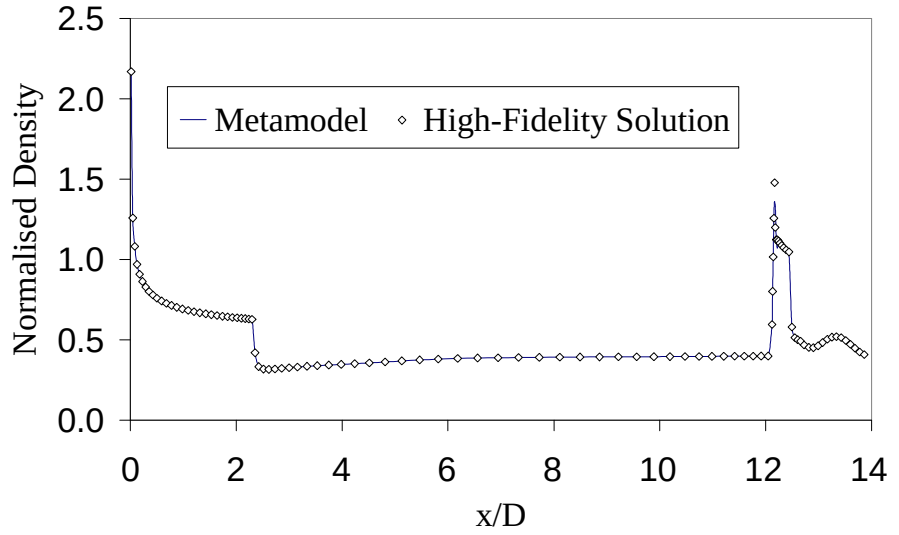
		α = 2.5º δ= 5º		
		Model	**CFD**	**% Error**
CSI	**Cx**	0.1901	0.1904	0.16
	Cz	0.4160	0.4162	0.05
	Cm	-4.5013	-4.5072	0.13
	Xcp	10.8197	10.8293	0.09
Tri-harmonic RBF	**Cx**	0.1901	0.1904	0.16
	Cz	0.4159	0.4162	0.07
	Cm	-4.5003	-4.5072	0.15
	Xcp	10.8212	10.8293	0.07
Gaussian RBF	**Cx**	0.1901	0.1904	0.16
	Cz	0.4161	0.4162	0.02
	Cm	-4.5074	-4.5072	0.00
	Xcp	10.8316	10.8293	0.02

Table (6-7) Comparison of model predictions with CFD simulations, δ = 5º and α = 2.5º

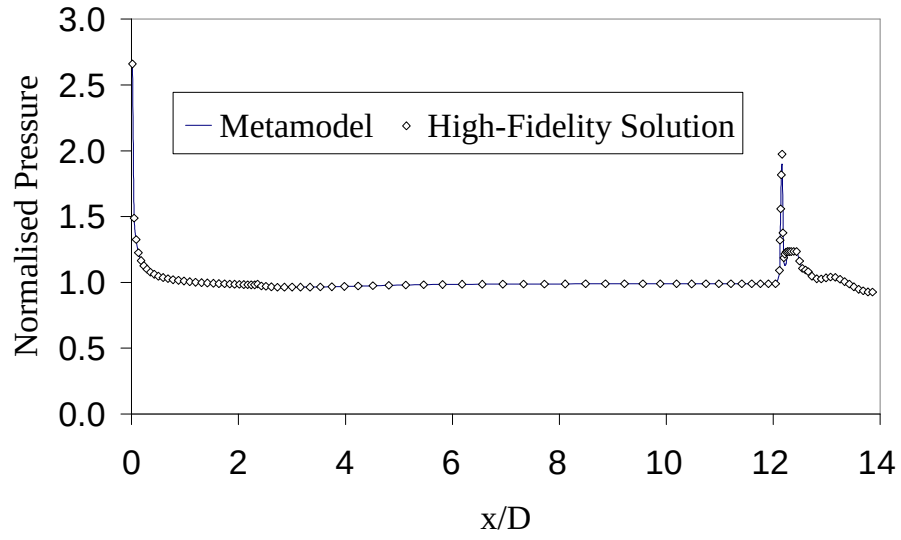
Figure (6-3) compares the axial pressure and density distributions obtained from the meta-model using cubic-splines interpolation with CFD simulations at 0º and 180º azimuth angles. The agreement is generally excellent with the exception of the shock intensity at the fin. Figure (6-4) compares the circumferential pressure distributions at stations located 1, 3 and 10.5 calibres downstream of the nose. The agreement is generally good over most of the body, but some small discrepancies are observed in the secondary flow separation and re-attachment at 10.5 calibres (Figure 6-4 (d)). The prediction is improved at the re-attachment by using Gaussian radial basis functions for interpolation (Figure 6-4 (e)) and so this indicates that the discrepancies result from the pseudo-continuous representation of the scalar coefficients. Predicted density and pressure contours compared with CFD simulations at two stream-wise locations are presented in figure (6-5). The first is located well downstream of the nose cylinder junction while the second is located at the trailing edge of the fins. This figure, together with Table (6-7) and figures (6-3) and (6-4) suggests that the meta-model provides an

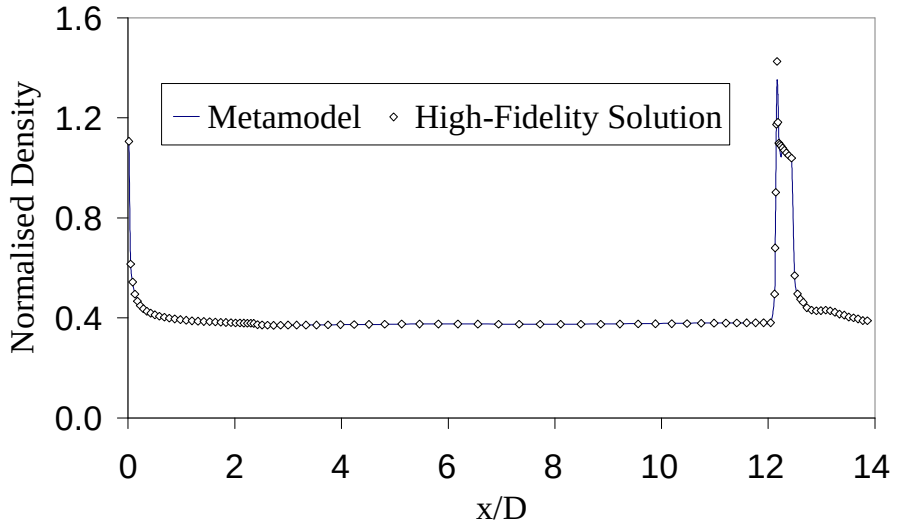
acceptable surrogate for the IMPNS flow solver and illustrates the potential of the method.



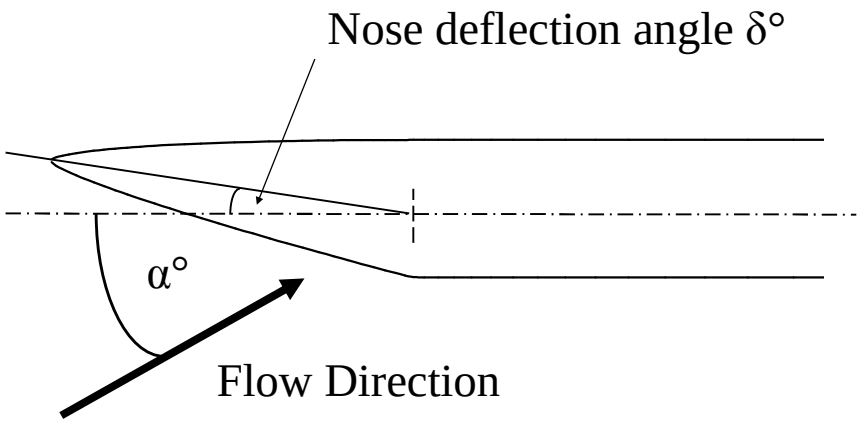


(a) Windward


Normalised Pressure
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Metamodel
High-Fidelity Solution
0
2
4
6
8
10
12
14
x/D

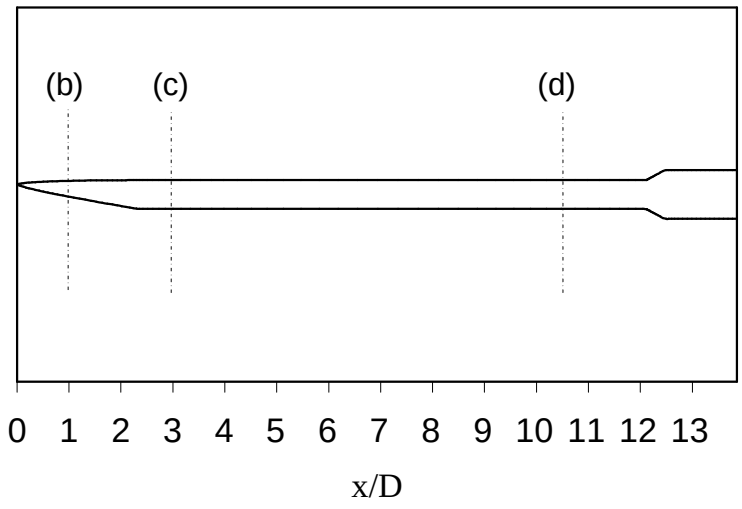

Normalised Density
1.6
1.2
0.8
0.4
0.0
Metamodel
High-Fidelity Solution
0
2
4
6
8
10
12
14
x/D

(b) Leeward

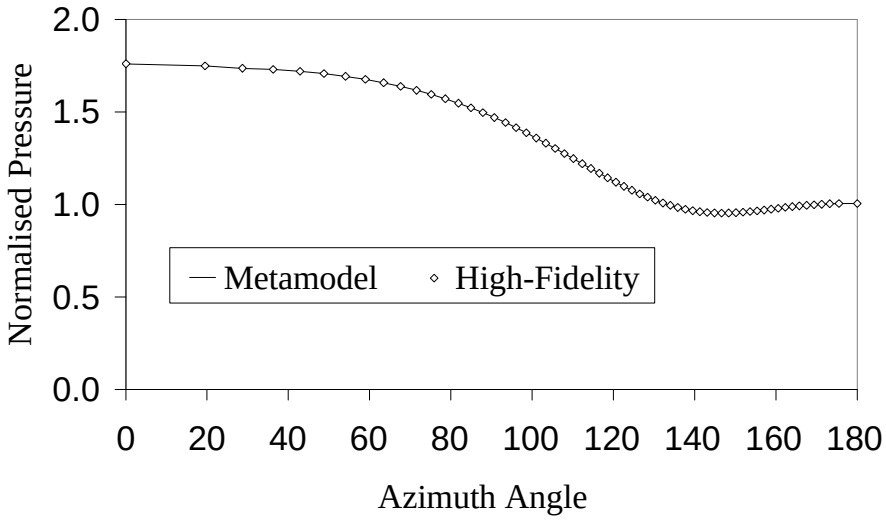


(c) Angles definition

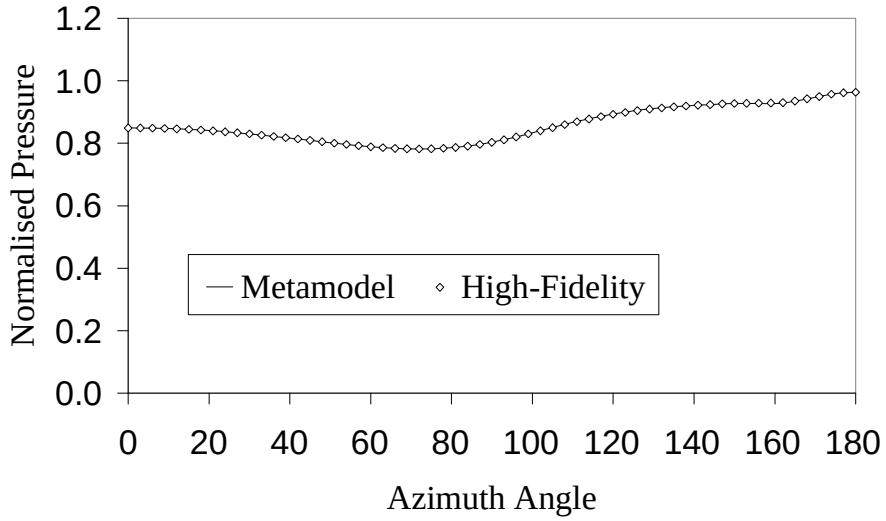
Figure (6-3) (a) and (b) Comparison of predicted axial distribution of pressure and density with CFD simulations, δ = 5º and α = 2.5º (c) angles definition



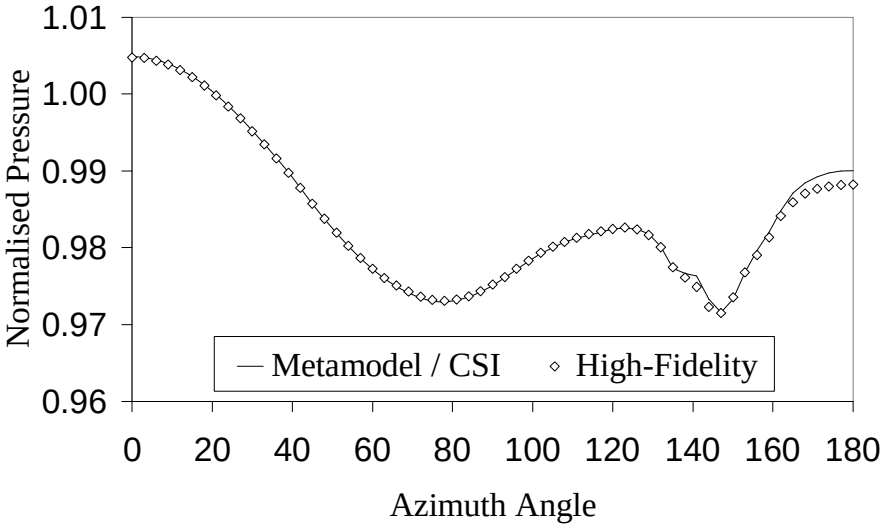
(a)


2.0
1.5
1.0
0.5
0.0
Normalised Pressure
Metamodel
High-Fidelity
0
20
40
60
80
100
120
140
160
180
Azimuth Angle

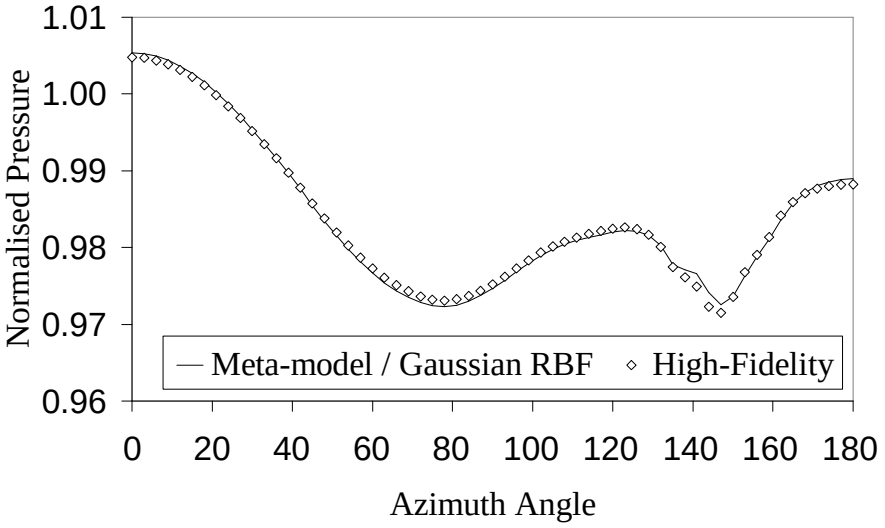
(b) x/D = 1.0


1.2
1.0
0.8
0.6
0.4
0.2
0.0
Normalised Pressure
Metamodel
High-Fidelity
0
20
40
60
80
100
120
140
160
180
Azimuth Angle

(c) x/D = 3.0

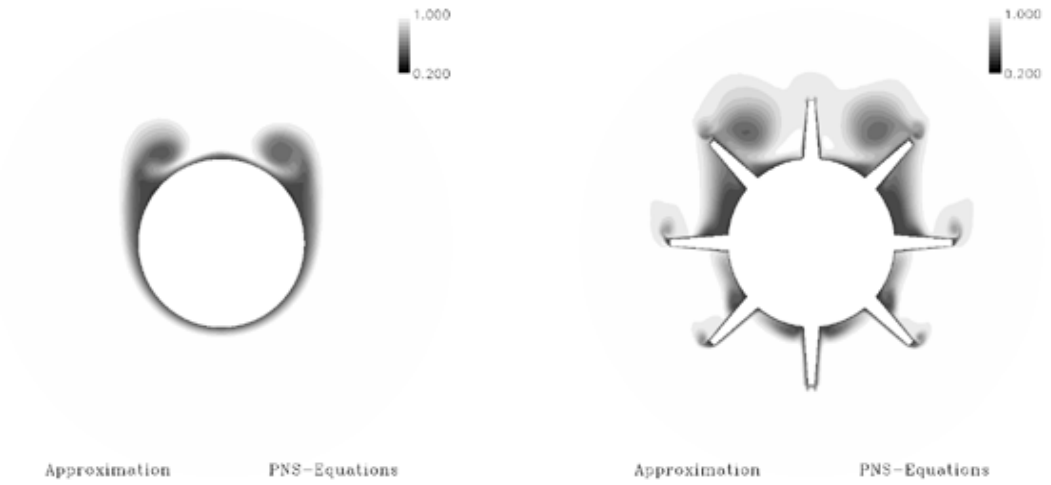


(d) x/D = 10.5



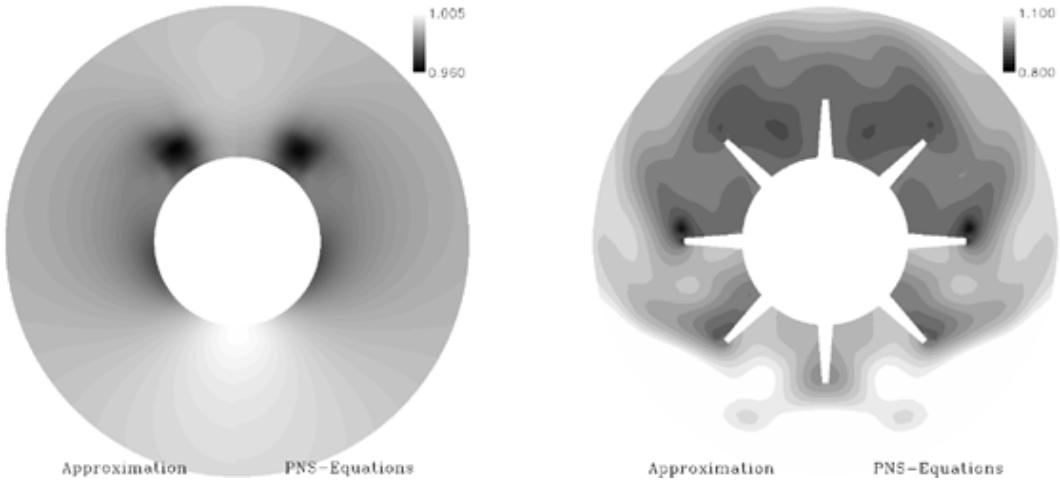
(e) x/D = 10.5

Figure (6-4) Circumferential pressure distributions, δ = 5º and α = 2.5º



x/D = 10. 5 **x/D = xmax**

(a) Density contours (left – meta-model, right – Navier-Stokes)



x/D = 10. 5 **x/D = xmax**

(b) Pressure Contours (left – meta-model, right – Navier-Stokes)

Figure (6-5) Comparison of pressure and density contours, δ = 5º and α = 2.5º

6.3.2 Computational time

In this particular example the full CFD output from the meta-model was obtained in 107 seconds of CPU time on a 3.3 GHz processor. This compares favourably to the 18,000 seconds of CPU time necessary for making one complete evaluation ie. one single computation of a snapshot using the CFD code. It must be emphasized that in this case, a space-marching code was used. In case the CFD code is a time marching one, the difference in CPU time would be considerably higher provided both are at par, for example both are parallelised. Consequently the meta-model becomes much more useful. In such case the output from the meta-model could also be used to restart the computation and hence it would assist in further reducing the overall computational time of the CFD code.

6.4 Comparison between the meta-model and the semi-empirical modelling tools

One possible application of this meta-modelling tool is as an alternative to the existing semi-empirical modelling tools, which tools are frequently used in industry to conduct trade and trajectory studies most especially. This section is aimed at making an objective comparison between these two different computational tools by exploring the relative advantages and disadvantages.

6.4.1 The semi-empirical aerodynamic codes

The U.S. Air Force Missile DATCOM and the Naval Surface Warfare Center Aeroprediction codes are two semi-empirical aerodynamic prediction codes that calculate aerodynamic forces, moments and stability derivatives as a function of the angle of attack and Mach number for a variety of axi-symmetric and non-axi-symmetric missile configurations. Both codes offer the facility to predict the static pressure and interference

factors as well as the capability for the user to easily change aerodynamic parameters to fit specific applications for a broad range of flight conditions. Flight conditions and aerodynamic parameters range from subsonic to hypersonic speeds, angles of attack up to 90 degrees and control deflections from -35 to +35 degrees. The output gives the aerodynamic forces and moments in addition to the location of the centre-of-pressure, interference factors and geometric data.

In particular, Missile DATCOM has trim capabilities and the ability to numerically model a configuration by inserting experimental data. DATCOM also has the ability to model aerofoils including both user defined types and NACA ones. In addition, it has the capability to develop aerodynamic data for air-breathing systems. The Aeroprediction code features include a plotting program for aerodynamic coefficients as well as a geometric sketch of the input configuration. The Aeroprediction code has a better axial force coefficient prediction capability than DATCOM and produces both structural loading and aerothermal output [114] [115].

Moore & Hymer [113] [114] claim that the AeroPrediction code predicts the static aerodynamic coefficients for missile configurations with flares and wings/tails within an average accuracy of +/- 10 % for axial and normal force, and +/- 4 % of the body length for the centre of pressure. The average accuracy for the pitch damping coefficients is within +/- 20 %. By average accuracy it is meant that enough AoAs or Mach numbers are considered to get a good statistical sample. In [115], Schmidt and Sooy compare AP98 and DATCOM 97 for various configurations. For these configurations it was reported that the predicted normal force for both codes have minimal error. The axial force is more difficult to predict and results within an error of +/- 10 % for the AP code and within +/- 12 % for DATCOM were achieved.

6.4.2 Trajectory studies

The evaluation of the flight performance of a missile or a projectile is typically a two step iterative process. In the first step, the aerodynamic coefficients for the airframe are

determined over the flight envelope. This may be accomplished through either experiment or by using some software which predicts the aerodynamics. Secondly, the aerodynamic coefficients are input into a trajectory model so that the aerodynamic forces acting upon the missile may be determined for any flight condition. The trajectory model is then executed and the results are analysed. If the flight performance as predicted by the trajectory model is not satisfactory, changes are made to the airframe. These changes will in turn affect the aerodynamics, thereby requiring a new set of aerodynamic data to be inserted into the trajectory model. This process continues until an airframe is found which optimizes some desired aspect of the flight performance.

Depending on the flight regime over which the aerodynamics are to be computed, a set of trim aerodynamics is normally generated in a matter of minutes using the AeroPrediction code version 2005 [116]. The term 'trim' implies that the aerodynamic coefficients correspond to a state in which the pitching moment coefficient C_m is equal to zero. These aerodynamic data is then input into the flight dynamics model for the optimization of the desired aspect with the actual amount of time it takes depends on the amount of aerodynamic data to be inserted as well as the experience level of the engineer.

The process featured in the preceding paragraph is for one iteration only. If the performance of the concept is adequate, this would probably complete the initial phase of the design, including the aerodynamic and performance assessment of the concept. However, in most cases, several design iterations are required to investigate the effect on the aerodynamics and performance of each design change. By combining the Aeroprediction code with the trajectory models for automatic trajectory generation of a given design concept, a large cost saving was reported in [114], [116].

Thus, it can be concluded that semi-empirical tools are highly efficient though not so accurate and the output information from these codes is limited. In the next section a comparison is made between the semi-empirical tools and the meta-modelling tool.

6.4.3 The meta-modelling tool versus the semi-empirical tool

From the previous section it is apparent that the semi-empirical tool is amply more efficient than the developed computational tool. This is being remarked on the premise that a whole trajectory study was conducted in 1 hour, which included 56 trim point evaluations. Possibly, comparable on-line computational time may be achieved between the two computational methods if faster algorithms are implemented in the meta-model which include also the parallelisation of the developed code. However, if the off-line computational time consumed in generating the snapshots for the meta-model is taken into consideration, there would be no comparison in time between the two methods.

Notwithstanding this drawback, the meta-model does offer significant advantages over the semi-empirical tools. Provided the reduced-basis is rich enough within the parametric space of interest and a suitable interpolation method is adopted, an accuracy close to the high-fidelity CFD solutions is normally attained. However, the accuracy of the meta-model will deteriorate as the reduced-basis is leaned. Moreover, the meta-model offers a complete solution for the whole flow-field just like in CFD while the semi-empirical methods provide only those properties defined in the tool. Furthermore, the meta-modelling tool has the capability of modelling various complex flow features and structures encountered in weapon aerodynamics such as

- flow separation and re-attachment with the subsequent heat flux generation,
- flow interaction between the main body flows and other parts such as wings and fins with the subsequent effects on control,
- flow interaction between multi-bodies,
- flow over stores in open bays and stores release etc.
- flow over bodies with any geometrical shape of interest including protuberances without limitations.

Finally, it must be remarked that the meta-modelling tool offers a more general computational tool whereas various kinds of problems can be simulated and predicted.

This makes it much more powerful than any other semi-empirical computational tool which caters for particular geometries only.

6.5 Concluding remarks

In this chapter a high-fidelity, low-cost aerodynamic model was presented for use in parametric studies of weapon aerodynamics. The model has been applied to two flow problems related to high-speed weapon aerodynamics; inviscid flow around a flare stabilized hypersonic projectile and supersonic turbulent flow around a fin stabilized projectile with drooping nose control.

The method employs a reduced-order model obtained from the proper orthogonal decomposition of an ensemble of CFD solutions. This decomposition produces an optimal linear set of orthogonal basis functions that best describe the ensemble of numerical solutions. These solutions are then projected onto this set of basis functions to provide a finite set of scalar coefficients that represent the solutions. A pseudo-continuous representation of these projection coefficients is constructed for each mode which allows predictions to be made of parameter combinations not in the original set of observations. It has been shown that by using generalized linear models for the pseudo-continuous representation of the projection coefficients describing the reduced-order model, predictions of parameter combinations not in the original set of observations are made efficiently.

Four different approaches to the construction of a response surface of the projection coefficients were investigated, a linear regression based method and three interpolation techniques employing linear spline, cubic spline and radial basis functions. The computed data suggest that interpolation based techniques provide a significant advantage over the regression method. This is attributed to the fact that the interpolation schemes pass through all of the sample points providing an improved representation of local minima

and maxima, while the global fit of the regression technique produces some unnecessary smoothing. This observation is supported by studies of data localization that suggest the regression method can provide similar accuracy to the interpolation schemes when employed over a reduced parameter space. In general, the radial basis function interpolation offered the most accurate prediction.

The use of Latin Hyper-Cube sampling methods was found to offer improved accuracy for a given number of sample points. For problems involving large numbers of parameters the LHCS may provide a practical approach to reduce the number of samples required to populate the design space. However, for problems involving geometric variation the LHCS requires a means of automatically generating high-quality grids. For this reason a more practical tool may incorporate a hybrid approach, a design-of-experiment technique in which the geometry variables are prescribed at specific levels and an LHCS technique which is used for other parameters.

The results of this study suggest that meta-models based upon POD of a small number of computational experiments together with response surface methods can provide a reliable low-cost high-fidelity prediction tool. In contrast to many of the modelling efforts reported in the literature, this model provides access to information about the whole flow-field. The current approach reduces the computational time by more than two orders of magnitude and potentially a further reduction can be made by implementing faster algorithms and parallelising various parts of the code.

Chapter 7

7.0 A hierarchical type model using POD-based ROMs

A significant disadvantage of POD-based reduced-order modelling is the up-front cost necessary to generate the dataset of observations from which subsequently an adequately rich set of basis and hence the surrogate model, are developed. One may even comment that the on-line computational efficiency of the POD-based ROM is offset to a large extent by the off-line computational time necessary to generate the original ensemble of data. As Beran and Silva remarked in [23], the trade-off can be favourable only when, after the initial computational investment, a compact ROM is constructed which can be used many times, and is valid over a useful range of parametric values. Consequently, some method which restricts the time duration in generating this training dataset without much loss of accuracy is of interest. In an effort to limit this initial computational encumbrance, a hierarchical type model which incorporates POD-based ROMs is proposed in this chapter so as to manage and control the up-front cost involved. Furthermore, this model provides a means for fusing computational data of variable-complexity while yielding solutions with the complete flow-field.

7.1 Some hierarchical type modelling concepts

The hierarchical type modelling concept was introduced by Alexandrov *et al.* [117] [118] [119] [120] [121] from the NASA Langley Research Center for solving

optimisation problems using an engineering approach rather than a rigorous mathematical one. In general, hierarchical type models (which are also referred to as multi-fidelity, variable-fidelity and variable-complexity) combine inexpensive low-fidelity analyses with more accurate but expensive high-fidelity solutions. This combination is sought as the evaluation of high-fidelity flow simulations can be very expensive and therefore it is of great interest to devise methodologies, most especially in design and analysis, involving as few high-fidelity analyses as possible. The hierarchical type modelling strategy generally uses multi-dimensional response surface technology to model the different fidelities. In this concept, a model of lower-fidelity such as coarser discretization, relaxed convergence tolerances, lower order of accuracy and omitted physics are used as the surrogate in place of the high-fidelity model. As a result, the original implicit problem is replaced by explicit approximations thus offering means by which the computational cost can be considerably reduced.

There are various strategies for reducing the number of expensive high-fidelity analyses. One strategy entails using low-fidelity models to reduce the region in the design space and once this is reduced a high-fidelity response surface is constructed over this reduced space [122]. Another possibility is to use low-fidelity models for identifying unimportant response surface terms or for identifying proper intervening variables that reduce the problem dimensionality [122].

Another strategy was proposed by Haftka [123]. In his work Haftka introduced the notion of employing a linearly varying scaling factor between models of variable-fidelity. Since then the concept of correction response surfaces has been applied by various people. In this, a high number of points are selected for the relatively inexpensive low-fidelity analyses and from these points a subset is chosen for high-fidelity analyses. The low-fidelity results are used to fit a response surface while the high-fidelity analyses are used to fit a linear correction response surface. For the common i^{th} design point considered between the two models, the ratio of evaluated response values γ is evaluated as follows:

$$\gamma_i = \left(\frac{y_{hf}}{y_{lf}} \right)_i \qquad \text{Eq. (7-1)}$$

The correction response surface $\hat{\gamma}(\mathbf{x})$ can then be used for establishing the following variable-fidelity approximation, which may then be used for various purposes.

$$\hat{y}_{vf}(\mathbf{x}) = \hat{y}_{lf}(\mathbf{x}).\hat{\gamma}(\mathbf{x}) \qquad \text{Eq. (7-2)}$$

Another approach which is related to this consists in using additive corrections, that is approximating the difference between low- and high-fidelity models and then adding this correction to the low-fidelity response surface constructed. A comparison between multiplicative and additive correction response surfaces was compiled by Toropov and Markine [124] who suggested that the former method leads to better approximations.

A different approach which utilises radial basis functions to fuse integrated experimental and computational data was proposed quite recently by Reisenthel *et al.* [125]. In the following section a review of Reisenthel's variable-fidelity model is made followed by a comprehensive study to understand its behaviour when using deterministic data such as a computational one.

7.2 Fusion between experimental and computational data

Validation of computational databases by experimental tests is widespread and of significant importance to develop a good level of confidence in the computed results. In general, the integration of experimental and computational data is essential to support decisions during the development phase of aerodynamic devices. An innovative data integration technique to fuse experimental and computational results has been proposed by Reisenthel *et al.* [125]. In this technique, instead of using the experimental data to validate the computational data directly, the point of view adopted is to recognize and accept that there will always be some differences between these two streams of data,

mainly due to approximations in the physical models used, as well as experimental limitations. Li *et al.* [126] defined data fusion as “the combination of a group of inputs with the objective of producing a single output of greater quality and reliability”. With this general notion in mind, Reisenthel *et al*. fused the two sets of data together to assist in developing a global understanding of it. In that work, a typical situation was considered where sparse experimental data is obtained from very expensive wind-tunnel tests at a limited number of configurations and conditions along with a large quantity of computational data which is relatively more affordable. A response surface methodology has been used to perform data fusion and subsequently enhance the effectiveness of the limited experimental test data with the aid of physics–based computational information. In addition, an important application of this technology is the use of sparse data points from limited wind tunnel tests to correct and fine tune or better still calibrate computational databases. Reisenthel *et al.* demonstrated the utility of this technique.

In this technique a dimensionality augmented response surface is developed which functions as a global interpolant for a heterogeneous set of data incorporating both computational and experimental information. The modelling of the experimental and computational data is achieved by adding an additional variable to the N dimensional problem. This auxiliary variable is binary in nature and is practically used to identify whether the data is either computational or else experimental. A single global response surface is then generated in the N+1 dimensions. Reisenthel *et al.* did not specifically state what type of RBF was used in his work but from the explanation conducted in his presentation it is evident that a local RBF similar to a Gaussian or an inverse multi-quadric was considered. By interrogating the response surface projected along the experimental data, a model is established which represents the correctness of the experimental data while at the same time inherits the essential features of the computational model.

As an illustration of this technique, Reisenthel *et al.* presented a case for the correction of computational aerodynamic databases using sparse experimental data of a generic body-tail missile configuration. Computational databases for integrated forces and

moments which varied over a wide range of angles of attack, roll angles and Mach numbers were considered. From this methodology an error database was developed.

In this research a simplified model of Reisenthel *et al.* was developed to study and understand his model. The simplification was achieved by assuming that the variable-fidelity data is all obtained deterministically. The following is an academic example which was considered and studied in this work.

7.3 An academic example showing a simplified Reisenthel's model

In this case two exponentially decaying sinusoidal sets of data were generated with relative phase shift, different amplitude and decay rate. In this example it is assumed that one set of data represents the low-fidelity training data while the other set of data represents the high-fidelity one. Figure (7-1) shows the two sets of data. The low-fidelity training data was calculated according to the following:

$$f(\theta) = 0.04\exp^{(-0.15\theta)}\, Sin(-2\theta) - 0.015$$

where θ is in radians. The high-fidelity data was generated with a slightly different amplitude and phase shift. Such data may be considered as a close representation of the static rolling moment of a body-tail missile.

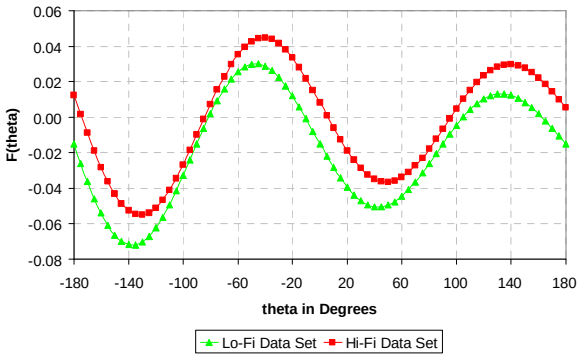


Figure (7-1) – Exponentially decaying sinusoidal sets of data with relative phase shift

Using the model described in the previous section with two variables namely, the angle of rotation in degrees and the auxiliary variable which denotes whether the training data is of low- or high-fidelity, a global response surface is generated using local radial basis functions. In this case Gaussian radial basis functions were considered with a shape parameter set at a constant value of 0.03. This value was arrived at after a few trials to set the most appropriate value using the low-fidelity training points. Thus, the equation used was $\Psi(r) = e^{-\frac{r^2}{0.03}}$.

Three predictions were made using all the low-fidelity data together with three or four high-fidelity data points. Figures (7-2) and (7-3) show the predictions obtained using this model with three and four high-fidelity data points respectively. In this case the high-fidelity points chosen correspond to the points at which the low-fidelity data curve reaches its peak value.

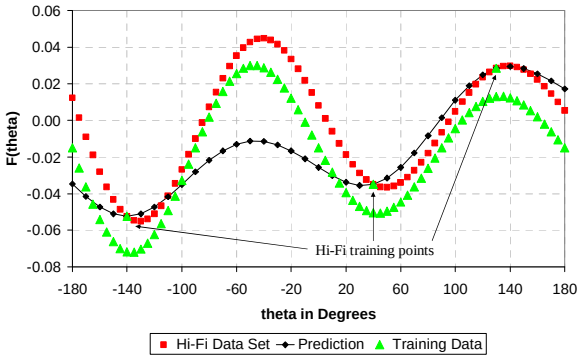


Figure (7-2) – Prediction using three high-fidelity data points

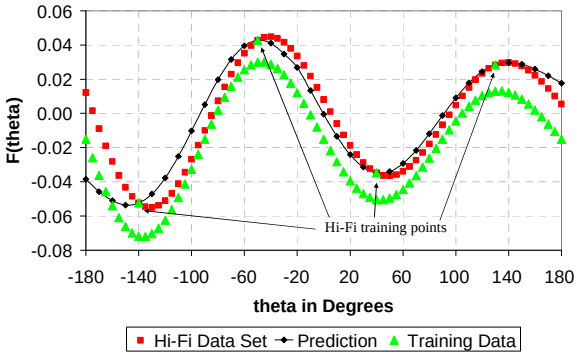


Figure (7-3) – Prediction using four high-fidelity data points

The prediction is considerably improved by considering the four high-fidelity training points very close to those points at which the high-fidelity data curve reaches its peak values. See figure (7-4). The occurrence of this feature is due to the phase shift between the two sets of data. In particular this feature emphasizes that the low-fidelity set of data must follow the same trend as the high-fidelity set of data for effective data fusion when using sparse high-fidelity data points. Otherwise, this can be compensated by considering a greater number of high-fidelity data points.

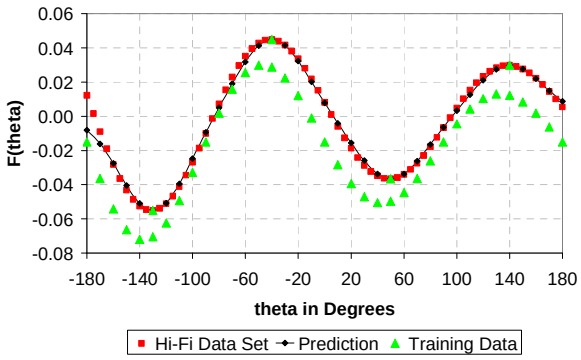


Figure (7-4) – Prediction using four high-fidelity data points close to peak values

In general, this technique offers an automatic way of fusing variable-fidelity data through the construction of interpolation and extrapolation schemes.

7.4 Fusion between experimental and computational data for the nose -controlled finned missile configuration

In this section an example of Reisenthel's variable-fidelity method is presented for a simple but realistic aerodynamic problem. In this example experimental and computational data are fused together for the nose-controlled finned missile configuration considered earlier, that is, with a nose deflection angle of 8º. The data considered is the integrated axial force coefficient as the greatest discrepancies between the experimental and computational data were recorded in this particular force coefficient. The force coefficients obtained experimentally for the nose-controlled finned missile configuration are used together with the integrated force coefficients obtained by computation using the

IMPNS software to populate a sub-space at two different fidelity levels. Multi-quadric radial basis functions (i.e. global RBFs) are used in this example to develop the pseudo-continuous representation through which high-fidelity predictions (based on the experimental results) are made. This differs from the work of Reisenthel *et al.* where local type RBFs were considered. It must be pointed out that in this section the high-fidelity training data is considered as a deterministic set of data as it was interpolated rather than regressed. Strictly, experimental data should be regressed.

