Volumetric Flow Prediction using Multiple Plane Particle Image Velocimetry

by

David J. Hunter

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

School of Engineering

September 2010

THE UNIVERSITY OF
WARWICK

Contents

Acknowledgments

Firstly, I would like to thank Prof. Peter Bryanston-Cross and Dr Brenda Timmerman for their invaluable supervision throughout my PhD, and for putting up with my occasional stubbornness, erratic working hours, and my constant inability to stick to deadlines.

The advice and support I have received from my friends and colleagues at Warwick has been overwhelming, and it would be impossible for me to thank everyone who has helped me. However, I am particularly indebted to Dr Andrew Skeen for his sharing his PIV expertise, Jop Vlaskamp for his help on all matters electronic, and Mark Brend, who has advised me on almost all aspect of my work. I should also thank Mark for tolerating my countless clumsy attempts to explain principal component analysis to him, although I expect his patience would have been considerably shortened had our conversations not taken place in the pub.

Finally, and most importantly, I wish to thank my family and my girlfriend Briony, whose love, and patience and encouragement have made this work possible.

Abstract

This thesis presents an approach to predicting a 3-dimensional, 3-component velocity field of a fluid flow that possesses a homogeneous dimension. At the core of this approach is the technique of stochastic estimation, which is commonly used to combine a small number of instantaneous measurements with previously acquired statistical data, to produce a prediction of the flow over a large number of locations. In the proposed technique, particle image velocimetry (PIV) is used to provide measurements for the stochastic estimation procedure, and the statistical stationarity along the homogeneous dimension of the flow is exploited to extend the use of stochastic estimation to provide a full volumetric prediction.

The first section concerns the prediction performance of stochastic estimation. It is shown how the traditional approach to stochastic estimation is equivalent to ordinary least squares (OLS) regression. The properties of OLS, previously unconsidered in stochastic estimation literature, are presented, and shown to have a number of practical uses in the design and implementation of stochastic estimation procedures. Several alternative approaches to flow prediction are selected for further study, and their performance is compared in a series of trials, based on data from a numerically simulated channel flow. The newly-introduced biased techniques are shown to outperform or equal the

accuracy of the stochastic estimation techniques across the entire range of parameters under investigation.

The second section introduces the proposed volumetric prediction technique. A proof of concept is obtained using volumetric data from the simulated channel flow, and the resulting predictions show excellent quantitative and qualitative agreement with the original data. The predicted vortex ring data compares favourably with previous theoretical and experimental studies, and visualisation of the volumetric data appears to show the existence of secondary vortical structures around the outside of the ring core, which have previously only been observed in numerical simulations.

Nomenclature

Abbreviations

3D-PTV volumetric particle tracking velocimetry, page 12

BLUE best linear unbiased estimator, page 66

CAV curl angular velocity, page 238

CRLB Cramér-Rao lower bound, page 68

CSE complimentary stochastic estimation, page 96

DNS direct numerical simulation, page 120

DOF degrees of freedom, page 195

EVD eigenvalue decomposition, or eigendecomposition, page 35

GLS generalised least squares, page 79

HPIV holographic particle image velocimetry, page 14

LDR low-dimensional response regression, page 96

LSE linear stochastic estimation, page 27

MLE maximum likelihood estimation, page 69

Experimental parameters

(u, v, w) velocity vector components in the (x, y, z) Cartesian coordinate system, page 207

(r', θ', z') local cylindrical coordinate system, aligned with an individual vortex ring, page 223

$(v_{t'}, v_{\theta'}, w')$ velocity vector components in the (r', θ', z') cylindrical coordinate system, page 223

(r, θ, z) cylindrical coordinate system, page 181

(v_r, v_θ, w) velocity vector components in the (r, θ, z) cylindrical coordinate system, page 207

δ_t time separation between the pulses of the light sheets for Q_1, Q_2 and Q_3, page 193

Δt PIV light sheet pulse separation time, page 9

Γ_0 average initial vortex ring circulation, page 179

h DNS channel half-height, page 121

ν kinematic viscosity, page 121

Ω MP-VSE prediction volume, page 154

P MP-VSE plane normal to homogeneous dimension, page 154

P' PIV measurement region used to provide P plane data, page 207

Q_i ith MP-VSE plane parallel to homogeneous dimension, page 155

R_i ith plane used to compare true and predicted flow field during MP-VSE proof-of-concept with simulated channel flow, page 164

e $1 \times p$ vector of linear regression model unexplained errors (for multivariate **y**), page 53

ϵ error term, represents the part of y unexplained by $\tilde{\mathbf{x}}$, page 59

$\boldsymbol{\gamma}_i$ ith population principal component, page 41

I   identity matrix, page 32

L   diagonal matrix of eigenvalues, page 35

l_i ith eigenvalue, page 35

λ_i ith population eigenvalue, page 41

$\hat{\mathbf{C}}$ LDR regression coefficient matrix for the prediction of **t** from **x**, estimated by $\hat{\mathbf{C}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{T}$, page 96

m number of observations, page 52

n number of independent variables in regression model, page 50

$\tilde{n}$ number of *directly observed* independent variables, page 55

p number of dependent variables in regression model, page 53

P   PLS coefficient matrix for the prediction of **x** from **t**, estimated by $\hat{\mathbf{P}} = (\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T\mathbf{X}$, page 93

Q   PLS coefficient matrix for the prediction of **y** from **t**, estimated by $\hat{\mathbf{Q}} = (\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T\mathbf{Y}$, page 91

R   PLS matrix of weights, page 90

$\mathbf{r}_i$ ith column of PLS matrix of weights **R**, page 93

T PLS/LDR matrix of score vectors, page 90

t PLS/LDR score vector, page 90

$\underline{\mathbf{P}}_{\xrightarrow{}\tilde{V}}$ orthogonal projection matrix; projects onto the subspace $\tilde{V}$, page 38

$\mathbb{R}^n$ n-dimensional vector space, page 30

r rank of the matrix $\hat{\mathbf{B}}_{OLS}$, which is at most $r = \min\left(\text{rank}(\mathbf{X}), \text{rank}(\mathbf{Y})\right) = \min(n, m, p)$, page 97

S diagonal matrix of singular values, page 38

s_i ith singular value, page 39

σ^2 unexplained variance, page 86

$\hat{\sigma}_\perp^2$ component of unexplained variance that resides in the null space of $\hat{\mathbf{Y}}_{OLS}$, page 103

$\hat{\sigma}_\perp^2$ component of unexplained variance contained in the subspace $\tilde{V}$, page 103

θ probability density function parameter vector, page 68

$\tilde{\mathbf{U}}$ matrix of truncated set of principal components, page 36

$\tilde{\mathbf{X}}$ approximation of $\mathbf{X}$, constructed from a truncated set of principal components, page 37

$\tilde{\mathbf{x}}$ approximation of $\mathbf{x}$ constructed from a truncated set principal components, page 36

$\tilde{\mathbf{Z}}$ matrix of truncated set of scores, page 36

$\tilde{\mathbf{z}}$ vector of truncated set of scores, page 36

U   matrix of principal components, page 33

$\mathbf{u}_i$ ith principal component, page 33

$\mathbf{u}_{L(i)}$ ith left singular vector, page 38

$\mathbf{U}_L$ matrix of left singular vectors, page 38

$\tilde{V}$ subspace, page 37

V_{CSE} subspace of $\mathbb{R}^p$ spanned by the CSE prediction $\hat{\mathbf{y}} = \mathbf{x}\hat{\mathbf{B}}_{CSE}$, page 99

$\tilde{V}_{\perp}$ null space of $\tilde{V}$, page 37

V_{OLS} subspace of $\mathbb{R}^p$ spanned by the OLS prediction $\hat{\mathbf{y}} = \mathbf{x}\hat{\mathbf{B}}_{OLS}$, page 98

$\mathbf{v}_{r(i)}$ ith right singular vector, page 38

$\mathbf{V}_R$ matrix of right singular vectors, page 38

V_{RRR} subspace of $\mathbb{R}^p$ spanned by the RRR prediction $\hat{\mathbf{y}} = \mathbf{x}\hat{\mathbf{B}}_{RRR}$, page 99

X $m \times n$ matrix, consisting of m observations of $\mathbf{x}$, page 52

x $1 \times n$ vector of independent variables, page 51

$\tilde{\mathbf{x}}$ $1 \times \tilde{n}$ vector of *directly observed* independent variables, page 55

Y $m \times p$ matrix, consisting of m observations of multivariate $\mathbf{y}$, page 53

y dependent variable (univariate), page 50

y $m \times 1$ vector, consisting of m observations of univariate y

        OR: $1 \times p$ vector of dependent variables (multivariate y), page 52

$\hat{y}$ prediction of y for a given observation of $\mathbf{x}$, page 51

Z matrix of scores, page 34

z vector of scores, page 33

z_i ith score, page 34

Stochastic estimation variables

$A_i, B_{ij}, \ldots$ stochastic estimation model coefficients, page 25

L vector of concatenated stochastic estimation model coefficients,
$\mathbf{L} = \{A_1, A_2, \ldots, A_{\tilde{n}}, B_{11}, B_{12}, \ldots, B_{\tilde{n}\tilde{n}}, C_{111}, \ldots\}$, page 25

E unconditional variable, page 21

e event vector, page 21

F_i ith element of concatenated unconditional variable vector **F**, page 25

L_i ith element of concatenated coefficient vector **L**, page 25

F vector of concatenated unconditional variable terms,
$\mathbf{F} = \{E_1, E_2, \ldots, E_{\tilde{n}}, E_1 E_1, E_1 E_2, \ldots, E_{\tilde{n}} E_{\tilde{n}}, E_1 E_1 E_1, \ldots\}$, page 25

G conditional variable, page 21

Chapter 1

Introduction

The experimental investigation of flow phenomena is an invaluable tool throughout many fields of inquiry. Its uses range from engineering applications, where experiments are employed to assess design decisions, through to research into the fundamental physics of fluids, where they provide support, and even inspiration, for new theories, models and concepts.

Although the history of fluids experimentation is as old as the field of fluid mechanics itself, the years surrounding the turn of the twentieth century were of particular significance. This era saw a number of groundbreaking flow visualisation experiments, most notably in the work conducted variously by Reynolds, Prandtl and Mach (reviewed in (Yang, 2001)). These experiments heralded a sustained period of development and adoption of a new class of experimental techniques, capable of determining the spatial and temporal dynamics of fluid phenomena.

In these early examples, the methods employed were purely qualitative in nature. However, in the years since, experimental techniques have progressed at an astonishing rate, embracing the development of photography and digital imaging, electronics, lasers, and computing along the way. Today, the ability to make accurate quantitative flow measurements is now well-established, and experimentalists have access to a number of techniques that can provide, for example, high temporal-resolution velocity point measurements (e.g. hot wire probes, laser Doppler anemometry), or high-spatial resolution planar measurements (e.g. particle image velocimetry). Much current research is now focused on the development of tools to provide full volumetric flow measurements.

It is perhaps ironic that advances in understanding of fluid dynamics, due in large part to the role played by experimentation, have fuelled the need for ever more detailed and complex measurements. Arguably, even with the

sophisticated tools currently available, it is often the case that the gap between what is required, and what can be measured, is as large as ever. Most flow phenomena of interest, particularly those of a turbulent nature, exhibit complex three-dimensional behaviour which manifests itself over a wide range of scales, both spatially and temporally. Hence, a complete description of a flow's behaviour requires the ability to make accurate, non-intrusive measurements, over a sufficiently large volume to capture the full region under investigation, yet with sufficient resolution in time and space to resolve the full range of scales present. Furthermore, such measurements are not limited solely to velocity, but also to all other relevant quantities, such as temperature, pressure and density. Until such techniques exist, certain measurements will continue to remain beyond the capabilities of fluid dynamicists.

In light of this problem, the technique of *stochastic estimation* provides a powerful means of bringing otherwise unobtainable experiments within reach. Underlying the approach is the idea that even when a property of a flow is not directly measured, some degree of its behaviour may still be apparent in the behaviour of other observed properties. If the relationship between the measured and unmeasured properties is known, then an estimate of the latter can be made. Using a purely statistical approach, stochastic estimation aims to determine the relationship between a potentially unlimited number of flow variables (e.g. point velocities), which can then be exploited to provide predictions of a scope beyond what could be obtained by measurement alone.

For example, before planar velocity measurements were commonplace, stochastic estimation provided a powerful means of estimating the instantaneous flow velocity at a large number of locations, using data from only a small number of hot-wire probes. Similarly, the disruption caused by the presence of physical probes can be avoided by predicting the flow field using probes

3

outside the region of interest. In recent years, stochastic estimation has allowed the benefits of planar and point measurements to be combined; first, the relationship between the planar and point measurements is obtained, which enables subsequent point measurements to be used to predict the planar flow field, yielding predictions with both high spatial and temporal resolution.

Since its introduction by Adrian (1977), stochastic estimation has acquired many different uses, both to address the changing needs of fluid dynamicists, and to incorporating further developments in measurement technology. The work presented in this thesis contributes a further application, which provides instantaneous volumetric predictions of a flow that possesses one or more homogeneous dimensions, i.e. a flow that, when expressed in an appropriate coordinate system, is statistically invariant along at least one axis. The proposed technique, from herein referred to as multiple plane volumetric stochastic estimation (MP-VSE) , uses the established technique of stereo particle image velocimetry (stereo PIV) to measure multiple parallel planes within the flow, and stochastic estimation is employed to predict the remaining volume. The proposed technique is flexible, modest in terms of its computational cost and experimental complexity, and can be used to predict volumes of a size and resolution currently well beyond the capability of conventional 3D-3C measurements techniques. Although the necessity of a homogeneous dimension limits the applicability of MP-VSE in many situations, there are still a significant number of commonly studied flow phenomena that fulfil this requirement. For example, flows over plates, wings, etc. often possess a strongly homogeneous dimension in a standard Cartesian coordinate system, and similarly for pipe flows, jets and vortex rings, when expressed in cylindrical coordinates.

The second contribution of this work relates to the theory of stochastic estimation itself. Although the technique is relatively modern, its mathemat-

ical underpinnings have been in existence for many years. As such, a wealth of relevant information exists in other areas, most notably from the field of linear regression analysis. A number of existing concepts are presented, along with some novel contributions, which form a detailed study of the practical performance of stochastic estimation. As well as bringing new understanding to the technique as it is conventionally applied, a number of alternative implementations are provided, which are shown to offer improved performance in many situations, including those likely to be encountered in applications of MP-VSE.

1.1   Thesis Outline

The thesis begins with a review of the relevant literature and mathematical concepts. An overview of full-field flow measurement techniques is given in chapter 2, which provides the opportunity to introduce the techniques that form the basis of the MP-VSE approach, and review the development and current status of approaches to volumetric flow measurement. Chapter 3 provides a description of the traditional mathematical treatment of stochastic estimation, and introduces the technique of principal component analysis, used extensively in later chapters.

The investigation into the practical performance of stochastic estimation begins in chapter 4, which introduces concepts from linear regression analysis. It is shown how and when stochastic estimation is equivalent to the technique *ordinary least squares* (OLS) regression. The properties of OLS are then presented, and their relevance to the task of flow prediction is discussed. Chapter 4 also introduces a selection of *biased* regression techniques, which are shown, theoretically at least, to outperform stochastic estimation in cer-

tain circumstances. This leads on to the work of chapter 5, which describes a series of simulated flow prediction experiments, used to quantitatively assess the performance of the biased techniques over a wide range of parameters.

The working principle of MP-VSE is introduced in chapter 6. The chapter closes with the description of a simulated MP-VSE experiment, which serves to establish the feasibility of the approach, and demonstrates that biased regression techniques are particularly suited to MP-VSE applications. Chapter 7 describes the design and implementation of an MP-VSE experiment, which has been used for the prediction of unstable vortex rings. The results of the experiment are compared with relevant experimental, numerical and theoretic work from literature.

Conclusions and avenues of future work are given in chapter 8.2.

Chapter 2

Full-Field Velocity Measurement Techniques

This chapter presents a review of several relevant full-field velocity measurement techniques, which are defined as those capable of the simultaneous measurement of velocity at multiple positions within a two or three dimensional region of a flow. In the following discussion, particular attention is given to the techniques of particle image velocimetry (PIV) and multiple plane PIV, which form the basis of the proposed MP-VSE approach, and which are subsequently employed in the experimental work in chapter 7. Although PIV forms the backbone of much of the work in this thesis, it is not the emphasis, and so a detailed discussion will not be undertaken here. A more detailed treatment of the concept can be found in the comprehensive book by Raffel *et al.* (2007), for example. The opportunity is also taken in this chapter to discuss the various approaches to volumetric flow measurement. The current strengths and limitations of these techniques are considered, which will highlight some of the advantages of the MP-VSE approach.

2.1 Particle Image Velocimetry

Particle image velocimetry (PIV) is one of a category of flow visualisation techniques which involve the observation of tracer particles in the flow. Such approaches have existed for over a century; first as a qualitative tool, as used by Prandtl as early as 1904 (described in (Yang, 2001)), and then later as the basis for several different quantitative techniques, e.g. particle tracking velocimetry and laser Doppler anemometry. PIV is a relatively recent addition to the field, having been established as a distinct technique in 1984, independently by both Adrian (1984) and Pickering and Halliwell (1984). Since then, PIV has arguably become the de facto technique for full-field planar velocity measurement.

The basic concept of PIV is appreciably simple. The flow in question is seeded with neutrally buoyant, flow-following tracer particles, and a planar region of the flow is illuminated using a light sheet of finite thickness. A measurement is obtained using one or more cameras, each making two recordings of the particle positions, separated in time by a short delay Δt. The velocity field within the viewing area is then determined from the spatial displacement of the particles.

Calculation of the particle displacement involves dividing the PIV images into small sub-regions, or windows, and finding the average movement of the particles within each interrogation area. This is typically achieved via cross-correlation, which in its basic form, involves spatially "shifting" a window from one image over the corresponding window from the other, and calculating the correlation between the two windows at each shifted position. This creates a two-dimensional correlation map, which provides the degree of correlation between the windows as a function of the spatial displacement between them. The displacement that yields the best correlation is indicated by a peak in the map, which is detected and subsequently used to calculate the average direction and magnitude of the velocity within the sub-region. In practice, the state-of-the-art cross-correlation routines are significantly more advanced than the simple procedure outlined here, and are capable of providing a much improved accuracy, reliability and spatial resolution (Scarano, 2002; Stanislas *et al.*, 2008).

The cross-correlation process results in a grid of spatially-arranged vectors, each describing the two in-plane velocity components of the velocity at that point. However, if all three velocity component are sought, then an extension to the PIV procedure, known as stereoscopic PIV (stereo PIV) can be employed. As the name suggests, stereo PIV employs two cameras which

view the same region of the light sheet from separate directions. Essentially, the standard cross-correlation approach is applied to the images from both cameras, producing two separate vector maps of the flow at the same instant. However, due to the different viewing locations, the perceived movement of the particles is different if the flow contains any out-of-plane movement. Using the disparity between the two velocity maps, it is possible to determine the out-of-plane velocity component. The process of recovering the third component places strict requirements on the calibration of the experimental setup. An accurate mapping from a position in the images to the corresponding position on the light sheet plane is required, so that the disparity between the two vectors is solely due to the out-of-plane velocity at that point. Any error in the mapping function means that the disparity may be calculated from the velocity at two separate points in the flow, which can result in an erroneous out-of-plane component. The task of minimising this calibration error has received much attention and sophisticated *self calibration* procedures exist to automatically detect and account for such problems, e.g. (Scarano *et al.*, 2005; Wieneke, 2005).

2.2 Volumetric Velocity Measurement

2.2.1 Multiple Plane PIV

Multiple plane PIV refers to the acquisition of instantaneous, or almost instantaneous, PIV measurements at two or more planes in the flow. A common use of this approach is volumetric flow measurement, where the volume is constructed from multiple closely spaced planar measurements. In this section, an overview of the many approaches to multiple plane PIV is presented. Tech-

nically, some of the following references do not relate specifically to volumetric measurements, but have been included here as potential candidates for use with the proposed MP-VSE technique.

The main challenge of multiple plane PIV is distinguishing between the separate light sheets, which can be achieved in several ways. A well established method of dual-plane measurements is through the use of orthogonally-polarised light sheets. Using polarisation-preserving seeding particles, discrimination of the two light sheets is possible by placing appropriate polarising filters on each camera system (Kähler and Kompenhans, 2000; Wernet *et al.*, 2005; Perret *et al.*, 2006). When more than two instantaneous PIV measurements are required, an alternative is to generate the light sheets using separate wavelengths. Separation of the measurement planes can then be obtained through the use of cameras with appropriate wavelength filters (Mullin and Dahm, 2006; Pfadler *et al.*, 2009) or a single camera system capable of accurate colour-capture (McGregor *et al.*, 2007; Pick and Lehmann, 2009). Generation of light sheets of different wavelength has been achieved variously by: multiple lasers of different wavelength (Mullin and Dahm, 2006; Pfadler *et al.*, 2009), separating the output of multiple-wavelength lasers (Post *et al.*, 1994) (and the subsequent mixing of wavelengths to produce multiple further colours (Ruck, 2009)), splitting white strobe light into multiple components (Pick and Lehmann, 2009), and through the use of stimulated Raman scattering (McGregor *et al.*, 2007). Many of these approaches theoretically provide the possibility of a large number of wavelength-separated light sheets (in the case of McGregor *et al.* (2007), 11 separate planes were generated). However, as the number of light sheets is increased, the wavelength separation of the sheets is necessarily reduced, and the increased level of cross-talk between neighbouring sheets ultimately places a limit on the depth of volume that can

be measured.

Light sheet discrimination can also be obtained by introducing a temporal separation between successive measurements, assuming that this separation is sufficiently small relative to the flow's time scales of interest. This can be achieved using a scanning light sheet and a single camera system (Brücker, 1996; Hori and Sakakibara, 2004). This approach is limited by the minimum achievable time separation between measurements which, depending on the equipment used, will be dictated by either the frame rate of the cameras, the mechanical scanning speed, or the laser repetition rate if a pulsed laser is employed. This ultimately restricts the approach to flows that have suitably large time scales, although improvements in technology will help to increase the applicability of the approach. To date, the most advanced application of scanning light sheet PIV is the work of Hori and Sakakibara (2004) who were able to measure 50 planes of a low Reynolds number ($Re \approx 1000$) water jet using 500Hz framing cameras.