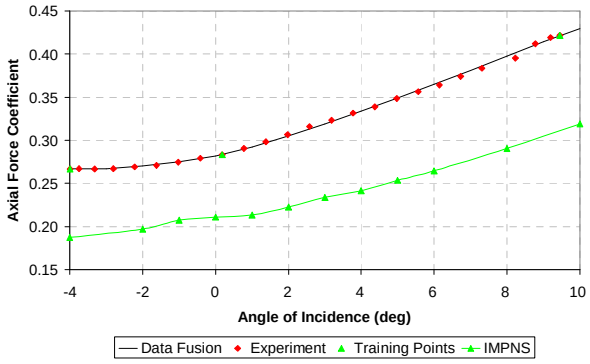


Figure (7-5) – Axial force coefficient prediction using three high-fidelity data points

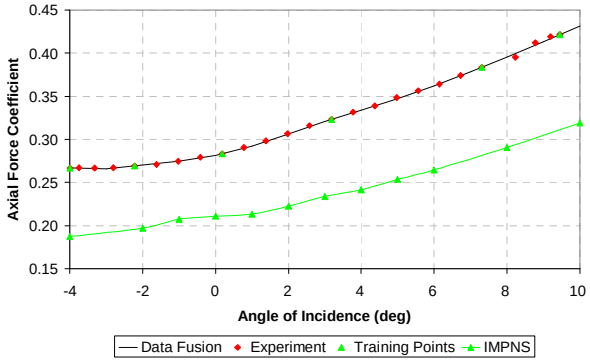


Figure (7-6) – Axial force coefficient prediction using six high-fidelity data points

Figures (7-5) and (7-6) show the predictions with a fixed number of low-fidelity (computational) data and with either three or six high-fidelity (experimental) training points. As expected, some slight improvement in the representation of the high-fidelity data is attained with an increasing number of training points as the predicted values follow closer to experiment. Interestingly, the global radial basis functions introduce

some favourable smoothing in the predictions as it can be observed from figures (7-5) and (7-6). In these figures the highlighted green data points are the training points while the red points are the actual experimental data with which the prediction is compared. In getting these responses from the data fusion technique, a numerical weight was introduced between the two levels so as to control the influence of the training dataset onto the prediction. In this analysis it has been observed that as the number of high-fidelity training points is increased, the weighting factor is reduced in magnitude. Figure (7-7) shows the Reisenthel's model response with three high-fidelity training data points corresponding with figure (7-5) and with different weighting factors.

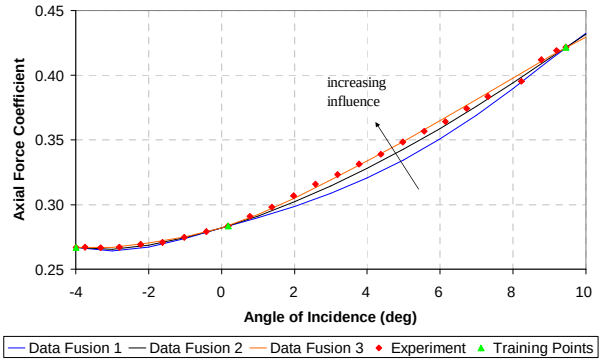


Figure (7-7) – Response from varying weight factors

From this section it can be concluded that variable-fidelity data fusion can be achieved by global radial basis functions as long as a faithful representation of the low-fidelity data is obtained by the radial basis function considered. In this case the response was relatively simple.

7.5 A variable-fidelity model incorporating POD-based ROMs

In this section POD-based ROMs created with snapshots at different fidelities are integrated together through the use of the variable-fidelity model of Reisenthel *et al.* with the objective of developing a new method by which variable-fidelity computational data is fused together and tuned high-fidelity predictions are made. This modified model has

the major advantage that while Reisenthel's model can be applied over individual or integrated properties at different fidelities, **this model can be applied over the whole CFD solutions at different fidelities such that the complete flow-field over the whole domain is output, just like the solution from a CFD calculation.** To the best of the author's knowledge, this is a new method which is being proposed for the first time in this thesis and is a contribution towards the body of knowledge.

Considering an ordered ensemble of heterogeneous data $\mathbf{A}$ where $\mathbf{A} \in \mathbf{R}^{mxn}$ is obtained from the solution vectors of low- and high-fidelity models at various parameter values with the total number of realizations or parameters combination $m = m_1 + m_2$, where m_1 is the number of snapshots obtained from the high-fidelity model and m_2 is the number of snapshots obtained from the low-fidelity model. It is assumed that $m_1 << m_2$ and *n* is the number of grid points over which the computational calculation is performed.

Let $\mathbf{A} \in \mathbf{R}^{mxn}$ denote the matrix whose rows are the modified snapshots as stated earlier in chapter 4 where the primed entries denote the high-fidelity solution vectors, from row 1 to row m_1 while the non-primed entries represent the low-fidelity solution vectors, from row $m_1 + 1$ to m where $m = m_1 + m_2$. In this case the parameters combination between the low and high-fidelity model solutions can be different and it is not necessary to have common snapshots between the variable-fidelities.

$$\mathbf{A} = \begin{pmatrix} a_1'^{(1)} & \cdot & \cdot & \cdot & a_n'^{(1)} \\ \cdot & & \cdot & \cdot & \cdot \\ a_1'^{(m_1)} & \cdot & \cdot & \cdot & a_n'^{(m_1)} \\ a_1^{(m_1+1)} & \cdot & \cdot & \cdot & a_n^{(m_1+1)} \\ \cdot & & \cdot & \cdot & \cdot \\ a_1^{(m_1+m_2)} & \cdot & \cdot & \cdot & a_n^{(m_1+m_2)} \end{pmatrix} \qquad \text{Eq. (7-3)}$$

The SVD of $\mathbf{A}$ which is equal to $\mathbf{U\Sigma V^T}$ can be written in reduced form as

$$\mathbf{A} = \begin{pmatrix} u'_{11} & . & u'_{1m_1} & u_{1(m_1+1)} & . & u_{1m} \\ . & . & . & . & . & . \\ u'_{m_1 1} & . & u'_{m_1 m_1} & u_{m_1(m_1+1)} & . & u_{m_1 m} \\ u'_{(m_1+1)1} & . & u'_{(m_1+1)m_1} & u_{(m_1+1)(m_1+1)} & . & u_{(m_1+1)m} \\ . & . & . & . & . & . \\ u'_{mm} & . & u'_{mm_1} & u_{m(m_1+1)} & . & u_{mm} \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & . & 0 & 0 & 0 & 0 \\ 0 & 0 & \sigma_{m_1} & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{m_1+1} & 0 & 0 \\ 0 & 0 & 0 & 0 & . & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_m \end{pmatrix} \begin{pmatrix} v'_{11} & . & v'_{1n} \\ . & . & . \\ v'_{m_1 1} & . & v'_{m_1 n} \\ v_{(m_1+1)1} & . & v_{(m_1+1)n} \\ . & . & . \\ v_{m1} & . & v_{mn} \end{pmatrix}$$

The rows of $\mathbf{V^T}$ are the proper orthogonal modes of the system. Multiplying the first two matrices on the RHS to obtain the projection coefficient matrix $[\alpha_{ij}]$

$$\mathbf{A} = \begin{pmatrix} \alpha'_{11} & . & \alpha'_{1m_1} & \alpha_{1(m_1+1)} & . & \alpha_{1m} \\ . & . & . & . & . & . \\ \alpha'_{m_1 1} & . & \alpha'_{m_1 m_1} & \alpha_{m_1(m_1+1)} & . & \alpha_{m_1 m} \\ \alpha'_{(m_1+1)1} & . & \alpha'_{(m_1+1)m_1} & \alpha_{(m_1+1)(m_1+1)} & . & \alpha_{(m_1+1)m} \\ . & . & . & . & . & . \\ \alpha'_{mm} & . & \alpha'_{mm_1} & \alpha_{m(m_1+1)} & . & \alpha_{mm} \end{pmatrix} . \begin{pmatrix} v'_{11} & . & v'_{1n} \\ . & . & . \\ v'_{m_1 1} & . & v'_{m_1 n} \\ v_{(m_1+1)1} & . & v_{(m_1+1)n} \\ . & . & . \\ v_{m1} & . & v_{mn} \end{pmatrix} \qquad \text{Eq. (7-4)}$$

From this matrix, the set of scalar coefficients $\alpha_{i,\, j=const}$ are considered in turn for j varying from 1 *to m* to develop multi-dimensional response surfaces for variable-fidelity data fusion. Note that the high-fidelity response surface is formed by the projection coefficients $\alpha_{i,\,j}$ for $1 \le j \le m_1$ while the low-fidelity response surface is formed by the projection coefficients $\alpha_{i,\,j}$ for $m_1 + 1 \le j \le m$. This is achieved by the introduction of an auxiliary variable $\varepsilon \equiv x_{N+1}$ to the N-dimensional problem with variables $(x_1, x_2, ..., x_N)$. This auxiliary variable simply denotes whether the data is of low-fidelity ($\varepsilon = 0$) or high-fidelity ($\varepsilon = 1$). From this a global response surface is then computed in the N+1 dimensions using radial basis functions. By interrogating the newly developed response surface projected along $\varepsilon = 1$ at any arbitrary parameter value not in the original set of observations, predictions can be made. The resulting model representation respects the

accuracy of the high-fidelity data while following the features of the low-fidelity data. In this model the introduction of the low-fidelity sub-space by making use of an additional variable allows us to perform an interpolation-based extrapolation. In other words, an extrapolation at high-fidelity based on interpolation at low-fidelity is conducted. Figure (7-8) shows a schematic diagram of the model aforementioned. It is to be noted that the influence of the low-fidelity data on the high-fidelity one can be controlled by a weighting factor. This weighting factor has the analogous effect of moving the two datasets closer together in which case ε is no longer equal to 1, ie. ($\varepsilon < 1$).

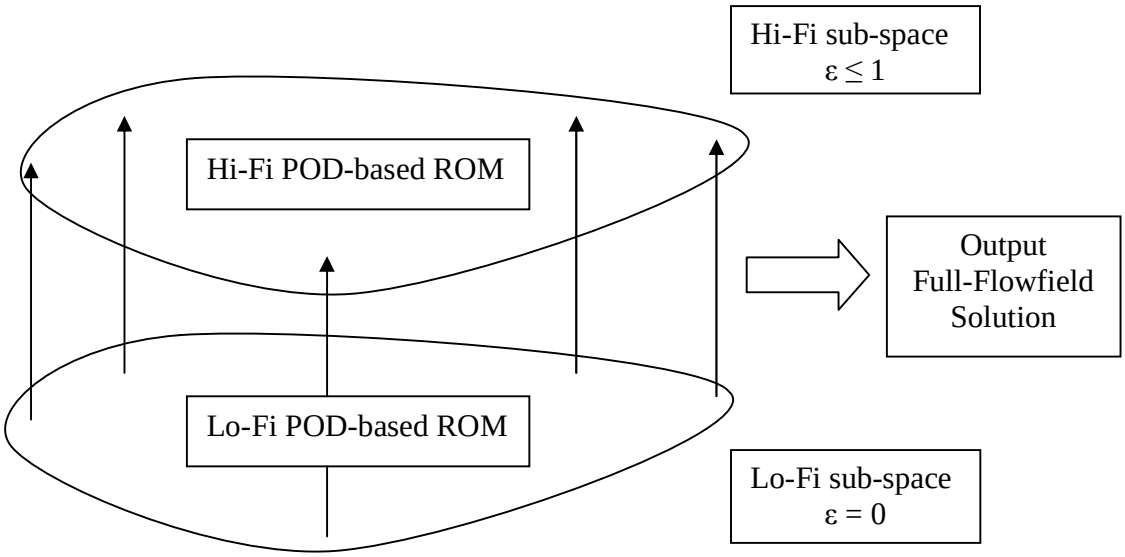


Figure (7-8) Pictorial representation of the variable-fidelity POD-based ROM method

Following this, the predicted solution vector $\hat{\mathbf{a}}$ is determined by

$$\hat{\mathbf{a}} = \bar{\mathbf{a}} + \sum_{k=1}^{m_1} \alpha'_{i=const,k} \mathrm{v}'_{k,\,j=const} + \sum_{k=1}^{q} \alpha_{i=const,k} \mathrm{v}_{k,\,j=const} \qquad \text{Eq. (7- 5)}$$

where $q < m_2$ and $\bar{\mathbf{a}}$ is the mean vector. This model gives access to the full flow-field data in contrast with the other variable-fidelity models reported in the literature.

7.6 Applications of the proposed model for fusing computational experiments

In this section the model proposed previously is applied to two high-speed aerodynamic flow problems with a fixed flow domain but having different boundary conditions. Firstly, the model is applied to a variable-fidelity problem where low- and high- order accurate viscous flow computations are considered. This particular problem is of interest because PNS flow solvers tend to become quite difficult to use at high angles of incidence at a high-order of accuracy. In this case the boundary conditions are everywhere the same between the low-fidelity and the high-fidelity analyses. In addition the flow physics is the same too. Secondly, the model is applied to a variable-fidelity problem where the flow physics and one of the boundary conditions between the low- and high- fidelities are different. In this latter case, the low-fidelity analyses are inviscid flow computations while the high-fidelity analyses are viscous flow computations. The different boundary condition at the wall between inviscid and viscous flow is obvious.

7.6.1 Low- and high-order accurate viscous flow computations

The application considered in this section is that of the cone-cylinder-flare looked at previously in chapters 5 and 6, although here a viscous flow problem is studied rather than an inviscid one. In this problem, the parameters were the angle of incidence which was varied within the range [0^o, 10^o] and the Mach number which was varied within the range [2, 6]. Here, the flare angle was maintained constant so that no errors are introduced in the model's outcome due to domain changes i.e., grid deformations. For this problem two Lagrange sub-spaces were generated by running the IMPNS software over the entire space at a constant Reynolds number, one with a set of low-fidelity snapshots and the other one with a set of high-fidelity snapshots. The low-fidelity snapshots were generated by computing nominally first-order accurate calculations, while the high-fidelity snapshots were calculated using a nominally third order in the cross-flow directions and second order in the streamwise. As explained in the previous section an auxiliary variable denoting whether the data is of low- or high- fidelity was added to the

problem, giving rise to a three-dimensional problem ie., the two parameters and the auxiliary variable. With this training data a global response surface of the projection coefficients was generated using Gaussian radial basis functions with a shape parameter of 1.0 from which a meta-model was developed in three-dimensions.

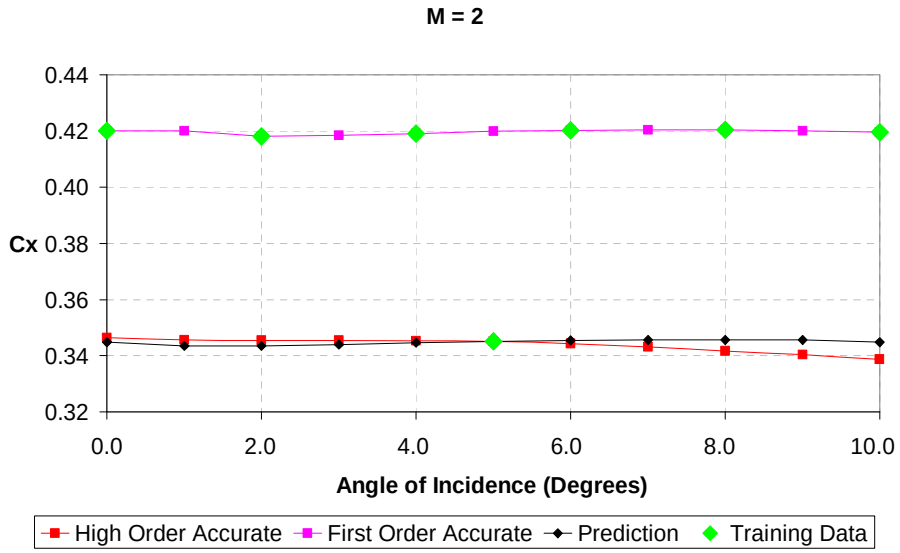
In this section two problems were considered with a different set of snapshots for the high-fidelity data. For the first case the training dataset was as follows:

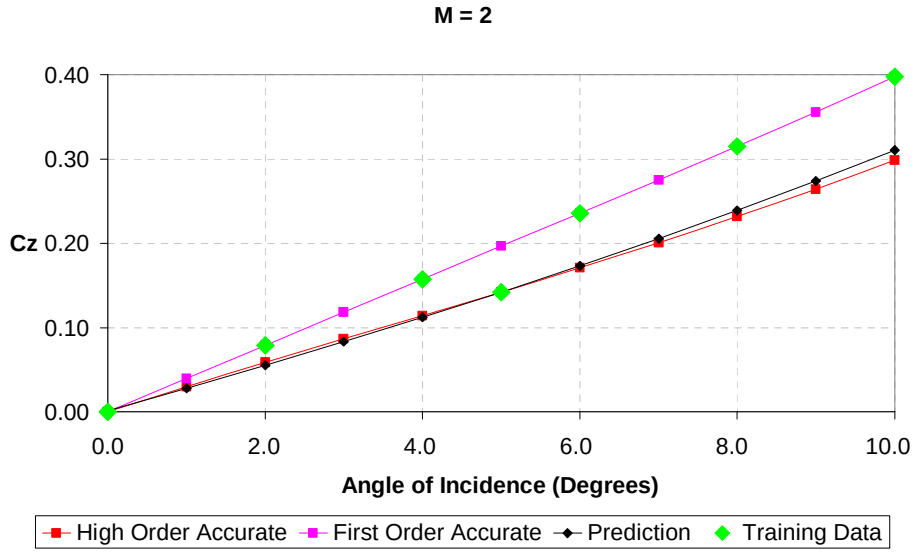
- For the low-fidelity dataset, the Mach number was varied at intervals of one across the range [2, 6], while the angle of incidence was varied at intervals of 2º across the range [0º, 10º]. A full-factorial design-of-experiment was set up with 30 snapshots in total.

- For the high-fidelity data, five snapshots were considered one at each Mach number. The five snapshots are at a constant angle of incidence of 5º.

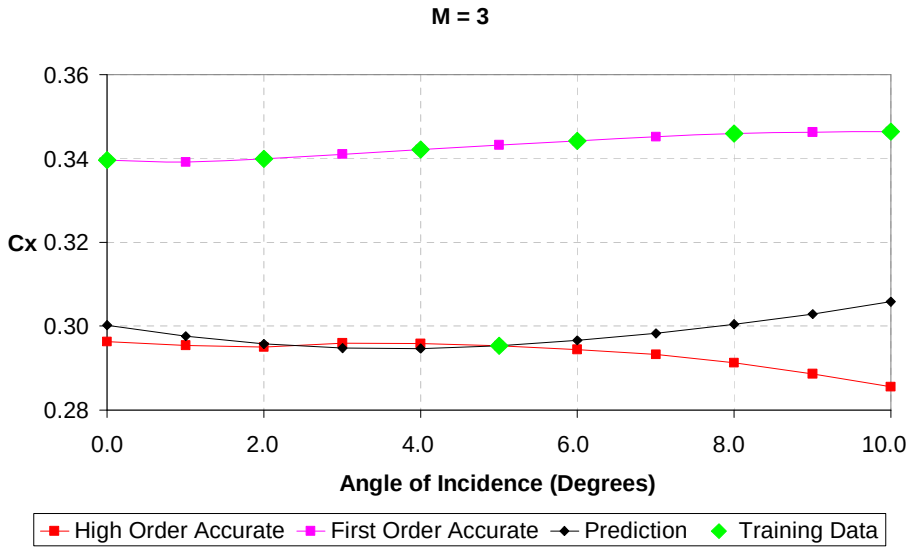
This set of observations was used to train a three-dimensional response surface which was used as a global interpolant along the three-dimensions. The following results were attained.

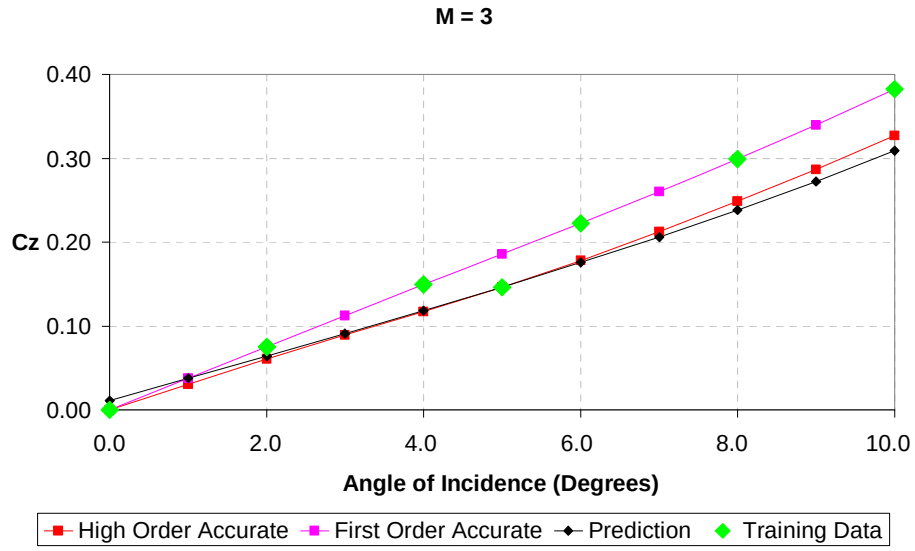
Figure (7-9) (a) to (e) shows the variation of the axial force coefficient C_x and the normal force coefficient C_z with the angle of incidence for the cone-cylinder-flare body at the different Mach numbers. The high-order accurate (high-fidelity) solution which is included for comparison is shown in red, the first-order accurate (low-fidelity) is shown in pink and the prediction achieved using the variable-fidelity/POD-based ROM method is shown in black. The training data which is input into the model is shown in light green.


M = 2
0.44
0.42
0.40
Cx 0.38
0.36
0.34
0.32
0.0
2.0
4.0
6.0
8.0
10.0
Angle of Incidence (Degrees)
High Order Accurate
First Order Accurate
Prediction
Training Data

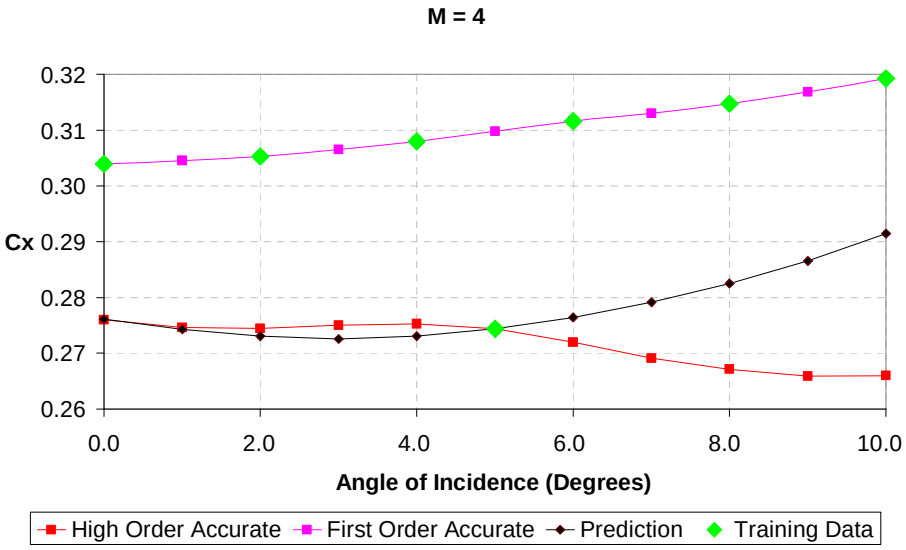

M = 2
0.40
0.30
Cz 0.20
0.10
0.00
0.0
2.0
4.0
6.0
8.0
10.0
Angle of Incidence (Degrees)
High Order Accurate
First Order Accurate
Prediction
Training Data

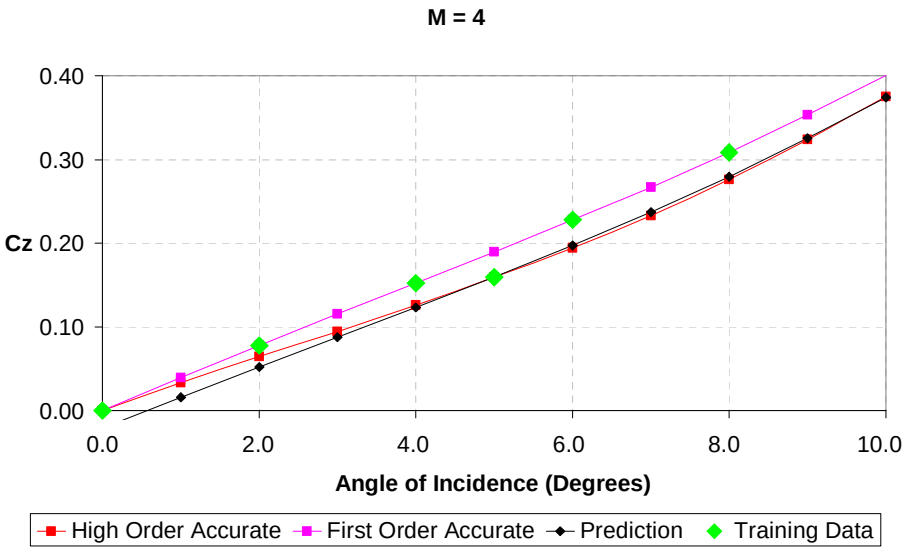
(a) M = 2


M = 3
0.36
0.34
Cx 0.32
0.30
0.28
0.0
2.0
4.0
6.0
8.0
10.0
Angle of Incidence (Degrees)
High Order Accurate
First Order Accurate
Prediction
Training Data

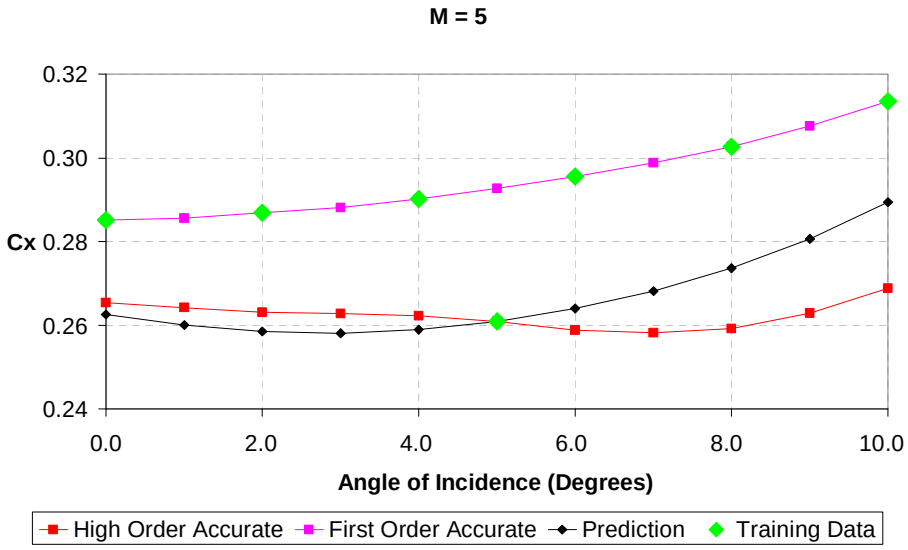

M = 3
0.40
0.30
Cz 0.20
0.10
0.00
0.0
2.0
4.0
6.0
8.0
10.0
Angle of Incidence (Degrees)
High Order Accurate
First Order Accurate
Prediction
Training Data

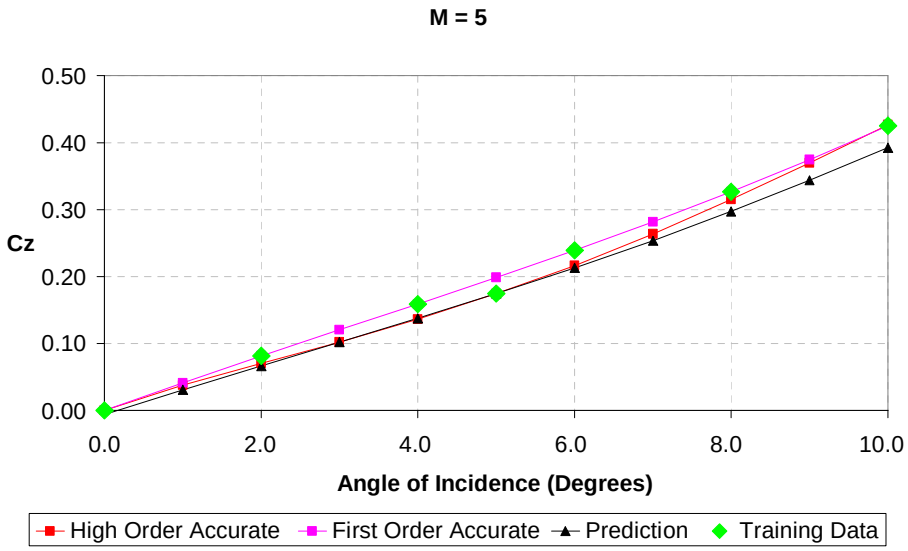
(b) M = 3


M = 4
0.32
0.31
0.30
Cx 0.29
0.28
0.27
0.26
0.0
2.0
4.0
6.0
8.0
10.0
Angle of Incidence (Degrees)
High Order Accurate
First Order Accurate
Prediction
Training Data

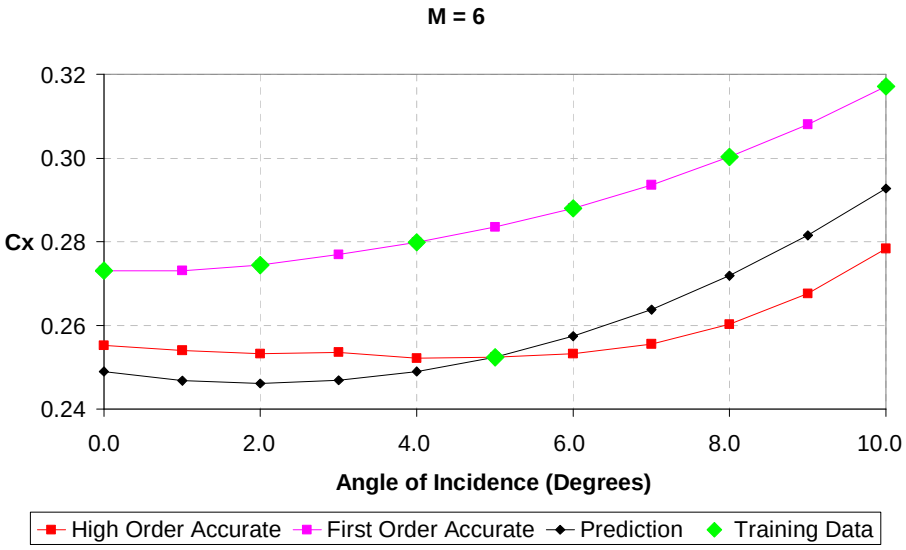

M = 4
0.40
0.30
Cz 0.20
0.10
0.00
0.0
2.0
4.0
6.0
8.0
10.0
Angle of Incidence (Degrees)
High Order Accurate
First Order Accurate
Prediction
Training Data

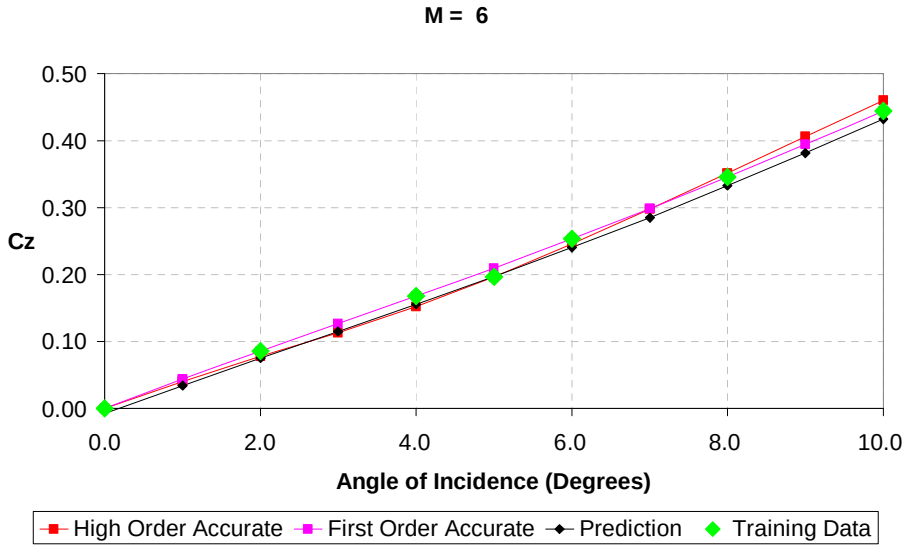
(c) M = 4


M = 5
0.32
0.30
Cx 0.28
0.26
0.24
0.0
2.0
4.0
6.0
8.0
10.0
Angle of Incidence (Degrees)
High Order Accurate
First Order Accurate
Prediction
Training Data


M = 5
0.50
0.40
0.30
Cz
0.20
0.10
0.00
0.0
2.0
4.0
6.0
8.0
10.0
Angle of Incidence (Degrees)
High Order Accurate
First Order Accurate
Prediction
Training Data

(d) M = 5





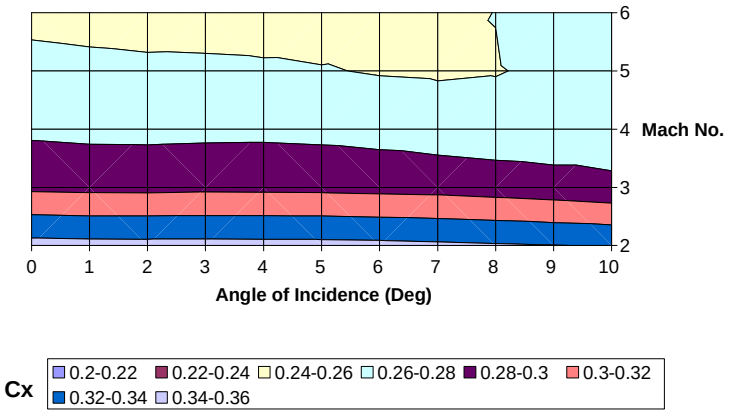
(e) M = 6

Figure (7-9) – Prediction of C_x & C_z

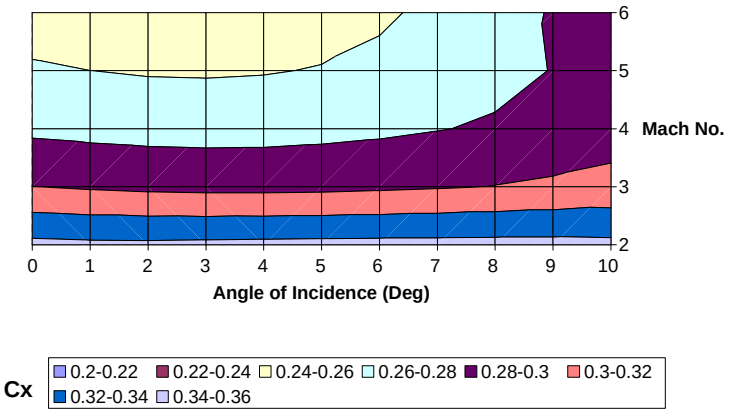
From figure (7-9) one can observe that the C_x predictions follow the features of the low-fidelity data while passing through the high-fidelity point used in the training dataset at an angle of incidence of 5º. In general, there is reasonably good prediction most especially over the range from 0º to 6º. Beyond an angle of attack of 6º the difference between the low- and high- fidelity data increases, thus causing a poorer prediction. For the normal force coefficient C_z prediction, coincidentally the Mach number of 2 and 3 cases show a very accurate prediction but do not strongly follow the features of the low-fidelity. The other cases at M = 4, M = 5 and M = 6 show a prediction where the features of the low-fidelity data are conspicuous. In the latter three cases, the prediction is superimposed on the high-fidelity training data point at an angle of incidence of 5º. In this work, it has been noticed that the closer together the low- and high-fidelity data are, the more the prediction follows the features of the low-fidelity. This effect can also be achieved numerically by controlling the magnitude of the introduced auxiliary variable, though one must be cautious not to ill-condition the radial basis function matrix mentioned in chapter 3.

These results once again emphasize the importance of using low-fidelity models which incorporate the correct trends, as the biggest errors were registered in those regions where the low- and high-fidelity follow different trends and have different features. This is most especially important when performing sparse data interpolation, since the reliance on the low-fidelity solutions is considerably greater. This observation was also made by Reisenthel *et al.* in their work.

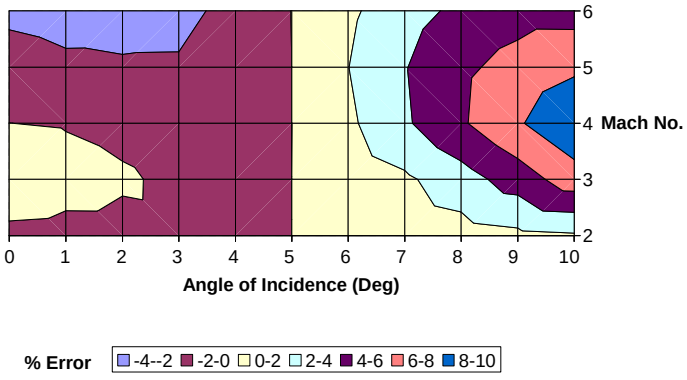
Figures (7-10) and (7-11) show the projected carpet plots of the force coefficients as predicted using the variable-fidelity/POD-based ROM method, the high-order accurate calculation using the IMPNS software and the percentage error of the prediction relative to the high-fidelity calculation. It is clear that the biggest errors occur towards the ends of the parametric ranges. In particular, the very high error recorded in the normal force coefficient at angles of incidence of 0º and 1º is due to the very small values rather than some flaw in the variable-fidelity/POD-based ROM method.