2.2.2 Particle Tracking Velocimetry

The technique of particle tracking velocimetry (PTV) is closely related to PIV, whereby velocity measurements are obtained from the displacement of tracer particles in images of an illuminated flow region. Rather than dividing the image(s) into interrogation areas, PTV instead measures the displacement of individual particles. PTV can be used for planar measurements in a similar manner to PIV, using a single camera to measure two velocity components of particles in a planar light sheet (Adamcyzk and Rimai, 1988). Arguably though, it has found most success when extended to volumetric measurements (3D-PTV), which use multiple cameras at known positions to image an illu-

minated volume. The 3D position of the tracer particles is calculated from the images using the *collinearity condition* (Maas *et al.*, 1993), which, assuming a pinhole camera model, states that an image point, the camera projective centre (i.e. pinhole) and the 3D object point all lie on the same line. In 3D-PTV, the positions of individual particle centres are found in each image, and the 3D locations are found from the intersections of the lines projected from the image points through the camera centres. A common extension of the technique is to make time-resolved measurements, by tracking individual particles over consecutive locations. This provides knowledge of particle trajectories through time and space, and makes PTV ideally suited for Lagrangian analysis of the flow. However, the process of detecting particles, obtaining 3D locations and tracking their motion is not trivial, and the particle images must be sufficiently sparse in order to ensure accuracy. This in turn, limits the seeding density and volume depth that can be measured by 3D-PTV techniques (Tropea *et al.*, 2007). A further issue is that the resulting vector maps are arbitrarily distributed within the region of interest, and vary between measurements.

2.2.3 Holographic PIV

Holography provides a means of recording three-dimensional information about a scene in a single image. The holographic process involves two coherent light beams; the scene under investigation is illuminated with one beam, and the second *reference* beam is directed onto the imaging medium. Scattered light from the scene (known as the object wave) reaches the recording medium where it interferes with the reference beam. The interference pattern is recorded as an image, and can later be used to extract a three-dimensional representation

of the scene.

In holographic PIV (HPIV) a holographic image of a particle field is obtained, and the vector field is then extracted from the reconstructed scene using 3D cross-correlation or particle tracking techniques (Hinsch, 2002). As an intrinsic three-dimensional imaging technique, holography is, in principal, ideally suited for volumetric measurement of particle fields. However, the practical success of HPIV is dependent on what Meng *et al.* (2004) describe as the "critical issues" of axial accuracy and information capacity. Axial accuracy refers to the ability to resolve the position of particles in the direction normal to the image plane (axial direction). The axial accuracy is generally lower than in the remaining two dimensions, and results in reconstructed particles that appear elongated in the axial direction. The information capacity sets a limit on the maximum measurement volume and particle density that can be measured before image aberration and reconstruction noise becomes too severe to correctly determine particle positions. A variety of HPIV configurations have been proposed (summarised in Meng *et al.* (2004)). Generally, improved performance in the aforementioned areas comes at the cost of (often vastly) increased complexity. Arguably, the largest trade-off involves the choice of recording medium. Currently, the best performance is provided with traditional film emulsion. However, the holographic film must be removed from the setup and wet processed to obtain the holographic image after exposure, and reconstruction of the particle field involves illuminating the hologram with the original reference beam (or by a reversed, or conjugate, version of the reference beam) to create an image of the particle field in space. A permanent record of the field is made by traversing an image sensor, plane-by-plane, through the volume. Overall, these steps render the film-based HPIV procedure an inconvenient and time consuming endeavour, which is far removed from the

14

near-instantaneous measurements available with conventional PIV measurements. Conversely, digital holography, which employs CCD or CMOS sensors to record the hologram, allows for a considerably simpler setup. Most importantly, access to the digital image of the interference pattern is instant, and reconstruction of the particle field can be computed in software. However, even state-of-the-art digital sensors have a size and resolution that is far lower than that of holographic film, which drastically reduces both the axial accuracy and information capacity. As such, digital HPIV is only capable of measuring small regions of sparsely seeded flows. For example, Raffel *et al.* (2007) quote a typical measurement volume of 1cm^3 for digital HPIV, as opposed to 5cm^3 for film-based HPIV. To date, digital HPIV has been employed with a seeding density of around 12 particles/mm^3 (Meng *et al.*, 2004), whereas upwards of 60 particles/mm^3 has been used for film-based holography (Pu and Meng, 2000). Furthermore, the decrease in axial accuracy results in reconstructed particles that are elongated by 10-100 times in the axial direction (Atkinson and Soria, 2008). In order to extract accurate velocity measurements from the reconstructed particle field, this effect must be accounted for. Suggested approaches include novel particle detection techniques (Meng *et al.*, 2004) software-based filtering (Shen and Wei, 2004) and a combination of holographic and tomographic techniques (Soria and Atkinson, 2008). Ultimately though, the development of digital HPIV is likely to be driven most strongly by improvements in digital sensor technology.

2.2.4 Tomographic PIV

Tomographic PIV is a recent approach to volumetric measurement (Elsinga, Scarano, Wieneke and van Oudheusden, 2006), which employs the principle

of tomography to reconstruct the most likely distribution of tracer particles within a three-dimensional region, based on multiple views of the particle field. A pulsed light source illuminates the volumetric particle field, which is recorded with multiple cameras (typically three to six). Under the tomographic approach, the illuminated particle field is interpreted as a three-dimensional intensity distribution. Each camera pixel has a line of sight through this distribution, and the measured intensity of the pixel is taken to be the integral of the light intensity along the line of sight. Based on observed pixel intensities and knowledge of the corresponding lines of sight, the goal is to calculate the most likely intensity distribution that gave rise to the observed data. This is achieved using an iterative algorithm, which reconstructs the intensity field onto a discrete grid of three-dimensional *voxels*. The velocity field is then calculated from the reconstructed field using a volumetric form of the standard PIV cross-correlation approach.

Although in its infancy, tomographic PIV has successfully been applied to a variety of different flows (Raffel *et al.*, 2007). It is noted that, to date, applications of the technique have employed a volume that is considerably smaller in depth than in width and height. This may arise because of the need to maintain focus throughout the entire volume. The volume depth may also ultimately be limited by the fact that the occurrence of false positives in the particle reconstruction process (*ghost particles*) has been shown to increase with the depth of the volume (Elsinga, van Oudheusden and Scarano, 2006).

2.2.5 Taylor's Frozen Field Hypothesis

Although not strictly a measurement technique, certain flows permit the use of Taylor's hypothesis (Taylor, 1938) to make a prediction of the velocity

16

in a volumetric region of the flow. Taylor's hypothesis, also known as the *frozen field* hypothesis, states that the temporal derivative of a given velocity component u_i is related to the spatial derivative of u_i along the mean direction of the flow x, simply by the velocity U:

$$\frac{\delta u_i}{\delta t} + U\frac{\delta u_i}{\delta x} = 0 \tag{2.1}$$

where U is conventionally taken to be the mean flow velocity. Essentially, this implies that velocity fluctuations at a fixed point are due to the passage of a "frozen" pattern of turbulence, which is convected along by the mean velocity (Warsi, 2006). Using this assumption, it is possible to infer the spatial gradient along x from knowledge of the temporal gradient at a single point in the flow. Taylor's hypothesis has been used extensively in turbulence research, primarily due to the fact that until the advent of full-field measurement techniques, it provided the most feasible way to investigate the spatial dynamics of the flow using the time-resolved measurements from hot-wire probes. Since its introduction, it has been applied in a wide variety of situations, commonly for the calculation of wavenumber (i.e. spatial frequency) spectra from temporal frequency spectra, and other mean quantities based on spatial derivatives (Townsend, 1980). Another possibility is to use the approximated spatial gradient to directly estimate the instantaneous velocity field along x using the velocity time history at a single point in the flow. This was conventionally applied using hot-wire "rakes", to allow the construction of a two dimensional velocity field from a one dimensional array of point measurements (e.g. (Bonnet *et al.*, 1998)).

With the advent of time-resolved PIV systems, the same approach can be used to provide a full volumetric reconstruction of the flow by placing the

17

measurement plane orthogonal to the mean flow direction, and approximating the third dimension from the time history of the measurements. This has been employed variously by Matsuda and Sakakibara (2005); van Doorne and Westerweel (2007); Ganapathisubramani and Lakshminarasimhan (2007).

The validity of Taylor's Hypothesis depends primarily how closely the flow resembles the aforementioned "frozen field". This is dependent on the turbulence intensity, defined as the root mean square of the velocity fluctuations relative to the magnitude of mean velocity, which should be small in order to for the hypothesis to hold. However, what constitutes a "valid" flow depends ultimately on what the hypothesis is used for. It is noted by Townsend (1980) that the approximation of mean quantities are generally much less sensitive to deviations from the frozen-field assumption than the approximation of exact spatial gradients. This is particularly relevant when applying Taylor's hypothesis as a means of predicting the instantaneous flow field, as it implies that although the reconstructed field may share many statistical properties with the real flow field, instantaneous realisations of the flow may only be a poor approximation to the truth. In practice, the error between the true and reconstructed field will be small at distances close to the measurement position, but will invariably increase with separation, due to the evolution of the velocity fluctuations and the entrance of new structures into the line of travel (Belmonte *et al.*, 2000).

2.3   Summary

A review of several full-field measurement techniques has been presented, which are relevant to the proposed MP-VSE technique, either as the basis of the technique itself (in the case of PIV and multiple plane PIV), or as al-

ternative ways of producing volumetric velocity data. A common limitation of all the volumetric techniques described here is the amount of data that can be measured, due to restrictions in terms of resolution, volume size, particle density, etc. This highlights a major advantage of MP-VSE, which can be used to predict a volume of a far greater size and resolution than currently available. As a prediction, MP-VSE will not be able to rival the accuracy of direct measurements, but it's ability to provide a more complete representation of large flows regions is likely to motivate it's use.

Chapter 3

Review of Mathematical Concepts

3.1 Introduction

This chapter serves as an introduction to some of the main mathematical concepts employed in the following work on stochastic estimation. Traditionally, stochastic estimation is viewed as a method of estimating *conditional averages* in turbulence, and so the chapter begins by describing the theory of conditional averaging, its role in fluid dynamics, and the practical difficulties of its use, which motivated introduction of stochastic estimation. This leads on to section 3.3, which provides the standard mathematical presentation of stochastic estimation, and briefly documents its history. Section 3.4 details the statistical technique of principal component analysis (PCA), which is employed in later chapters as a tool to investigate a variety of implementations of the stochastic estimation procedure. Furthermore, PCA, under the name of *proper orthogonal decomposition* (POD), is used extensively in fluid dynamics, and often in conjunction with stochastic estimation. This aspect of the technique is considered in section 3.5.

3.2 Conditional Averaging

Given the (possibly multivariate) random variables $\mathbf{G}$ and $\mathbf{E}$, the conditional average $\langle \mathbf{G}|\mathbf{E} = \mathbf{e}\rangle$ describes the average behaviour of $\mathbf{G}$ during the occurrence of the event $\mathbf{E} = \mathbf{e}$. Here, the angled brackets denote an ensemble average over statistically independent observations. For convenience, the conditional average is often written as $\langle \mathbf{G}|\mathbf{E}\rangle$. The variable $\mathbf{G}$ is referred to as the *conditional* variable, and $\mathbf{E}$ as the *unconditional* variable.

For fluid dynamics applications, the conditional variable usually consists of velocity components at points within the flow. The unconditional variable

may also be velocity, although many other measures, such as temperature, pressure, or quantities derived from velocity have been used extensively. As an introductory example, consider the case where $\mathbf{G}$ consists of the velocity components, u, v, w at the points $\mathbf{x}_1', \mathbf{x}_2', \ldots$, and $\mathbf{E}$ is the velocity at the points $\mathbf{x}_1, \mathbf{x}_2, \ldots$, i.e.:

$$\mathbf{G} = \{u(\mathbf{x}_1', t), v(\mathbf{x}_1', t), w(\mathbf{x}_1', t), u(\mathbf{x}_2', t), v(\mathbf{x}_2', t), w(\mathbf{x}_2', t), \ldots\} \qquad (3.1)$$

and:

$$\mathbf{E} = \{u(\mathbf{x}_1, t), v(\mathbf{x}_1, t), w(\mathbf{x}_1, t), u(\mathbf{x}_2, t), v(\mathbf{x}_2, t), w(\mathbf{x}_2, t), \ldots\} \qquad (3.2)$$

In this example, the event $\mathbf{e}$ is defined by the vector:

$$\mathbf{e} = \{e_1 \pm \Delta e, e_2 \pm \Delta e, e_3 \pm \Delta e, e_4 \pm \Delta e, e_5 \pm \Delta e, e_6 \pm \Delta e \ldots\} \qquad (3.3)$$

which specifies a set of velocities that define the event of interest, where the quantity Δe is chosen to ensure that the event occurs with sufficient frequency. The conditional average $\langle \mathbf{G} | \mathbf{E} = \mathbf{e} \rangle$ therefore describes the average behaviour

of the velocity field at $\mathbf{x}_1', \mathbf{x}_2', \ldots$, given the occurrence of:

$$u(\mathbf{x_1}, t) = e_1 \pm \Delta e$$

$$\text{and} \ \ v(\mathbf{x_1}, t) = e_2 \pm \Delta e$$

$$\text{and} \ \ w(\mathbf{x_1}, t) = e_3 \pm \Delta e$$

$$\text{and} \ \ u(\mathbf{x_2}, t) = e_4 \pm \Delta e \qquad (3.4)$$

$$\text{and} \ \ v(\mathbf{x_2}, t) = e_5 \pm \Delta e$$

$$\text{and} \ \ w(\mathbf{x_2}, t) = e_6 \pm \Delta e$$

$$\text{and} \ \ \ldots$$

The traditional role of conditional averages in turbulence is to investigate coherent structure. Here, the event is chosen to be a phenomenon of perceived importance to the flow, which may be the occurrence of high velocity fluctuations, acceleration or deceleration of velocity, high vorticity or particular types of Reynolds stress (Antonia, 1981; George *et al.*, 1989). The resulting field of conditional averages can then provide information about the state of the flow that caused the event in question. One of the strengths of conditional averaging is that it can provide information about the flow at a large (technically unlimited) number of locations, using only limited measurement equipment; all that is required is the means to make simultaneous measurements of the unconditional data, $\mathbf{E}$, and at least one element of the conditional vector, $\mathbf{G}$. The conditional average for each element of $\mathbf{G}$ can be obtained separately, with the full field of conditional averages obtained by traversing the conditional measurement location to all the points in turn. In practice, the process of acquiring conditional average measurements usually employs some form of conditional sampling, whereby the occurrence of the event $\mathbf{E} = \mathbf{e}$ is monitored

in real-time, and one or more elements of the conditional quantity $\mathbf{G}$ are sampled whenever $\mathbf{E} = \mathbf{e}$. This approach has several drawbacks. As evidenced by equation (3.4), as the size of the event vector grows, the likelihood of the event occurring reduces, and the time taken to acquire enough data for convergence of the average can become prohibitively large. This problem is made worse if the size of $\mathbf{G}$ is also large, and the conditional average is to be constructed in a piece-wise manner, using the traversal method outlined above. Given the amount of experimental effort required, a second problem is that the measured data is only valid for the chosen event; if other events are to be investigated, then the process must be carried out again.

3.3   Stochastic Estimation

Stochastic estimation was introduced by Adrian (1977) as a means of estimating conditional averages, without the problems associated with their direct measurement. Underlying the approach of stochastic estimation is the idea that the conditional average $\langle \mathbf{G}|\mathbf{E} \rangle$ is a function of the unconditional variable $\mathbf{E}$, i.e. $\langle \mathbf{G}|\mathbf{E} \rangle = f(\mathbf{E})$. Stochastic estimation aims to model this function, which can then be used to estimate the conditional average for every possible choice of the event $\mathbf{e}$, without the need for repeat experiments. Furthermore, stochastic estimation does not necessitate the experimental expense of traditional conditional average measurements. To show this, the derivation of the stochastic estimation procedure will be presented. In the following, it is assumed that the conditional variable $\mathbf{G}$ contains $\tilde{n}$ elements, and for simplicity, $\mathbf{G}$ is univariate (which is herein notated without boldface, i.e. G). For multivariate $\mathbf{G}$, a separate stochastic estimation model must be built for each element. It is also assumed that both $\mathbf{E}$ and G have a mean of zero. The

approach here is based on the descriptions given in Adrian (1994) and Tropea *et al.* (2007).

Stochastic estimation approximates the conditional average function using a power series expansion, centred around $\mathbf{E} = 0$ and truncated at some order:

$$\langle G|\mathbf{E}\rangle = f(\mathbf{E}) = \sum_{i=1}^{\tilde{n}} A_i E_i + \sum_{i=1}^{\tilde{n}} \sum_{j=1}^{\tilde{n}} B_{ij} E_i E_j + \sum_{i=1}^{\tilde{n}} \sum_{j=1}^{\tilde{n}} \sum_{k=1}^{\tilde{n}} C_{ijk} E_i E_j E_k + \ldots + \epsilon \tag{3.5}$$

where ϵ is the truncation error. For simplicity, it is helpful to group the coefficients, $A_i, B_{ij}, \ldots$ and the unconditional variable terms into two separate variables of length n:

$$\mathbf{L} = \{A_1, A_2, \ldots, A_{\tilde{n}}, B_{11}, B_{12}, \ldots, B_{\tilde{n}\tilde{n}}, C_{111}, \ldots\} \tag{3.6}$$

and:

$$\mathbf{F} = \{E_1, E_2, \ldots, E_{\tilde{n}}, E_1 E_1, E_1 E_2, \ldots, E_{\tilde{n}} E_{\tilde{n}}, E_1 E_1 E_1, \ldots\} \tag{3.7}$$

The truncated series expansion of the conditional average is then simply:

$$\langle G|\mathbf{E}\rangle = \sum_{i=1}^{n} L_i F_i + \epsilon \tag{3.8}$$

Where L_i and F_i are the ith elements of $\mathbf{L}$ and $\mathbf{F}$, respectively. Stochastic estimation estimates the coefficients $\mathbf{L}$ using the criteria of minimising the mean squared error between the true and estimated conditional average:

$$\operatorname*{argmin}_{L} \left\langle \left(\langle G|\mathbf{E}\rangle - \sum_{i=1}^{n} L_i F_i \right)^2 \right\rangle \tag{3.9}$$

The necessary condition for this minimisation is the *orthogonality principle*, which states that the error, $\langle G|\mathbf{E}\rangle - \sum_{i=1}^{n} L_i F_i$, must be uncorrelated with the variable $\mathbf{F}$:

$$\left\langle \left(\langle G|\mathbf{E}\rangle - \sum_{i=1}^{n} L_i F_i \right) F_j \right\rangle = 0 \tag{3.10}$$

for $j = 1, \ldots, n$. This can be rearranged to yield a system of n simultaneous equations:

$$\sum_{i=1}^{n} \langle F_i F_1 \rangle L_i = \langle F_1 G \rangle$$

$$\sum_{i=1}^{n} \langle F_i F_2 \rangle L_i = \langle F_2 G \rangle \tag{3.11}$$

$$\ldots$$

$$\sum_{i=1}^{n} \langle F_i F_n \rangle L_i = \langle F_n G \rangle$$

or simply:

$$\sum_{i=1}^{n} \langle F_i F_j \rangle L_i = \langle F_j G \rangle \tag{3.12}$$

for $j = 1, 2, \ldots n$. The solution to this system of equations requires knowledge of the average quantities, $\langle F_i F_j \rangle$ and $\langle F_j G \rangle$. Note that both these averages are unconditional; as such, they can be acquired experimentally without the need for conditional sampling. Furthermore, all the required quantities can be measured independently, which means that the stochastic estimation model can be constructed with a small number of measurement probes[1]. Contrast this with approach of directly measuring conditional averages, which requires each element of the conditional variable to be measured in conjunction with the *full* unconditional variable $\mathbf{E}$.

---

[1]technically, the number of probes required is determined by the order at which the power series is truncated; two probes are required for a linear model, three for a second order, etc

Several comparisons between stochastic estimation and directly measured conditional averages have been made, and stochastic estimation has been shown to provide an excellent approximation in a wide number of flows (Adrian *et al.*, 1989; Naguib *et al.*, 2001; Guezennec, 1989). Commonly, the expansion of $\langle G|\mathbf{E}\rangle$ is truncated at the first or second order, leading to *linear* stochastic estimation (LSE), and *quadratic* stochastic estimation (QSE). The use of higher orders has been investigated, but only a negligible improvement resulted from their use in the situations under investigation (Tung and Adrian, 1980).

An important new application for stochastic estimation was introduced by Cole *et al.* (1991), who proposed that, rather than using a user-specified event for the conditional average, the instantaneous measurements of $\mathbf{E}$ could be used instead. The strength of this approach arises from the fact that the best estimate (in terms of minimum squared error) of the instantaneous value of a random variable, y, given knowledge of the variable x, is the conditional average $\langle y|x\rangle$. Hence, a prediction of the instantaneous value of $\mathbf{G}$ can be obtained whenever a measurement of the event $\mathbf{E}$ is made.

This approach has been exploited to overcome several limitations of conventional measurement techniques. For example, a small number of probes can be used to build stochastic estimation models using the piece-wise approach described above, which can then be used to predict a high resolution velocity field using only a limited number of point measurements (Cole *et al.*, 1991; Geiseke and Guezennec, 1994; Cole and Glauser, 1998). In recent years, stochastic estimation has combined the benefits of planar and point measurements, to provide flow predictions with both high spatial and temporal resolution. By measuring PIV frames in conjunction with high frequency point measurements (commonly pressure probes), a stochastic estimation model can

be used to predict the PIV velocity field at the same frequency as the point measurements. This approach has a further benefit in that the disruption caused by the presence of physical probes can be avoided by predicting the flow field using probes outside the region of interest. This has included the investigation of boundary layer separation/reattachment using wall-mounted pressure probes to predict the velocity field (Taylor and Glauser, 2002; Ukeiley and Murray, 2005; Hudy *et al.*, 2007), and the use of near field pressure measurements to predict the flow in jets (Picard and Delville, 2000; Tinney *et al.*, 2008).

Further developments of the stochastic estimation procedure have seen the technique combined with other mathematical techniques. It is now common to carry out the stochastic estimation in conjunction with *proper orthogonal decomposition* (POD), which has enjoyed great success as a tool to investigate the existence and dynamics of coherent structures in turbulent flows. A discussion of POD, and its use in stochastic estimation, is provided in the following section.

Another extension of the technique, known as *spectral* stochastic estimation, carries out the estimation in the frequency domain by using the Fourier coefficients of the unconditional vector to predict those of the conditional vector (Ewing and Citriniti, 1999; Tinney *et al.*, 2006, 2008). Whereas conventional stochastic estimation uses a measurement of the unconditional vector at a single point in time to make the prediction, spectral stochastic estimation can improve the prediction accuracy by incorporating unconditional data measured over a range of times (Ewing and Citriniti, 1999). A further benefit is that the technique automatically accounts for the time delays between the measurements when the conditional and unconditional locations are separated by a large distance.