(a) high-fidelity solution (high-order accurate)

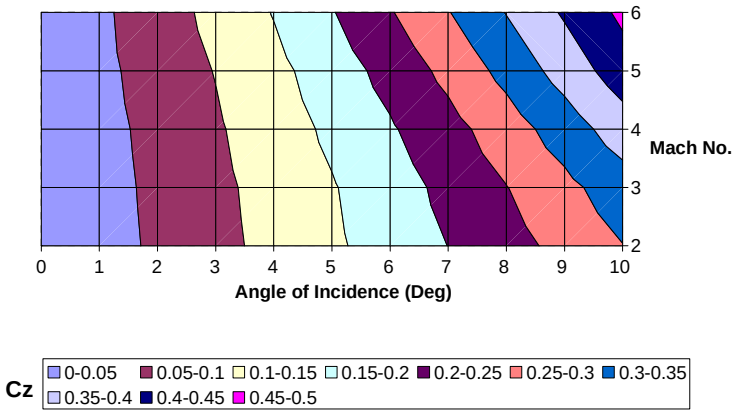


(b) predicted from the variable-fidelity/POD-based ROM model

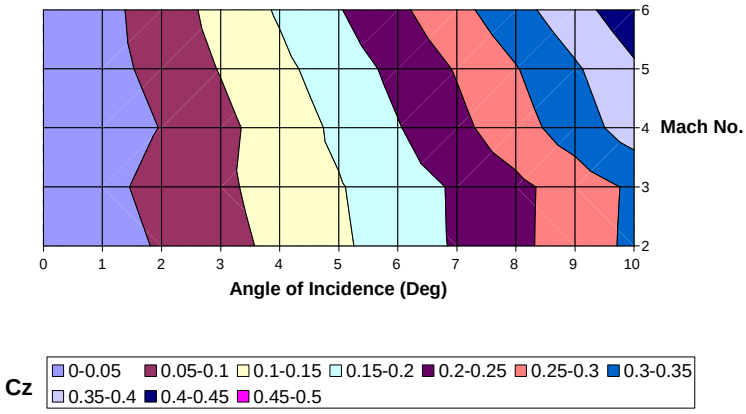


(c) percentage error of the predicted solution

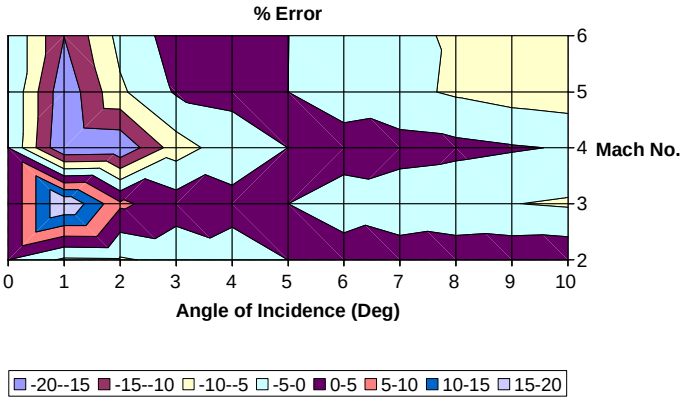
Figure (7-10) – Projected carpet plots of the axial force coefficient with five high-fidelity data points



(a) high-fidelity solution (high-order accurate)



(b) predicted from the variable-fidelity/POD-based ROM model



(c) percentage error of the predicted solution

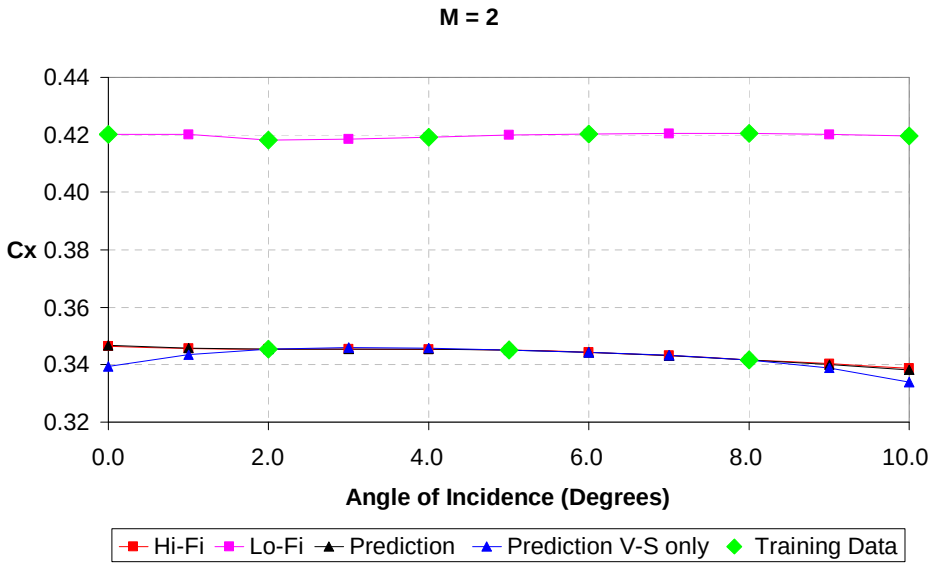
Figure (7-11) – Projected carpet plots of the normal force coefficient with five high-fidelity data points

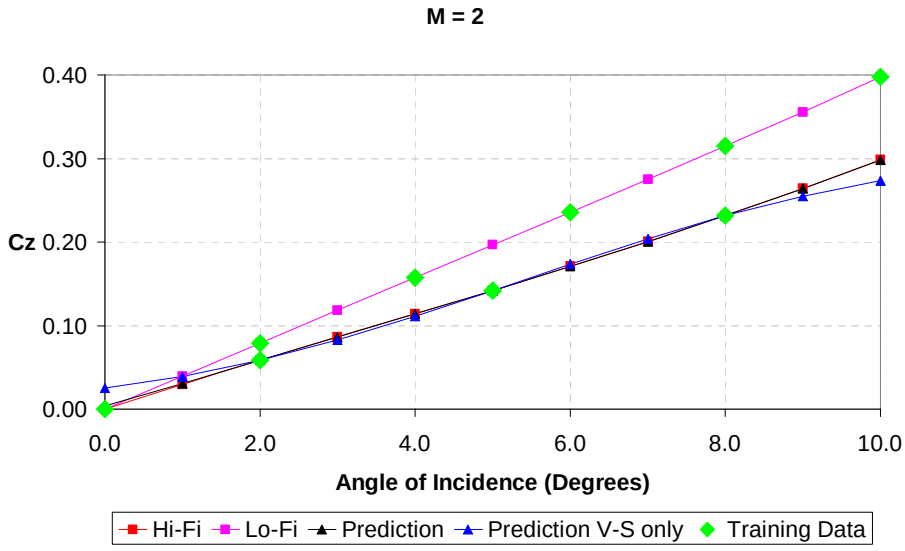
In the second case the training dataset considered is as follows:

- The low-fidelity dataset considered in the previous case was considered once again.

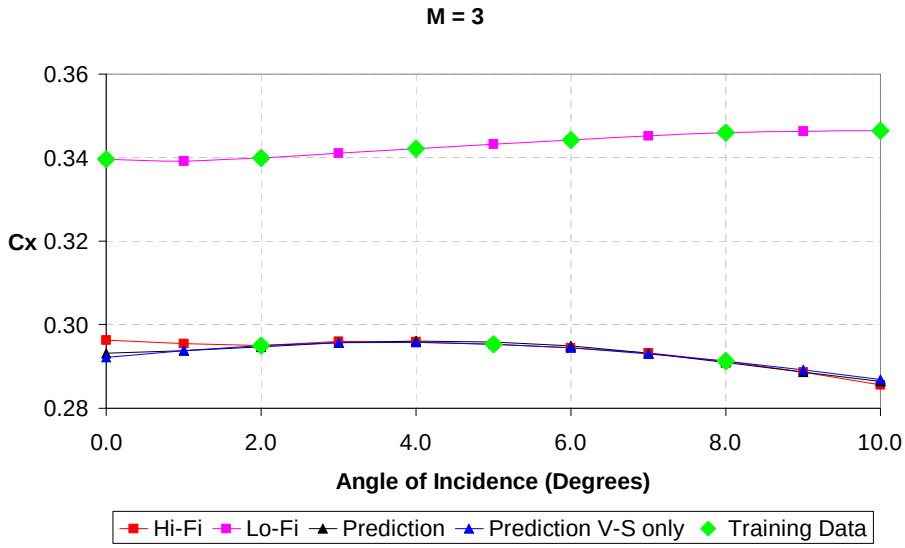
- For the high-fidelity data, fifteen snapshots in total were considered. A full-factorial design-of-experiment was set up with Mach numbers at 2,3,4,5,6 and an angle of incidence of 2^o, 5^o and 8^o. These angles of incidence were chosen so that the high-fidelity Lagrangian sub-space will only partially cover the whole parametric space. Therefore, this would provide a problem where some regions would be predicted by interpolation and some others by extrapolation.

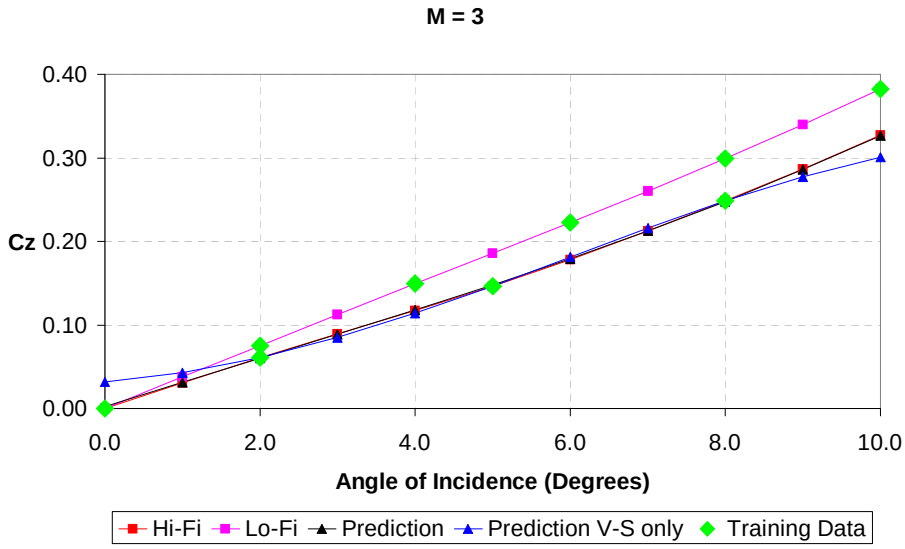
This dataset of observations was used to train a three-dimensional response surface which was used as a global interpolant along the three-dimensions. The following are the results obtained by using this training dataset.


M = 2
0.44
0.42
0.40
Cx 0.38
0.36
0.34
0.32
0.0
2.0
4.0
6.0
8.0
10.0
Angle of Incidence (Degrees)
Hi-Fi
Lo-Fi
Prediction
Prediction V-S only
Training Data

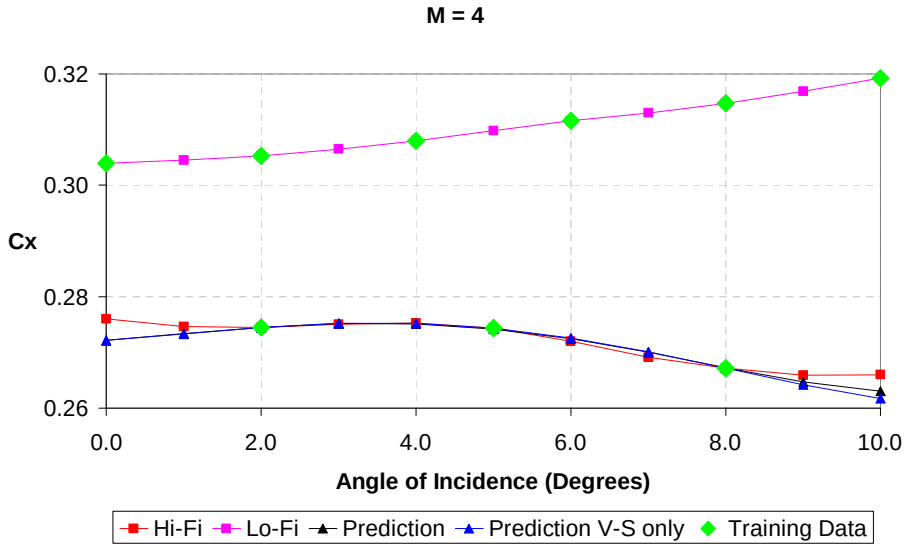

M = 2
0.40
0.30
Cz 0.20
0.10
0.00
0.0
2.0
4.0
6.0
8.0
10.0
Angle of Incidence (Degrees)
Hi-Fi
Lo-Fi
Prediction
Prediction V-S only
Training Data

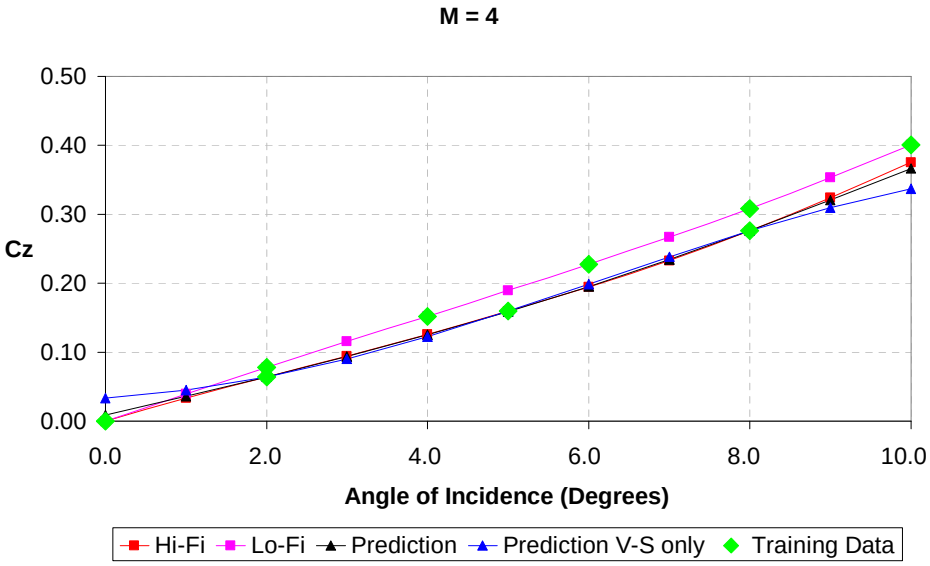
(a) M = 2


M = 3
0.36
0.34
Cx 0.32
0.30
0.28
0.0
2.0
4.0
6.0
8.0
10.0
Angle of Incidence (Degrees)
Hi-Fi
Lo-Fi
Prediction
Prediction V-S only
Training Data

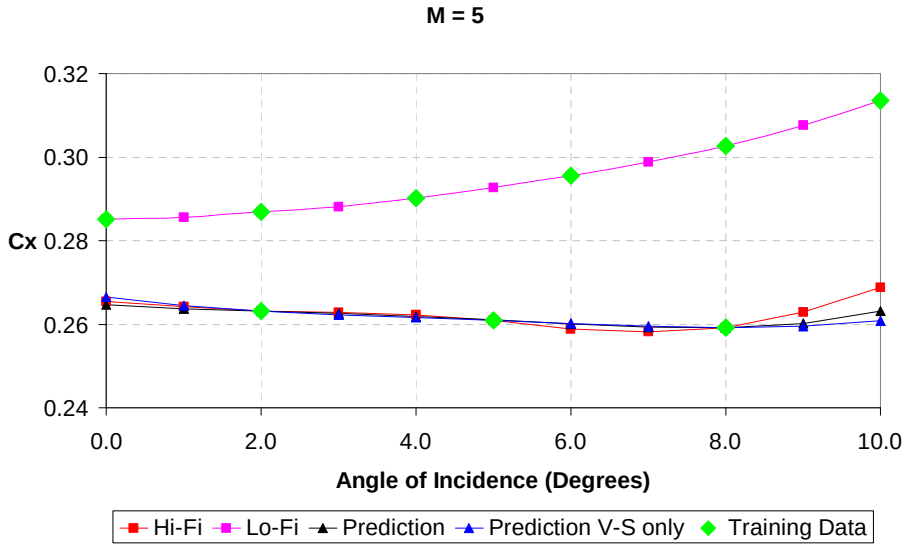

M = 3
0.40
0.30
Cz 0.20
0.10
0.00
0.0
2.0
4.0
6.0
8.0
10.0
Angle of Incidence (Degrees)
Hi-Fi
Lo-Fi
Prediction
Prediction V-S only
Training Data

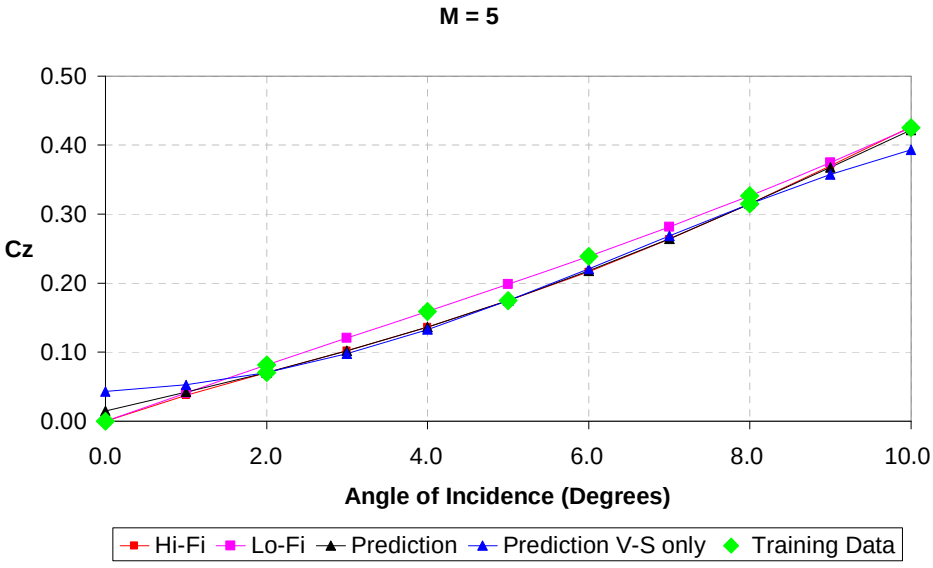
(b) M = 3


M = 4
0.32
0.30
0.28
0.26
Cx
0.0
2.0
4.0
6.0
8.0
10.0
Angle of Incidence (Degrees)
Hi-Fi
Lo-Fi
Prediction
Prediction V-S only
Training Data

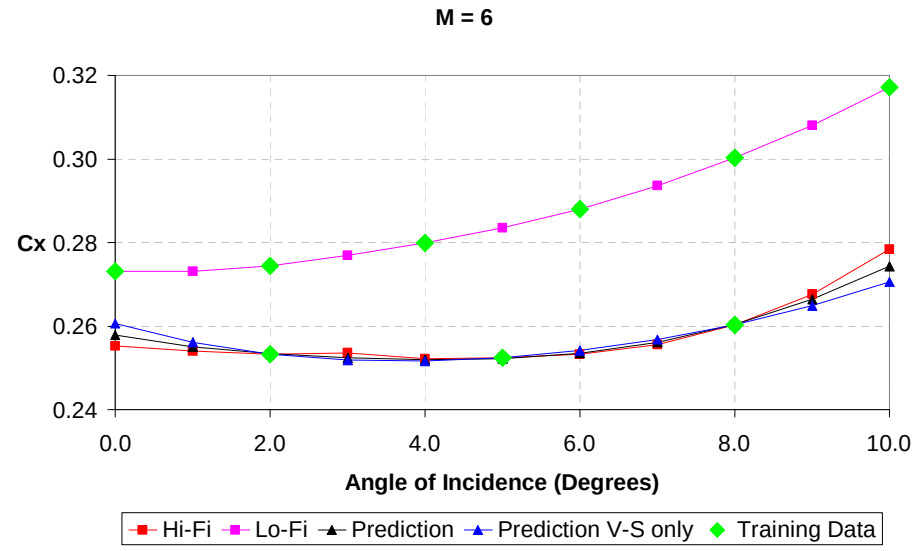

M = 4
0.50
0.40
0.30
0.20
0.10
0.00
Cz
0.0
2.0
4.0
6.0
8.0
10.0
Angle of Incidence (Degrees)
Hi-Fi
Lo-Fi
Prediction
Prediction V-S only
Training Data

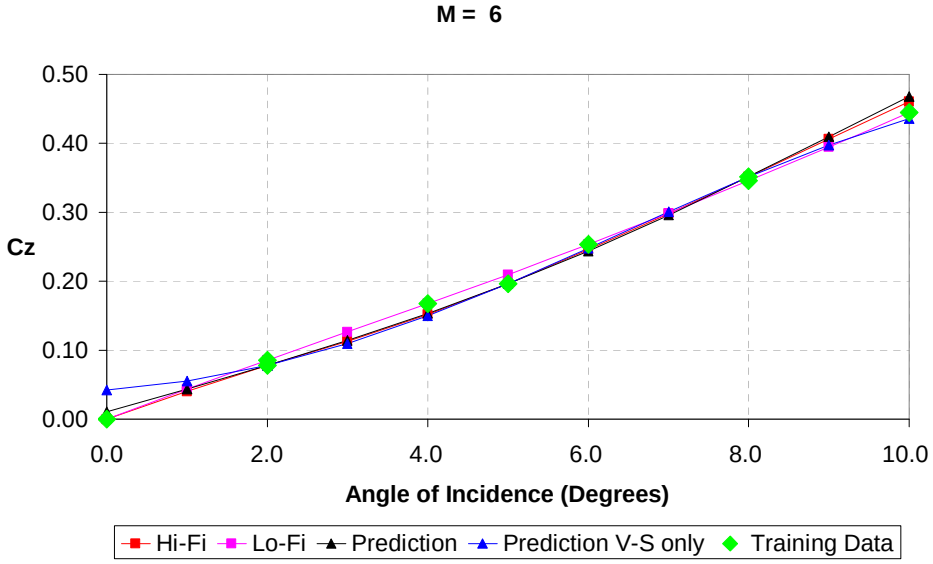
(c) M = 4


M = 5
0.32
0.30
Cx 0.28
0.26
0.24
0.0
2.0
4.0
6.0
8.0
10.0
Angle of Incidence (Degrees)
Hi-Fi
Lo-Fi
Prediction
Prediction V-S only
Training Data


M = 5
0.50
0.40
0.30
Cz
0.20
0.10
0.00
0.0
2.0
4.0
6.0
8.0
10.0
Angle of Incidence (Degrees)
Hi-Fi
Lo-Fi
Prediction
Prediction V-S only
Training Data

(d) M = 5



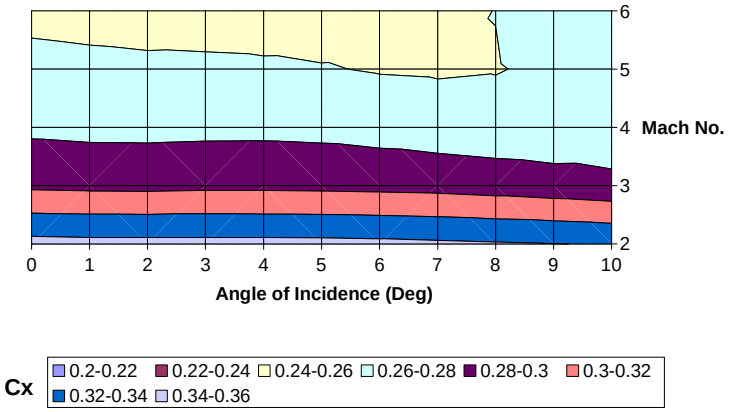


(e) M = 6

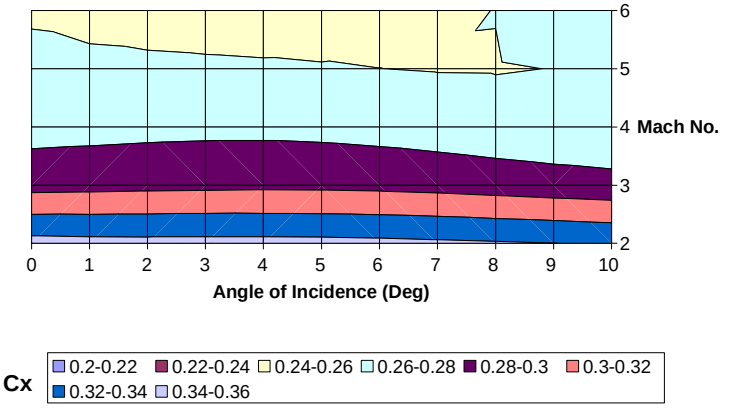
Figure (7-12) – Prediction of Cx & Cz

Figure (7-12) shows the variation of the axial and normal force coefficients with the angle of attack at different Mach numbers. While general significant improvement is achieved when using more high-fidelity data points over a wider area as expected, from these figures it is clear that by using the variable-fidelity/POD-based ROM model, a reasonable improvement is registered at regions beyond the high-fidelity Lagrangian sub-space. This improvement is achieved due to the introduction of the low-fidelity sub-space to the problem which allows us to conduct an extrapolation of the high-fidelity sub-space based on an interpolation of the low-fidelity sub-space. The resulting dimensionality augmented response surface is ultimately globally interpolated. In addition, some minor improvement is noticeable most especially from the normal force coefficient within the high-fidelity data points range.

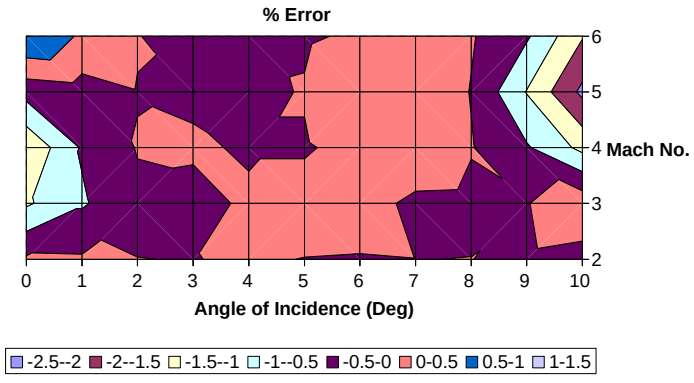
Figures (7-13) and (7-14) show the carpet plots of the force coefficients. In this latter example there exists a very good agreement between the ones predicted and the ones calculated using CFD (high-fidelity). This observation is reinforced by the percentage error plots, wherein the error magnitude is relatively small compared with the previous example. The error is generally approximately within +/- 2.0 % for both force coefficients except for a few small regions. Once again it must be remarked that the high error values in the normal force coefficient at low angles of incidence is due to the very small numerical values calculated.



(a) high-fidelity solution (high-order accurate)

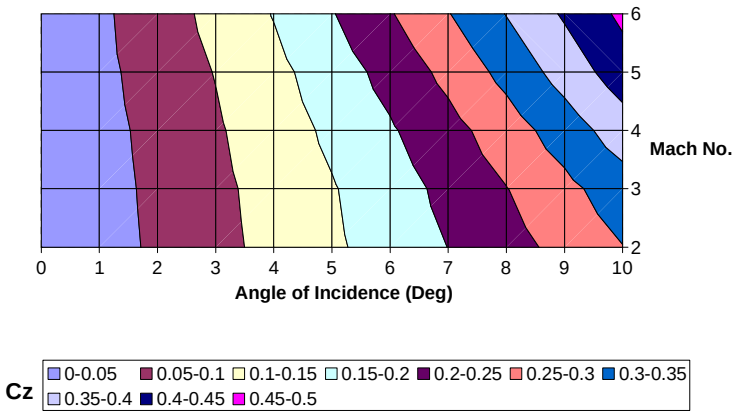


(b) predicted from the variable-fidelity/POD-based ROM model

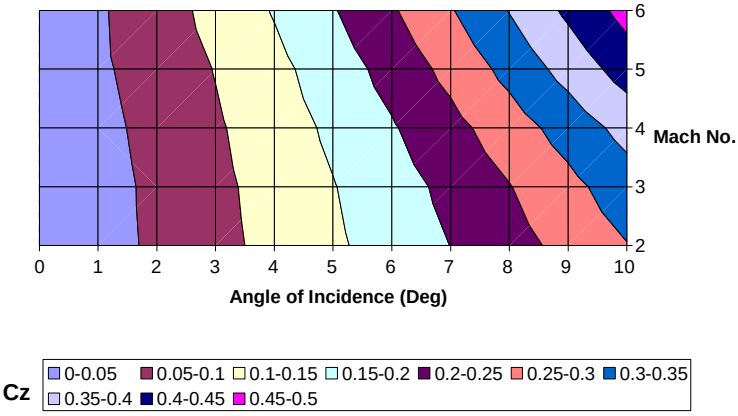


(c) percentage error of the predicted solution

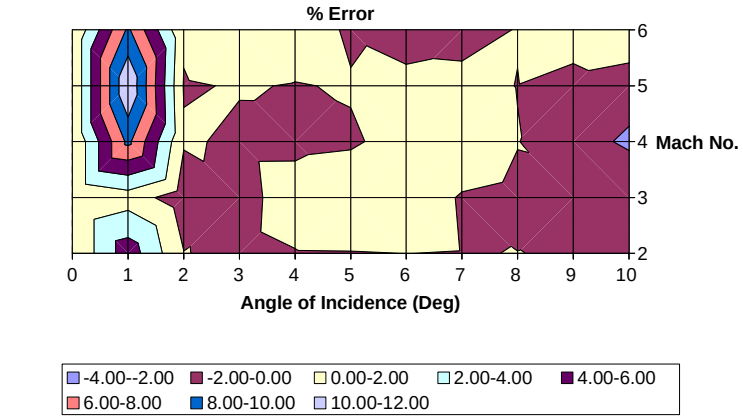
Figure (7-13) – Projected carpet plots of the axial force coefficient with fifteen high-fidelity data points



(a) high-fidelity solution (high-order accurate)



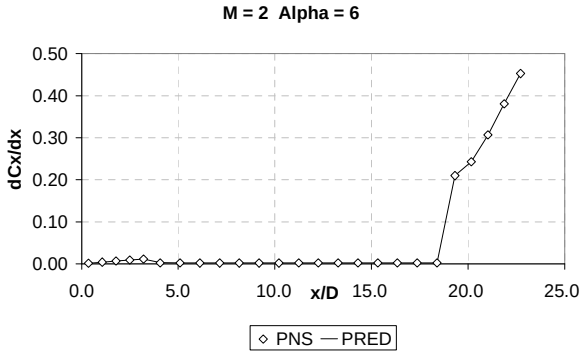
(b) predicted from the variable-fidelity POD-based ROM model

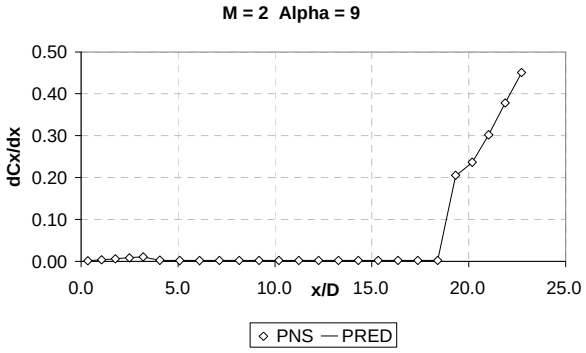


(c) percentage error of the predicted solution

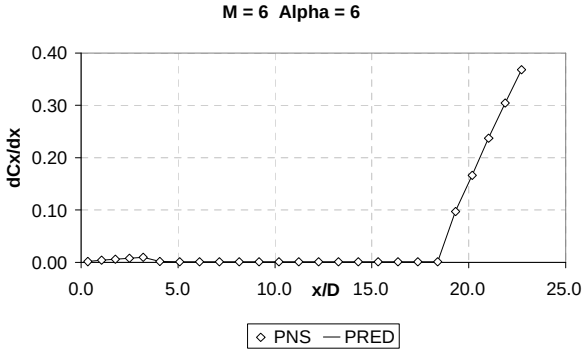
Figure (7-14) – Projected carpet plots of the normal force coefficient with fifteen high-fidelity data points

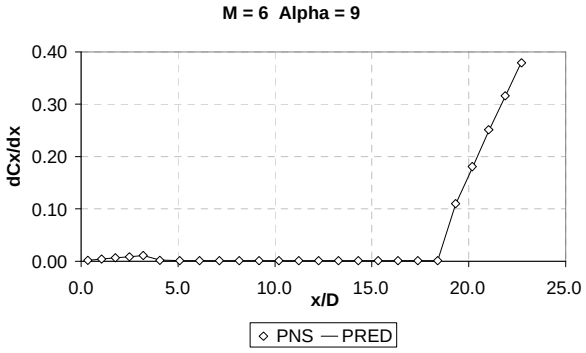
Further plots of the derivatives of the forces with fifteen high-fidelity data points are presented in figures (7-15) and (7-16). Predictions are presented at angles of incidence of 6° and 9° for the most extreme Mach number values considered ie. 2 and 6. The agreement between the ones predicted using the variable-fidelity/POD-based ROM method (solid line in figures) and the ones calculated using CFD (diamonds) is excellent for both the axial and normal force derivatives.





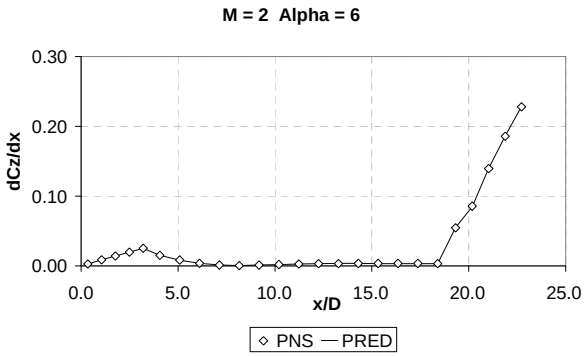
(a)

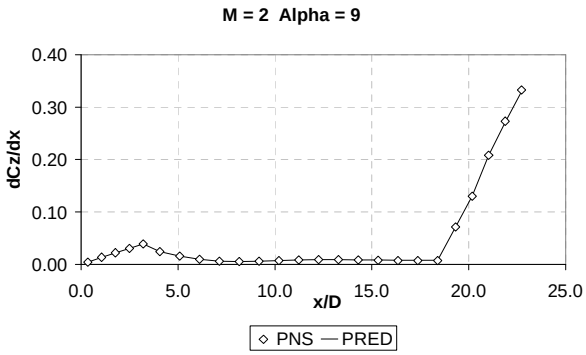




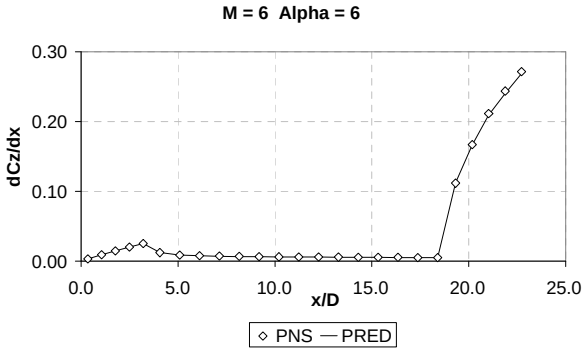
(b)

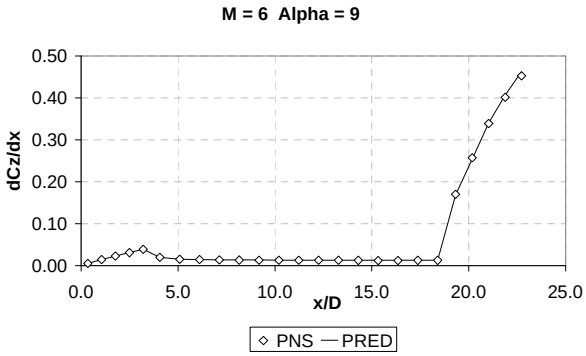
Figure (7-15) Plots of the derivative of the axial force coefficient versus the number of calibres





(a)

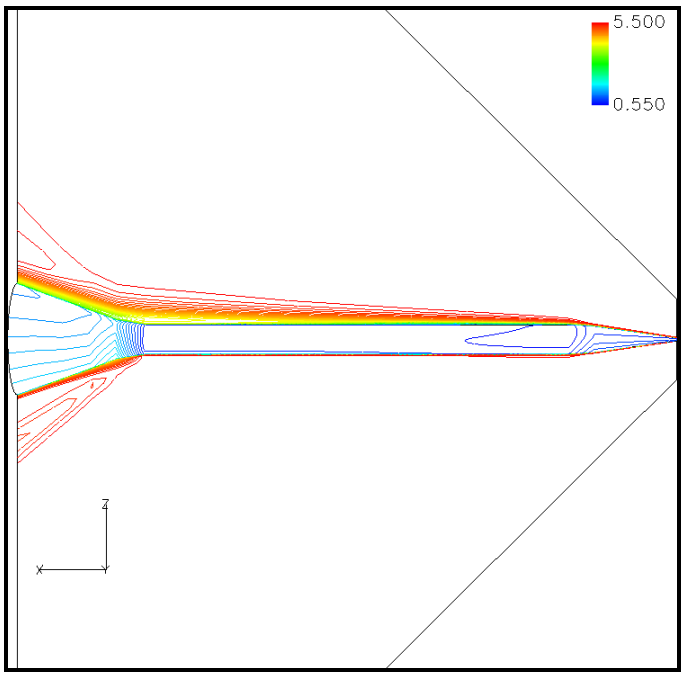




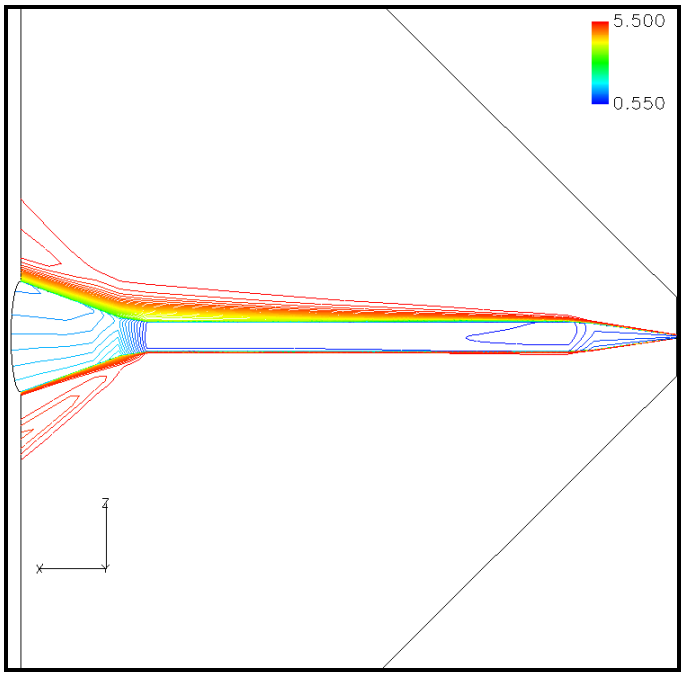
(b)

Figure (7-16) Plots of the derivative of the normal force coefficient versus the number of calibres

Moreover for the latter case, flow-field comparisons are presented between the predictions obtained by the variable-fidelity/POD-based ROM method and the high-fidelity simulations. Comparisons of the total pressure contours were made at the two most extreme Mach number values ie., 2 and 6 and at angles of incidence of 6° and 9°. The former angle of incidence is situated within the high-fidelity training dataset while the latter angle of incidence is situated beyond. At these parametric values, reasonably accurate axial and normal force coefficients and their derivatives were attained and so it was interesting to observe the flow-field predictions. In this case the total pressure was considered since it is a sensible indication of the accuracy by which all the primitive variables are predicted. The comparisons show that the predicted total pressure contours are only very slightly different from the ones calculated over certain regions of the flow-field, hence the predictions are satisfactory. In the following figures (7-17) to (7-21), the upper part (a) represent the prediction from the variable-fidelity/POD-based ROM method and lower part (b) is obtained from the full-order CFD.