Although stochastic estimation provides the means of predicting an unlimited number of locations from a finite set of unconditional measurements, it is rarely used for the prediction of volumetric regions. This is likely a result of the practicalities of undertaking such an endeavour; before any prediction can be made, each conditional variable must first be measured in unison with each of the unconditional variables, in order to establish the stochastic estimation model at that point. Using standard point, line or planar measurements to build the stochastic estimation model, the task of producing a volumetric prediction requires the measurement position to be traversed through the volume, to account for every point of interest. As such, attempting to predict a volume of a reasonable size or resolution quickly becomes prohibitively complex and time consuming. Nonetheless, there have been a few notable instances of volumetric predictions. For example, Geiseke and Guezennec (1994) and Druault, Delville and Bonnet (2005a) have both used stochastic estimation to predict a time-resolved velocity field in a plane normal to the mean flow direction. By applying Taylor's hypothesis, the third spatial dimension was then inferred from the time history of the predicted plane. To the author's knowledge, the only volumetric predictions to be obtained without resorting to Taylor's hypothesis were carried out by Tinney *et al.* (2008). In this work, the velocity field of an axisymmetric jet was predicted using near-field pressure measurements. A volumetric prediction was achieved by the piece-wise mapping of the velocity field with a stereo PIV system, which was traversed to 21 separate positions along the streamwise axis of the jet. To improve the grid resolution in the streamwise axis, a sub-grid interpolation scheme was subsequently employed.

3.4 Principal Component Analysis

This section introduces the technique of principal component analysis (PCA). The following treatment of PCA is similar to that found in most textbooks on the subject e.g. (Kirby, 2001; Joilliffe, 2004; Jackson, 1991). However, in preparation for the work in following chapters, PCA is described primarily in terms of vector and matrix algebra, and particular emphasis is placed on the geometric properties of the technique. Also, throughout the discussion, the opportunity is taken to introduce several other important concepts that will be referred to throughout the rest of the thesis.

As a necessary prelude to the discussion of PCA, the concept of *vector spaces* will be introduced. Consider the multivariate random variable $\mathbf{x}$, which consists of a set of n real variables, i.e. $\mathbf{x} = [x_1 \ x_2 \ \ldots \ x_n]$. Geometrically, an observation of $\mathbf{x}$ can be considered as a single point in n-dimensional vector space, $\mathbb{R}^n$. The coordinate system within the vector space is determined by the *basis*. A basis, $\mathcal{B}$, for $\mathbb{R}^n$ consists of a set of n orthogonal vectors, i.e.:

$$
\begin{aligned}
\mathbf{v}_1 &= \begin{bmatrix} v_{11} \ v_{12} \ \ldots \ v_{1n} \end{bmatrix} \\
\mathbf{v}_2 &= \begin{bmatrix} v_{21} \ v_{22} \ \ldots \ v_{2n} \end{bmatrix} \\
&\quad\quad\vdots \\
\mathbf{v}_n &= \begin{bmatrix} v_{n1} \ v_{n2} \ \ldots \ v_{nn} \end{bmatrix}
\end{aligned}
\tag{3.13}
$$

which allow a point $\mathbf{x}$ to be uniquely expressed as a linear combination of the vectors:

$$
\mathbf{x} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \ldots + a_n\mathbf{v}_n
\tag{3.14}
$$

where $a_1, a_2, \ldots, a_n$ are the coordinates of $\mathbf{x}$ in terms of the basis, $\mathcal{B}$. To illustrate this concept, the simple example in figure 3.1 shows how the coordinates
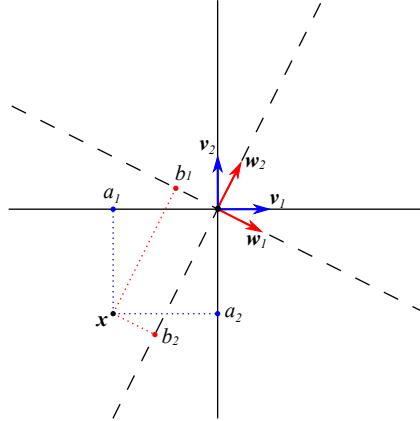
Figure 3.1: The coordinates of $\mathbf{x}$ relative to the two different bases. The point $\mathbf{x}$ has the coordinates (a_1, a_2) in terms of the basis defined by the vectors $\mathbf{v}_1$, $\mathbf{v}_2$, and the coordinates (b_1, b_2) in terms of the basis defined by the vectors $\mathbf{w}_1$, $\mathbf{w}_2$

of an observation of a 2-dimensional variable $\mathbf{x}$ differ depending on the basis. In each case, the coordinates arise from the orthogonal projection of the point $\mathbf{x}$ onto the basis vectors.

For practical purposes, it is helpful to express the vector space operations in matrix form. Here, the basis vectors form the column vectors of the matrix $\mathbf{V} = [\mathbf{v}_1^T \ \mathbf{v}_2^T \ \ldots \ \mathbf{v}_n^T]$ and the coordinates are represented as $\mathbf{a} = [a_1 \ a_2 \ \ldots \ a_n]$. The point $\mathbf{x}$ is then given by:

$$\mathbf{x} = \mathbf{a}\mathbf{V}^T \tag{3.15}$$

An important basis of the vector space $\mathbb{R}^n$ is the *standard* basis, $\mathcal{B}_{std}$, which consists of a set of unit vectors, where the ith element of the ith vector is one,

and all other elements are zero:

$$\mathbf{e}_1 = [1 \ 0 \ \dots \ 0]$$
$$\mathbf{e}_2 = [0 \ 1 \ \dots \ 0]$$
$$\vdots$$
$$\mathbf{e}_n = [0 \ 0 \ \dots \ 1]$$

(3.16)

In matrix form, the standard basis is the identity matrix $\mathbf{I}$, and the coordinate vector $\mathbf{a}$ is simply $\mathbf{x}$, i.e. $\mathbf{x} = \mathbf{xI}$. Only orthonormal bases will be considered in the following work, which are defined by vectors that are both orthogonal, and of unit length. Geometrically, any orthonormal basis of $\mathbb{R}^n$ can be considered a rotation and/or reflection of the standard basis about the origin.

It is often necessary to change from one basis to another. Given two bases $\mathcal{B}_1$ and $\mathcal{B}_2$, defined by the vectors $\mathbf{V}$ and $\mathbf{W}$, the point $\mathbf{x}$ is given by:

$$\mathbf{x} = \mathbf{a}\mathbf{V}^T$$

(3.17)

and:

$$\mathbf{x} = \mathbf{b}\mathbf{W}^T$$

(3.18)

Hence, with knowledge of the coordinates $\mathbf{a}$, and the basis vectors $\mathbf{V}$ and $\mathbf{W}$ the coordinates $\mathbf{b}$ can be found by:

$$\mathbf{b} = \mathbf{a}\mathbf{V}^T(\mathbf{W}^T)^{-1}$$
$$= \mathbf{a}\mathbf{V}^T\mathbf{W}$$

(3.19)

which arises from the fact that the inverse of an orthonormal matrix is identical

to its transpose. Similarly:

$$\mathbf{a} = \mathbf{b} \mathbf{W}^T \mathbf{V} \tag{3.20}$$

In this context, PCA can be viewed as a means of defining an "optimal" basis for a set of m independent observations of the n-dimensional variable $\mathbf{x}$, i.e. $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots, \mathbf{x}^{(m)}$. These observations are represented by the $m \times n$ matrix, $\mathbf{X}$:

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}^{(1)} \\ \mathbf{x}^{(2)} \\ \vdots \\ \mathbf{x}^{(m)} \end{bmatrix} \tag{3.21}$$

In the following section, it is assumed that the data is mean-centred, such that $\frac{1}{m} \sum_{j=1}^{m} \mathbf{x}^{(j)} = [0 \; 0 \; \dots \; 0]$. As such, the set of m observations of x forms a "cloud" of points in $\mathbb{R}^n$, centred at the origin. The aim of PCA is to choose a basis for $\mathbb{R}^n$, so that each consecutive basis vector passes through the cloud of points in the direction of highest variance, subject to the constraint that the vector must be orthogonal to all preceding vectors. These vectors, denoted, $\mathbf{U} = [\mathbf{u}_1^T \; \mathbf{u}_2^T \; \dots \; \mathbf{u}_n^T]$, are known as the principal components of $\mathbf{X}$. From equation (3.15), a point $\mathbf{x}$ can be represented as:

$$\mathbf{x} = \mathbf{z} \mathbf{U}^T \tag{3.22}$$

where $\mathbf{z}$ are the coordinates, or *scores*, of $\mathbf{x}$ in terms of the basis defined by the principal components $\mathbf{U}$. As $\mathbf{U}$ is orthonormal, the scores can be calculated by:

$$\mathbf{z} = \mathbf{x} \mathbf{U} \tag{3.23}$$

which is commonly described as the *projection* of $\mathbf{x}$ onto the principal compo-

(a) Scatter plot of the original data set $\mathbf{X}$. The red arrows represent the principal components $\mathbf{u}_1$ and $\mathbf{u}_2$.

(b) Scatter plot of the resulting scores $\mathbf{Z} = \mathbf{X}\mathbf{U}^T$
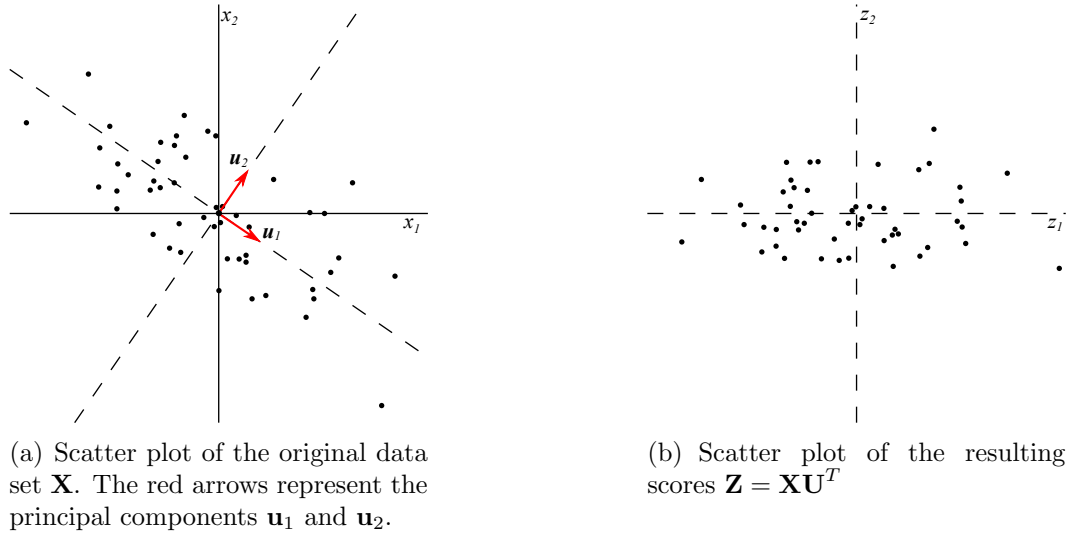
Figure 3.2: The application of PCA to a 2-dimensional data set

nents. Similarly, the ith score is given by:

$$z_i = \mathbf{x}\mathbf{u}_i^T \tag{3.24}$$

Also, the full set of n scores for all m observations can be calculated from $\mathbf{X}$ as:

$$\mathbf{Z} = \mathbf{X}\mathbf{U} \tag{3.25}$$

where:

$$\mathbf{Z} = \begin{bmatrix} \mathbf{z}^{(1)} \\ \mathbf{z}^{(2)} \\ \vdots \\ \mathbf{z}_{(m)} \end{bmatrix} \tag{3.26}$$

Figure 3.2 shows the application of the PCA procedure to a 2-dimensional data set.

From the above description, it is evident that the calculation of the

principal components is a maximisation problem, which must be carried out subject to certain constraints. The quantity being maximised is the variance of $\mathbf{x}$ along the direction of the principal component, $\mathbf{u}_i$, which is defined as $\text{Var}[\mathbf{x}\mathbf{u}_i^T]$. Therefore:

$$
\begin{aligned}
\mathbf{u}_i &= \underset{\mathbf{u}_i}{\operatorname{argmax}} \ \text{Var}[\mathbf{x}\mathbf{u}_i^T] \\
&= \underset{\mathbf{u}_i}{\operatorname{argmax}} \ \frac{1}{m} \sum_{j=1}^{m} \left(\mathbf{x}^{(j)} \mathbf{u}_i^T \right)^2 \\
&= \underset{\mathbf{u}_i}{\operatorname{argmax}} \ \frac{1}{m} \left(\mathbf{X}\mathbf{u}_i^T \right)^T \left(\mathbf{X}\mathbf{u}_i^T \right) \\
&= \underset{\mathbf{u}_i}{\operatorname{argmax}} \ \mathbf{u}_i \mathbf{C} \mathbf{u}_i^T
\end{aligned}
\tag{3.27}
$$

where $\mathbf{C} = \frac{1}{m}\mathbf{X}^T\mathbf{X}$ is the covariance matrix of $\mathbf{X}$. For all principal components, the maximisation is carried out subject to the constraint that $\mathbf{u}_i$ is a unit vector, e.g. $\|\mathbf{u}_1\| = \mathbf{u}_1^T\mathbf{u}_1 = 1$. Also, for $i = 2, \dots, n$, it is required that the $\mathbf{u}_i$ is orthogonal to all preceding vectors, i.e. $\mathbf{u}_i^T\mathbf{u}_j = 0$ for all $j < i$. The solution to this problem may be found by the technique of *Lagrange multipliers*. The working is quite involved, especially for an arbitrary number of components, n, and so will not be undertaken here. The full derivation can instead be found in Giri (2004), for example. The solution is given by the eigendecomposition (EVD) of the covariance matrix $\mathbf{C} = \frac{1}{m}\mathbf{X}^T\mathbf{X}$:

$$
\mathbf{C} = \mathbf{U}\mathbf{L}\mathbf{U}^T
\tag{3.28}
$$

where the matrix of principal components $\mathbf{U} = [\mathbf{u}_1^T \ \mathbf{u}_2^T \ \dots \ \mathbf{u}_n^T]$ are the eigenvectors of $\mathbf{C}$, and $\mathbf{L}$ is a diagonal matrix consisting of the corresponding eigenvalues, $l_1, l_2, \dots, l_n$, ordered in decreasing size. The eigenvalue l_i corresponds to the variance accounted for by the ith principal component, i.e. $l_i = \text{Var}[\mathbf{x}\mathbf{u}_i^T]$.

A property of the scores $\mathbf{z}$ is that the individual components, z_i, are uncorrelated. Therefore, $\mathrm{E}[z_i z_j] = 0$ for all $i \neq j$, and $\mathrm{E}[z_i z_i] = l_i$, i.e. the variance of $\mathbf{X}$ along the direction of the ith principal component.

For the purposes of this work, the most important property of PCA concerns its optimality. This can be defined formally by considering an approximation of the variable $\mathbf{x}$, built from only the first k principal components and corresponding scores:

$$\tilde{\mathbf{x}} = z_1 \mathbf{u}_1 + z_2 \mathbf{u}_2 + \ldots + z_k \mathbf{u}_k$$
$$= \tilde{\mathbf{z}} \tilde{\mathbf{U}}^T \tag{3.29}$$

where $\tilde{\mathbf{U}} = [\mathbf{u}_1^T, \mathbf{u}_2^T, \ldots, \mathbf{u}_k^T]$ and $\tilde{\mathbf{z}} = [z_1 \; z_2 \; \ldots \; z_k]$. The mean square error (MSE) between the true and approximate values of $\mathbf{x}$ over the set of m observations is therefore:

$$\mathrm{MSE} = \frac{1}{m} \sum_{i=1}^{m} \left(\mathbf{x}^{(i)} - \tilde{\mathbf{z}}^{(i)} \tilde{\mathbf{U}}^T \right)^2 \tag{3.30}$$

The principal components are an optimal basis in the sense that for any k, there is no other choice of k orthonormal vectors that can approximate the m observations of $\mathbf{x}$ with a lower MSE.

This property can be exploited in situations where the majority of the observed n-dimensional data $\mathbf{X}$ is contained within a small number of dimensions, k (otherwise known as a k-dimensional *subspace* of $\mathbb{R}^n$). By calculating the scores $\tilde{\mathbf{Z}} = \mathbf{X}\tilde{\mathbf{U}}$ using the first k principal components, the observed data is mapped from $\mathbb{R}^n$ to $\mathbb{R}^k$ space, while still preserving the important information contained within $\mathbf{X}$.

From equation (3.29), it is possible to reconstruct the approximation

36
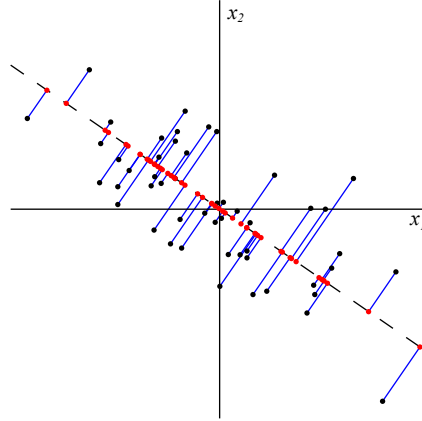
Figure 3.3: An approximation to the data set $\mathbf{X}$ using the first principal component. The red points indicate the locations of the approximated points $\tilde{\mathbf{x}}$, which corresponds to the orthogonal projection of the original points onto the first principal component. Although the approximated points still exist in two-dimensional space, they are limited to a one-dimensional subspace.

to $\mathbf{X}$ as $\tilde{\mathbf{X}} = \tilde{\mathbf{Z}}\tilde{\mathbf{U}}$ which restores the data back to $\mathbb{R}^n$ space. However, all information within the discarded dimensions is lost, which means that $\tilde{\mathbf{X}}$ is now limited to a k-dimensional subspace within $\mathbb{R}^n$, which is defined by the k principal components. In the following chapters, this subspace is referred to by the term $\tilde{V}$. The concept is illustrated in figure 3.3, which shows the result of approximating $\mathbf{X}$ with the largest principal component. It can be seen that the resulting locations of the points in $\tilde{\mathbf{X}}$ are found by the orthogonal projection of the original points onto the first principal component. Hence, in the general case $\tilde{\mathbf{X}}$ can be thought of as the projection of $\mathbf{X}$ onto the subspace defined by the vectors $\tilde{\mathbf{U}}$. The unused dimensions of $\mathbb{R}^n$ are subsequently referred to as the *null space* of $\tilde{\mathbf{X}}$, which is denoted $\tilde{V}_\perp$.

Thus far, the process of approximating $\mathbf{X}$ has been described as a two-stage process, although the final result can be achieved in a single operation.

Given that $\tilde{\mathbf{Z}} = \mathbf{X}\tilde{\mathbf{U}}$, the approximation is simply:

$$\tilde{\mathbf{X}} = \tilde{\mathbf{Z}}\tilde{\mathbf{U}}^T$$
$$= \mathbf{X}\tilde{\mathbf{U}}\tilde{\mathbf{U}}^T \tag{3.31}$$
$$= \mathbf{X}\underline{\mathbf{P}}_{\overrightarrow{\tilde{V}}}$$

The matrix $\underline{\mathbf{P}}_{\overrightarrow{\tilde{V}}}$ is known as an orthogonal projection matrix, which results when any matrix of orthonormal column vectors is multiplied by its transpose. The main property of such matrices is already apparent from the above discussion; namely, a projection matrix has the effect of projecting a vector or matrix onto a subspace, which in this case is $\tilde{V}$. A technique closely related to PCA is the singular value decomposition (SVD). The SVD can be considered to be a form of the EVD extended to rectangular matrices, such that the $m \times n$ matrix $\mathbf{X}$ can be decomposed as:

$$\mathbf{X} = \mathbf{U}_L \mathbf{S} \mathbf{V}_R^T \tag{3.32}$$

where $\mathbf{U}_L$ is an $m \times m$ matrix of orthonormal *left singular vectors* $\mathbf{u}_{L(i)}$[2] i.e. $\mathbf{U}_L = [\mathbf{u}_{L(1)}\ \mathbf{u}_{L(2)}\ \ldots \mathbf{u}_{L(m)}]$, and $\mathbf{V}_R$ is an $n \times n$ matrix of orthonormal *right singular vectors*, i.e. $\mathbf{V}_R = [\mathbf{v}_{R(1)}\ \mathbf{v}_{R(2)}\ \ldots \mathbf{v}_{R(m)}]$. $\mathbf{S}$ is an $m \times n$ rectangular

---

[2]the slight change in naming convention here is necessary to distinguish between the left singular values $\mathbf{u}_{L(i)}$ and the principal components $\mathbf{u}_i$

diagonal matrix of singular values, which for $m > n$ is:

$$\mathbf{S} = \begin{bmatrix} s_1 & 0 & \ldots & 0 \\ 0 & s_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & s_n \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & 0 \end{bmatrix} \tag{3.33}$$

and for $m < n$ is:

$$\mathbf{S} = \begin{bmatrix} s_1 & 0 & \ldots & 0 & \ldots & 0 \\ 0 & s_2 & \ldots & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ 0 & 0 & \ldots & s_m & \ldots & 0 \end{bmatrix} \tag{3.34}$$

Hence, there are a total of $r = \min(n, m)$ singular values.

The SVD is closely related to PCA, which, as shown in equation (3.28), is performed by computing the EVD of the covariance matrix $\mathbf{C} = \frac{1}{m}\mathbf{X}^T\mathbf{X}$, i.e.:

$$\frac{1}{m}\mathbf{X}^T\mathbf{X} = \mathbf{U}\mathbf{L}\mathbf{U}^T \tag{3.35}$$

Consider the SVD of the matrix $\frac{1}{\sqrt{m}}\mathbf{X}$:

$$\frac{1}{\sqrt{m}}\mathbf{X} = \mathbf{U}_L\mathbf{S}\mathbf{V}_R^T \tag{3.36}$$

The matrix $\mathbf{C}$ can therefore be expressed as:

$$\begin{aligned}
\mathbf{C} &= \frac{1}{m}\mathbf{X}^T\mathbf{X} \\
&= \left(\mathbf{U}_L\mathbf{S}\mathbf{V}_R^T\right)^T\mathbf{U}_L\mathbf{S}\mathbf{V}_R^T \\
&= \mathbf{V}_R\mathbf{S}\mathbf{U}_L^T\mathbf{U}_L\mathbf{S}\mathbf{V}_R^T
\end{aligned} \tag{3.37}$$

Multiplying an orthonormal matrix by its transpose produces the identity matrix, $\mathbf{I}$, hence:

$$\begin{aligned}
\frac{1}{m}\mathbf{X}^T\mathbf{X} &= \mathbf{V}_R\mathbf{S}^T\mathbf{I}\mathbf{S}\mathbf{V}_R^T \\
&= \mathbf{V}_R\mathbf{S}^2\mathbf{V}_R^T
\end{aligned} \tag{3.38}$$

where $\mathbf{S}^2$ is a diagonal matrix of the squared singular values, $s_1^2, s_2^2, \ldots, s_n^2$. This expression is identical in form to the EVD of $\frac{1}{m}\mathbf{X}^T\mathbf{X}$, which shows that the principal components $\mathbf{u}$, and eigenvalues l, of $\mathbf{X}$ are the right singular vectors $\mathbf{v}_R$, and squared singular values s^2, of $\frac{1}{\sqrt{m}}\mathbf{X}$.