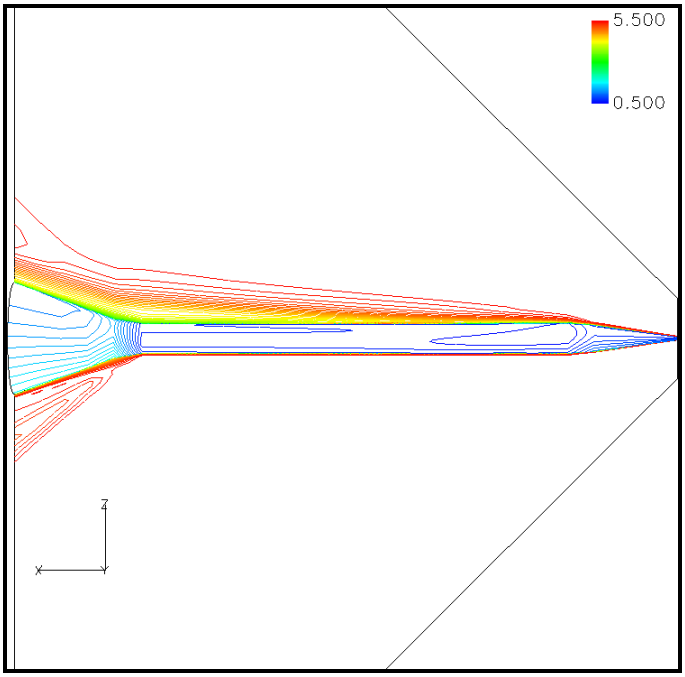


(a) Predicted total pressure contours at M = 2 and alpha = 6°

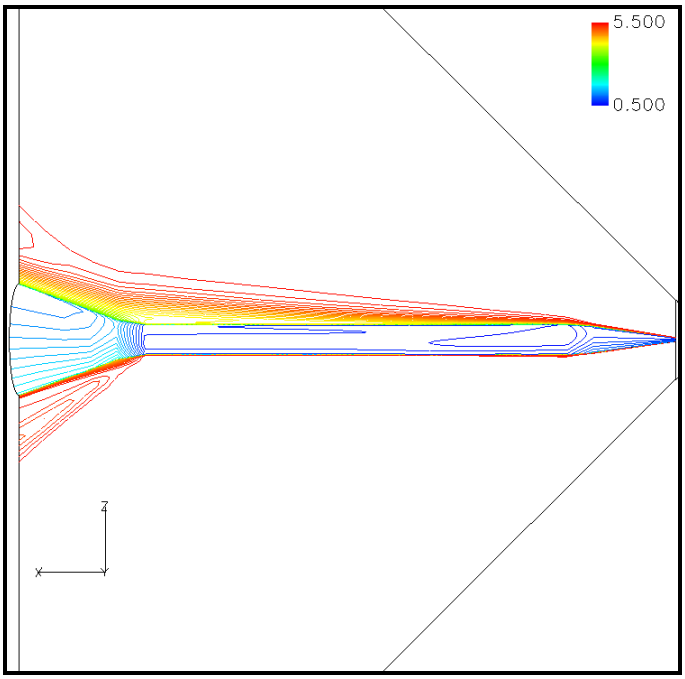


(b) PNS total pressure contours at M = 2 and alpha = 6°

Figure (7-17) Comparison of the total pressure contours at M = 2 and alpha = 6°

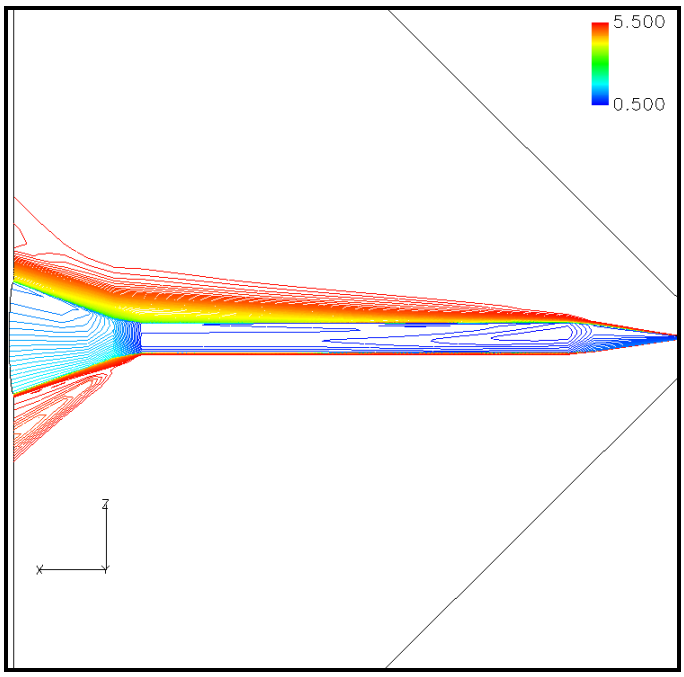


(a) Predicted total pressure contours at M = 2 and alpha = 9°

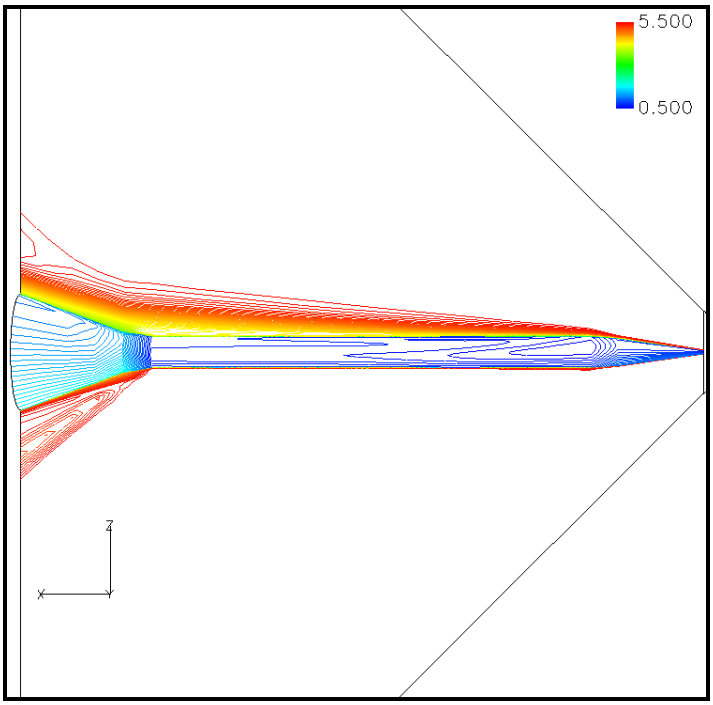


(b) PNS total pressure contours at M = 2 and alpha = 9°

Figure (7-18) Comparison of the total pressure contours (64) at M = 2 and alpha = 9°

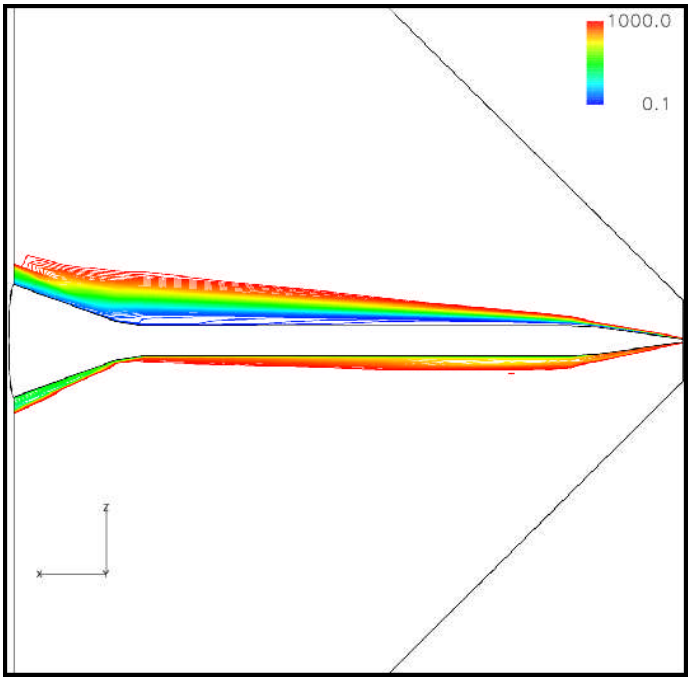


(a) Predicted total pressure contours at M = 2 and alpha = 9°

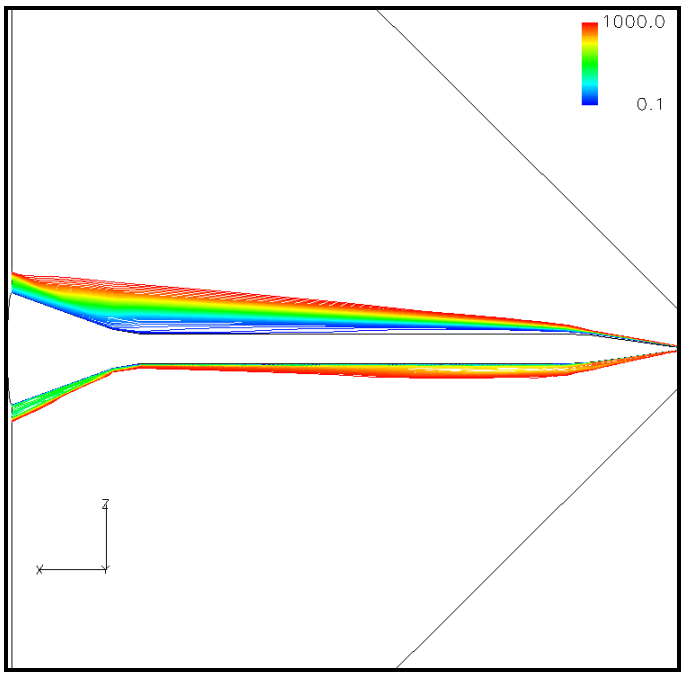


(b) PNS total pressure contours at M = 2 and alpha = 9°

Figure (7-19) Comparison of the total pressure contours (128) at M = 2 and alpha = 9°

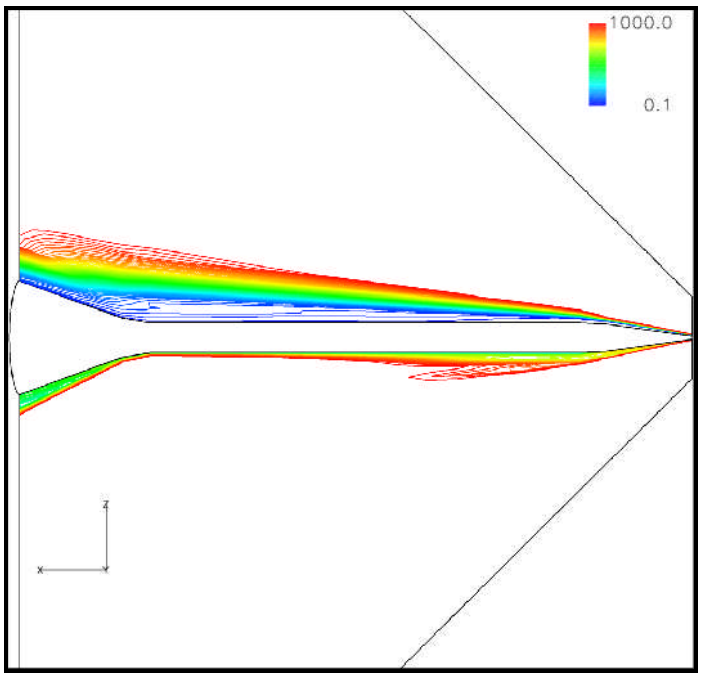


(a) Predicted total pressure contours at M = 6 and alpha = 6°

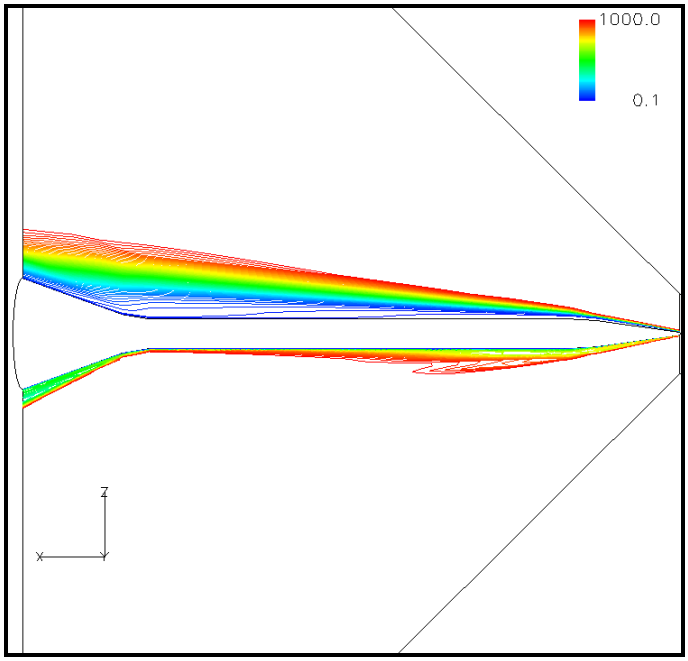


(b) PNS total pressure contours at M = 6 and alpha = 6°

Figure (7-20) Comparison of the total pressure contours at M = 6 and alpha = 6°



(a) Predicted total pressure contours at M = 6 and alpha = 9°



(b) PNS total pressure contours at M = 6 and alpha = 9°

Figure (7-21) Comparison of the total pressure contours at M = 6 and alpha = 9°

In general, this new method is composed of two models ie. the variable-fidelity model of Reisenthel *et al*. and POD-based ROMs. Being so, it inherits all the advantages and disadvantages from both its parents. No beneficial interaction was observed between the two models. For example, the POD-based reduced-order modelling limitations are still present within the proposed method. In particular, it has been observed that this model behaves pretty much like the variable-fidelity model of Reisenthel *et al*., that is, the prediction follows very closely the trend of the low-fidelity training data points and so it relies to a great extent on it, most especially when the high-fidelity training data is sparse. However, the influence of the low-fidelity data upon the high-fidelity one drops down as the number of high-fidelity training data points is increased. Possibly, a way to verify whether the model is adequately representing a particular application is by increasing in steps the number of high-fidelity training data and monitoring the discrepancies from its output. As the number of the high-fidelity training data is increased, the response from the model must tend towards some limiting value, which will be within some approximation error. This bounded error results from the radial basis function network.

It is well known that POD-based ROMs work well in an interpolatory setting. This was confirmed in chapters 5 and 6 of this work. However, it is not clear whether this ROM technique works well in an extrapolatory setting unless the physics of the problem change linearly beyond the parametric range, even because there is no published research work about this to the author's knowledge [1]. Therefore, by using this variable-fidelity/POD-based ROM method, a transformation of the problem is made from an extrapolatory setting into an interpolatory one by the introduction of the auxiliary variable, hence making it possible for the POD-based ROM to work well. Thus, this technique may be considered as a first step towards achieving a POD-based ROM to work well for a linear or non-linear problem in an extrapolatory setting.

7.6.2 Inviscid and viscous flow computations

Fusing inviscid and viscous flow computational data is more challenging than the type of fusion considered earlier due to the different problem formulation where the physics and boundary conditions at the wall are different, ie. the slip and non-slip condition for inviscid and viscous flows respectively. Although it would still be possible to fuse inviscid and viscous data in the way suggested in the previous section, the problem would not be well posed and so the outcome is generally less accurate than what is being suggested in this section. Consequently, the data fusion technique adopted earlier can be applied for the pressure and the density variables but strictly not to the three components of velocity. Therefore, some way of getting around this problem must be found before attempting to fuse data with different physics. The following method was adopted to conduct data fusion in this case.

7.6.2.1 Description of the method

In this method a variable-fidelity dataset of observations is generated by conducting inviscid and viscous flow computations over the same computational mesh of a particular problem of interest. The same computational mesh is considered so that the modes or basis vectors which are a function of the position in space will be at fixed places across the whole set of snapshots for both fidelities. This will ensure that the locations of the POD modes are the same and hence no errors will be introduced due to different locations. A POD of this ensemble of CFD solutions is performed from which the basis vectors are obtained. By projecting the CFD solutions onto the basis vectors, a set of projection coefficients is achieved. This set of projection coefficients is a set of discrete data at two different levels of fidelity (one with scalar coefficients for inviscid data and the other one with scalar coefficients for viscous data). A pseudo-continuous representation of these coefficients is generated by radial basis functions to provide means by which high-fidelity predictions are made at parameter combinations not in the original set of observations.

The method adopted to fuse the primitive variables of the inviscid and viscous computations is as follows;

1. A POD of the inviscid and viscous snapshots is accomplished for the density and pressure field variables only. The derived inviscid and viscous projection coefficients are then modelled using radial basis function networks so that two response surfaces are generated at two different levels representing the inviscid and viscous projection coefficients. The various fidelity levels in the radial basis function networks are represented by the inclusion of another auxiliary variable in addition to the physical or geometric parameters.

2. Another POD of the velocities of the viscous flow field only is performed. In this case the velocity fields are not fused together as in the case of the pressure and density due to the different physics of the problem. Predictions were made using a multi-quadric radial basis function with a shift parameter equal to zero. A bi-linear spline interpolation can also be applied. It is suggested that if a one- or a two-equation turbulence model is used, the output parameters from the turbulence model are treated in the same way the velocities are predicted. In this example, an algebraic turbulence model was used.

7.6.2.2 Application of the proposed method

The application considered in this section is again that of the cone-cylinder-flare looked at previously. Once again the parameters were the angle of incidence which was varied within the range [0º, 10º] and the Mach number which was varied within the range [2, 6]. The flare angle was maintained constant so that no errors are introduced in the model's outcome due to domain changes i.e., grid deformations. Two Lagrange sub-spaces were generated by running the IMPNS software over the entire space, one with a set of low-fidelity snapshots and the other one with a set of high-fidelity snapshots. The

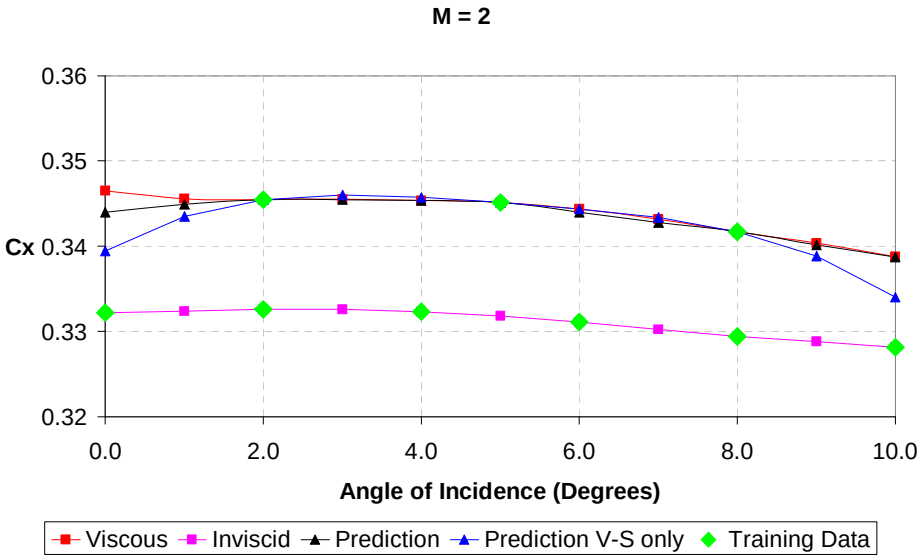
low-fidelity snapshots were generated by computing high-order accurate inviscid flow calculations, while the high-fidelity snapshots were calculated using high-order accurate viscous flow computations. In both sets of computations the order of space-wise discretization in the streamwise flux was calculated to a second order of accuracy while the order of spacewise discretization in each of the cross-flow fluxes was calculated to a third order of accuracy. As explained in the previous section an auxiliary variable denoting whether the data is of low- or high- fidelity was augmented with the other input variables, giving rise to a three-dimensional problem ie., the two parameters mentioned earlier and the auxiliary variable. With this training data a global response surface of the projection coefficients was generated using Gaussian radial basis functions with a shape parameter of 1.0.

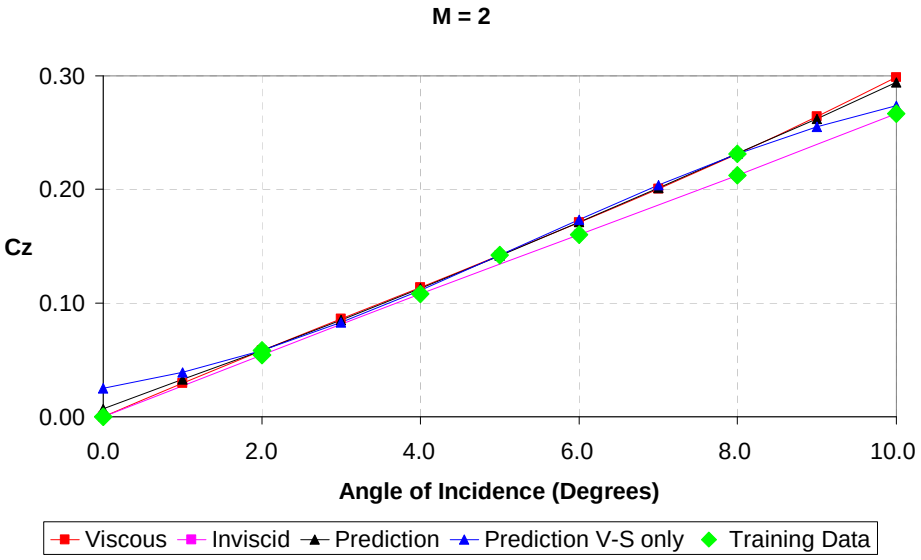
For the problem considered here a training dataset was generated as follows:

- For the low-fidelity dataset, the Mach number was varied at intervals of one across the range [2, 6], while the angle of incidence was varied at intervals of 2º across the range [0º, 10º]. A full-factorial design-of-experiment was set up with 30 snapshots in total.

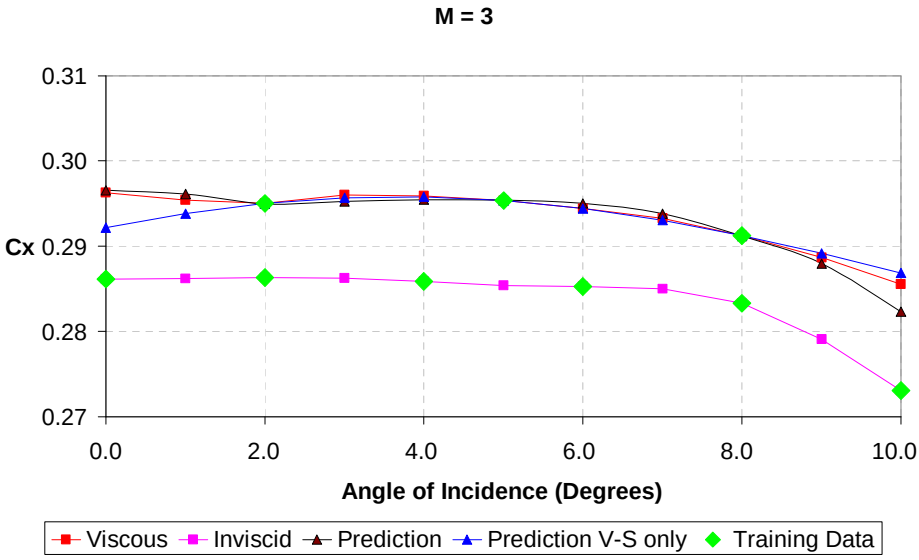
- For the high-fidelity data, fifteen snapshots in total were considered. A full-factorial design-of-experiment was set up with Mach numbers at 2, 3, 4, 5, 6 and an angle of incidence of 2º, 5º and 8º. These angles of incidence were chosen so that the high-fidelity Lagrangian sub-space will only partially cover the low-fidelity parametric space. Thus, this would provide a problem where some regions would be predicted by interpolation and some others by extrapolation.

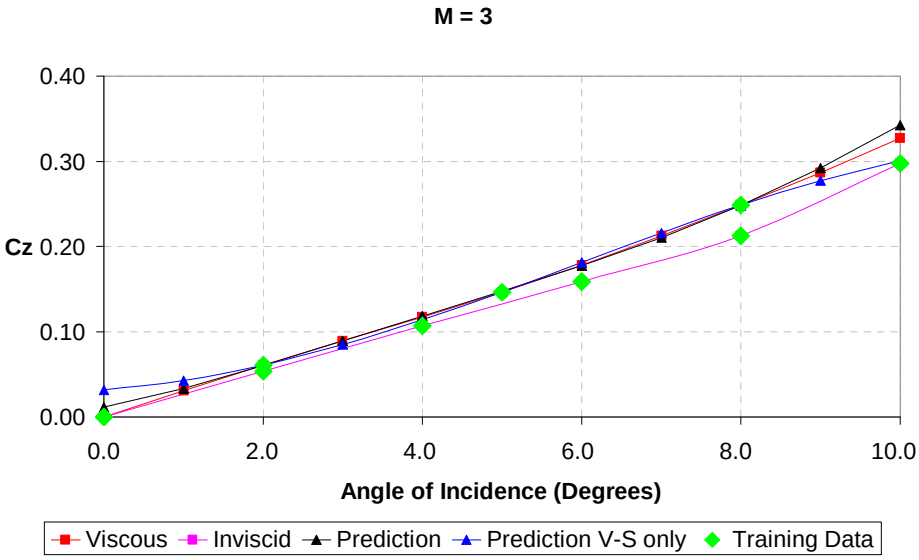
This set of observations was used to train a three-dimensional response surface. The following results shown in figure (7-22) were obtained.


M = 2
0.36
0.35
Cx 0.34
0.33
0.32
0.0
2.0
4.0
6.0
8.0
10.0
Angle of Incidence (Degrees)
Viscous
Inviscid
Prediction
Prediction V-S only
Training Data

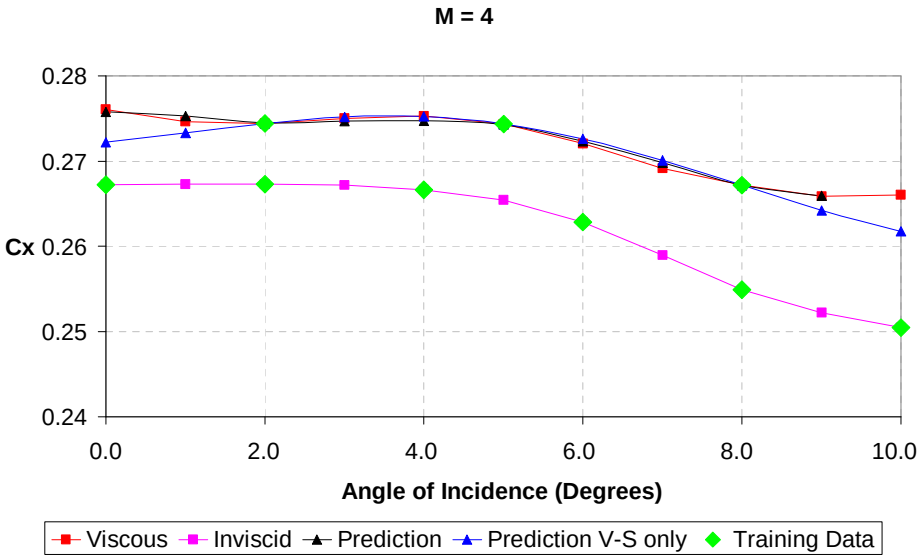

M = 2
0.30
0.20
Cz
0.10
0.00
0.0
2.0
4.0
6.0
8.0
10.0
Angle of Incidence (Degrees)
Viscous
Inviscid
Prediction
Prediction V-S only
Training Data

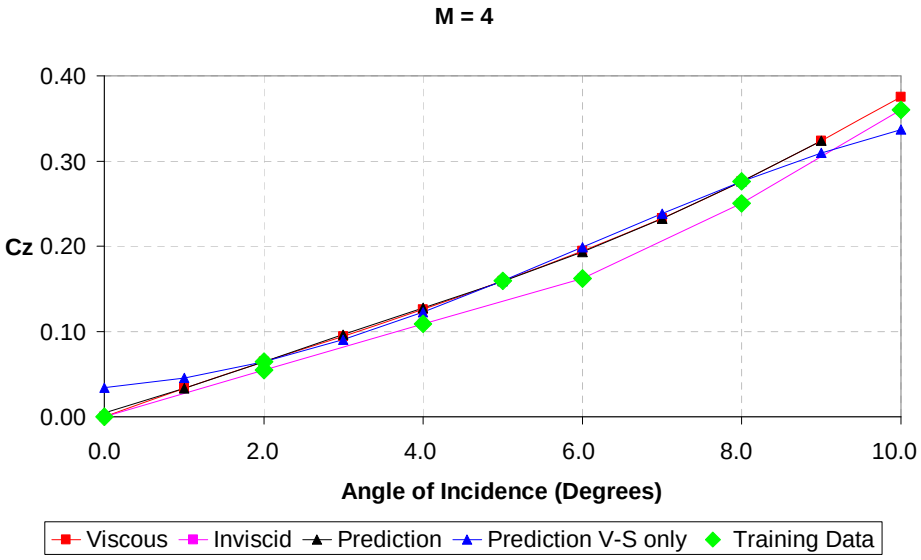
(a)


M = 3
0.31
0.30
Cx 0.29
0.28
0.27
0.0
2.0
4.0
6.0
8.0
10.0
Angle of Incidence (Degrees)
Viscous
Inviscid
Prediction
Prediction V-S only
Training Data

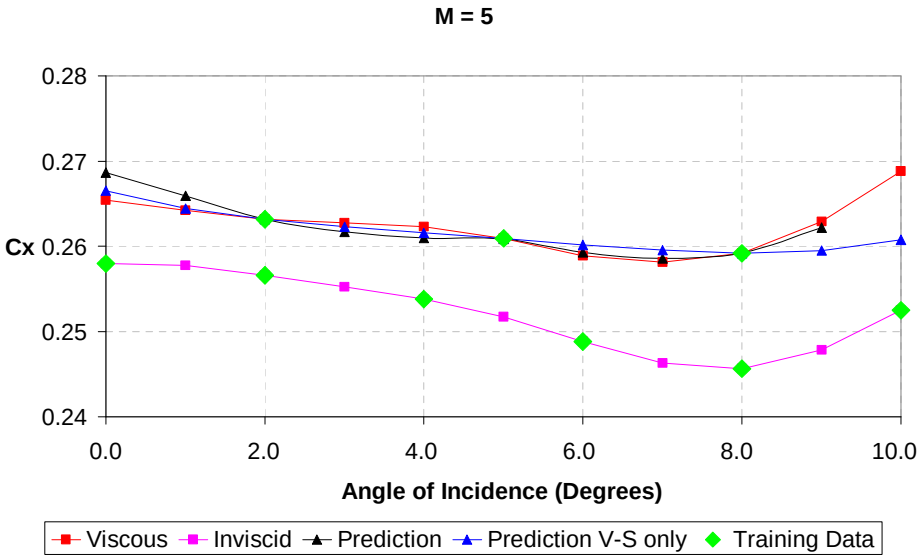

M = 3
0.40
0.30
Cz 0.20
0.10
0.00
0.0
2.0
4.0
6.0
8.0
10.0
Angle of Incidence (Degrees)
Viscous
Inviscid
Prediction
Prediction V-S only
Training Data

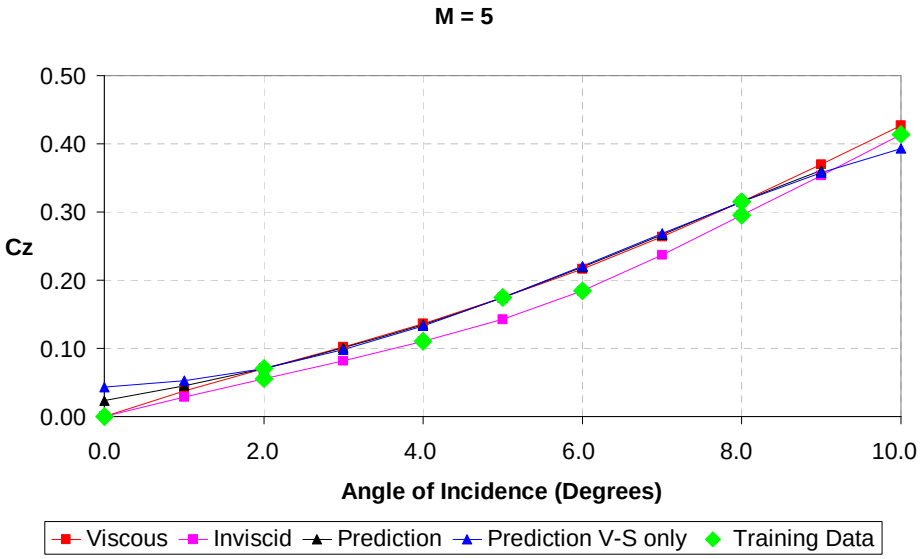
(b)


M = 4
0.28
0.27
Cx 0.26
0.25
0.24
0.0
2.0
4.0
6.0
8.0
10.0
Angle of Incidence (Degrees)
Viscous
Inviscid
Prediction
Prediction V-S only
Training Data

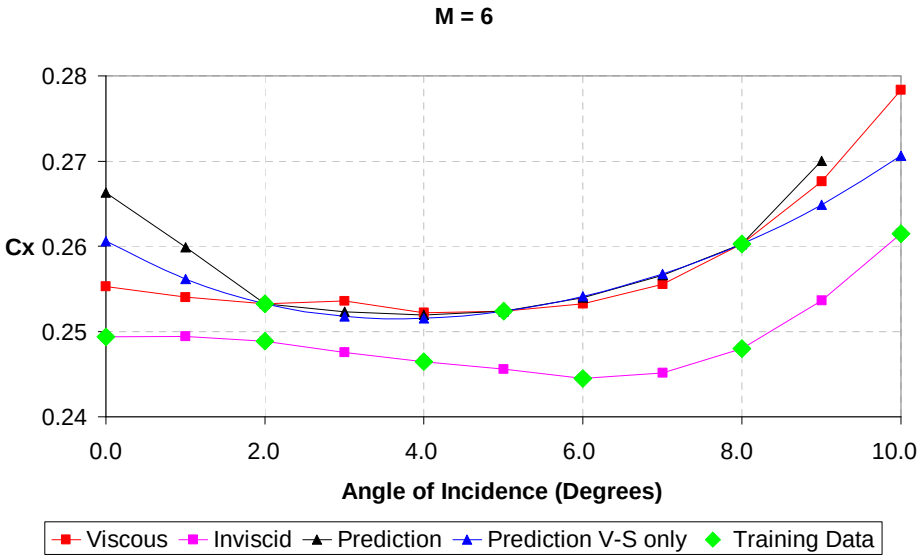

M = 4
0.40
0.30
Cz 0.20
0.10
0.00
0.0
2.0
4.0
6.0
8.0
10.0
Angle of Incidence (Degrees)
Viscous
Inviscid
Prediction
Prediction V-S only
Training Data

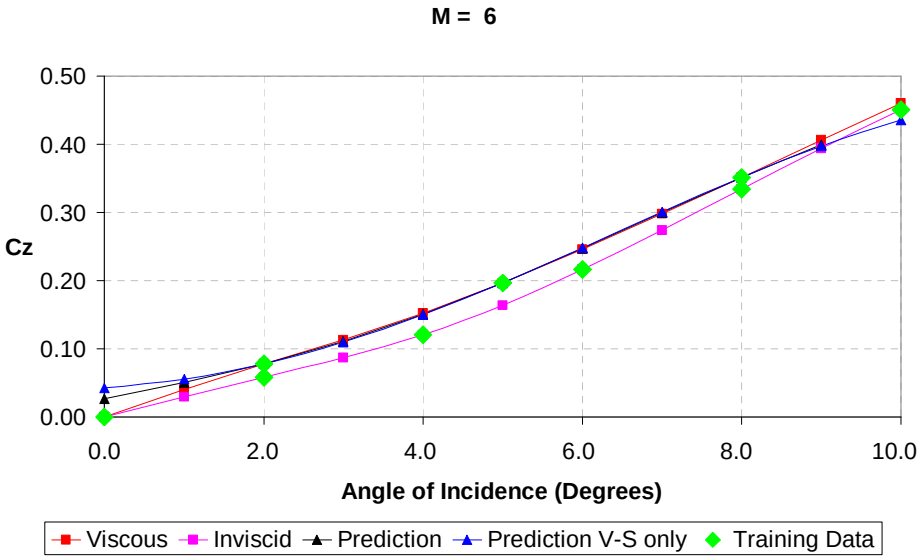
(c)


M = 5
0.28
0.27
Cx 0.26
0.25
0.24
0.0
2.0
4.0
6.0
8.0
10.0
Angle of Incidence (Degrees)
Viscous
Inviscid
Prediction
Prediction V-S only
Training Data


M = 5
0.50
0.40
0.30
Cz
0.20
0.10
0.00
0.0
2.0
4.0
6.0
8.0
10.0
Angle of Incidence (Degrees)
Viscous
Inviscid
Prediction
Prediction V-S only
Training Data

(d)



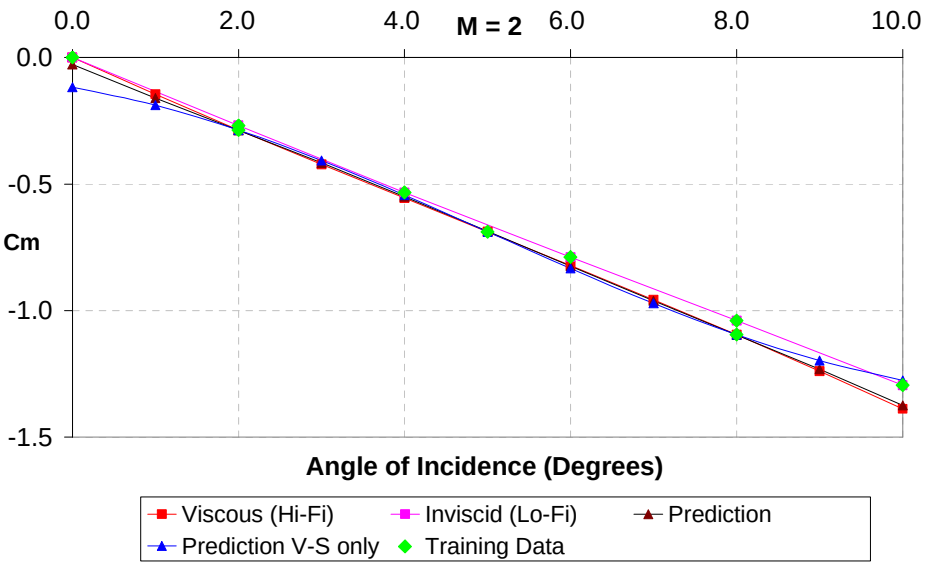


(e)

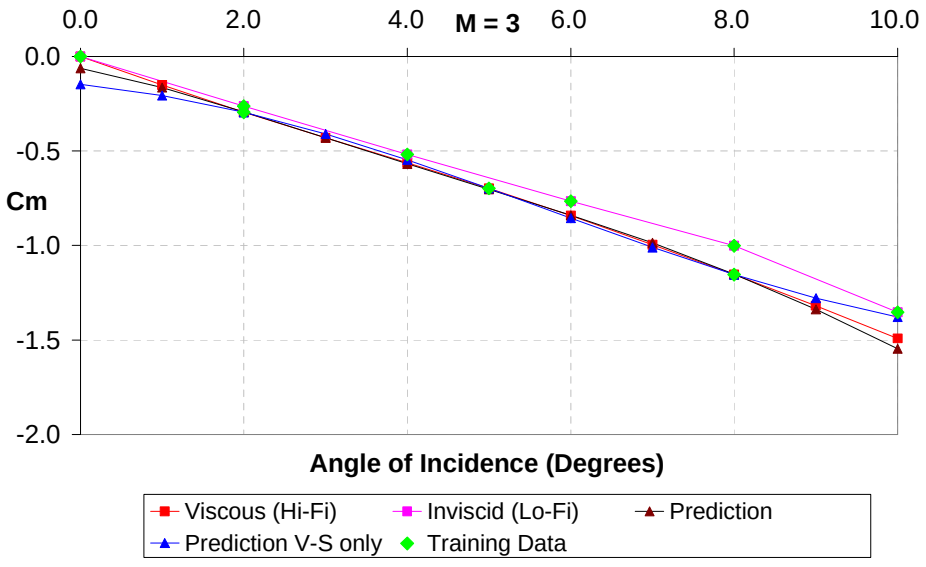
Figure (7-22) – Prediction of C_x and C_z

Plots of the axial and normal force coefficients are displayed in figure (7-22). In the figures the highlighted green points are the training data points (input) which are used by the variable-fidelity/POD-based reduced-order modelling method to effect the prediction. The predicted values (output) from the model are in black while the high-fidelity set of data which is in red are included so that a comparison can be made between the prediction (black line) and the high-fidelity data (red line). The blue line is the prediction obtained by a POD-based ROM derived from the viscous snapshots of the same problem, ie. from the fifteen viscous snapshots only. Generally, the predictions from the variable-fidelity/POD-based ROM method are accurate and better than the predictions from the POD-based ROM, except for two small regions as revealed by the axial force coefficient. One of these regions is at M = 3 and at an angle of incidence greater than 8º. Here the prediction follows the trend of the low-fidelity data which diverts away from the high-fidelity set of data. The other region is at M = 6 and an angle of incidence of less than 2º. Here the main source of error is the viscous part of the axial force coefficient which was determined very poorly. Note that the predicted normal force coefficient does not pass through the point (0,0) since this technique is only an approximation. However, it can be forced to go through (0,0) by including a snapshot at this point.