Hence, the SVD provides an alternative approach to performing PCA. Also, the SVD can be used to establish a third approach, known as the method of *snapshots* (Sirovich, 1987). This time, the EVD of the matrix $\frac{1}{m}\mathbf{X}\mathbf{X}^T$ is found. Using the same approach as before, the EVD of this can be expressed in terms of the SVD of $\frac{1}{\sqrt{m}}\mathbf{X}$:

$$\frac{1}{m}\mathbf{X}\mathbf{X}^T = \mathbf{U}_L\mathbf{S}^2\mathbf{U}_L^T \tag{3.39}$$

Therefore, the EVD of this matrix can be used to find the left singular vectors $\mathbf{U}_L$ and the singular values $\mathbf{S}$. The corresponding right singular vectors can

then be found by rearranging (3.32):

$$\mathbf{V}_R^T = \mathbf{U}_L^T \mathbf{S}^{-1} \mathbf{X} \qquad (3.40)$$

From, (3.38), $\mathbf{V}_R$ corresponds to $\mathbf{U}$, the principal components of $\mathbf{X}$. Notice that the snapshot method is performed on the $m \times m$ matrix $\frac{1}{m}\mathbf{X}\mathbf{X}^T$, whereas the standard approach is performed on the $n \times n$ matrix $\frac{1}{m}\mathbf{X}^T\mathbf{X}$. Due to the computational cost of computing the EVD, the snapshot method is preferable when $m < n$. A full discussion of the relative merits of these three approaches is saved for section 5.3.

So far, PCA has been described as an operation on finite sets of data. However, for many applications, PCA is used to make inferences about the underlying population from which the data was sampled. Assuming that the observed variables $\mathbf{x}$ are drawn from the same statistical population, the *sample* principal components, $\mathbf{u}_i$ and eigenvalues, l_i can be used as estimates of the *population* principal components $\boldsymbol{\gamma}_i$ and eigenvalues λ_i. These population quantities can be can be viewed as those which would arise from the PCA procedure as the number of samples $m \rightarrow \infty$.

3.5 Proper Orthogonal Decomposition

PCA, under its many guises, is ubiquitous throughout many fields, although its treatment is expressed using a number of different terminologies, conventions and derivations. In fluid dynamics, it has the title of *Proper Orthogonal Decomposition* (POD). The final part of this section presents a brief summary of the use of POD in fluids applications, and its relationship with stochastic estimation. Before continuing, it is first necessary to address some pertinent

differences between POD and PCA. A superficial difference relates to terminology. For example, the principal components and scores are referred to as POD *modes* and *expansion coefficients*. For consistency with POD literature, this convention will be adopted for the duration of this section. A more fundamental deviation between the techniques arises from the interpretation of the calculated quantities. It was mentioned at the end of the previous section that PCA is often used to infer properties of the underlying population. In POD applications, this concept is taken a stage further, Here, the data $\mathbf{x}$ is taken to be the state of the flow within a domain in space (and possibly time). The resulting sample principal components are not only used as an estimate of the underlying population variables, but, as will be shown, it is also common to attribute a degree of physical importance to the principal components.

Furthermore, POD is viewed, and subsequently derived, as an operation on continuous functions, rather than on finite observations of discrete data as is the case with PCA. In practice, however, the implementation of the two techniques is identical save for a few notable exceptions. For example, in flows with homogeneous or periodic spatial dimensions (or stationary temporal dimension), the application of POD can be shown to be equivalent to the Fourier decomposition along these directions. In such cases, it is standard procedure to first perform a Fourier transform of the data along the homogeneous directions, and then for each frequency, perform POD on the complex Fourier coefficients along the remaining inhomogeneous dimensions. A final difference between the two approaches is that POD permits the use of different inner products to define the optimality of the decomposition. For PCA, the principal components are chosen to maximise $\mathrm{Var}[\mathbf{x}\mathbf{u}_i^T]$, where $\mathbf{x}\mathbf{u}_i^T$ is the dot product of $\mathbf{x}$ and $\mathbf{u}_i$. Generally, the dot product is also used in POD applications, although in some instances other inner products may be chosen

(usually in production of low order dynamical systems using the Galerkin projection, which will not be considered here). The reader is referred to Holmes *et al.* (1996) for a full mathematical description of POD. Chatterdee (2000) and Tropea *et al.* (2007) are also recommended for their comparison between POD and the SVD.

POD was originally introduced by Lumley (1967) as an objective approach to identifying coherent structures in turbulence. Today, the majority of its uses still fall into this broad category, although it is applied in a wide variety of ways.

In many cases, coherent structure can be inferred directly from the modes themselves. When a particular mode accounts for a sufficiently large amount of energy (i.e. variance) of the flow under investigation, the shape of the mode is often assumed to correspond to a physically occurring coherent structure in the flow (Lumley, 1981)[3]. If the sole interest is the POD modes, then only knowledge of the covariance matrix $\mathbf{C}$ is required. The elements of $\mathbf{C}$ can be constructed in a piecewise manner, using two-point correlations acquired with a minimum of two probes, and as with stochastic estimation, this allows the construction of high resolution reconstructions with minimal experimental effort.

This approach constituted some of the earliest uses of POD, and has been used to extract the prominent coherent structure in turbulent pipe flow (Bakewell and Lumley, 1967) the turbulent wake behind a cylinder (Payne and Lumley, 1967), mixing layers (Delville *et al.*, 1999), and rectangular jets (Sakai *et al.*, 2006). Even where more sophisticated measurement techniques

---

[3]In cases where POD has been applied in the Fourier domain (due to the presence of homogeneous/periodic/stationary dimensions), the modes are also in the Fourier domain, and so have no direct physical meaning. In order to produce a spatial representation of the modes, the *shot noise decomposition* is usually performed (Holmes *et al.*, 1996)

enable further uses of POD, the presentation and analysis of the largest modes is standard.

However, if the direct and full measurement of the variable x is possible, then the observations can be projected onto a reduced set of modes to produce a partial reconstruction $\tilde{\mathbf{x}}$. Generally, the partial reconstruction is not simply considered to be an approximation to the true state of the flow, but rather a representation of the large scale, coherent structure. Similarly, projection onto the lower modes can be thought of as capturing the small scale, incoherent structure in the flow. This approach has been used extensively in the study of a wide variety of flows (Citriniti and George, 2000; Druault, Delville and Bonnet, 2005b; Graftieaux *et al.*, 2001).

Given the availability of time-resolved measurements of $\mathbf{x}$, POD can be applied in two ways; the simplest approach is to perform POD solely on the spatial dimensions. Then, the spatial modes can be projected onto each instantaneous measurement. The time variation of the resulting temporal expansion coefficients can be used to investigate the temporal dynamics of the flow structure (Pederson and Meyer, 2002; Pastur *et al.*, 2005). The second option is to include both spatial and temporal information in the decomposition. However, when the flow is stationary in time it is necessary to perform the Fourier transform along the time dimension, yielding a set of Fourier modes. Such modes cannot be directly interpreted as physical structures, but normally have the benefit of providing a more efficient decomposition of the flow in terms of the energy captured (Bonnet *et al.*, 1998; Picard and Delville, 2000).

From this brief survey, it is evident that certain applications of POD require more sophisticated experimental setups than others. However, there are several ways of overcoming experimental limitations, which allow the application of POD in ways that would not otherwise be possible. For example,

44

Druault, Guibert and Alizon (2005) calculated the spatial POD modes from time-resolved, yet relatively low-speed, PIV measurements of a flow in an engine cylinder, and then interpolated the time-varying expansion coefficients to produce a low-dimensional description of the flow that was effectively continuous in time. In an investigation into the periodic vortex shedding of a flow over a square-section cylinder, van Oudheusden *et al.* (2005) showed that the phase of non-time-resolved PIV measurements could be identified from the relationship between the first two temporal expansion coefficients. This consequently enabled the construction of a low-order, phase resolved reconstruction of the flow.

The most popular approach of extending the capabilities of POD is through the use of stochastic estimation. As described in the previous section, stochastic estimation can be used to extend the capabilities of measurement techniques to provide high resolution, time-resolved flow predictions. This makes it ideally suited for use with POD, and has the further benefit that both stochastic estimation and POD are derived from the same measured quantities, namely the covariances between the point locations in the flow, also referred to as two-point correlations. The combination of the two techniques was first proposed by Bonnet *et al.* (1994), under the name of the *complimentary technique*. Two-point correlations are used to obtain both the principal components and the stochastic estimation model. Stochastic estimation is then used to predict the behaviour of the flow over a large number of locations, using only a small number of point measurements. The predicted field is then projected onto the POD modes to produce a low dimensional reconstruction. Employing a similar approach, Taylor and Glauser (2002) showed that if instantaneous measurements of $\mathbf{x}$ are available, stochastic estimation can be used to predict the principal components directly, rather than predicting the full instantaneous

velocity field. This involves finding the spatial principal components and calculating the most important scores for each measurement, which are then used as the conditional variables in the stochastic estimation process. The benefit of this approach is to vastly reduce the amount of data involved in the predictions, which reduces the computational complexity required. As such, the technique is particularly suited for applications of real-time flow sensing and control (Taylor and Glauser, 2002; Ausseur *et al.*, 2006). The combined use of stochastic estimation and POD is often applied for purposes other than simply overcoming experimental limitations for conventional POD analysis. Rather than restricting the investigation to the behaviour of the predicted flow only, it is increasingly used to analyse the *relationship* between the predicted quantity and the quantity used to predict it. The approach introduced by Picard and Delville (2000) is commonly used for this purpose, which involves performing POD on the unconditional data instead, and using a reduced set of resulting scores to predict the flow field. This was used to study the relationship between the velocity within the shear layer of a subsonic jet and the near-field pressure signature. The scores from the largest pressure principal components were used to predict the velocity field in turn, which revealed the presence of vortical structure in the velocity field that directly related to the dominant pressure principal components. Taking the opposite approach, Ukeiley and Murray (2005) performed POD on the velocity field within a mixing layer and partitioned the instantaneous measurements into large scale, small scale and quasi-Gaussian fluctuations, which were used to predict the far field sound pressure. This approach revealed that the small scale coherent structure was responsible for the majority of the noise emission.

Chapter 4

Stochastic Estimation and

Linear Regression

The concept of exploiting the statistical relationship between variables for the purposes of prediction and analysis is common across the entire spectrum of scientific inquiry. Arguably the most familiar and well-understood approach is that of linear regression, which encompasses a broad range of techniques with the shared goal of fitting a linear model to a set of dependent and independent variables. In the following work, several linear regression concepts are introduced to the field of stochastic estimation, which relate to the performance that can be expected in practical situations, when the stochastic estimation model is built from a finite number of observations. As described in section 3.3, the aim of stochastic estimation is to find a set of model coefficients that are optimal, in the sense that they minimise the mean square error between the resulting model and the true conditional average. However, while the resulting model will always be optimal for the sample set it is built from, this is generally not the case for the population at large. Unfortunately, it is invariably optimality in terms of the whole population which is sought. For example, when used for the instantaneous prediction of either an observable property of the flow, or a quantity derived from such, this can be thought of as the prediction of the conditional variable G, using observations of $\mathbf{E}$ separate from the data set used to build the model. In such cases it is required that the performance of the stochastic estimation model extends beyond the original data, and is capable of yielding accurate predictions for any future observations. Even when stochastic estimation is used in its traditional role of estimating conditional averages with a user defined event, is still important to obtain a conditional average that is representative of the true conditional average, rather than one which is specific only to the data it is built from.