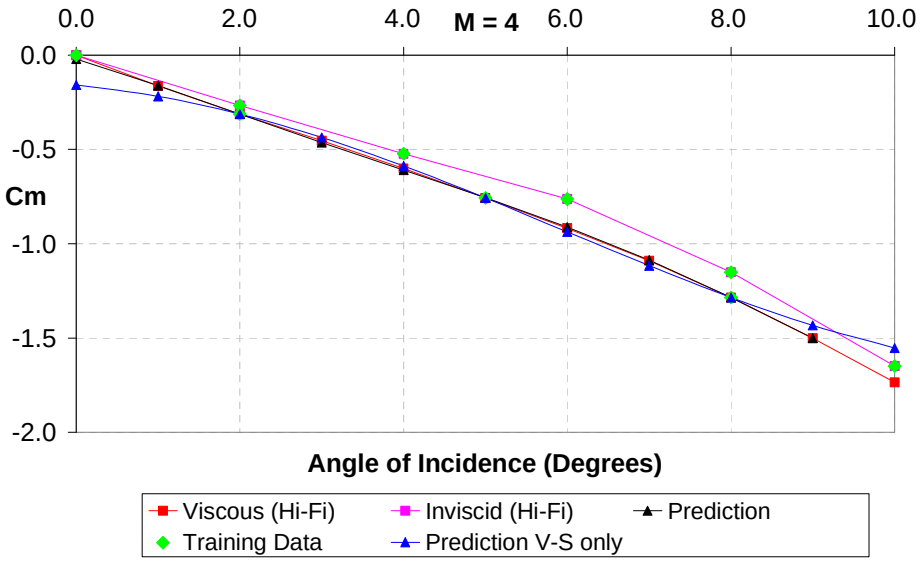
Figure (7-23) shows the variation of the pitching moment with angle of incidence at the various Mach numbers considered. The agreement between the variable-fidelity/POD-based ROM and the high-fidelity set of data is good. Moreover, the prediction from the variable-fidelity/POD-based ROM (black line) is better than that of the POD-based ROM developed from the viscous snapshots only (blue line) both within the region of the high-fidelity set of data and beyond it.


0.0
2.0
4.0
M = 2
6.0
8.0
10.0
0.0
-0.5
Cm
-1.0
-1.5
Angle of Incidence (Degrees)
Viscous (Hi-Fi)
Inviscid (Lo-Fi)
Prediction
Prediction V-S only
Training Data

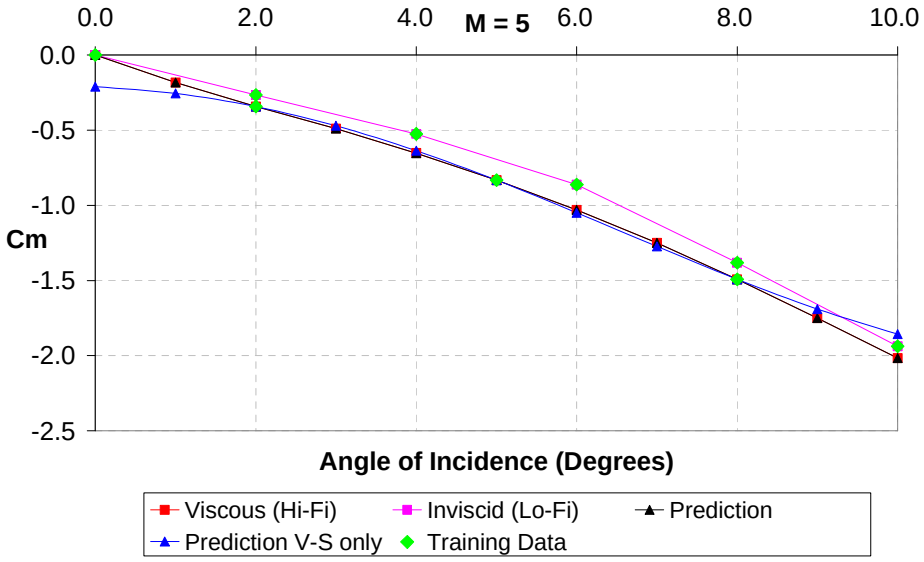
(a)


0.0
2.0
4.0
M = 3
6.0
8.0
10.0
0.0
-0.5
Cm
-1.0
-1.5
-2.0
Angle of Incidence (Degrees)
Viscous (Hi-Fi)
Inviscid (Lo-Fi)
Prediction
Prediction V-S only
Training Data

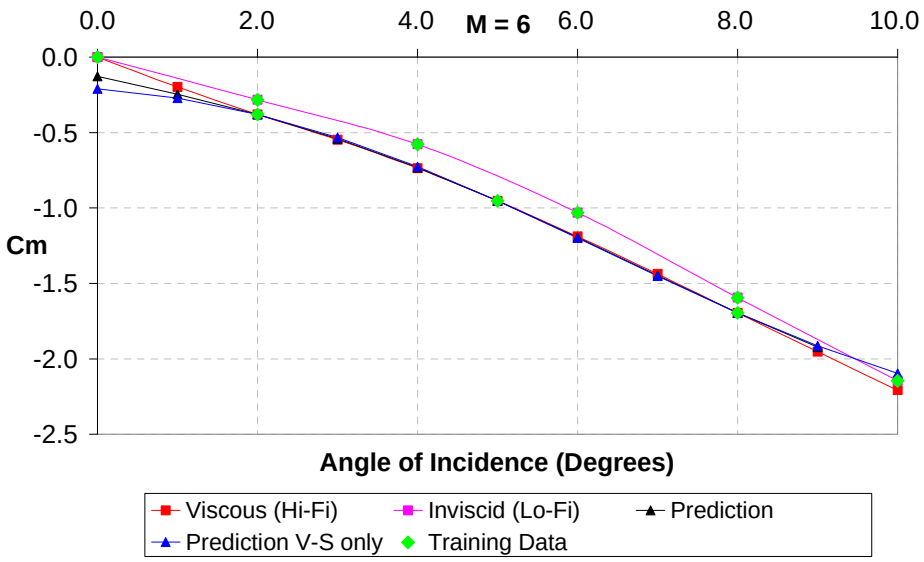
(b)


0.0
2.0
4.0
M = 4
6.0
8.0
10.0
0.0
-0.5
Cm
-1.0
-1.5
-2.0
Angle of Incidence (Degrees)
Viscous (Hi-Fi)
Inviscid (Hi-Fi)
Prediction
Training Data
Prediction V-S only

(c)


0.0
2.0
4.0
M = 5
6.0
8.0
10.0
0.0
-0.5
-1.0
Cm
-1.5
-2.0
-2.5
Angle of Incidence (Degrees)
Viscous (Hi-Fi)
Inviscid (Lo-Fi)
Prediction
Prediction V-S only
Training Data

(d)

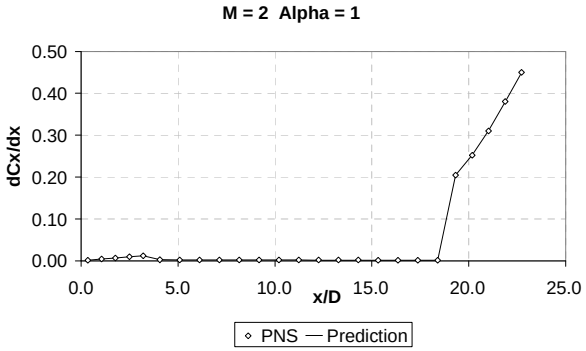


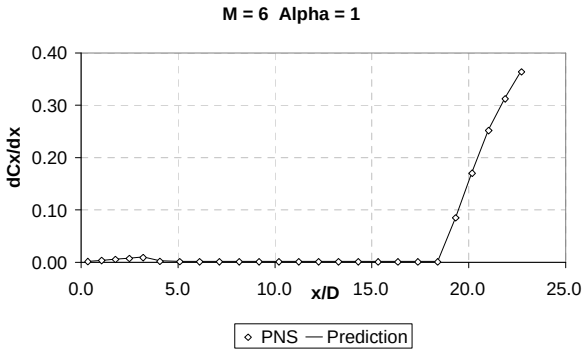
(e)

Figure (7-23) – Prediction of C_m

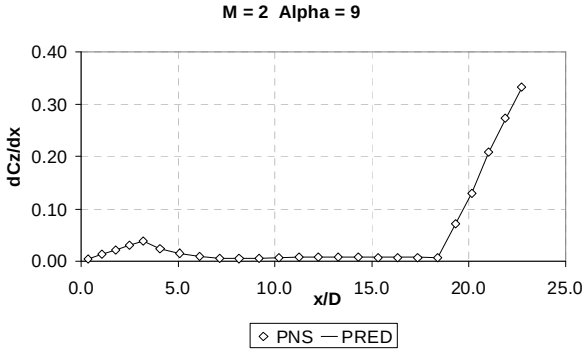
In this section the same observations were perceived as in the previous one. The technique offers a general improvement in the prediction. In the region beyond the high-fidelity data points, the prediction is normally better than that performed using the viscous snapshots only. Where this is not the case it is due to the fact that the prediction follows the same trend of the inviscid computations which have a trend which is slightly different from the high fidelity one. This trait is evident most especially from the axial force coefficients. From the normal force coefficients it is noticeable that the proposed methodology offers improvement both beyond and within the high-fidelity data. It is evident that the prediction within the high-fidelity data points region is damped, ie. one could not observe the oscillations predicted using the viscous snapshots only. So it is quite conspicuous that an improvement is obtainable by using this technique.

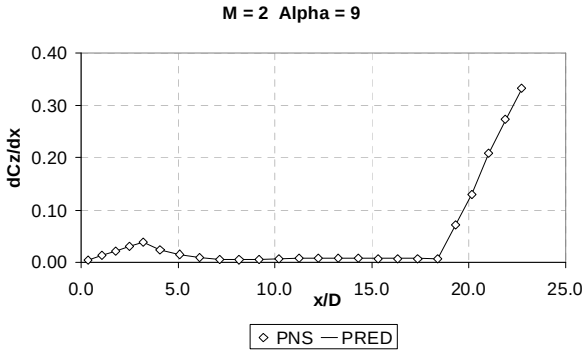
Plots of the axial and normal force derivatives are shown in figures (7-24) and (7-25) at two different Mach numbers which correspond to the extreme values. The angles of incidence were chosen such that one set will be enclosed by the high fidelity data while the other sets will be beyond the high-fidelity data but will be enclosed by the low-fidelity (inviscid) solutions. Again the agreement between the predicted derivatives and the IMPNS computed ones is good.


M = 2 Alpha = 1
dCx/dx
0.50
0.40
0.30
0.20
0.10
0.00
0.0
5.0
10.0
15.0
20.0
25.0
x/D
PNS
Prediction

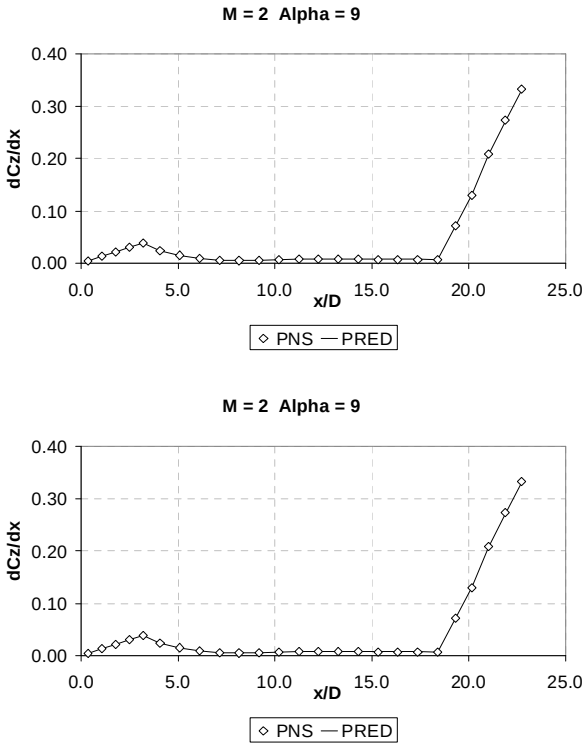

M = 6 Alpha = 1
dCx/dx
0.40
0.30
0.20
0.10
0.00
0.0
5.0
10.0
15.0
20.0
25.0
x/D
PNS
Prediction

(a)


M = 2 Alpha = 9
dCz/dx
0.40
0.30
0.20
0.10
0.00
0.0
5.0
10.0
15.0
20.0
25.0
x/D
PNS
PRED

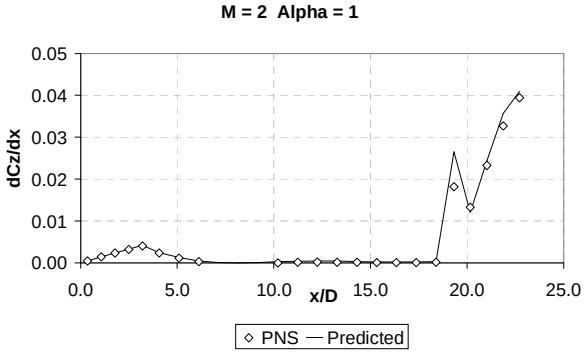

M = 2 Alpha = 9
dCz/dx
0.40
0.30
0.20
0.10
0.00
0.0
5.0
10.0
15.0
20.0
25.0
x/D
PNS
PRED

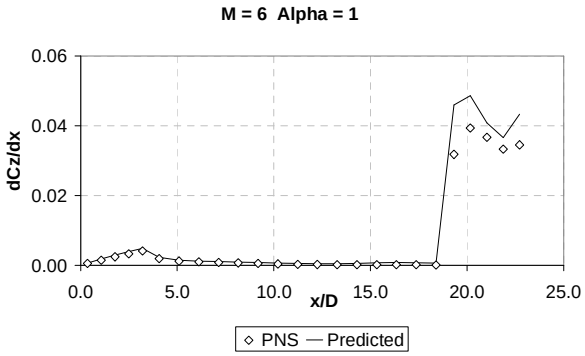
(b)



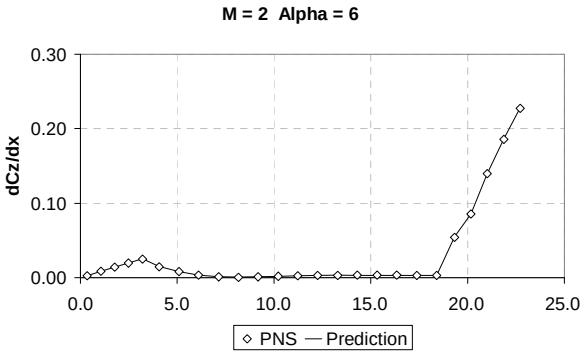
(c)

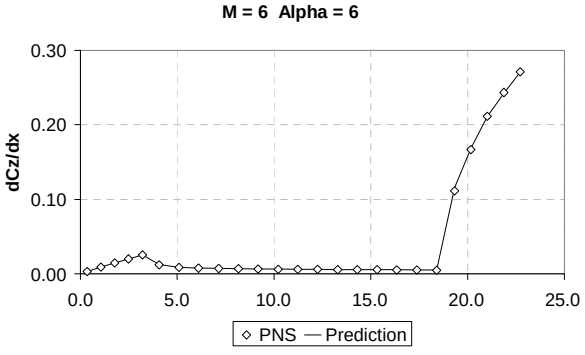
Figure (7-24) Plots of the axial force derivative against axial distance in calibres


M = 2 Alpha = 1
dCz/dx
0.05
0.04
0.03
0.02
0.01
0.00
0.0
5.0
10.0
15.0
20.0
25.0
x/D
PNS
Predicted

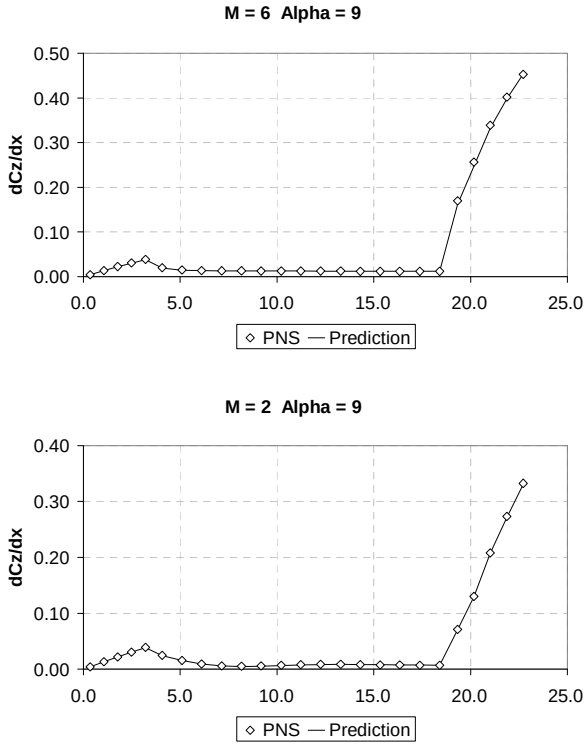

M = 6 Alpha = 1
dCz/dx
0.06
0.04
0.02
0.00
0.0
5.0
10.0
15.0
20.0
25.0
x/D
PNS
Predicted

(a)


M = 2 Alpha = 6
dCz/dx
0.30
0.20
0.10
0.00
0.0
5.0
10.0
15.0
20.0
25.0
PNS
Prediction


M = 6 Alpha = 6
dCz/dx
0.30
0.20
0.10
0.00
0.0
5.0
10.0
15.0
20.0
25.0
PNS
Prediction

(b)

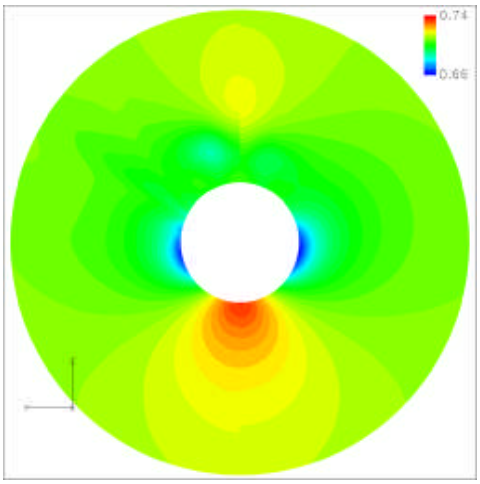


(c)

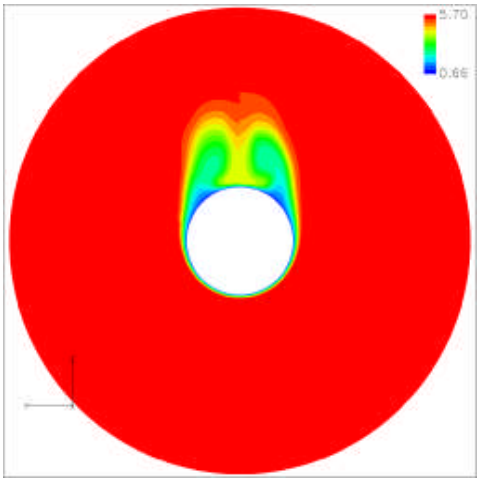
Figure (7-25) Plots of the normal force derivative versus the axial distance in calibres

In particular, the prevailing advantage from this method is that the entire flow-field can be predicted in contrast with other variable-fidelity models which are available in the literature. Consequently, flow-field comparisons are once again presented between the predictions obtained by the variable-fidelity/POD-based ROMs method and the high-fidelity simulations. In this case comparisons of the static pressure, total pressure and entropy at different axial locations are made. The axial locations considered were (i) at the end of the cylindrical part that is x/D = 18.4 where the vortical flow is well developed and (ii) at the very end of the projectile where one could observe the vortical flow and the flare's shock interact. The comparisons were made at the two extreme Mach number values and again at incidences of 6° and 9°. In all the figures, the left-hand side part of the figure is obtained from the approximation method while the right-hand side is

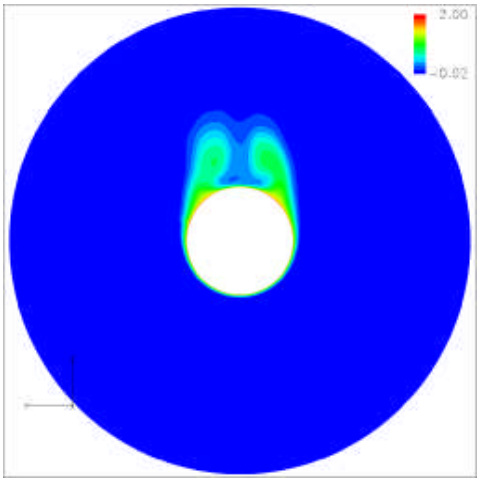
obtained from CFD. The comparisons between the predicted flooded contours and the high-fidelity CFD calculated ones vary from reasonably comparable to relatively poor. Refer to figures (7-26) to (7-31). These figures together with the previous force coefficients' charts suggest that the main source of error in these predictions is due to inaccurate prediction of the velocities.



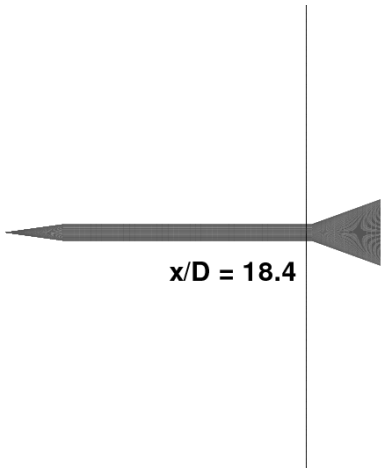
(a) Static pressure contours at an axial location of 18.4



(b) Total pressure contours at an axial location of 18.4

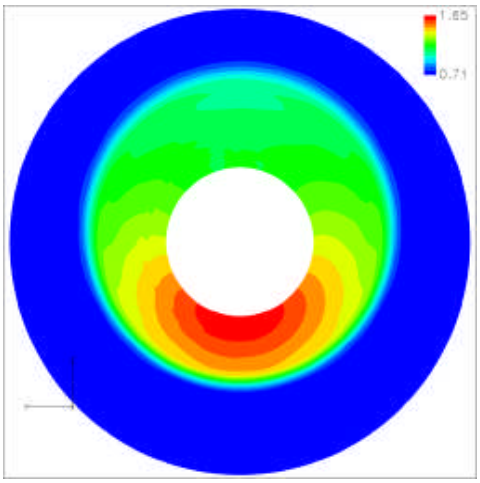


(c) Entropy contours at an axial location of 18.4

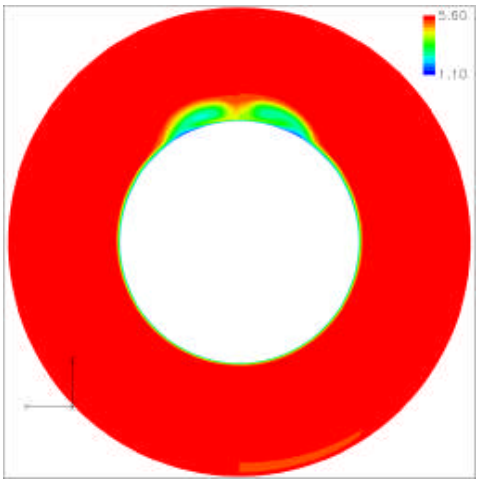


(d) x/D = 18.4

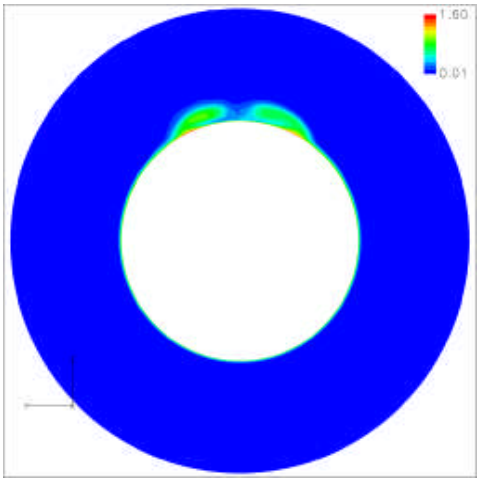
Figure (7-26) Comparison at M = 2 and Alpha = 6° between approximate method on the left-hand half part of the figure and PNS solution on the right-hand part



(a) Static pressure contours at the end of the projectile

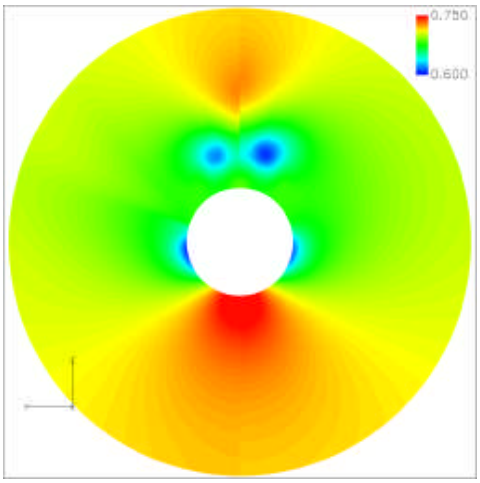


(b) Total pressure contours at the end of the projectile

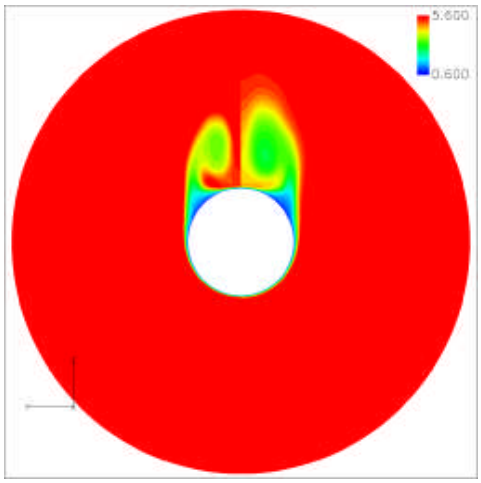


(c) Entropy contours at the end of the projectile

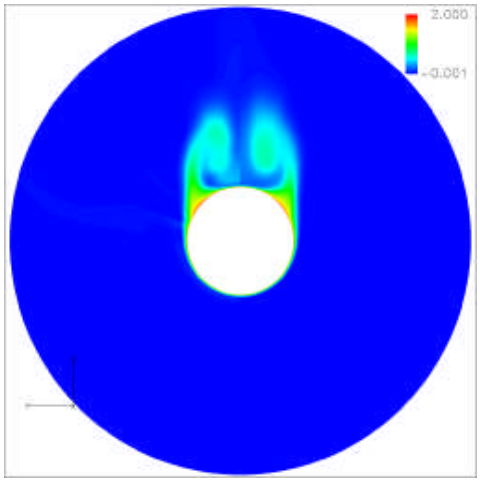
Figure (7-27) Comparison at M = 2 and Alpha = 6° between approximate method on the left-hand half part of the figure and PNS solution on the right-hand part



(a) Static pressure contours at an axial location of 18.4

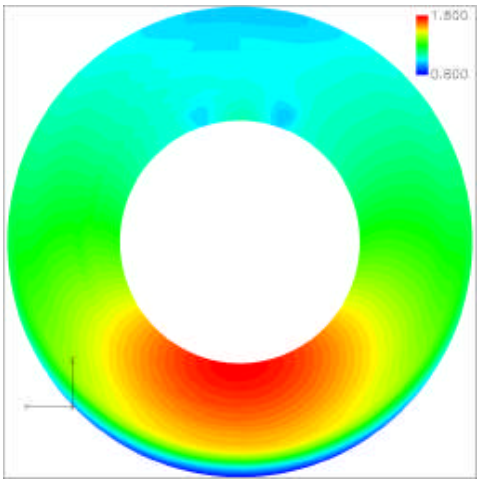


(b) Total pressure contours at an axial location of 18.4

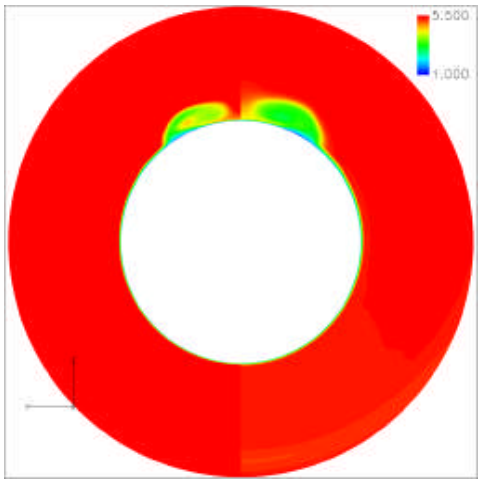


(c) Entropy contours at an axial location of 18.4

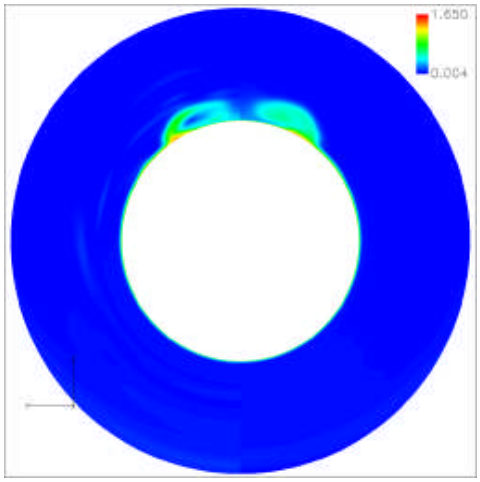
Figure (7-28) Comparison at M = 2 and Alpha = 9° between approximate method on the left-hand half part of the figure and PNS solution on the right-hand part



(a) Static pressure contours at the end of the projectile

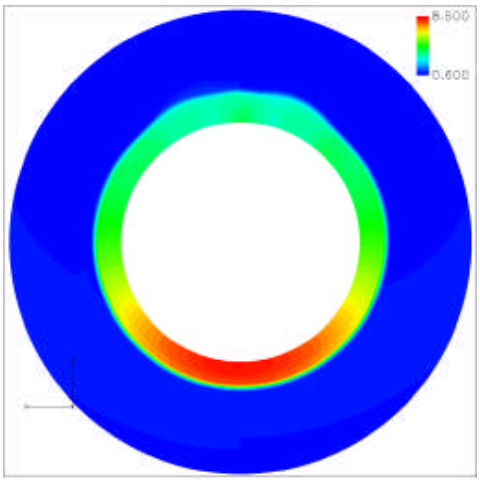


(b) Total pressure contours at the end of the projectile

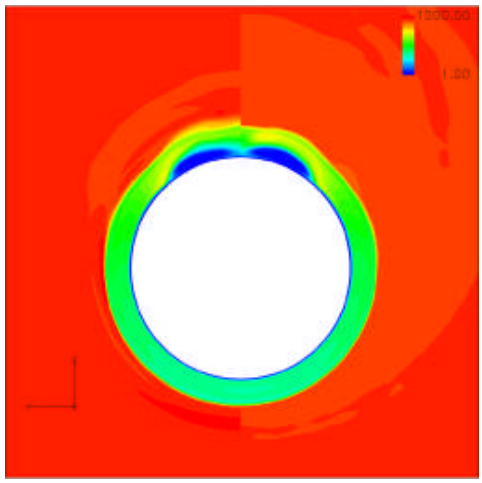


(c) Entropy contours at the end of the projectile

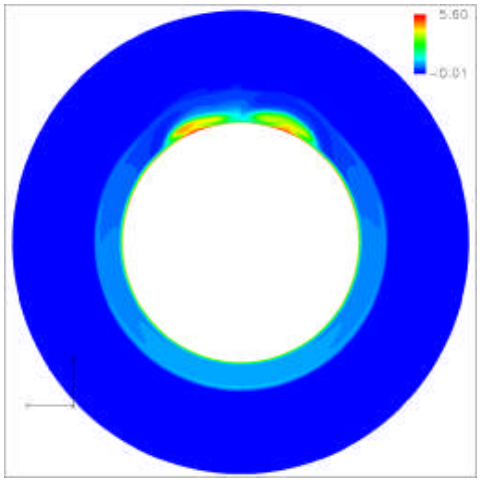
Figure (7-29) Comparison at M = 2 and Alpha = 9° between approximate method on the left-hand half part of the figure and PNS solution on the right-hand part



(a) Static pressure contours at the end of the projectile

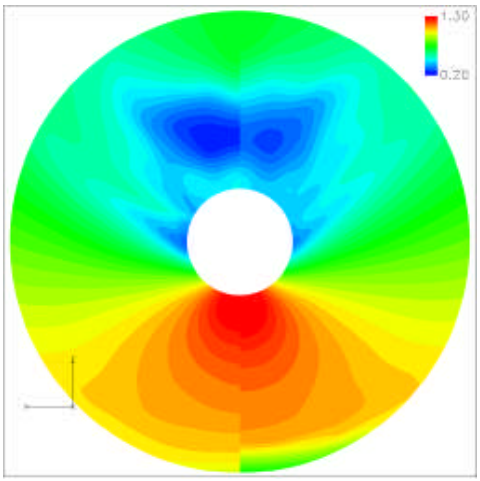


(b) Total pressure contours at the end of the projectile

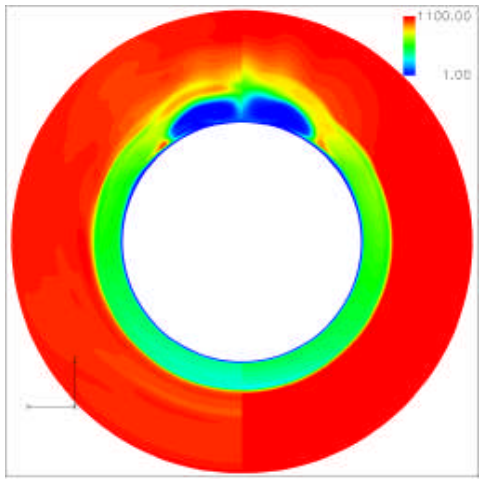


(c)Entropy contours at the end of the projectile

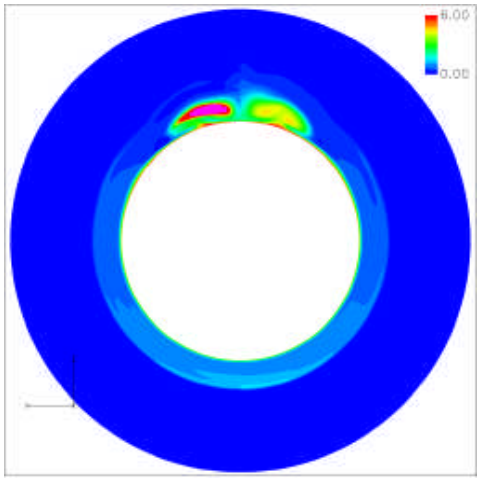
Figure (7-30) Comparison at M = 6 and Alpha = 6° between approximate method on the left-hand half part of the figure and PNS solution on the right-hand part



(a) Static pressure contours at the end of the projectile



(b) Total pressure contours at the end of the projectile



(c)Entropy contours at the end of the projectile

Figure (7-31) Comparison at M = 6 and Alpha = 9° between approximate method on the left-hand half part of the figure and PNS solution on the right-hand part

In general, this procedure offers the significant advantage that the whole prediction process is automatic. This implies that instead of fitting or tuning a radial basis function to represent as accurately as possible the response surface of the projection coefficients, the problem is addressed in terms of the inviscid calculations. The response surface is generated by considering a number of inviscid snapshots while the tuning is achieved by the few viscous snapshots considered. However, a drawback that this technique may have is that in order to increase the influence of the trend of the low-fidelity data upon the high-fidelity one, the separation between the two response surfaces may require some form of adjustment by controlling the magnitude of ε. For example in this case the auxiliary variable ε was varied from 0.2 to 1.0 as the Mach number was increased from 2 to 6.

In addition this technique is on the one hand less robust when compared with that of the previous task, ie. fusing low-order and high-order accurate data, where the projection coefficients of all the primitive variables are fused together. In fact some integrated coefficients at the extreme ends of the parametric space of the inviscid/viscous data fusion were not evaluated. On the other hand this technique is more accurate than if the method of section 7.6.1 was adopted.

7.6.3 Investigating the behaviour of the predicted response when the number of inviscid computations is increased

A study was conducted to investigate what would happen on increasing the number of inviscid snapshots. From this investigation it was observed that when the inviscid snapshots were increased from 30 to 55 in total, such that the inviscid snapshots are at intervals of 1º within the range [0 º, 10 º] for the angle of incidence and the Mach number is again at intervals of 1 within the range [2, 6], there was a very slight improvement almost insignificant in the prediction. This implies that when the pseudo-representation is already offering a good representation with the original number of snapshots, then no significant further improvement can be achieved from considering a greater number of inviscid snapshots.

7.7 Comparison with Robinson's POD mapping variable-fidelity model

In the literature, variable-fidelity models using POD are not common but one in particular was suggested by Robinson *et al.* in [127] and [128]. This variable-fidelity model is based on the missing point or gappy POD method for the reconstruction of incomplete datasets. A detailed description of this POD mapping variable-fidelity model is made in Appendix C. In this section a qualitative comparison between the method proposed in this thesis and the one proposed by Robinson *et al*. is made.

A drawback of Robinson's *et al.* model is that it requires datasets be generated in pairs ie., the low-fidelity and high-fidelity training points must be at the same input parameters, which is not a necessity for the proposed model in this thesis. Moreover, the method of Robinson does not allow for problems with different boundary conditions. Therefore, it follows that the proposed model in this thesis offers more flexibility. Robinson's model will start to be accurate when the number of high-fidelity training data is considerably higher than the model proposed in this thesis, thus making it less practical. Although it must be emphasized that the accuracy level reached by Robinson's method is higher than the one proposed here, as the accuracy of the latter would be determined by the response surface approximation. In addition, the prediction from the variable-fidelity POD-based ROM method follows very closely the trend of the low-fidelity training data points and so it relies to a great extent on it, most especially when the high-fidelity training data is sparse.

7.8 Concluding remarks

In this chapter a hierarchical type/POD-based ROM model is proposed and applied to two high-speed weapon aerodynamics problems. In this method a hierarchical type model which incorporates POD-based ROMs is constructed which makes use of interpolation schemes for dealing with extrapolatory problems. In general, this new method is composed of two models ie. the variable-fidelity model of Reisenthel *et al.* and two POD-based ROMs, one is constructed with the high-fidelity data while the other is constructed with the low-fidelity one. In particular it was observed that this model behaves similarly to the variable-fidelity model of Reisenthel *et al*., that is, the prediction follows very closely the trend of the low-fidelity training data points and so it relies to a great extent on it, most especially when the high-fidelity training data is sparse. However, the predictions get more and more accurate as the number of high-fidelity data points is increased. Besides this, the POD-based reduced-order modelling limitations do persist within the proposed methodology.

In section 7.6.1 the variable-fidelity problem considered was set-up between low-order and high-order accurate solutions. Therefore, in this case both the physics and the boundary conditions of the variable-fidelity data were similar. The model was applied directly on all the primitive variables and very good predictions were achieved. Thus, it was demonstrated that this technique works well when dealing with the same physical conditions throughout the whole set of data. Similarly, the same method should work well for problems between coarse and fine grids. In such cases some form of interpolation would be necessary to evaluate the variables at those fine grid points which fall at intermediate places. But this should not impose any problem at all as linear interpolation will do the job.

In section 7.6.2 the variable-fidelity problem considered was set-up between inviscid and viscous solutions. Consequently, both the physics and the boundary condition at the wall of the variable-fidelity solutions are different in this case. Although it would still be possible to fuse inviscid and viscous data in the way suggested in section 7.6.1, the

problem would not be well posed. Therefore, the data fusion technique adopted in section 7.6.1 was applied on the pressure and the density variables only. The velocities were then predicted using the POD-based ROM for the viscous snapshots only. This method predicts accurate force coefficients but flow-fields with relatively low accuracy.

In general, this procedure offers the significant advantage that the prediction process is automatic. This implies that instead of fitting or tuning a radial basis function to represent as accurately as possible the response surface of the projection coefficients, the problem is addressed in terms of the low-cost and low-fidelity calculations. The response surface is generated by considering a considerable number of low-fidelity snapshots while the calibration is then achieved by a few viscous snapshots. Potentially this method offers a reduction in the up-front cost required to generate the training dataset for a POD-based ROM.

Chapter 8

8.0 The kernel POD

The preceding chapters dealt with the common and most popular form of POD. This is essentially a linear reduced basis method which is suitable for non-linear problems provided a suitably rich set of basis vectors is available. This technique is analogous with the many numerical algorithms used for analyzing non-linear systems which are based on the solution of a sequence of linear sub-problems. However, for strong non-linear problems such reduced basis method would require the generation of a large number of snapshots to create a rich set of basis vectors which capture and model the strong non-linearity appropriately. In this chapter a different approach for modelling non-linear aerodynamic problems is considered and investigated. An attempt is made to model both the weak and strong non-linear features, particularly in the transonic flow regime. In this approach a non-linear kernel is integrated into the linear POD algorithm to transform the linear algorithm into a non-linear one, hence making it possible to model the non-linear features in the input space with a smaller number of basis vectors. The basic idea behind this method is to map the data in the input space to a feature space via some non-linear map, and then apply the linear method in the feature space for further analysis. In this chapter, an aerodynamic application of this technique is presented. It must be pointed out that this chapter deals with the relatively recent research niche of kernel methods in machine learning. In the literature this technique is generally referred to as kernel PCA, however in this work the term kernel POD has been introduced and used interchangeably with the former one. Interestingly, the term PCA is normally used by mathematicians, while the term POD is used by engineers.

8.1 Introduction to kernels and kernel methods

In the field of data classification, data is separated into different classes by a plane in the input space. Sometimes it is an easy task to correctly separate the data when it is linearly separable. However, it may be impossible to separate the data without errors if the data is non-linearly separable as shown in figure (8-1). In this case, some curve such as an ellipse is required to correctly separate the crosses and circles in $\Re^2$ that is, in two-dimensions.

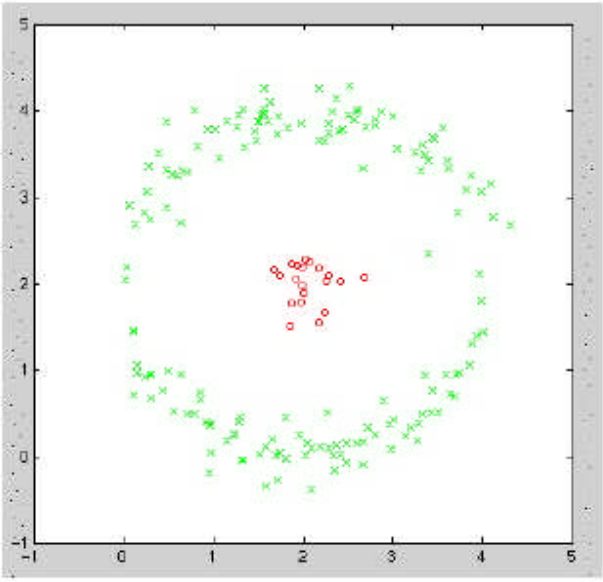


Figure (8-1) Data in $\Re^2$

It is obvious that the circles and the crosses cannot be correctly separated by a straight line but they can be separated by a plane when the data is mapped into a higher dimension $\Re^3$ by some mapping function as shown in figure (8-2). In this example, the mapping is performed according to the following:

$$\Phi : \boldsymbol{X} = \Re^2 \rightarrow \boldsymbol{F} = \Re^3$$
$$(x, y) \mapsto (x^2, y^2, x^2 + y^2)$$

Thus, a non-linearly separable case in the input space is converted into a linearly separable case in the feature space. The feature space is the space a mapping function maps to.

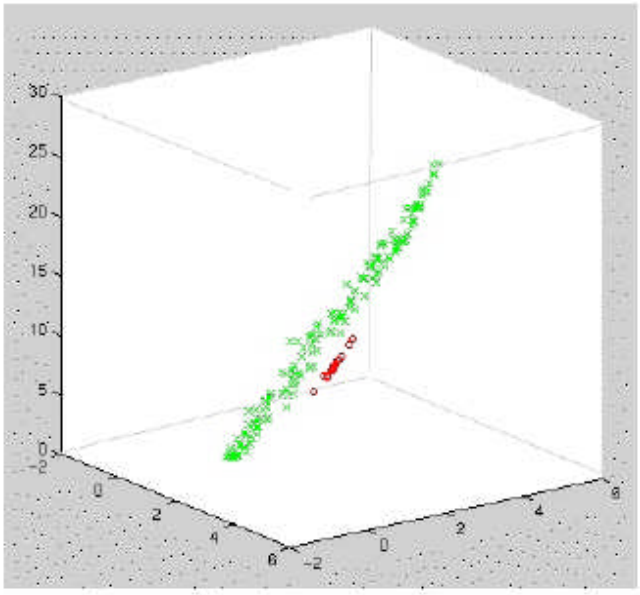


Figure (8-2) Data mapped into $\Re^3$

In multi-dimensional datasets, a hyper-plane is always required to separate the data and according to the Vapnik-Chervonenkis [129] dimension theorem, theoretically $n+1$ dimensional data can be separated in $\Re^n$. Therefore, the higher the dimension some data is mapped to, the more data can be correctly separated in the feature space. Theoretically, it is possible to map data into infinite dimensions to correctly separate infinite data. From this, a question crops up. How can such a mapping function be implemented? Definitely, it is impossible to do any calculation on a vector with infinite elements and store this in memory. Luckily, in many classifiers, a dot product between two data vectors can be implemented. But still if the vectors are infinitely long, this cannot be performed. Consequently a function is needed to handle this situation and take advantage of dot products. Kernel functions are functions which make it possible to calculate infinite or high–dimensional inner (dot) products [130][131]. Kernels are defined as follows:

A kernel is a function K such that for all $\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{X}$

$$K(\boldsymbol{x}, \boldsymbol{y}) = \Phi(\boldsymbol{x}).\Phi(\boldsymbol{y}), \qquad \text{Eq.(8-1)}$$

where Φ is a mapping from $\boldsymbol{X}$ to an inner product feature space $\boldsymbol{F}$, also referred to as dot product space.

From this function an infinite or highly-dimensional inner product of the right-hand side of Eq. (8-1) is evaluated by some n times inner products of the left-hand side of Eq. (8-1) if $\boldsymbol{x}, \boldsymbol{y} \in \Re^n$. Therefore, the kernel function is a very convenient form for the implementation of a mapping function.

In recent years there has been a lot of interest in the study of kernel methods [130] [131]. The basic idea is to map the data in the input space $\boldsymbol{X}$ to a feature space $\boldsymbol{F}$ via some non-linear map function Φ, and after that a linear method is applied in the feature space. This computational procedure depends on the inner products $\Phi(\boldsymbol{x}).\Phi(\boldsymbol{y})$ in the feature space (where $\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{X}$), which are obtained efficiently from a suitable kernel function. Besides this kernel methods have the important computational advantage that non-convex non-linear optimization is not involved, thus provide an elegant non-linear generalization of many existing linear algorithms. Therefore, kernels offer a powerful tool to extend the use of linear algorithms to non-linear problems and there are some very good performing linear algorithms which can also be used in the non-linear case by the application of kernels. Some well-known examples are support vector machines in supervised learning and kernel PCA in unsupervised learning. [132]

Another important advantage of kernels is that they simplify the representation problem. Kernels simplify the representation because a direct map to the feature space is not required but only a positive definite similarity measure is required. Another advantage is that data from different sources are represented in one framework and consequently will be easier to combine the data.

8.2 Various types of kernels

An important issue in kernel POD lies in choosing the kernel function $K(\mathbf{x},\mathbf{y})$ or otherwise specifying the kernel matrix K_{ij}. Some widely used kernels are the linear which is essentially POD, polynomial, Gaussian and sigmoidal kernels, [131] given by

i. The linear kernel

$$K(\mathbf{x},\mathbf{y}) = \mathbf{x}.\mathbf{y} \qquad \text{Eq.(8-2)}$$

which simply identifies the feature space with the input space;

ii. The polynomial kernel is defined as

$$K(\mathbf{x},\mathbf{y}) = ((\mathbf{x}.\mathbf{y}))^{p_1} \qquad \text{Eq.(8-3)}$$

or

$$K(\mathbf{x},\mathbf{y}) = ((\mathbf{x}.\mathbf{y}) + p_2)^{p_1} \qquad \text{Eq.(8-4)}$$

where p_1 and p_2 are both integers. p_1 is the degree of the polynomial. The former polynomial kernel of degree 2 corresponds to a feature space spanned by all products of two variables, that is, (x^2, xy, y^2). The other polynomial kernel of degree 2 corresponds to a feature space spanned by all products of at most 2 variables, that is, $(1, x, y, x^2, xy, y^2)$. More generally the former kernel corresponds to a feature space spanned by all products of exactly p_1 variables, while the latter corresponds to a feature space spanned by all products of at most p_1 variables. The following is an example of a polynomial kernel. Assuming $p_1 = 2$ and $p_2 = 0$ in Eq. (8-4), then

$$(\mathbf{x},\mathbf{y})^2 = \left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}\right)^2 = (x_1 y_1 + x_2 y_2)^2 = (x_1^2 y_1^2 + 2x_1 x_2 y_1 y_2 + x_2^2 y_2^2)$$

$$(\mathbf{x}.\mathbf{y})^2 = \left(\begin{bmatrix} x_1^2 \\ x_2^2 \\ \sqrt{2}x_1x_2 \end{bmatrix}, \begin{bmatrix} y_1^2 \\ y_2^2 \\ \sqrt{2}y_1y_2 \end{bmatrix} \right) = \Phi(\mathbf{x}).\Phi(\mathbf{y}) = \left(x_1^2y_1^2 + 2x_1x_2y_1y_2 + x_2^2y_2^2 \right)$$

iii. The radial basis function RBF kernel is defined as

$$K(\mathbf{x},\mathbf{y}) = e^{\frac{-\|\mathbf{x}-\mathbf{y}\|^2}{2\sigma^2}} \qquad \text{Eq.(8-5)}$$

which maps the inputs onto the surface of an infinite-dimensional sphere. The following is a proof showing that the Gaussian kernels are indeed infinitely dimensional.

Proof:

$$K(\mathbf{x},\mathbf{y}) = e^{\frac{-\|\mathbf{x}-\mathbf{y}\|^2}{2\sigma^2}} = e^{\frac{-\left(\mathbf{x}^2-2\mathbf{x}\mathbf{y}+\mathbf{y}^2\right)}{2\sigma^2}} = e^{\frac{-\mathbf{x}^2}{2\sigma^2}} e^{\frac{2\mathbf{x}\mathbf{y}}{2\sigma^2}} e^{\frac{-\mathbf{y}^2}{2\sigma^2}}$$

$$= e^{\frac{-\mathbf{x}^2}{2\sigma^2}} \left(1 + \frac{2xy/2\sigma^2}{1!} + \frac{\left(2xy/2\sigma^2\right)^2}{2!} + \ldots\ldots + \frac{\left(2xy/2\sigma^2\right)^n}{n!} + \ldots \right) e^{\frac{-\mathbf{y}^2}{2\sigma^2}}$$

$$= e^{\frac{-\mathbf{x}^2}{2\sigma^2}} \left(1 + \frac{\sqrt{\left(1/\sigma^2\right)}x}{\sqrt{1!}} + \ldots + \frac{\sqrt{\left(1/\sigma^2\right)^n}x^n}{\sqrt{n!}} \right) \left(1 + \frac{\sqrt{\left(1/\sigma^2\right)}y}{\sqrt{1!}} + \ldots + \frac{\sqrt{\left(1/\sigma^2\right)^n}y^n}{\sqrt{n!}} \right) e^{\frac{-\mathbf{y}^2}{2\sigma^2}}$$

$$= \Phi(\mathbf{x})\,\Phi(\mathbf{y})$$

This type of kernel is considered as a general purpose kernel in the machine learning field.

iv. The sigmoid kernel is defined by

$$K(\mathbf{x},\mathbf{y}) = \text{Tanh}\left(\kappa\,(\mathbf{x}.\mathbf{y}) + \theta\right) \qquad \text{Eq.(8-6)}$$

for suitable values of gain $\kappa > 0$ and threshold $\theta < 0$. The sigmoid kernel is not always positive definite, but it is still used in practice.

8.3 The Kernel POD: Theory and Explanation

Schoelkopf, Smola and Mueller [132] introduced the kernel PCA as a non-linear generalization of PCA and therefore somehow it can be used in the reconstruction, compression and de-noising of data, which are common applications of PCA. Linear PCA assumes a Gaussian distribution for the variations in the training data set but in reality the distribution could be more complex and so it is better captured using non-linear methods like kernel PCA.

The generalization is achieved by mapping the original input space into a higher and possibly infinite dimensional feature space ***F*** before extracting the linear basis functions. In particular, consider snapshots $\mathbf{x}_1,.....,\mathbf{x}_N \in R^D$ and features $\Phi(\mathbf{x}_1),\ldots.,\Phi(\mathbf{x}_N) \in \boldsymbol{F}$ computed by some mapping $\Phi: R^D \rightarrow \boldsymbol{F}$. Kernel PCA is based on the insight that the basis functions in ***F*** can be computed for mappings $\Phi(\mathbf{x})$ that are only implicitly defined by specifying the inner product in feature space, that is the kernel function $K(\mathbf{x},\mathbf{y}) = \Phi(\mathbf{x}).\Phi(\mathbf{y})$.

To perform the POD in feature space, the eigenvalues $\lambda > 0$ and eigenvectors $\mathbf{V} \neq \mathbf{0} \in \boldsymbol{F}$ must satisfy $\lambda\mathbf{V} = \overline{C}\,\mathbf{V}$ where $\overline{C} = \Phi(\mathbf{x}_k).\Phi(\mathbf{x}_k)^T$. For simplicity it is assumed that the mapped data is centered in ***F***. When substituting $\overline{C}$ into the eigenvector equation this shows that all solutions $\mathbf{V}$ must lie in the span of Φ - mappings of the training data. This implies that we can consider the equivalent system

$$\lambda\left(\Phi(\mathbf{x}_k).\mathbf{V}\right) = \left(\Phi(\mathbf{x}_k).\overline{C}\mathbf{V}\right) \text{ for all } k = 1, \ldots\ldots, l \qquad \text{Eq.(8-7)}$$

and that there exist coefficients $\alpha_1,\alpha_2,.....,\alpha_l$ such that

$$\mathbf{V} = \sum_{i=1}^{l} \alpha_i \Phi(\mathbf{x}_i) \qquad \text{Eq.(8-8)}$$

Substituting $\overline{C}$ and Eq. (8-8) into Eq. (8-7), and defining an *l* x *l* matrix *K* by $K_{ij} \equiv \left(\Phi(\mathbf{x}_i).\Phi(\mathbf{x}_j)\right) = k(\mathbf{x}_i, \mathbf{x}_j)$, then a problem which is cast in terms of dot products is formed;

solve $l\lambda \boldsymbol{\alpha} = K\boldsymbol{\alpha}$ Eq.(8-9)

where $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_l)^T$. Normalizing the solutions $\mathbf{V}^k$, i.e. $\left(\mathbf{V}^k.\mathbf{V}^k\right) = 1$, translates into $\lambda_k\left(\boldsymbol{\alpha}^k.\boldsymbol{\alpha}^k\right) = 1$.

To extract non-linear basis functions for the Φ-mapping of a test point $\mathbf{x}$, the projection onto the *k*-th basis function is computed by

$$\beta_k \equiv \left(\mathbf{V}^k.\Phi(\mathbf{x})\right) = \sum_{i=1}^{l} \alpha_i^k k\left(\mathbf{x}, \mathbf{x}_i\right).$$

For feature extraction, *l* kernel functions are evaluated instead of a dot product in ***F*** which is expensive if ***F*** is high-dimensional (or, as for Gaussian kernels, infinite-dimensional). To reconstruct the Φ-image of a vector $\mathbf{x}$ from its projections β_k onto the first *n* basis functions in ***F*** (assuming that the eigenvectors are ordered by decreasing eigenvalue size), a projection operator P_n is defined as

$$P_n\Phi(\mathbf{x}) = \sum_{i=1}^{n} \beta_k V^k \qquad \text{Eq. (8-10)}$$

If *n* is large enough to take into account all directions belonging to eigenvectors with non-zero eigenvalue, then $P_n\Phi(\mathbf{x}_i) = \Phi(\mathbf{x}_i)$. Otherwise kernel POD still satisfies (i) that the overall squared reconstruction error $\sum_i \left\| P_n\Phi(\mathbf{x}_i) - \Phi(\mathbf{x}_i) \right\|^2$ is minimal and (ii) the retained variance is maximal among all projections onto orthogonal directions in ***F***. In most applications one is interested in reconstructing data in the input space rather than in

F. This can be achieved by computing a vector **z** satisfying $\Phi(\mathbf{z}) = P_n\Phi(\mathbf{x}_i)$ with the hope that for the kernel used, such a **z** will be a good approximation of **x** in the input space. However, such a solution vector **z** will not always exist and if it exists, it may not be unique.

When the vector $P_n\Phi(\mathbf{x}_i)$ has no inverse mapped solution vector, an approximation can be made by minimizing

$$\rho(\mathbf{z}) = \left\|\Phi(\mathbf{z}) - P_n\Phi(\mathbf{x})\right\|^2 \qquad \text{Eq. (8-11)}$$

This can be expanded into

$$\rho(\mathbf{z}) = \left\|\Phi(\mathbf{z})\right\|^2 - 2\left(\Phi(\mathbf{z}).P_n\Phi(\mathbf{x})\right) + \left\|P_n\Phi(\mathbf{x})\right\|^2 \qquad \text{Eq. (8-12)}$$

Substituting Eq. (8-8) and Eq. (8-10) into Eq. (8-12), we get an expression which is written in terms of dot products. Consequently, a kernel can be introduced to obtain a formula for ρ and therefore for $\nabla_{\mathbf{z}}\rho$ which does not rely on carrying out Φ explicitly

$$\rho(\mathbf{z}) = k(\mathbf{z},\mathbf{z}) - 2\sum_{k=1}^{n}\beta_k\sum_{i=1}^{l}\alpha_i^k k(\mathbf{z},\mathbf{x}_i) + \left\|P_n\Phi(\mathbf{x})\right\|^2 \qquad \text{Eq. (8-13)}$$

In this section it has been assumed that the features are centred on the origin and therefore the dominant eigenvalues of the kernel matrix K_{ij} measure the variance along the principal components in feature space. The features can always be centred by subtracting out their mean, namely, by the transformation $\Phi(\mathbf{x}_i) - \frac{1}{N}\sum_j \Phi(\mathbf{x}_j) \to \Phi(\mathbf{x}_i)$. When the mapping $\mathbf{\Phi}(\mathbf{x})$ is only implicitly specified by the kernel function, the centering transformation can be applied directly to the kernel matrix as shown in section 8.3.5.

For a centred kernel matrix, the relative weight of the leading d eigenvalues, obtained by dividing their sum by the trace measures the relative variance captured by the leading d eigenvectors. When this ratio is nearly unity, the data can be viewed as inhabiting a d-dimensional sub-space of the feature space or equivalently a d-dimensional manifold of the input space.

8.3.1 Inverse mapping for Gaussian Kernels

To optimize Eq. (8-13) standard gradient descent methods can be employed [133]. Considering kernels of the form $k(\mathbf{x},\mathbf{y}) = k\left(\|\mathbf{x}-\mathbf{y}\|^2\right)$ and therefore $k(\mathbf{x},\mathbf{x}) = \text{constant}$ for all **x**, an optimal **z** can be determined as follows;

From Eq.(8-12) and Eq.(8-13) it can be deduced that

$$\rho(\mathbf{z}) = \left(\Phi(\mathbf{z}).P_n\Phi(\mathbf{x})\right) + \Omega \;=\; \sum_{i=1}^{l} \gamma_i k(\mathbf{z},\mathbf{x}_i) + \Omega \qquad \text{Eq. (8-14)}$$

where $\gamma_i = \sum_{k=1}^{n} \beta_k \alpha_i^k$. For an extremum, the gradient with respect to **z** vanishes, ie. $\nabla_{\mathbf{z}}\,\rho(\mathbf{z}) = \sum_{i=1}^{l} \boldsymbol{\gamma}_i k'\left(\|\mathbf{z}-\mathbf{x}_i\|^2\right)(\mathbf{z}-\mathbf{x}_i) = 0$. This leads to a necessary condition for the extremum: $\mathbf{z} = \sum_i \delta_i \mathbf{x}_i / \sum_j \delta_j$, with $\delta_i = \gamma_i k'\left(\|\mathbf{z}-\mathbf{x}_i\|^2\right)$. For a Gaussian kernel $k(\mathbf{x},\mathbf{y}) = \exp\left(-\|\mathbf{x}-\mathbf{y}\|^2 / c\right)$ we get

$$\mathbf{z} = \frac{\sum_{i=1}^{l} \gamma_i \exp\left(-\|\mathbf{z}-\mathbf{x}_i\|^2 / c\right)\mathbf{x}_i}{\sum_{i=1}^{l} \gamma_i \exp\left(-\|\mathbf{z}-\mathbf{x}_i\|^2 / c\right)}. \qquad \text{Eq. (8-15)}$$

Note that the denominator is equal to $\Phi(\mathbf{z}).P_n\Phi(\mathbf{x})$. Assuming that $P_n\Phi(\mathbf{x}) \neq 0$ then, $\left(\Phi(\mathbf{x}).P_n\Phi(\mathbf{x})\right) = \left(P_n\Phi(\mathbf{x}).P_n\Phi(\mathbf{x})\right)$ > 0. Since k is smooth, then there exists a

neighbourhood of the extremum of Eq. (8-14) in which the denominator of Eq. (8-15) is $\neq 0$. Thus, an iteration scheme for **z** can be devised as follows

$$\mathbf{z}_{t+1} = \frac{\sum_{i=1}^{l} \gamma_i \exp(-\|\mathbf{z}_t - \mathbf{x}_i\|^2 / c)\mathbf{x}_i}{\sum_{i=1}^{l} \gamma_i \exp(-\|\mathbf{z}_t - \mathbf{x}_i\|^2 / c)} \qquad \text{Eq. (8-16)}$$

Numerical instabilities related to $\left(\Phi(\mathbf{z}).P_n\Phi(\mathbf{x})\right)$ being small can be dealt with by restarting the iteration with a different starting value. Furthermore, it can be noticed that any fixed-point of Eq.(8-16) will be a linear combination of the kernel PCA training data $\mathbf{x}_i$.

8.3.2 Inverse mapping for polynomial kernels

Following the previous exposition, an iterative scheme for polynomial kernels can be devised. Eq. (8-13) can now be written as

$$\rho(\mathbf{z}) = (\mathbf{z}\,\mathbf{z}+1)^d - 2\sum_{i=1}^{N} \gamma_i (\mathbf{z}\,\mathbf{x}_i + 1)^d + \Omega$$

On putting the derivative w.r.t. **z** equal to 0 for an extremum, an iterative formula for the inverse mapping of a polynomial kernel is achieved as follows

$$\mathbf{z}_{t+1} = \sum_{i=1}^{N} \gamma_i \left(\frac{\mathbf{z}_t \mathbf{x}_i + 1}{\mathbf{z}_t \mathbf{z}_t + 1} \right)^{d-1} \mathbf{x}_i \qquad \text{Eq. (8-17)}$$

8.3.3 Inverse mapping for sigmoid kernels

Similarly, an iterative scheme for sigmoid kernels can be devised as follows. Eq. (8-13) can be written as

$$\rho(\mathbf{z}) = \mathrm{Tanh}\left(\kappa(\mathbf{z},\mathbf{z})+\theta\right) - 2\sum_{i=1}^{N}\gamma_i \mathrm{Tanh}\left(\kappa(\mathbf{z},\mathbf{x}_i)+\theta\right) + \Omega$$

Again, on putting the derivative w.r.t. **z** equal to 0 for an extremum, an iterative formula for an inverse mapping of a sigmoid kernel is obtained as follows :

$$\mathbf{z}_{t+1} = \sum_{i=1}^{N}\gamma_i \frac{\left(1-\mathrm{Tanh}^2\left(\kappa\left(\mathbf{z}_t,\mathbf{x}_i\right)+\theta\right)\right)}{\left(1-\mathrm{Tanh}^2\left(\kappa\left(\mathbf{z}_t,\mathbf{z}_t\right)+\theta\right)\right)}\mathbf{x}_i \qquad \text{Eq. (8-18)}$$

8.3.4 Inverse mapping using non-iterative methods

The kPOD is a non-linear feature extractor and when linear POD is performed in the feature space, it is equivalent to making non-linear POD in the input-space. While the mapping from the input space to feature space is of primary importance in kernel based methods, the reverse mapping from the feature space to the input space is almost always necessary. As demonstrated by Mika *et al.* [133], the exact 'pre-image' typically does not exist and one can only settle for an approximate solution. But even this may not be trivial as the dimensionality of the feature space can be infinite. For certain invertible kernels, this non-linear problem can be solved using a fixed point iteration method as described in the previous sections 8.3.1, 8.3.2 and 8.3.3 for different kernels. However this method is dependent on the initial starting point and is highly susceptible to local minima. To circumvent this problem, two algorithms were proposed, one by Kwok *et al.* [134] and the other by Rathi *et al.* [135]. Both algorithms are based on the relationship between the feature-space and the input-space distances. These are both non-iterative, involve only linear algebra and do not suffer from numerical instability or local minimum problems which are sometimes experienced by the fixed point iteration method. The algorithm proposed by Kwok *et al.* uses multi-dimensional scaling and an expensive computation in the input space to obtain the co-ordinates of the pre-image. Rathi *et al.* reconstruct the pre-image of a vector by first finding the distance of this point to all training points in the feature space. Then a new co-ordinate system to represent the data in the input space by performing SVD on a matrix of n nearest neighbours is found and multi-dimensional scaling is used to project the solution onto this new co-ordinate

system. The eigenvectors of this new co-ordinate system are then used to find the approximate pre-image in the original input space. It is reported that although the method uses a couple of approximations for reconstruction, the results are impressive. In this work these techniques were not implemented since the fixed point iteration method was found to be more than adequate. However, it may be the case that such algorithms are useful for more involving problems.

8.3.5 kPOD Algorithm outline

The kPOD algorithm implemented in this work is the following:

1. Given a set of *m*-dimensional training data $\{\mathbf{x}_k\}$, $k = 1,\ldots.. l$, the kernel matrix is computed $K \in \Re^{lxl} = k(\mathbf{x}_i, \mathbf{x}_j)$.

2. Centering is carried out in the feature space for $\sum_{k=1}^{l} \Phi(\mathbf{x}_k) = 0$ according to

   $\tilde{K} = K - 1_l K - K 1_l + 1_l K 1_l$

   where $1_l = \frac{1}{l}\begin{bmatrix} 1 & . & . & . & 1 \\ . & . & & & . \\ . & & . & & . \\ . & & & . & . \\ 1 & . & . & . & 1 \end{bmatrix}$

3. Solve the eigenvalue problem $l\lambda\boldsymbol{\alpha} = K\boldsymbol{\alpha}$ and normalize $\boldsymbol{\alpha}_k$ such that

   $\langle \alpha_k, \alpha_k \rangle = \frac{1}{\lambda_k}$.

4. For a test pattern **x,** the nonlinear basis functions are extracted via

   $\langle \mathbf{V}_k, \Phi(\mathbf{x}) \rangle = \sum_{i=1}^{l} \alpha_i^k k(\mathbf{x}_i, \mathbf{x})$

5. An inverse mapping procedure mentioned earlier (either section 8.3.1 or 8.3.2 or 8.3.3) is adopted to map the solution from the feature space back to the input space.

8.4 Kernels derived from the training data

Weinberger and Saul introduced a technique called semi-definite embedding (SDE) [136] which is essentially a variation on kernel PCA in which the kernel matrix is also learned from the data. This is in contrast with the classical kernel PCA in which one has to choose a kernel function a priori. To derive this technique, Weinberger and Saul formulated the problem of learning the kernel matrix as an instance of semi-definite programming. Since the kernel matrix K represents inner products of vectors in a Hilbert space it must be positive semi-definite. Also the kernel should be centred, i.e., $\sum_{ij} K_{ij} = 0$. In addition, the technique imposes constraints on the kernel matrix to ensure that the distances and angles between points and their neighbours are preserved under the neighbourhood graph η. That is, if both x_i and x_j are neighbours that is $\eta_{ij} = 1$ or are common neighbours of another input that is $(\eta^T \eta)_{ij} \rangle 0$, then the distance should be preserved

$$\left\| \Phi(x_i) - \Phi(x_j) \right\|^2 = \left\| x_i - x_j \right\|^2 \qquad \text{Eq. (8-19)}$$

In terms of the kernel matrix, this constraint can be written as:

$$K_{ii} - 2K_{ij} + K_{jj} = \left\| x_i - x_j \right\|^2 \qquad \text{Eq. (8-20)}$$

By adding an objective function to maximize the trace of the kernel matrix $Tr(K)$ which represents the variance of the data points in the learned feature space, the technique constructs a semi-definite program for learning the kernel matrix K. The last detail of SDE is the construction of the neighbourhood graph η_{ij}. This graph is constructed by connecting the *k* nearest neighbours using a similarity function over the data, $\left\| x_i - x_j \right\|$. In its last step, SDE runs kernel PCA on the learned kernel *K*. The algorithm of SDE is summarised in the following section.

8.4.1 SDE Algorithm outline

1. Construct neighbours, η, using k-nearest neighbours.
2. Maximize Tr(K) subject to $K \geq 0, \sum_{ij} K_{ij} = 0,$ and

   $\forall \text{ij} \quad \eta_{\text{ij}} > 0 \quad \vee \left(\eta^T \eta\right)_{ij} > 0 \Rightarrow K_{ii} - 2K_{ij} + K_{jj} = \left\| x_i - x_j \right\|^2$
3. Perform the kernel PCA with the learned kernel K.

In the next section the techniques described in section 8.3 were implemented into a FORTRAN code and used to model the transonic flow over the RAE 2822 aerofoil.

8.5 Application of kPOD-based reduced-order modelling

8.5.1 Transonic flow over the RAE 2822 aerofoil

One of the significant challenges in the use of POD for high-speed flow fields is to capture moving shock waves as the flow parameters or boundary conditions change. Techniques that generate POD-based reduced-order models (POD/ROM) for subsonic and supersonic flows do not generate a useful ROM for a high-speed case with moving shocks. Excessive data and time are required for modelling this accurately. As it has already been stated in the literature review, for this problem Lucia [44] proposed and used a technique to exploit POD for accurately treating moving shock waves. This technique involves the decomposition of the solution domain to isolate regions that contain shocks. The main idea behind this is to use the POD modes for a global approximation of the flow-field and to use a standard finite volume scheme in the region where shocks occur. A reduced-order model for each region is developed independently and the solution for the entire domain is formed through a linking of the boundaries of each region. This technique was applied to a one-dimensional quasi-steady nozzle flow-field by Lucia [44] for demonstration. However, LeGresley and Alonso [45] applied this technique for the shape optimisation of a two-dimensional aerofoil. The results attained are good though some discrepancy could still be detected between the high-fidelity solution and the POD/ROM with domain decomposition POD/ROM/DD.

It is thought that by combining POD-based ROMs with conventional spatial discretization schemes, more efficient multi-scale algorithms for the analysis of some parameter dependent PDEs can be developed. The idea of multi-scale modelling is to use two models at least – one to represent the global features of the solution and a detailed model to capture the local features. The main issues involved in designing a multi-scale scheme are (1) the numerical method employed to identify the localized regions where a full-order model should be used and (2) the approach used for handling the coupling between the two models in order to arrive at consistent values for the field variables in the interface region. Thus, while it is recognised that this domain-decomposition

technique is reasonably good, it does offer some challenging issues. It would be very interesting if a comparison is made between the computational time of the technique proposed by Lucia, that is, the POD/ROM/DD and one which does not involve any domain decomposition methods (so avoiding the subsequent challenging issues) but simply a technique that globally approximates the solution using the POD methodology, which solution is in turn used to initialize the high-fidelity computation. It would be expected that this latter technique would entail some more computing time but definitely would be as accurate as the high-fidelity solution. The computational time should not be excessively large compared with the previous one, even because the global approximation would be very accurate outside the shock region and consequently would require hardly any iterations to reach the converged solution outside the shock region.

In this section an investigation is conducted to establish whether it is possible to model this highly challenging problem by a more practical method, because it is much more straightforward to implement. This is the kernel POD method. In this case an attempt is made to model both the weak and strong non-linear features in the fluid flow by kPOD-based reduced-order modelling.

8.5.1.1 Snapshots generation

Snapshots of viscous steady flow over the RAE 2822 aerofoil at an angle of incidence of 3.19º were generated at various Mach numbers within the range [0.30, 0.72] to create a Lagrangian subspace. In this case, the commercial software Fluent [53] was used to generate the snapshots. The turbulence modelling was conducted by using the Spalart-Allmaras turbulence model [57]. A wall function was considered.

8.5.1.1.1 Computational grid

One of the computational meshes used is shown in figures (8-3) and (8-4). An O-type mesh with 291 x 61 grid points was utilised in this case.

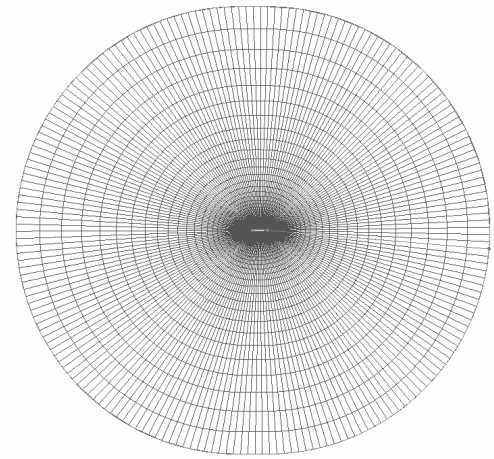


Figure (8-3) – O-type mesh with 201 x 61 grid points used to compute the flow over the RAE 2822 aerofoil

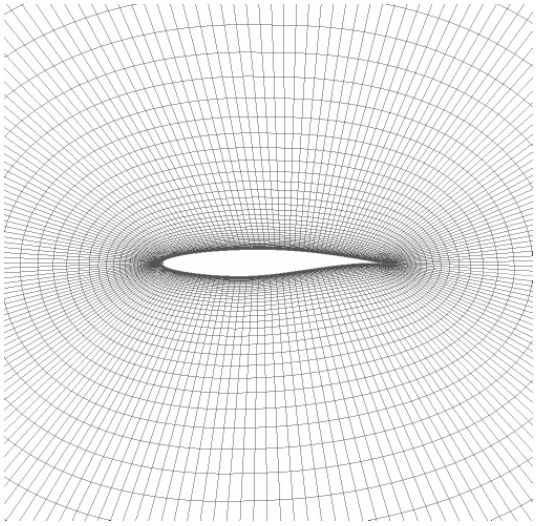


Figure (8-4) – A close-up of the O-type mesh with 201 x 61 grid points used to compute the flow over the RAE 2822 aerofoil

8.5.1.1.2 Flow solutions

Initially, snapshots were generated within the Mach number range [0.30, 0.70] with uniform intervals of 0.1. Note that the square brackets imply that both boundary values are included. Therefore a total of five snapshots were considered. This data set was considered as this is the same set of snapshots LeGresley and Alonso [45] have used to demonstrate their POD-based ROM with domain decomposition method. An example of the pressure flow-field at a Mach number of 0.4 is shown in figure (8-5). The surface static pressure distributions of all the snapshots under consideration are shown in figure (8-6). The absolute values of the static pressure were plotted so that one can survey the precise relative changes between one snapshot and the other. In the latter figure a profile of the aerofoil is included for a better comprehension of the changes that are taking place and where these changes are occurring.

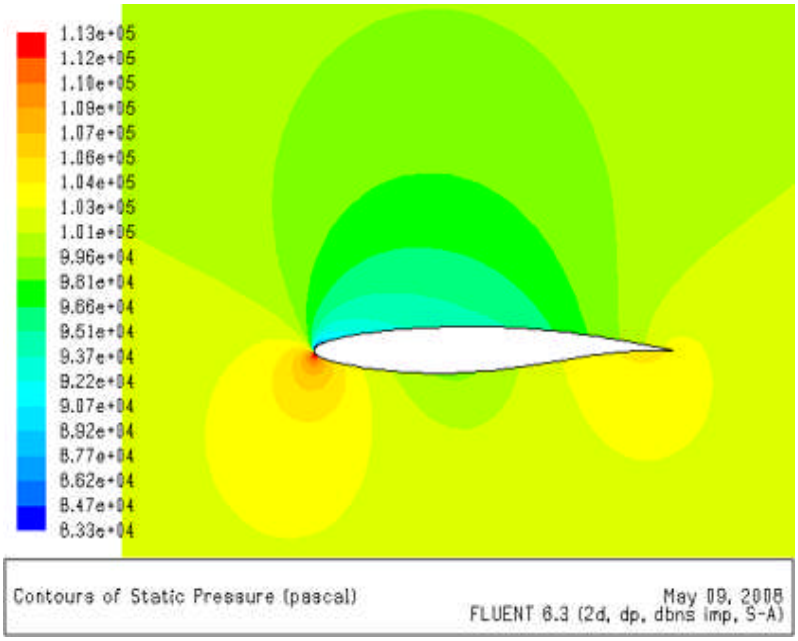


Figure (8-5) – Pressure contours over the RAE 2822 aerofoil at M = 0.4 and an angle of incidence of 3.19°

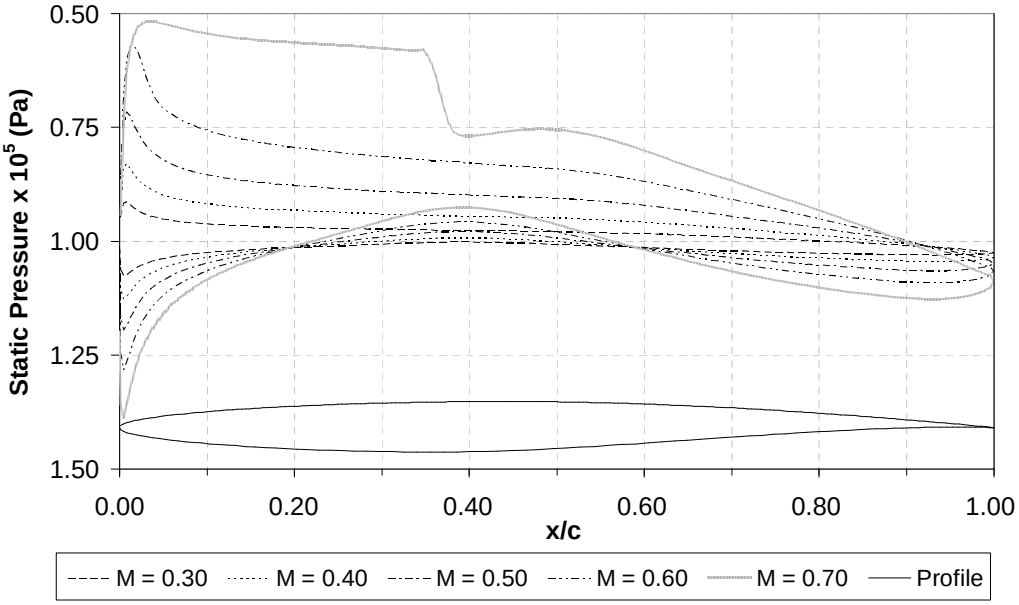


Figure (8-6) – Surface pressure distributions over the RAE 2822 aerofoil at an angle of incidence of 3.19º and at various Mach numbers

8.5.1.2 POD-based ROM predictions with 5 snapshots only

In the following cases, linear interpolation was considered for prediction purposes. POD predictions at two different Mach numbers one at M = 0.35 and one at M = 0.67 were considered in this work, just like in [45]. The former one was considered mainly to investigate how the weak non-linearity in the aerodynamic flow is modelled, while the latter one was considered to investigate how the strong non-linearity in the aerodynamic flow is modelled ie. the moving shock wave over the surface of the aerofoil.

The POD predictions at M = 0.35 and M = 0.67 are shown in figures (8-7) and (8-8). The POD method predicts a pressure which is reasonably accurate at M = 0.35 with only two POD modes. One may observe however some discrepancy between the POD predicted and the CFD result over the upper surface. At M = 0.67, the prediction is accurate over the bottom surface of the aerofoil and over the trailing one-third of the chord of the upper surface. However, over the leading two-thirds of the chord of the

upper surface, the prediction is bad due to the development of the shock wave which moves over the surface of the aerofoil with changes in the flow condition. This is a strong non-linear feature which definitely makes it difficult for the POD method to model this characteristic with a few number of snapshots. In this case all the POD modes are used.

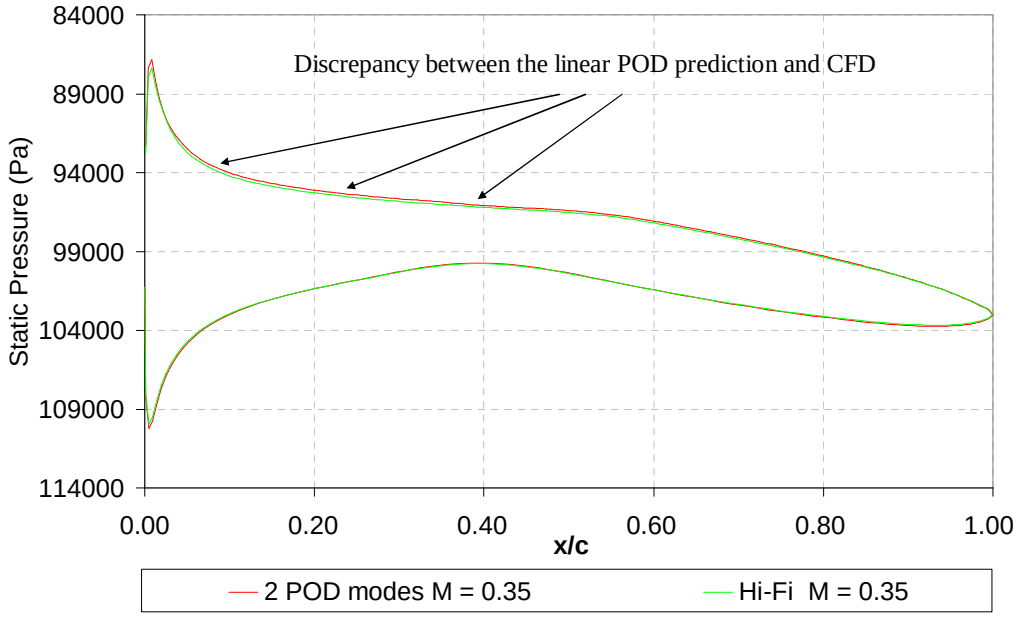


Figure (8-7) – Comparison of the surface pressure distributions at M = 0.35 and an angle of incidence of 3.19º between the POD predicted and the CFD solution.

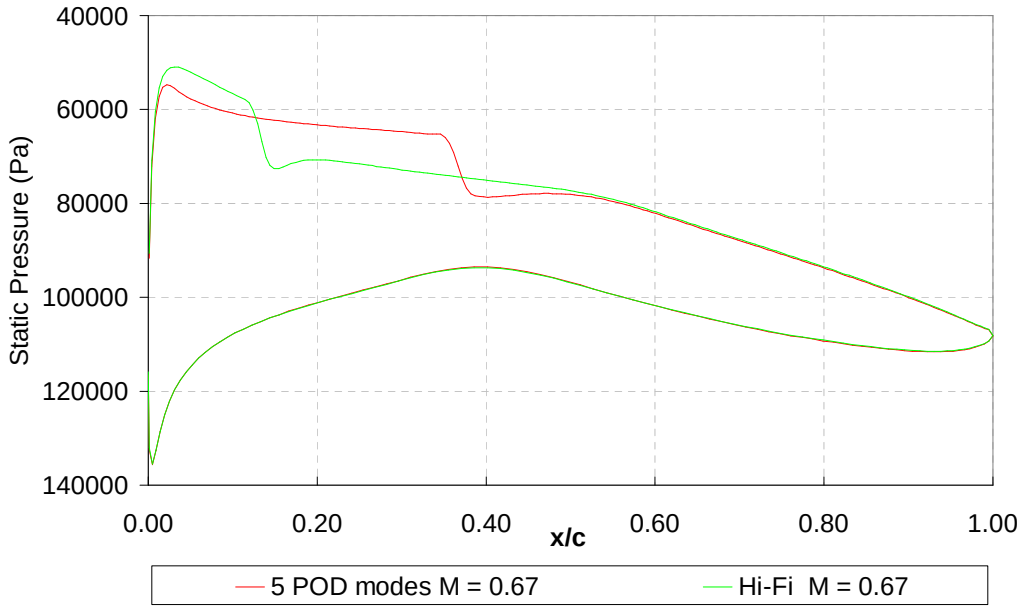


Figure (8-8) – Comparison of the surface pressure distributions at M = 0.67 and an angle of incidence of 3.19º between the POD predicted and the CFD solution

Figure (8-9) shows plots of the static pressure at various places on the surface of the aerofoil (as marked by the arrows) against the Mach number of each snapshot. Each node is denoted by a number. For nodes 223 and 264 it is evident that the pressure variation with the Mach number is non-linear, while for nodes 19, 38, 64 and 137 the variation is very close to a linear relationship throughout. The static pressure values determined by CFD at M = 0.35 and M = 0.67 are also indicated as red and green diamonds respectively. This suggests that the POD predicts the static pressure very well at M = 0.35, in fact all red diamonds are located on the prediction line. Moreover, the prediction at M = 0.67 is fine at nodes 19, 38, 64 and 137, while it is not so adequate for prediction at nodes 223 and 264. By coincidence the green diamond for node 264 happens to be very close to the point where the POD predicted and CFD static pressures coincide, and so this explains why for node 264 the CFD value is close to the POD prediction. In view of this, an investigation is conducted to study the effects that it would have by kernelizing the linear POD method.

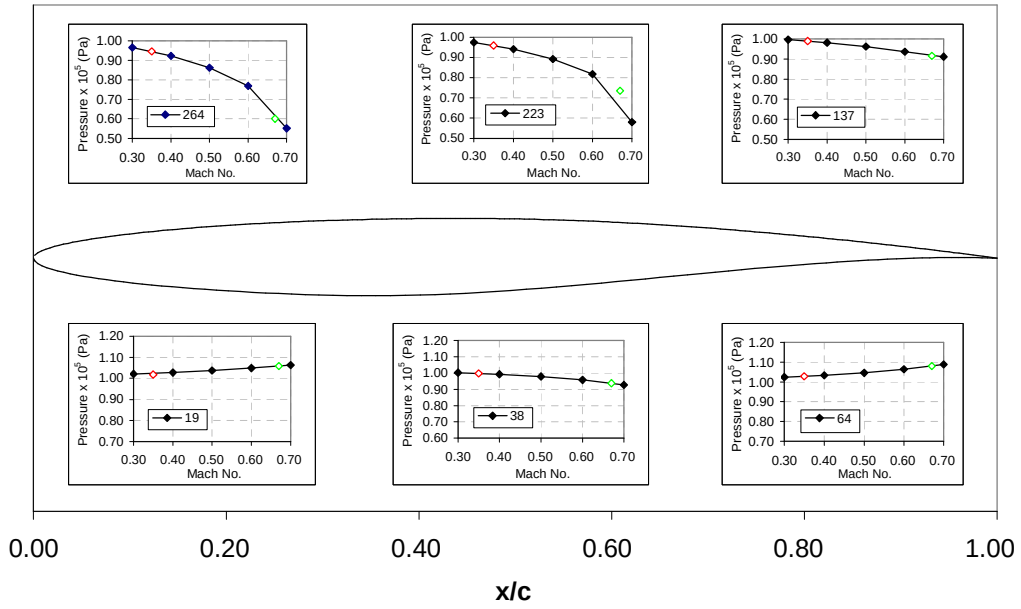


Figure (8-9) – Surface pressure distributions over the RAE 2822 aerofoil at an angle of incidence of 3.19º and at the Mach numbers considered. The CFD results at M = 0.35 and M = 0.67 are indicated as red and green diamonds respectively

8.5.1.3 kPOD-based ROM predictions with 5 snapshots only

Predictions using various types of kernels, namely, polynomial, Gaussian RBF and Sigmoid kernels were considered to model this non-linear flow problem. The same number of snapshots (ie. 5) considered earlier was used. The results obtained are the following:

8.5.1.3.1 Polynomial kernel

In this case a polynomial kernel with $p_1 = 3$ and $p_2 = 1.0$ in Eq. (8-4) was considered. The following results were obtained.

At M = 0.35, the same result was achieved with the POD and the kPOD although the former is with 2 POD modes while the latter is with 3 kPOD modes. See figure (8-10) This observation is explained by looking at the variances captured by the respective

modes. The total variance captured with the 2 POD modes is 99.97 %, while the total variance captured with the 3 kPOD modes is 99.31 %. The variance captured with 2 kPOD modes is 91.65 %. In this case the linear POD result converges with fewer POD modes for the simple reason that the problem is pretty much linear.

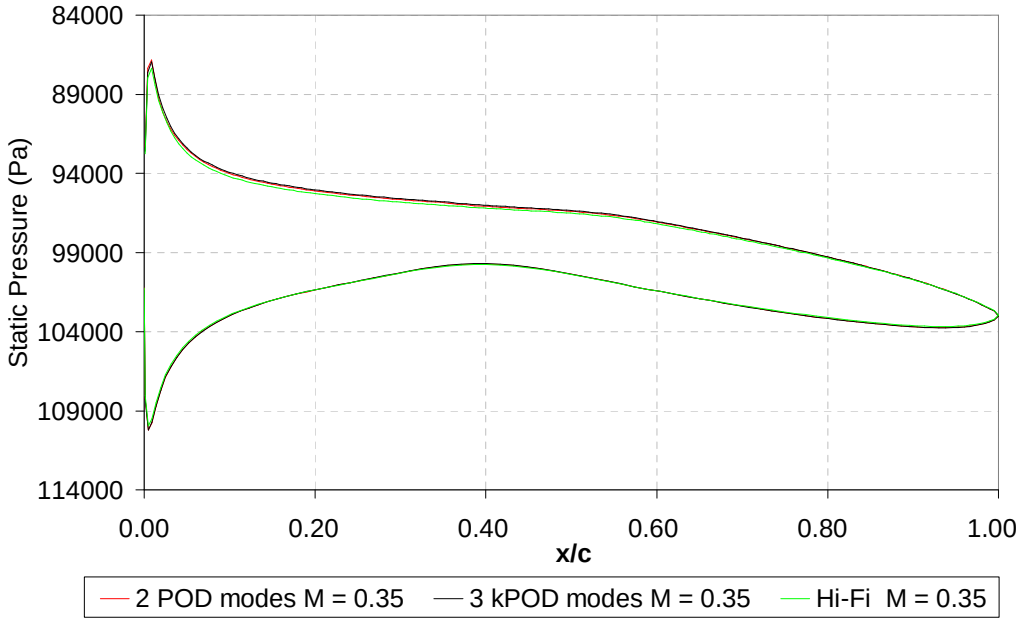


Figure (8-10) – Comparison of the surface pressure distributions over the RAE 2822 aerofoil at M = 0.35 and an angle of incidence of 3.19º

At M = 0.67, the results obtained by the two methods are identical with 5 modes although it is very far away from the high-fidelity CFD solution over the upper half of the aerofoil surface where the shock development occurs. See figure (8-11).

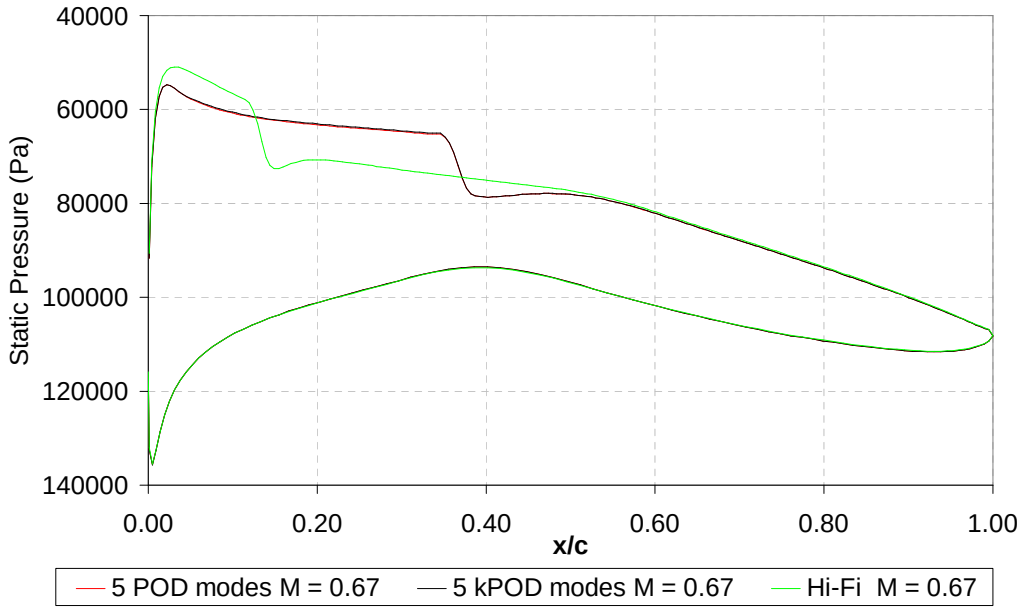


Figure (8-11) – Comparison of the surface pressure distributions over the RAE 2822 aerofoil at M = 0.67 and an angle of incidence of 3.19º

No significant changes in the pressure distribution were observed by changing p_1 and p_2 in the polynomial kernel.

8.5.1.3.2 Gaussian RBF kernel

The results obtained from a Gaussian RBF kernel are shown in figures (8-12) and (8-13). In this case the constant in the RBF kernel which is equivalent to twice the variance, was considered to be equal to 0.125.

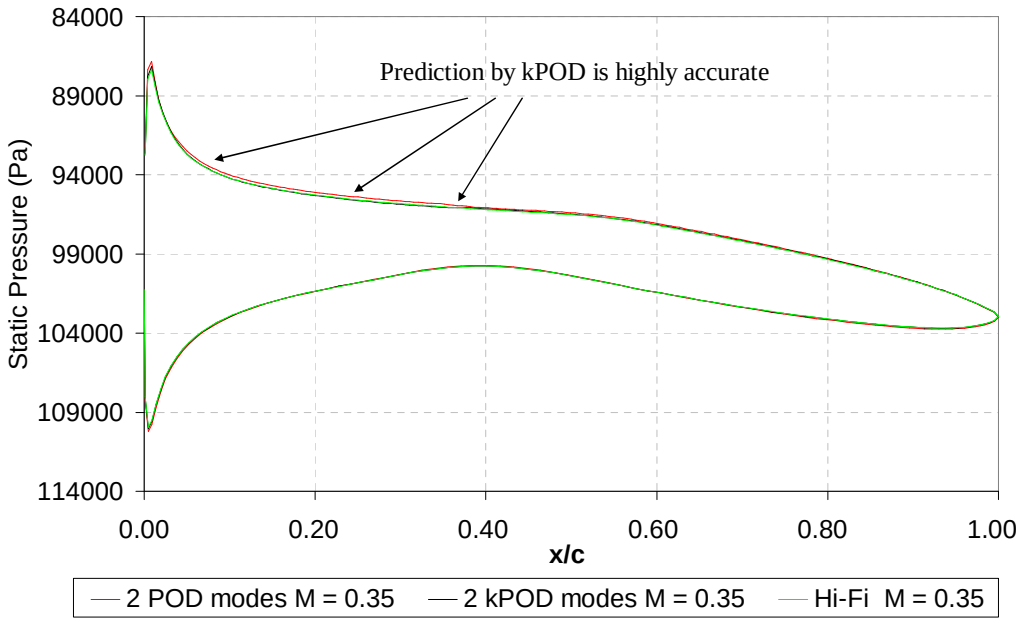


Figure (8-12) – Comparison of the surface pressure distributions over the RAE 2822 aerofoil at M = 0.35 and an angle of incidence of 3.19º

In this case, with 2 kPOD modes a result which is more accurate than with 2 POD modes was observed. Refer to figure (8-12). In fact, the total percentage RMS error in pressure of the kPOD prediction is 0.039 % while that of the linear POD is 0.134 %. Note that these RMS values are representative of the term $\Delta p/p$ at a chord length of 0.2. Note that no significant further improvement was noticed with the linear method when more modes were taken into consideration, after all the variance or energy captured with 2 POD modes was over 99.9 %. It was also noticed that further kPOD modes worsen the result and consequently these can be considered as noise. So in practice a question arises how can one decide upon stopping with a result depending on certain kPOD modes while neglecting the rest. In reality this is a problem of contemporary research in machine learning [137].

The discrepancy between the POD-based ROM method and CFD is due to the global nature of the linear POD method with its result being contaminated by snapshots at high Mach number. In kPOD with a Gaussian kernel at this value of variance this does not happen due to the fact that this kernel essentially classifies the snapshots and the

interpolation is therefore conducted with respect to the two most important kPOD modes. In fact, there exists very slight difference in the variance captured by the kPOD modes in the feature space. The variance captured by 2 kPOD modes is 40.53 %.

On the other hand, again the results attained at M = 0.67 with the two different methods are identical with 5 modes.

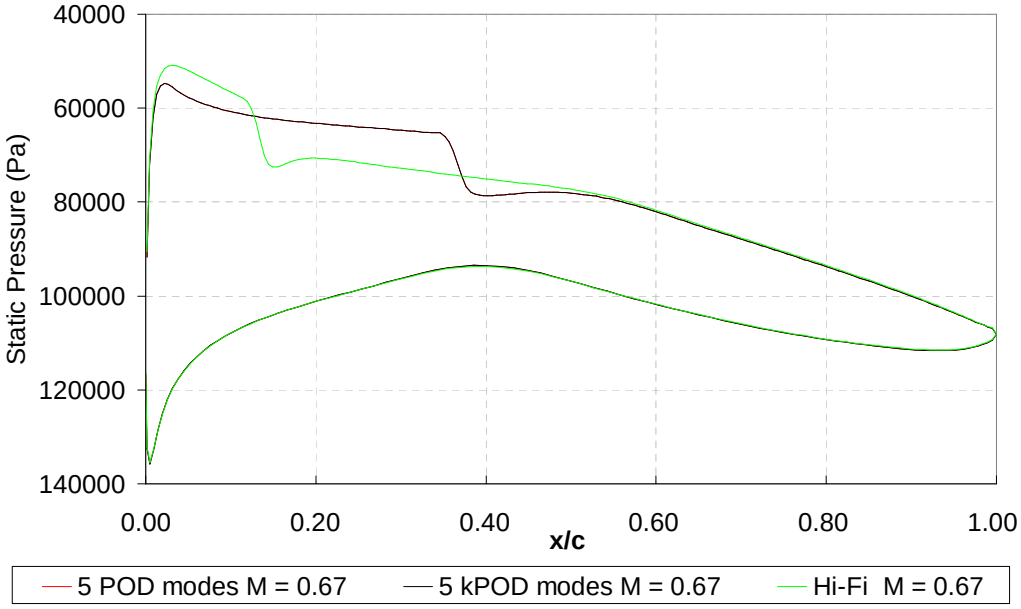


Figure (8-13) – Comparison of the surface pressure distributions over the RAE 2822 aerofoil at M = 0.67 and an angle of incidence of 3.19º

8.5.1.3.3 Sigmoid kernel

The transonic flow over the RAE 2822 aerofoil was also modelled using the sigmoid kernel and the previous number of snapshots. In this investigation it was observed that the kPOD results are very similar to the linear POD results. Up to a certain extent this observation was expected since the hyperbolic tangent function ie. the sigmoid can be considered as a piecewise-linear function and therefore weak non-linearities are not captured appropriately in contrast with the Gaussian kernel.

8.5.1.4 kPOD-based ROM predictions with more snapshots

In the previous section it has been noticed that the kPOD-based ROM with a Gaussian kernel modelled the weak non-linearities better than the linear POD method. Nevertheless, the strong non-linear feature such as the moving shock wave at M = 0.67 was not modelled appropriately. Consequently, a similar investigation was conducted using more snapshots over a smaller Mach number range to create a richer set of basis so that the behaviour of the model is studied.

In this example eight snapshots were generated within the Mach number range [0.64, 0.71] with uniform intervals of 0.01. The surface static pressure distributions of all the snapshots under consideration are shown in figure (8-14). The absolute values of the static pressure were plotted so that one can survey the relative changes between one snapshot and the other, particularly at the shock. A profile of the aerofoil is included for a better visualization of the changes that are taking place and where these changes are occurring.

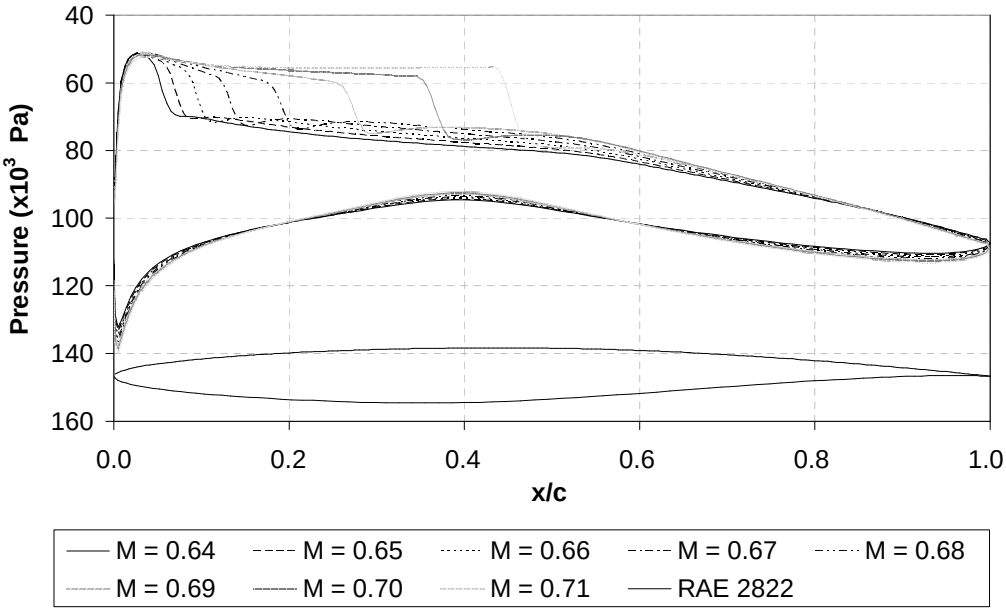


Figure (8-14) – Surface pressure distributions over the RAE 2822 aerofoil at an angle of incidence of 3.19º and at various Mach numbers

With this set of snapshots, two predictions were made one at M = 0.645 and the other at M = 0.675. The results acquired are shown in figures (8-15) and (8-16) respectively. The CFD results are also included in the same figure for a comparison.

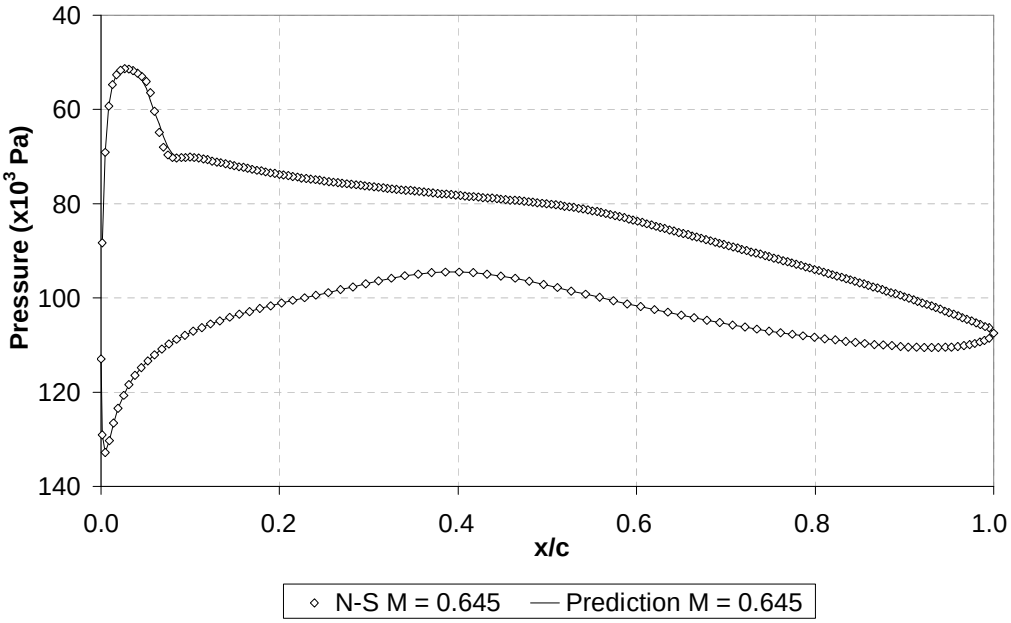


Figure (8-15) – Comparison between the kPOD-based ROM prediction and CFD at M = 0.645 and an angle of incidence of 3.19º

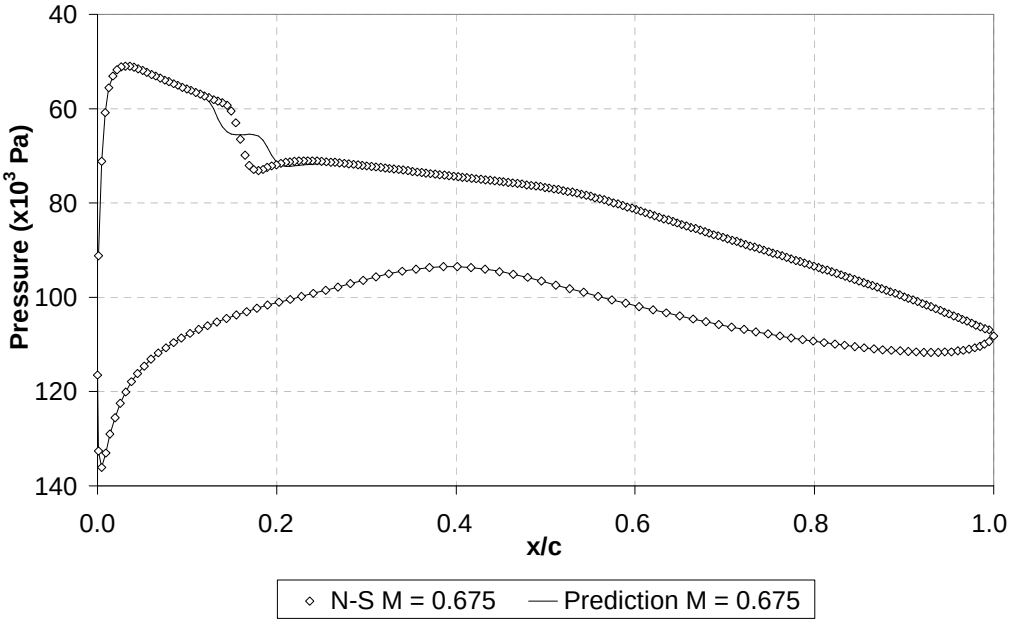


Figure (8-16) – Comparison between the kPOD-based ROM prediction and CFD at M = 0.675 and an angle of incidence of 3.19º

Thus, while the shock prediction at M = 0.645 is reasonable, the shock prediction at M = 0.675 is poor, since the model predicts two smaller shocks in series rather than a strong one. The reason behind this is explained by observing the series of snapshots considered. The shock at M = 0.65 is slightly downstream of the shock at M = 0.64, thus the shock movement between the Mach number variation is small. Consequently, the shock at M = 0.645 is sensibly predicted by the snapshots' shock in that vicinity. At M = 0.68, the shock occurs almost 10 % of the chord way down from the shock at M = 0.67. Thus, the shock movement between these two snapshots is considerable and therefore this explains why the prediction using the kPOD-based ROM is so poor. For a reasonable prediction at a Mach number of 0.675, further snapshots are required within the Mach number range $0.67 < M < 0.68$. Predictions at Mach numbers over 0.68 require further snapshots as the shock movement is considerably larger, which makes the applicability of the ROM in such a circumstance almost impractical. It is to be pointed out that no significant difference was noticeable between the predictions obtained from the linear POD-based reduced-order model and the kernel POD-based one with these number of snapshots.

8.6 Concluding remarks

In this work three types of apriori chosen kernels were considered, namely the Gaussian kernel, the polynomial kernel and the sigmoid kernel. These were used to kernelize the linear POD algorithm.

In signal and image processing, it has been reported by various researchers in the field that the kPOD produces some significant improvement over the linear POD method in image reconstruction [138] [139]. In this work it has been observed that the kPOD-based ROM method using a general purpose kernel like the Gaussian, modelled adequately the weak non-linear features in the flow over the RAE 2822 aerofoil. This shows that the weak non-linear features in the flow are analogous with the non-linearity observed in image processing. However, it must be accentuated that the highly non-linear features such as the moving shock wave on the surface of the aerofoil is not modelled by few snapshots using kPOD-based ROM.

Therefore, weak non-linearity is indeed modelled better with the kPOD method. It was also observed that fewer kPOD modes are sometimes sufficient to describe some problem of interest compared with the number of linear POD modes required. However, this depends to a great extent on various factors such as problem formulation, kernel types and their various parameters.

Also, it has been shown that the kPOD although it is a generalization of the POD method, it does not really inherit all the abilities of the linear technique because (i) the inverse mapping process is quite intricate when compared with data reconstruction from the linear method as it strictly requires to solve an optimization problem to get back the inverse (ii) the ROM procedure is much more complex.

From this work it is also evident that the kPOD with the apriori chosen kernels considered is far from adequate to model the strong non-linear features such as a moving shock wave. A considerable number of snapshots are required for an accurate prediction

and so the technique is not an alternative to multi-scale modelling. Notwithstanding, these apriori chosen kernels failed to effectively and efficiently model strong non-linear features in aerodynamic flows, it is thought that it would still be worthwhile to investigate the behaviour of kernels derived from the data itself rather than choosing a kernel beforehand and using it to map the data from the input space to the feature space. In particular, the method proposed by Weinberger and Saul [136] mentioned in section 8.4, deserves some attention.

Chapter 9

9.0 Conclusions

In this research work a high-fidelity low-cost surrogate of a computational fluid dynamics analysis tool namely the IMPNS software was developed and its results presented. This computational tool is composed of general and physics-based approximation methods by which three dimensional high-speed aerodynamic flow-field predictions are made with high efficiency and an accuracy which is comparable with that of CFD. This makes the tool potentially suitable for analysis of complex aerodynamics flow problems.

The tool makes use of reduced-order modelling and hence reduced-basis methods that are suitable for both linear and non-linear problems, whereby the basis vectors are computed via the proper orthogonal decomposition (POD) of a set of observations or snapshots. The decomposition produces an optimal linear set of orthogonal basis vectors or POD modes that best describe the ensemble of snapshots. These solutions are then projected onto this set of basis vectors to provide a finite set of scalar coefficients that represent the solutions. A pseudo-continuous representation of these projection coefficients is constructed for each of the basis vectors, which allow predictions to be made of parameter combinations not in the original set of observations. Response surface construction methods based on parametric and non-parametric models for the pseudo-continuous representation of the projection coefficients were evaluated. Also, an

exploration of the performance of a few design-of-experiment approaches for the generation of the initial ensemble of computational experiments was accomplished.

The model was applied to two flow problems related to high-speed weapon aerodynamics;

- inviscid and viscous flow about a flare stabilized hypersonic projectile and
- supersonic turbulent flow around a fin stabilized projectile with drooping nose control.

Comparisons of the solutions derived from the surrogate model with high-fidelity CFD simulations suggest that the POD provides a reliable and robust approach to the construction of reduced-order models (ROMs). The utility of POD-based ROMs was demonstrated by showing that the complete solutions for the snapshots which formed part of the original dataset of observations were reconstructed by using a few POD modes. This procedure provides a data compression methodology by retaining the most energetic modes (those capturing the most variance) while discarding the rest without any significant loss of detail, thus offering an effective data management system for both storage and handling of data.

Comparisons of surrogate model predictions with high-fidelity CFD simulations suggest that POD-based reduced-order modelling together with response surface methods provide a reliable and robust approach for efficient and accurate predictions. In contrast to the many modelling efforts reported in the literature, this surrogate model provides access to information about the whole flow-field.

A number of different approaches to the construction of a response surface of the projection coefficients were investigated, a linear regression based method and three interpolation techniques employing linear spline, cubic spline and radial basis functions. The computed data suggest that interpolation based techniques provide a significant advantage over the regression method. This is attributed to the fact that the interpolation schemes pass through all of the sample points providing an improved representation of

local minima and maxima, while the global fit of the regression technique produces some unnecessary smoothing. This observation was supported by studies of data localization that suggest the regression method can provide similar accuracy to the interpolation schemes when employed over a reduced parameter space. In general, the radial basis function scattered interpolation method offered the most accurate predictions.

Moreover, the use of optimised sampling methods was found to offer improved accuracy for a given number of sample points. For problems involving large numbers of parameters the latin-hypercube sampling (LHCS) may provide a practical approach to reduce the number of samples required to populate the design space. However, for problems involving geometric variation the LHCS requires a means of automatically generating high-quality grids. For this reason a more practical tool may incorporate a hybrid approach, a design-of-experiment technique in which the geometry variables are prescribed at specific levels and an LHCS technique which is used for other parameters.

In the POD analysis, the training period comes at a computational cost. This up-front cost is a significant disadvantage of the POD-based reduced-order modelling and the trade-off can be favourable only when, after the initial computational investment, a compact ROM is constructed which can be used many times, and is valid over a useful range of parametric values. In this work, a methodology which limits the time duration in generating the training dataset has been proposed. In this methodology a hierarchical type model which fuses together variable-fidelity computational data while using POD-based ROMs is proposed for the first time. In this model, the scalar coefficients which are obtained by projecting the solution vectors onto the basis vectors, are mapped between spaces of low and high fidelities, to achieve predictions with complete flow-field information. In general, this technique offers an automatic way of fusing variable-fidelity data through interpolation and extrapolation schemes together with reduced-order modelling (ROM). The model was applied to a cone-cylinder-flare projectile at supersonic and hypersonic speeds. Two problems with different boundary conditions were considered. In particular, it has been observed that this model behaves much alike the variable-fidelity model of Reisenthel *et al*. [125], that is, the prediction follows very

closely the trend of the low-fidelity training data points and so it relies to a great extent on it, most especially when the high-fidelity training data is sparse. However, the predictions tend to follow more the high-fidelity training data points as their number is increased. Besides this, the POD-based reduced-order modelling limitations do subsist within the proposed methodology. Nevertheless, it was shown that this model works reasonably well for the high-speed aerodynamics problems considered.

In this work a study of the behaviour of a kernel POD algorithm to model the flow about an aerofoil in the transonic flow regime was also undertaken. Three types of apriori chosen kernels were considered, namely the Gaussian kernel, the polynomial kernel and the sigmoid kernel. In signal and image processing, it has been consistently reported by various researchers in the field that the kernel POD produces some improvement over the linear POD method in image reconstruction [138] [139]. In this work it has been observed that the kPOD-based ROM method using the general purpose Gaussian kernel, modelled accurately the weak non-linear features in the flow over the RAE 2822 aerofoil with a small number of snapshots. It therefore seems that the non-linearity encountered in images is analogous to the weak non-linear features of the transonic flow over the aerofoil. Furthermore, it was noticed that the behaviour of kPOD-based ROM and POD-based ROM is similar for modelling strong non-linear features such as the moving shock wave on the surface of the aerofoil as the boundary conditions change. Both require a large number of snapshots for the adequate and appropriate modelling of moving shock waves.

9.1 Future work

- In this work no big effort was spent in developing a very efficient computational tool in terms of computational time and consequently further significant improvements from this perspective can be made. In particular, the code can be parallelised and faster numerical algorithms implemented. For example, the preconditioning matrix **P** mentioned in section 3.2.2.2.1 can be

chosen conveniently by using the Householder transformation or Givens rotation [10] for instance to generate orthogonal matrices instead of using direct methods. On the same line of thought the tool can be further developed for industry's use, that is, for use in a productive environment.

- Potentially, such a computational tool can be used for both design and analysis. If the computational tool is intended to be used in a design environment then it may be appropriate that a Hermitian subspace is considered instead of a Lagrangian one. This should considerably enhance the response surface representation of the projection coefficients. In a design setting, it is of utmost importance to accurately calculate the sensitivities with respect to parametric changes for an accurate approximation of the path between states, hence the importance of a Hermitian subspace.

- The variable-fidelity/POD-based ROM method was used to fuse (i) low-order accurate and high-order accurate viscous flow data and (ii) inviscid and viscous flow data. This modelling tool can be further developed to model and fuse computational data between coarse and fine grids. This problem can be very easily implemented. In this case the data fusion can be applied for all variables at common grid points between the coarse and fine grids. Intermediate grid points in the fine grid can be determined by some linear form of interpolation. Furthermore, other ways of fusing inviscid and viscous data may be sought.

- In this work it was deduced that the kernel POD method with the apriori chosen kernels taken into consideration, is far from adequate to model the strong non-linear features in an aerodynamic flow such as a moving shock wave with varying boundary conditions. Notwithstanding, these apriori chosen kernels failed to effectively and efficiently model highly non-linear features in aerodynamic flows, it is thought that it would still be worthwhile to investigate the behaviour of kernels derived from the data itself rather than choosing a kernel beforehand and using it. In particular, the method proposed by

Weinberger and Saul [136] which was mentioned in chapter 8 deserves some attention.

References

[1] Burkardt, J., Du, Q., Gunzburger, M., Lee, H.C., 'Reduced order modelling of complex systems', Proceedings NA03, Dundee, 2003

[2] Smith, L.I., 'A tutorial on Principal Components Analysis', available at http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf

[3] Antoulas, A.C., Sorenson, D.C. and Gugercin, S. 'A survey of model reduction methods for large scale systems', Contemporary Mathematics, AMS Publications, **280** : 193-219, 2001

[4] Karhunen, K., 'Zur Spektraltheorie Stochastischer Prozesse', Ann. Acad. Sci. Fennicae, **37**, 1946

[5] Loeve, M.M., 'Probability Theory', Princeton, N.J., Van Nostrand, 1955

[6] Hotelling, H., 'Analysis of a complex system of statistical variables into principal components ', Journal of Educational Psychology, **24** : 417-441, 488-520, 1933

[7] Lorenz, E.N., 'Empirical Orthogonal Functions and Statistical Weather Prediction', Cambridge, M.I.T., Department of Meteorology, Statistical Forecasting Project, 1956

[8] Brooks, C.L., Karplus, M., and Pettitt, B.M., 'Proteins: A theoretical perspective of dynamics, structure and thermodynamics', New York, Wiley, 1988

[9] Lumley, J.L., 'The structure of inhomogenous turbulent flows', In: A.M. Yanglom and V.I. Tatarski, Editors, *Atmospheric turbulence and radio wave propagation*, Nauka, Moscow,166-178, 1967

[10] Golub, G.H., Van Loan, C.F., 'Matrix Computations', The Johns Hopkins University Press, 3rd edition, 1996

[11] Sirovich, L., 'Turbulence and the Dynamics of Coherent Structure: I, II and III', Quarterly Applied Mathematics, **45**: 561, 1987

[12] Stewart, G.W., 'On the Early History of the Singular Value Decomposition', SIAM Review, **35**: 551-566, 1993

[13] Schmidt, E., 'Zur theorie der linearen und nichtlinearen Integralgleichungen I Teil: Entwicklung willkulicher Funktion nach Systemen vorgesschriebener', Mathematische Annalen, **63**: 433-476, 1907

[14] LeGresley, P.A. and Alonso, J.J., "Investigation of Non-Linear Projection for POD Based Reduced Order Models for Aerodynamics", AIAA Paper 2001-0926, presented at 39th Aerospace Sciences Meeting and Exhibit, Reno, NV, 2001

[15] Epureanu, B.I., Dowell, E.H., and Hall K., 'A Parametric Analysis of Reduced Order Models of Potential Flows in Turbomachinery using Proper Orthogonal Decomposition' , 2001-GT-0434, Proceedings of ASME Turbo Expo 2001, New Orleans, Louisiana, 2001

[16] Bui-Thanh, T., Damodaran, M., and Willcox, K., 'Proper Orthogonal Decomposition Extensions for Parametric Applications in Compressible Aerodynamics', AIAA 2003-4213, 21st Applied Aerodynamics Conference, Orlando, Florida, 2003

[17] Bui-Thanh, T., Damodaran, M., and Willcox, K., 'Aerodynamic Data Reconstruction and Inverse Design using Proper Orthogonal Decomposition', AIAA Journal, **42** (8), 2004

[18] Sirovich, L., 'Turbulence and the Dynamics of Coherent Structure, Part I: Coherent structures', Quarterly Applied Mathematics, **45** (3): 561-571, 1987

[19] Holmes, P., Lumley, J. L., and Berkooz, G., "Turbulence, Coherent structures, Dynamical Systems and Symmetry" Cambridge University Press, 1996

[20] Dowell, E.H., Hall, K.C., Thomas, J.P., Florea, R., Epureanu, B.I. and Heeg, J., "Reduced Order Models in Unsteady Aerodynamics", AIAA Paper 99-1261, 1999

[21] Hall, K. C., Thomas, J. P., and Dowell, E. H., "Reduced-Order Modeling of Unsteady Small-Disturbance Flows Using a Frequency-Domain Proper Orthogonal Decomposition Technique", AIAA Paper 99-0655, 1999

[22] Romanowski, M.C., 'Reduced Order Unsteady Aerodynamic and Aeroelastic Models using Karhunen-Loéve Eigenmodes', AIAA Paper 96-194, 1996

[23] Beran,P., and Silva,W., 'Reduced-Order Modeling: New Approaches for Computational Physics', AIAA Paper 2001-0853, 2001

[24] Kim, T., 'Frequency-Domain Karhunen-Loéve Method and Its Application to Linear Dynamic Systems', AIAA Journal, **36** (11) : 2117-2123, 1998

[25] Willcox, K.E., Paduano, J.D., and Peraire, J., and Hall, K.C., 'Low Order Aerodynamic Models for Aeroelastic Control of Turbomachines', AIAA Paper 99-1467, 1999

[26] Thomas, J.P., Dowell, E.H., and Hall, K.C., 'Using automatic differentiation to create a non-linear reduced-order model of a computational fluid dynamic solver', AIAA Paper 2006-7115, 11th AIAA Multidisciplinary Analysis and Optimization Conference, 2006

[27] Kirby, M., and Sirovich, L., 'Application of the Karhunen-Loéve Procedure for the Characterization of Human faces', IEEE Transactions on Pattern Analysis and Machine Intelligence, **12** (1), 1990

[28] Everson, R., and Sirovich, L., "The Karhunen-Loéve for Gappy Data", J. Opt. Soc. Am., **12** : 1657-1664, 1995

[29] Cohen, K., Siegel, S., and McLaughlin,T., 'Sensor Placement Based on Proper Orthogonal Decomposition Modeling of a Cylinder Wake', AIAA 2003-4259, June 2003

[30] Kirby, M., Boris, J.P., Sirovich, L., 'A proper orthogonal decomposition of a simulated supersonic shear layer', Intl. J. for Numerical Methods in Fluids Vol. 10, p. 411-428, 1990

[31] Baker, C., 'Orthogonal Decomposition of surface pressure fields', Proceedings of the 10th Inter. Conference on Wind Engineering, June 21-24, Copenhagen, Denmark.

[32] Dunyak, J., Gilliam, X., Smith, D., Wu, F., 'Proper orthogonal decomposition modeling for full-scale pressure fields', Proceedings of the 14th Engineering Mechanics Conference, ASC, May 21-24, Austin, Texas, 2000

[33] Ly, H.V. and Tran, H.T., 'Modeling and Control of Physical Processes using Proper Orthogonal Decomposition' Journal of Mathematical and Computer Modelling, 1999

[34] Tang, L., Chen, P., Liu, D.D., Gao, X., Shyy, W., Utturkar, Y., Zhang, B., 'Proper Orthogonal Decomposition and Response Surface method for TPS/RLV Structural Design and Optimization: X-34 Case Study', 43rd AIAA Aerospace Sciences Meeting and Exhibit, Reno, Nevada, 2005

[35] Lighthill, M.J., 'A New Method of Two-dimensional Aerodynamic Design', Aeronautical Research Council, London, 1945

[36] McFadden, G.B., 'An Artificial Viscosity Method for the Design of Supercritical Airfoils', New York University Report No. C00-3077-158.

[37] Hicks, R.M. and Henne, P.A., 'Wing Design by Numerical Optimization', Journal of Aircraft, **15** : 407-412, 1978

[38] Jameson, A., 'Aerodynamic Design via Control Theory', Journal of Scientific Computing, **3** : 233-260, 1988

[39] Reuther, J., Jameson, A., Farmer, J., Martinelli, L. and Saunders, D., 'Aerodynamic shape optimization of complex aircraft configurations via an adjoint formulation', AIAA paper 96-0094, AIAA 34th Aerospace Sciences Meeting and Exhibit, Reno, NV, 1996

[40] Pettit, C.L. and Beran, P.S., 'Reduced-order modelling for flutter prediction', AIAA 2000-1446, 41st AIAA/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference, Atlanta, GA, 2000

[41] Thomas, J. P., Dowell, E. H. and Hall, K.C., 'Three-dimensional transonic aeroelasticity using proper orthogonal decomposition based reduced order models', AIAA paper 2001-1526, 42nd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference and Exhibit, , Seattle, WA, 2001

[42] Cai, X., Ladeinde, F. 'A comparison of two POD methods for airfoil design optimization', AIAA Paper 2005-4912, 35th AIAA Fluid Dynamics Conference and Exhibit, 2005

[43] Anttonen, J.S.R., 'Techniques for reduced order modelling of aeroelastic structures with deforming grids' , Ph.D. Dissertation, Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio, 2001

[44] Lucia, D.J., 'Reduced Order Modeling for High Speed Flows with moving shocks', Ph.D. Dissertation, Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio, 2001

[45] LeGresley, P.A., Alonso, J.J., 'Dynamic Domain Decomposition and Error Correction for Reduced Order Models', 41st Aerospace Sciences Meeting and Exhibit, Reno, Nevada, 2003

[46] Lieu, T., and Farhat, C., 'Adaptation of POD-based Aeroelastic ROMs for varying mach number and angle of attack: Application to a complete F-16 configuration', AIAA Paper 2005-7666, U.S. Air Force T&E Days, 2005

[47] Lieu, T., Farhat, C., and Lesoinne, M., 'POD-based Aeroelastic analysis of a complete F-16 configuration: ROM adaptation and demonstration', AIAA Paper 2005-2295, 46th Structures, Structural dynamics & Materials Conference, 2005

[48] Du, Q., Faber, V., and Gunzburger, M., 'Centroidal Voronoi tessellations: Applications and algorithms', SIAM Review, **41** : 637-676, 1999

[49] Du, Q., and Gunzburger, M., 'Model reduction by proper orthogonal decomposition coupled with centroidal Voronoi tessellations', Proceedings of ASME FEDSM 2002, Fluids Engineering Division, Montreal, Quebec, Canada, July 2002

[51] Ludlow, DK., 'IMPNS Theory Guide', Cranfield Univesity, CoA Report NFP-0112, 2001

[52] Ludlow, DK., 'IMPNS User's Manual', Cranfield University, CoA Report NFP-0113, 2001

[53] Fluent 6.3 User's Guide, Fluent Inc. 2008

[54] Baldwin, B., Lomax, H., 'Thin layer approximation and algebraic model for separated turbulent flow', AIAA 78-257, 1978

[55] Degani., D., Schiff, L.B., 'Computation of turbulent supersonic flows around pointed bodies having cross-flow separation', Journal of Computational Physics, **66** (3): 173-196, 1986

[56] Qin, N., Jayatunga, C., 'Algebraic turbulence modelling for vortical flows around slender bodies', NATO RTO-MP-5, Missile Aerodynamics, Paper 20, 1998

[57] Spalart., P.R., Allmaras, S.R., 'A One-Equation Turbulence Model for Aerodynamic Flow', AIAA 92-0439, 1992

[58] Qin. N., Ludlow, D.K., 'A cure for anomalies of Osher and AUSM+ schemes for hypersonic viscous flows around swept cylinders', Proceedings of the 22nd International Symposium on Shock Waves, Imperial College, London, UK, 1999; 635-640 (Editors: Ball GJ, Hillier R and Roberts GT)

[59] Osher., S., Solomon, F., 'Upwind Difference Schemes for Hyperbolic Systems of Conservative Laws', Math. of Comp., **38** : 339–374, 1992

[60] Shaw, S., Qin, N., 'A matrix-free preconditioned Krylov subspace method for the PNS equations', AIAA 98-111, 1998

[61] Qin, N., Ludlow, D.K., Zhong, B., Shaw, S.T., Birch, T.J., 'Multigrid acceleration of a preconditioned GMRES implicit PNS solver', AIAA 99-0779, 1999

[62] Chang, C.L., Merkle, C.L., 'The relation between flux vector splitting and parabolised schemes', Journal of Computational Physics, **80** : 344-361, 1989

[63] Saad, Y., Schultz, M.H., 'GMRES: A generalised minimal residual algorithm for solving non-symmetric linear systems', SIAM J. Sci. Stat. Comput., **7**: 856–869, 1986

[64] Brandt, A., 'Multi-Level Adaptive Solutions to Boundary-Value Problems', Math. of Comp., **31** : 333–390, 1977

[65] Qin, N., Richards, B.E., 'Finite volume 3DNS and PNS solutions of hypersonic viscous flows around a delta wing using Osher's flux difference splitting', Proceedings of a Workshop on Hypersonic Flows for Re-entry Problems, 1990

[66] Birch, T.J., Qin, N., Jin, X., 'Computation of supersonic viscous flows around a slender body at incidence', AIAA 94-1938, 1994

[67] Birch, T.J., Ludlow, D.K., Qin, N., 'Towards an efficient, robust and accurate solver for supersonic viscous flows', Proceedings of the ICAS 2000 Congress, Harrogate, UK, 2000

[68] Birch, T.J., Prince, S.A., Ludlow, D.K., Qin, N., 'The application of a parabolized Navier-Stokes solver to some hypersonic flow problems', AIAA 2001-1753, 2001

[69] Mifsud, M.J., Shaw, S.T., 'Credible CFD Simulations of Complex Supersonic Weapon Configurations using the IMPNS Software', CCA Report 2005-02, Centre for Computational Aerodynamics, Department of Aerospace Science, School of Engineering, Cranfield University, 2005

[70] Mifsud, M.J., Chaplin, R.A., MacManus, D.G., Stollery, J.L., 'Heat-flux computations of the HSTDV launch vehicle with isothermal wall conditions', College of Aeronautics Report no. NFP-2007, Cranfield University, UK, 2007

[71] Qin, N., Lecture notes on computation of high-speed flows, College of Aeronautics, Cranfield University, UK, 1999

[72] Vigneron, Y.C., Rakich, J.V., Tannehill, J.C., 'Calculation of supersonic viscous flows over delta wings with sharp leading edges', AIAA paper 78-1137, 1978

[73] Steger, J.L., Warming, R.F., 'Flux vector splitting of the inviscid gasdynamic equations with application to finite difference methods', J. Comp. Phys., **40** : 263-293, 1981

[74] Anderson, W.K., Thomas, J.L. Van Leer, B., 'Comparison of finite volume flux vector splittings for the Euler equations', AIAA Journal, **24** : 1453-1460, 1986

[75] Box, G.E.P. and Draper, N.R., 'Empirical Model Building and Response Surfaces', John Wiley & Sons, 1987

[76] Myers, R.H. and Montgomery, D.C., 'Response Surface Methodology: Process and product optimization using designed experiments', John Wiley & Sons, 1995

[77] Forrester, A., Bressloff, N., Keane, A., 'Response Surface Model Evolution', In: *16th AIAA Computational Fluid Dynamics Conference*, Orlando , Florida, 2003

[78] Giunta, A.A., 'Aircraft Multidisciplinary Design Optimization using Design of Experiments Theory and Response Surface Modeling Methods', Ph.D. thesis, Virginia Polythecnic Institute and State University, 1997

[79] Balabanov, V.O., 'Development of Approximations for HSCT Wing Bending Material Weight using Response Surface Methodology, Ph.D. thesis, Virginia Polytechnic Institute and State University, 1997

[80] Venter, G., Non-dimensional Response Surfaces for Structural Optimization with Uncertainty', Ph.D. thesis, University of Florida, Gainsville, 1998

[81] Vavalle, A., 'Response surface aerodynamic optimization for blended wing body aircraft', Ph.D. thesis, Cranfield University, UK, 2005

[82] Powell, M.J.D., 'Radial basis functions for multivariable interpolation: a review', In: J.C.Mason and M.G.Cox, Editors, *Algorithms for Approximation*, Oxford University Press, 1987

[83] Buhmann, M.D., 'Radial Basis Functions: Theory and Implementations', Cambridge Monographs on Applied and Computational Mathematics, 2003

[84] Franke, R. and Nielson, G., 'Smooth interpolation of large sets of scattered data', International Journal for Numerical Methods in Engineering, **15** (11) : 1691-1704, 1980

[85] Carlson R.E. and Foley T.A., 'The Parameter R^2 in Multiquadric Interpolation', *Computers Math. Applic.*, **21**(9) : 29-42, 1991

[86] Applied Research Associates NZ Ltd. http://www.aranz.com/about/

[87] Liao, Y., Fang, S.C. and Nuttle, H.L., 'Relaxed conditions for radial basis networks to be universal approximators', Neural networks, **16**: 1019-1028, 2003

[88] Franke, R., 'Scattered data interpolation: tests of some methods', Mathematics of Computation, **38**: 181-200, 1982

[89] Carr, J.C., Fright, W.R. and Beatson, R.K., 'Surface interpolation with radial basis functions for medical imaging', IEEE Transactions on Medical Imaging, **16** : 96-107, 1997

[90] Kansa, EJ., 'Multiquadrics – A scattered data approximation scheme with applications to computational fluid-dynamics – Surface Approximations and Partial Derivative Estimates', *Computers Math. Applic.*, **19** (8/9) : 127-145, 1990

[91] Michelli, C.A., 'Interpolation of scattered data: distance matrices and conditionally positive definite functions', *Constructive Approximation*, **2**: 11-22, 1986

[92] Wang, B.P., 'Parameter optimization in multiquadric response surface approximations', *J. of Structural and Multidisciplinary Optimization*, **26** : 219-223, 2004

[93] Keane, A.J., and Nair, P.B., 'Computational Approaches for Aerospace Design: the pursuit of excellence', John Wiley & Sons, 2005

[94] Krishnamurthy, T., 'Comparison of response surface construction methods for derivative estimation using moving least squares, kriging and radial basis functions', In: *46th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference* , Austin, Texas, 2005

[95] Kirby, M., 'Geometrical Data Analysis: An Empirical Approach to Dimensionality Reduction and the Study of Patterns', John Wiley & Sons, 2000

[96] Zhongmin, W., 'Hermite-Birkhoff interpolation of scattered data by radial basis functions', *Approximation Theory and Applications*, **8**: 1-10, 1992

[97] Balmes, E., 'Parametric families of reduced finite element models: Theory and Applications', *Mechanical Systems and Signal Processing*, **10**:381-394, 1996

[98] Ito, K. and Ravindran, S.S., 'A reduced-order method for simulation and control of fluid flows', Journal of Computational Physics, **143**: 403-425, 1998

[99] Prud'homme, C., Rovas, D.V. and Veroy, K., 'Reliable real-time solution of parametrized partial differential equations: reduced-basis output bound methods', ASME Journal of Fluids Engineering, **124**: 70-80, 2002

[100] Montgomery, D.C., 'Design and Analysis of Experiments', John Wiley & Sons, 4th Edition, 1997

[101] McKay, M.D., Conover, W.J., and Beckman, R.J., 'A comparison of three methods for selecting values of input variables in the analysis of output from a computer code', *Technometrics*, **21**: 239-245, 1979

[102] Audze, P. and Englais, V., 'New approach to planning out of experiments', *Problem of Dynamics and Strength*, **35** : 104-107, 1977

[103] Lumley, J.L., 'Stochastic tools in turbulence', Academic Press, New York, 1971

[104] Vulikh, B.Z., 'Introduction to Functional Analysis for Scientists and Technologists', *Pergamon Press*, 1963

[105] Chatterjee, A., 'An introduction to the proper orthogonal decomposition', *Current Science*, 1999

[106] Sirovich, L., 'Chaotic dynamics of coherent structures', Physica D, **37**:126-145, 1989

[107] Pettit, C.L., and Beran, P.S., 'Reduced-order modelling for flutter prediction', AIAA paper 00-1446, *AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, Atlanta, Georgia, 2000

[108] Schmidt, E., Held, B. and Savick, D., 'Hypervelocity Launch Dynamics', AIAA paper 93 – 0502, 31st Aerospace Sciences Meeting & Exhibit, Reno NV, 1993

[109] Plostins, P., Silton, S., Schmidt, E. and Soencksen, K., 'Aerodynamic Jump at Hypervelocity', AIAA paper 2005 – 438, 43rd AIAA Aerospace Sciences Meeting and Exhibit, Reno NV, 2005

[110] Landers, M.G., Hall, L.H., Auman, L.M. and Vaughn, M.E., 'Deflectable Nose and Canard Controls for a fin stabilized projectile at supersonic and hypersonic speeds', AIAA paper 2003 – 3805, 21st Applied Aerodynamics Conference, Orlando FL, 2003

[111] Meunier, M.,M., 'Viscous CFD analysis of a deflectable nose, axi-symmetric projectile at high-speeds', M.Sc. thesis, Cranfield University, UK, 2005

[112] Shoesmith, B., Birch, T., Meunier, M., Mifsud, M. and Shaw, S., 'CFD analysis of a supersonic projectile with deflectable nose control', AIAA paper 2006-3200, 3rd AIAA Flow Control Conference, San Francisco CA, 2006

[113] Moore, F.G. and Hymer, T.C., 'Semi-empirical prediction of pitch damping moments for configurations with flares', AIAA paper 2001-0101, 39th Aerospace Sciences Meeting and Exhibit, Reno, NV, 2001

[114] Moore, F. and Hymer, T., 'The 2002 Version of the Aeroprediction Code (AP02)', AIAA paper 2003-26, 41st Aerospace Sciences Meeting and Exhibit, Reno, NV, 2003

[115] Schmidt, R. and Sooy, T., 'Aerodynamic Predictions, Comparisons and Validations using Missile DATCOM (97) and Aeroprediction 98 (AP98)', Journal of Spacecraft and Rockets, **22**(2) : 257-265, 2005

[116] Moore, F. and Hymer, T., '2005 Version of the Aeroprediction Code (AP05)', Journal of Spacecraft and Rockets, **42** (2) : 240-256, 2005

[117] Alexandrov, N.M., 'Robustness properties of a trust region framework for managing approximations in engineering optimization', 6th AIAA/NASA/ISSMO symposium on multidisciplinary analysis and optimization, **2** : 1056-1059, AIAA-96-4102-CP, Bellevue, WA, 1996

[118] Alexandrov, N.M., Dennis, J.E., Lewis, R.M., and Torczon V., 'A trust region framework for managing the use of approximation models in optimization', NASA/CR-201745, ICASE Report No. 97-50, 1997

[119] Alexandrov, N.M., Lewis, R.M., Gumbert, C.R., Green, L.L. and Newman P.A., 'Optimization with variable-fidelity models applied to wing design', AIAA paper 2000-0841, 2000

[120] Alexandrov, N.M., Nielsen, E.J., Lewis, R.M, and Anderson, W.K., 'First-order model management with variable-fidelity physics applied to multi-element airfoil optimization', *8th AIAA/USAF/NASA/ISSMO symposium on multidisciplinary analysis and optimization*, AIAA paper 2000-4886, Long Beach, CA, 2000

[121] Alexandrov, N.M., Lewis, R.M., Gumbert, C.R., Green, L.L., and Newman P.A., 'Approximation and model management in aerodynamic optimization with variable-fidelity models', *Journal of Aircraft*, **38** (6), 1093-1101, 2001

[122] Kaufmann, M., Balabanov, V., Burgee, S.L., Giunta, A.A., Grossman, B., Haftka, R.T., Mason, W.H. and Watson, L.T., 'Variable-complexity response surface approximations for wing structural weight in HSCT design', *Computational Mechanics*, **18** (2), 112-126, 1996

[123] Haftka, R.T., 'Combining global and local approximations', *AIAA Journal*, **29** (9), 1523-1525, 1991

[124] Toropov, V.V. and Markine, V.L., 'Use of simplified numerical models as approximations: Application to a dynamical optimal design problem', In: *ISSMO/NASA First Internet Conference on Approximation and Fast Reanalysis Techniques in Engineering Optimization*, 1998

[125] Reisenthel, P.H., Love, J.F., Lesieutre, D.J. and Dillenius, M.F.E., 'Innovative Fusion of Experiment and Analysis for Missile Design and Flight Simulation', *RTO Symposium on Innovative Missile Systems*, Amsterdam, 2006, (Restricted paper. Presentation is accessible from the world wide web: http://www.nearinc.com/)

[126] Li, H., Manjunath, B.S. and Mitra, S.K., 'Multisensor Image Fusion using the Wavelet Transform Graphical Models and Image Processing', **57** : 235-245, 1993

[127] Robinson, T.D., Eldred, M.S., Willcox, K.E. and Haimes, R., 'Strategies for multifidelity optimization with variable dimensional hierarchical models', In: *47th*

AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, Newport, Rhode Island, 2006

[128] Robinson, T.D., Willcox, K.E., Eldred, M.S. and Haimes, R., 'Multifidelity optimization for variable-complexity design', In: *11th AIAA/ISSMO/ Multidisciplinary Analysis and Optimization Conference*, Portsmouth, Virginia, 2006

[129] Vapnik, V. and Chervonenkis, A., 'Theory of pattern recognition; statistical learning problems', Nauka, Moscow, 1974

[130] Shawe-Taylor, J. and Cristianini, N., 'Kernel methods for pattern analysis', Cambridge University Press, 2004

[131] Schoelkopf, B. and Smola, A., 'Learning with Kernels: Support Vector Machines, Regularization, Optimization and Beyond', MIT Press, 2001

[132] Schoelkopf, B., Smola, A. and Mueller, K.R., 'Non-linear component analysis as a kernel eigenvalue problem', *Neural Computation*, **10**: 1299-1319, 1998

[133] Mika, S., Schoelkopf, B., Smola, A., Mueller, K.R., Scholz, M. and Rätsch G., 'Kernel PCA and de-noising in the feature space', *Advances in Neural Information Processing Systems*, **11**: 536-542, MIT Press, 1999

[134] Kwok,J.T.-Y and Tsang, I.W.-H, 'The pre-image problem in kernel methods', IEEE Transactions on Neural Networks, **15**(6), 1517-1525, 2004

[135] Rathi, Y., Dambreville, S. and Tannenbaum, A., 'Statistical shape analysis using kernel PCA', IS&T, SPIE Symposium on Electronic Imaging, 2006

[136] Weinberger, K.Q. and Saul, L.K., 'Unsupervised learning of image manifolds by semi-definite programming', Int. J. Comput. Vision **70** (1) : 77–90, 2006

[137] Schölkopf, B., Mika, S., Smola, A.J., Rätsch, G. and Müller, K.R., 'Kernel PCA pattern reconstruction *via* approximate pre-images', In: L. Niklasson, M. Bodén, and T. Ziemke, Editors, *Proceedings of the 8th International Conference on Artificial Neural Networks*, Perspectives in Neural Computing, 147 -152, Springer Verlag, Berlin, 1998

[138] Liu, Q., Cheng, J., Lu, H. and Ma, S., 'Distance based Kernel PCA Image reconstruction', *IEEE Proceedings of the 17th International Conference on Pattern Recognition*, 2004

[139] Kim, K.I., Jung, K. and Kim, H.J., 'Face recognition using kernel principal component analysis', *IEEE Signal Processing Letters*, **9**(2): 40-42, 2002

Appendix A

A.1 Shift parameter optimisation in multi-quadric response surface approximation

In this appendix, the shift parameter h of a multi-quadric RBF is introduced and considered instead of the shape parameter mentioned earlier. It is to be noted that the difference between the two parameters is only in the place where they appear in the multi-quadric RBF.

An efficient method of computing the optimal shift parameter based on the leave-one-out cross validation technique is proposed by Wang [92]. In the same work Wang also proved that the condition number of the multi-quadric coefficient matrix is an increasing function of the shift parameter h. This same technique proposed by Wang has been adopted and implemented in this work.

Some results from theory for multi-quadric RBFs :

When a set of data (x_j, F_j) is composed of distinct data points, the multi-quadric coefficient matrix $\boldsymbol{A}$

$$\boldsymbol{A} = \left[A_{ij}\right] = \sqrt{\left\|x_i - x_j\right\|^2 + h} \quad , h \geq 0 \qquad \text{Eq. (A-1)}$$

has the following properties:

1. Matrix $\boldsymbol{A}$ is always invertible, that is, it has a rank N. This was proved by Michelli [91]. This property guarantees the existence and uniqueness of multi-quadric approximation for a given set of data and a shift parameter h.

2. Matrix $\boldsymbol{A}$ has one positive eigenvalue and N-1 negative eigenvalues. Again this was proved by Michelli [91].

3. Matrix $\boldsymbol{B} = \left[1 / A_{ij} \right]$ that is the elements of the matrix $\boldsymbol{B}$ are the reciprocal of the elements in matrix $\boldsymbol{A}$, is strictly positive definite. Again from Michelli [91].

4. The condition number of matrix $\boldsymbol{A}$ increases monotonically with the parameter h. This implies that an upper bound for the value of h must be imposed. Wang [92] showed that the derivative of the eigenvalue of $\boldsymbol{A}$ with respect to the shift parameter h is always positive.

The proof that $\frac{d\lambda_i}{dh} \geq 0$ is as follows:

Considering the eigenvalue problem

$$\boldsymbol{A}\boldsymbol{v} = \lambda \boldsymbol{v} \qquad \text{Eq. (A-2)}$$

Since the matrix $\boldsymbol{A}$ is symmetric all the eigenvalues λ_i are real. Choosing a real orthonormal eigenvector v_i, i.e., $v_i^T v_i = 1$.

Appendix A – Shift parameter optimisation in multi-quadric response surface approx.

Let $\lambda(A) = \lambda_1, \lambda_2, \lambda_3, ..., \lambda_n$ = Eigenvalues of $\boldsymbol{A}$ in ascending order of $|\lambda_i|$ and property (2) guarantees that $\lambda_n > 0$, all other $\lambda_i \leq 0$. Then the eigenvalue derivative can be calculated using the formula

$$\frac{d\lambda_i}{dh} = v_i^T \frac{dA}{dh} v_i \qquad \text{Eq. (A-3)}$$

From Eq. (A-1) we have

$$\left(\frac{dA}{dh}\right)_{ij} = \frac{1}{2}\frac{1}{A_{ij}} \qquad \text{Eq. (A-4)}$$

Thus, from property (3) we have

$$\left(\frac{dA}{dh}\right)_{ij} = \frac{1}{2}B = \text{ a positive definite matrix}$$

It follows the quadratic form

$v_i^T \frac{dA}{dh} v_i$ is positive definite

Thus, $\frac{d\lambda_i}{dh} > 0$

Now, proceeding with the proof that the condition number increases monotonically with the parameter *h*, for a symmetric matrix $\boldsymbol{A}$, the condition number C can be defined as

$$C = \frac{\max(|\lambda_A|)}{\min(|\lambda_A|)} = \frac{\lambda_n}{-\lambda_1},$$

where λ_1 and λ_n are the minimum and maximum eigenvalues of $\boldsymbol{A}$ respectively.

Thus we have,

$$\frac{dC}{dh} = \frac{1}{\lambda_1^{\,2}}\left(\lambda_n \lambda_1^{'} - \lambda_n^{'} \lambda_1\right)$$

Recalling that the eigenvalues are ordered in ascending order of their magnitude; thus

$$\lambda_n > 0,\ \ \lambda_n^{'} > 0,\ \ \lambda_1 < 0,\ \ \lambda_1^{'} > 0\,.$$

Therefore, we have

$$\frac{dC}{dh} > 0$$

This property explains why the multi-quadric approximation deteriorates rapidly when h becomes too large. This is simply due to the increase in the condition number of the multi-quadric coefficient matrix.

The shift parameter can be optimized using the leave-one-out cross validation technique. The problem can be formulated as:

Find h to minimize E(h), where

$$E(h) = \sum_{i=1}^{N}\left(\bar{f}_i - f_i\right)^2 \qquad \text{Eq. (A-5)}$$

and $\bar{f}_i$ is the function value predicted at the i-th data point using multi-quadric approximation based on the database that excludes the i-th data.

Note that the direct evaluation of Eq. (A-5) would require solving the multi-quadric approximation N times for each trial value of h.

Alternatively, E(h) can be calculated from the following efficient method

$$E = \sum_{i=1}^{n} (C_i / B_{ii})^2 \qquad \text{Eq. (A-6)}$$

where

$C_i =$ coefficient of the multi-quadric approximation for all the data,

$\boldsymbol{B} = \boldsymbol{A}^{-1}$,

$\boldsymbol{A} =$ multi-quadric coefficient matrix,

$B_{ii} =$ the i-th diagonal element of $\boldsymbol{B}$

A proof for Eq. (A-6) can be found in [92].

Appendix B

B.1 Hermite interpolation using RBFs

To illustrate the idea of Hermite interpolation, consider and denote a training dataset by

$$[\mathbf{x}^{(i)}, \mathrm{y}(\mathbf{x}^{(i)}), \nabla \mathrm{y}(\mathbf{x}^{(i)})], \qquad i=1,2,.....,n$$

where $\nabla \mathrm{y}(\mathbf{x}) = [\partial y / \partial x_1, \partial y / \partial x_2, ..., \partial y / \partial x_p] \in R^p$

which denotes the first derivative of the output y(**x**) with respect to the components of the input vector. Then, a Hermite interpolant can be written in terms of a set of RBFs as follows:

$$\hat{y}(x) = \sum_{i=1}^{n} \alpha_i \psi\left(\left\| x - x^{(i)} \right\|\right) + \sum_{i=1}^{n} \sum_{j=1}^{p} \alpha_{ij} \frac{\partial \psi\left(\left\| x - x^{(i)} \right\|\right)}{\partial x_j} \qquad \text{Eq. (B-1)}$$

where α_i and α_{ij}, i = 1,2, ..., n, j = 1,2,....., p are undetermined weights.

Since the training dataset contains $y(\mathbf{x}) \in R$ and $\nabla y(\mathbf{x}) \in R^p$ at n points, a total of $n(p + 1)$ linear algebraic equations are required to ensure a unique solution. The first set of n equations using the function values at the points $\mathbf{x}^{(i)}$, $i = 1, 2,....., n$ can be written as

$$\sum_{i=1}^{n} \alpha_i \psi\left(\left\|x^{(k)} - x^{(i)}\right\|\right) + \sum_{i=1}^{n}\sum_{j=1}^{p} \alpha_{ij} \frac{\partial \psi\left(\left\|x^{(k)} - x^{(i)}\right\|\right)}{\partial x_j} = y(x^{(k)}), \qquad \text{Eq. (B-2)}$$

where $k = 1, 2,....., n$.

An additional set of np equations can be derived by matching the surrogate derivatives with the derivative information available in the training dataset as follows:

$$\nabla \hat{y}(\mathbf{x}^{(i)}) = \nabla y(\mathbf{x}^{(i)}), \quad i = 1, 2,....., n \qquad \text{Eq. (B-3)}$$

Note that to apply Eq.(B-3) the Hermite interpolant in Eq. (B-1) needs to be differentiated with respect to x_l, which yields

$$\frac{\partial \hat{y}(\mathbf{x})}{\partial x_l} = \sum_{i=1}^{n} \alpha_i \frac{\partial \psi\left(\left\|\mathbf{x} - \mathbf{x}^{(i)}\right\|\right)}{\partial x_l} + \sum_{i=1}^{n}\sum_{j=1}^{p} \alpha_{ij} \frac{\partial^2 \psi\left(\left\|\mathbf{x} - \mathbf{x}^{(i)}\right\|\right)}{\partial x_l \partial x_j} \qquad \text{Eq. (B-4)}$$

It can be noted from the preceding equation that in order to construct a Hermite interpolation model, the radial basis function must be differentiable at least twice. The system of linear algebraic equations arising from Eq.(B-2) and Eq.(B-3) can be compactly written as $\mathbf{K}_g \boldsymbol{\alpha}_g = \mathbf{y}_g$ where

$$\boldsymbol{\alpha}_g = \left\{\alpha_1, \alpha_{11}, \alpha_{13},...., \alpha_{1p},...., \alpha_n, \alpha_{n1}, \alpha_{n2},...., \alpha_{np}\right\}^T \in R^{n(p+1)} \qquad \text{Eq. (B-5)}$$

and

$$\boldsymbol{y}_g = \left\{ \begin{matrix} y(x^{(1)}), \dfrac{\partial y(x^{(1)})}{\partial x_1}, \dfrac{\partial y(x^{(1)})}{\partial x_2}, ..., \dfrac{\partial y(x^{(1)})}{\partial x_p}, ... \\ ...., y(x^{(n)}), \dfrac{\partial y(x^{(n)})}{\partial x_1}, \dfrac{\partial y(x^{(n)})}{\partial x_2}, ..., \dfrac{\partial y(x^{(n)})}{\partial x_p} \end{matrix} \right\}^T \in R^{n(p+1)} \qquad \text{Eq. (B-6)}$$

The coefficient matrix $\mathbf{K} \in R^{n(p+1)n(p+1)}$ can be written in partitioned form as follows

$$\mathbf{K}_g = \begin{bmatrix} \Theta_{11} & \Theta_{12} & ... & \Theta_{1n} \\ \Theta_{21} & \Theta_{22} & ... & \Theta_{2n} \\ ... & ... & ... & ... \\ \Theta_{n1} & \Theta_{n2} & ... & \Theta_{nn} \end{bmatrix} \qquad \text{Eq. (B-7)}$$

where

$$\boldsymbol{\Theta}_{ij} = \begin{bmatrix} \psi(\left\|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\right\|) & \dfrac{\partial \psi(\left\|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\right\|)}{\partial x_1} & ... & \dfrac{\partial \psi(\left\|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\right\|)}{\partial x_p} \\ \dfrac{\partial \psi(\left\|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\right\|)}{\partial x_1} & \dfrac{\partial^2 \psi(\left\|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\right\|)}{\partial x_1^2} & ... & \dfrac{\partial^2 \psi(\left\|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\right\|)}{\partial x_1 \partial x_p} \\ ... & ... & ... & ... \\ \dfrac{\partial \psi(\left\|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\right\|)}{\partial x_p} & \dfrac{\partial^2 \psi(\left\|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\right\|)}{\partial x_p \partial x_1} & ... & \dfrac{\partial^2 \psi(\left\|\mathbf{x}^{(i)} - \mathbf{x}^{(j)}\right\|)}{\partial x_p^2} \end{bmatrix} \qquad \text{Eq. (B-8)}$$

This set of equations can be solved to compute the weights of the Hermite RBF interpolant. In comparison to the standard RBF approximation technique presented earlier, the size of the system of equations to be solved in the Hermite interpolation approach is a function of *p* since $\mathbf{K} \in R^{n(p+1)n(p+1)}$. As a result, the computational cost and memory requirements of Hermite interpolation become significant when the number of training points and design variables is increased. However, one way to reduce the computational effort and memory requirements is to use sensitivity information only at those training points that are of particular interest.

Appendix C

C.1 Robinson's variable-fidelity model using POD mapping

In this method a set of q snapshots or column vectors $\mathbf{x}^1$, $\mathbf{x}^2$, $\mathbf{x}^3$,…., $\mathbf{x}^q$ each with different solutions in time or space is computed. A matrix **X** is then formed as follows

$$\mathbf{X} = \left[\left(\mathbf{x}^1 - \bar{\mathbf{x}}\right) \ \left(\mathbf{x}^2 - \bar{\mathbf{x}}\right) \ \text{.....} \ \left(\mathbf{x}^q - \bar{\mathbf{x}}\right) \right], \qquad \text{Eq. (C-1)}$$

where $\bar{\mathbf{x}}$ is the mean of the snapshots. From the SVD of matrix **X,** the left singular vectors correspond to the POD modes or basis vectors φ^j , j = 1, 2,…., q and a low dimensional representation of a solution **x** is given by

$$\mathbf{x} \approx \bar{\mathbf{x}} + \sum_{j=1}^{r} \alpha_i \varphi^i \qquad \text{Eq. (C-2)}$$

where α_i is the coefficient which quantifies the contribution of the i^{th} POD mode to the solution **x**.

The gappy POD method makes it possible to reconstruct data from a 'gappy dataset' that is a set of data in which some of the data is missing. As a first step towards the

solution of this problem, a mask vector is defined which describes for a particular solution vector where data is available and where data is missing. For the solution **x,** the corresponding mask vector **n** is defined as follows:

$$n_i = 0 \text{ if } x_i \text{ is unknown}$$
$$n_i = 1 \text{ if } x_i \text{ is known}$$

where x_i denotes the i^{th} element of the vector **x**. The dot product multiplication is defined as $(\mathbf{n},\mathbf{x})_i = n_i x_i$ and the gappy product multiplication is defined as $(\mathbf{u},\mathbf{v})_n = ((\mathbf{n},\mathbf{u}),(\mathbf{n},\mathbf{v}))$, and the induced norm is $\left(\|\mathbf{v}\|_n\right)^2 = (\mathbf{v},\mathbf{v})_n$.

Now for some vector **x** that has some unknown elements**,** it is assumed that the repaired vector $\breve{\mathbf{x}}$ can be represented by Eq. (C-2). In this, the projection coefficients are chosen to minimize the error between the available and reconstructed data. This error can be defined as

$$\varepsilon = \left\|\mathbf{x} - \breve{\mathbf{x}}\right\|_n^2 \qquad \text{Eq. (C-3)}$$

where only the original existing data elements in **x** are compared. The coefficients which minimize the error ε can be found by differentiating Eq. (C-3) with respect to each of the α_i in turn. This will then lead to the linear system of equations

$$\mathbf{M}\boldsymbol{\alpha} = \boldsymbol{f} \qquad \text{Eq. (C-4)}$$

where the ij^{th} component of **M** is given by

$$M_{ij} = \left(\varphi^i, \varphi^j\right)_n \qquad \text{Eq. (C-5)}$$

and the i^{th} component of $\boldsymbol{f}$ is given by

$$f_i = \left(\mathbf{x}, \varphi^i\right)_n \qquad \text{Eq. (C-6)}$$

Solving Eq. (C-4) for α, the missing elements of **x** can be obtained using the expansion of Eq. (C-2).

This method can be used to map data between high- and low-fidelity space data, with the high-fidelity treated as the unknown data and the low-fidelity as the known. In this mapping application, the basis vectors must span both low- and high-fidelity spaces and this is achieved by generating a set of r training pairs, for which the low- and the high-fidelity vectors describe the same physical system. These training pairs are combined in the following way to form the snapshot matrix:

$$\mathbf{X} = \begin{bmatrix} \left(\hat{\mathrm{x}}_1 - \bar{\hat{\mathrm{x}}}\right) & \left(\hat{\mathrm{x}}_2 - \bar{\hat{\mathrm{x}}}\right) & \ldots & \left(\hat{\mathrm{x}}_\mathrm{r} - \bar{\hat{\mathrm{x}}}\right) \\ - & - & \ldots & - \\ \left(\mathrm{x}_1 - \bar{\mathrm{x}}\right) & \left(\mathrm{x}_2 - \bar{\mathrm{x}}\right) & \ldots & \left(\mathrm{x}_\mathrm{r} - \bar{\mathrm{x}}\right) \end{bmatrix} \qquad \text{Eq. (C-7)}$$

where the i^{th} column of **X** contains both the i^{th} low- and the i^{th} high-fidelity snapshots, and $\bar{\hat{\mathbf{x}}}$ denotes the mean of the low-fidelity snapshot set. Again the left singular vectors of this matrix give rise to the POD basis vectors, which are partitioned in the same way as the snapshot vectors. Thus, Eq. (C-2) can be decomposed into two equations

$$\mathbf{x} = \bar{\mathbf{x}} + \sum_{i=1}^{q} \alpha_i \varphi^i \qquad \text{Eq. (C-8)}$$

$$\hat{\mathbf{x}} = \bar{\hat{\mathbf{x}}} + \sum_{i=1}^{q} \alpha_i \hat{\varphi}^i \qquad \text{Eq. (C-9)}$$

Eq. (C-9) can be solved in a least squares sense to find the coefficients α that best represent the low-fidelity vector. These coefficients are then used in Eq. (C-8) to calculate the high-fidelity vector.

Appendix C – Robinson's variable-fidelity model using POD mapping