The fact that the stochastic estimation model does not yield an optimum model for the whole population may come as no surprise. What is

more interesting is the fact that, given a finite set of observations, stochastic estimation may not even yield the best *estimate* of the optimum population model. In other words, for a given set of observations, there are techniques that may be able to produce a model that has a smaller MSE of prediction for future observations, even though the model does not minimise the MSE for the sample set it was built from, as is the case with stochastic estimation. This can lead to situations where the benefits of using an alternative approach to stochastic estimation are considerable. The questions of why and when these situations occur, and which techniques may be more appropriate when they do, form the motivation of the work in this chapter.

The chapter itself is split into two main sections. The first part is concerned with the properties of one particular method of linear regression, known as ordinary lest squares (OLS) estimation. This technique is shown to provide an identical solution to the stochastic estimation procedure in many situations. The properties of the OLS estimator are presented in a manner compatible with stochastic estimation theory, which, to the author's knowledge, have yet to be considered in fluids literature. The implication that certain regression techniques are theoretically capable of outperforming OLS is considered, and the OLS properties are also shown to have a number of uses when determining the choice of terms to include in a stochastic estimation model for a given application. In the second part of this chapter, the concept of biased regression is introduced. As well as offering the possibility of improved performance over the traditional form of stochastic estimation, such techniques are shown to be closely related to some of the POD-based stochastic estimation procedures detailed in section 3.4. For the techniques under consideration, expressions are derived to show how the introduction of bias into the model is able reduce prediction error. These expressions are used to show how, and in which situa-

tions, the techniques should able to outperform stochastic estimation. These conclusions are subsequently tested quantitatively in chapter 5.

The chapter begins by introducing the *linear regression model*, which will be used extensively throughout this work.

4.0.1 The Linear Regression Model

The aim of linear regression is to model the relationship between a dependent variable, y, and a set of n independent variables, $x_1, x_2, \ldots, x_n$, using the general form: $_ii$th independent variable

$$y = b_c + b_1 x_1 + b_2 x_2 + \ldots + b_n x_n + e \tag{4.1}$$

where the model coefficients, $b_c, b_1 \ldots, b_n$ are fixed scalar values, and the error term e represents the component of y that is not explained by the model. Using matrix notation, the model can be expressed as:

$$y = b_c + \mathbf{x}\mathbf{b} + e \tag{4.2}$$

where $x = [x_1 \ x_2 \ \ldots \ x_n]$ and $b = [b_1 \ b_2 \ \ldots \ b_n]^T$. To provide further simplification, the first independent variable, can be set to one, $x_1 = 1;$, and the model can be written as:

$$y = b_1 x_1 + b_2 x_2 + \ldots + b_n x_n + e \tag{4.3}$$

where $b_1 = b_c$. Here, in vector notation, the model is simply:

$$y = \mathbf{x}\mathbf{b} + e \tag{4.4}$$

In contrast to Linear Stochastic Estimation, the term *linear* in the context of linear regression refers to a model which is linear in terms of the model coefficients, $\mathbf{b}$, rather than in terms of the original variables. This provides the ability to fit nonlinear functions to the data using the standard linear regression model. For example, given the independent variable $\tilde{x}$,

$$y = b_0 + b_1\tilde{x} + b_2\tilde{x}^2 + b_3\tilde{x}^3 \ldots + b_n\tilde{x}^n + e \tag{4.5}$$

is still a valid linear model of the form $y = \mathbf{x}\mathbf{b} + e$, where $\mathbf{x} = [1 \ \tilde{x} \ \tilde{x}^2 \ \tilde{x}^3 \ \ldots \ \tilde{x}^n]$.

Given a finite set of observations of $\mathbf{x}$ and y, linear regression seeks to calculate the coefficients $\mathbf{b}$ so that the resulting model is optimal in terms of some criteria, which is usually defined as a maximisation or minimisation problem. The resulting estimate of the coefficients, denoted $\hat{\mathbf{b}}$, has a wide variety of uses, but in this work the focus is on the subsequent prediction of y, given further observations of $\mathbf{x}$, which is achieved using the estimated model:

$$\hat{y} = \mathbf{x}\hat{\mathbf{b}} \tag{4.6}$$

where $\hat{y}$ is the prediction of y for the given observation $\mathbf{x}$.

4.1   Ordinary Least Squares Estimation

The standard method of estimating the coefficients $\mathbf{b}$ is the technique of *ordinary least squares* (OLS) estimation. OLS estimation uses the criterion of squared error loss to derive an estimate from a finite set of observed data. Hence, given a set of m independent observations of $(y, \mathbf{x})$, the estimated co-

efficients $\hat{\mathbf{b}}_{OLS}$ are determined by the minimisation of:

$$\hat{\mathbf{b}}_{OLS} = \underset{b}{\operatorname{argmin}} \left(\sum_{i=1}^{m} ((y^{(i)} - \mathbf{x}^{(i)}\mathbf{b})^2) \right) \tag{4.7}$$

where $\mathbf{x}^{(i)}$ and $y^{(i)}$ are the ith observations of y and $\mathbf{x}$. Representing the m observations of $\mathbf{x}$ as the $m \times n$ matrix $\mathbf{X}$:

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}^{(1)} \\ \mathbf{x}^{(2)} \\ \vdots \\ \mathbf{x}^{(m)} \end{bmatrix} \tag{4.8}$$

and the observations of y as the $m \times 1$ vector $\mathbf{y}$:

$$\mathbf{y} = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(m)} \end{bmatrix} \tag{4.9}$$

Equation (4.7) can then be written in matrix form:

$$\hat{\mathbf{b}}_{OLS} = \underset{b}{\operatorname{argmin}} \left((\mathbf{y} - \mathbf{X}\mathbf{b})^T (\mathbf{y} - \mathbf{X}\mathbf{b}) \right) \tag{4.10}$$

The minimum is found by setting the first derivative, with respect to $\mathbf{b}$, to zero:

$$\frac{\delta}{\delta \mathbf{b}} (\mathbf{y} - \mathbf{X}\mathbf{b})^T (\mathbf{y} - \mathbf{X}\mathbf{b}) = -2\mathbf{X}^T \mathbf{y} + 2\mathbf{X}^T \mathbf{X}\mathbf{b} = 0 \tag{4.11}$$

The second derivative is:

$$\frac{\delta^2}{\delta\mathbf{b}\delta\mathbf{b}^T}(\mathbf{y} - \mathbf{X}\mathbf{b})^T(\mathbf{y} - \mathbf{X}\mathbf{b}) = 2\mathbf{X}^T\mathbf{X} \qquad (4.12)$$

which is non-negative for all $\mathbf{X}$, indicating that equation (4.11) is a minimum. Rearranging (4.11) leads to:

$$\mathbf{X}^T\mathbf{X}\mathbf{b} = \mathbf{X}^T\mathbf{y} \qquad (4.13)$$

This system of equations is known as the *normal equations*, which, assuming $\mathbf{X}^T\mathbf{X}$ is invertible, is solved by:

$$\hat{\mathbf{b}}_{OLS} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \qquad (4.14)$$

which is known as the Ordinary Least Squares (OLS) estimator. Until now, the dependent variable y has been considered univariate. The OLS estimator is also applicable in cases where the dependent variable consists of p separate variables, i.e. $\mathbf{y} = [y_1 \ y_2 \ \dots \ y_p]$, leading to the linear model:

$$\mathbf{y} = \mathbf{x}\mathbf{B} + \mathbf{e} \qquad (4.15)$$

where $\mathbf{e} = [e_1 \ e_2 \ \dots \ e_p]$. Given the finite matrices of observations, now denoted $\mathbf{X}$ and $\mathbf{Y}$, the OLS estimator is:

$$\hat{\mathbf{B}}_{OLS} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y} \qquad (4.16)$$

The OLS estimator for multivariate $\mathbf{y}$ produces the same results as applying the univariate OLS estimator to each element of $\mathbf{y}$ separately. For simplicity,

only the univariate case will be considered in the rest of this work, although in practical implementations, the use of the multivariate estimator can provide a more computationally efficient approach.

4.1.1 Equivalence of Stochastic Estimation and Ordinary Least Squares Regression

Superficially, OLS regression and stochastic estimation have separate goals, in that stochastic estimation provides an estimate of a conditional average, whereas linear regression aims to model the relationship between instantaneous values of a set dependent and independent variables. However, in many cases, the difference is purely one of interpretation, as both OLS and stochastic estimation can be shown to yield identical solutions. Recall from chapter 3.3 that stochastic estimation gives an estimate of the conditional average, $\langle G|\mathbf{E}\rangle$, where G is the conditional variable and $\mathbf{E}$ is a vector of $\tilde{n}$ unconditional variables. Stochastic estimation approximates the conditional average with a truncated power series expansion:

$$\langle G|\mathbf{E}\rangle = \sum_{i=1}^{\tilde{n}} A_i E_i + \sum_{i=1}^{\tilde{n}}\sum_{j=1}^{\tilde{n}} B_{ij} E_i E_j + \sum_{i=1}^{\tilde{n}}\sum_{j=1}^{\tilde{n}}\sum_{k=1}^{\tilde{n}} C_{ijk} E_i E_j E_k + \ldots + \epsilon \quad (4.17)$$

which can be written as:

$$\langle G|\mathbf{E}\rangle = \sum_{i=1}^{n} L_i F_i + \epsilon \quad (4.18)$$

where:

$$\mathbf{L} = \{A_1, A_2, \ldots, A_{\tilde{n}}, B_{11}, B_{12}, \ldots, B_{\tilde{n}\tilde{n}}, C_{111}, \ldots\} \quad (4.19)$$

and:

$$\mathbf{F} = \{E_1, E_2, \ldots, E_{\tilde{n}}, E_1 E_1, E_1 E_2, \ldots, E_{\tilde{n}} E_{\tilde{n}}, E_1 E_1 E_1, \ldots\} \qquad (4.20)$$

where $\mathbf{L}$ and $\mathbf{F}$ are of length n.

The equivalence of stochastic estimation and OLS occurs when the stochastic estimation model is built from joint observations of G and the *full* event vector $\mathbf{E}$. This will always be the case when the model is to be used for the instantaneous prediction of G (see section 3.3), which is the focus of the work in this thesis. However, in cases where the model is built piecewise, i.e. from successive joint observations of G and subsets of $\mathbf{E}$, the equivalence does not apply, and will not be considered here. Given m observations of G and $\mathbf{E}$ (and hence $\mathbf{F}$), it is possible to build a linear regression model of the form $y = \mathbf{x}\mathbf{b} + e$, that is equivalent to the stochastic estimation model. The corresponding regression model involves the following equivalences:

$$y \equiv G \qquad (4.21)$$

and the n-element vector:

$$\mathbf{x} \equiv \mathbf{F} \qquad (4.22)$$

Although an equivalent for the $\tilde{n}$ element event vector $\mathbf{E}$ is not strictly necessary, for completeness it is defined as:

$$\tilde{\mathbf{x}} \equiv \mathbf{E} \qquad (4.23)$$

which is introduced to make explicit the difference between the directly observed variables, $\tilde{\mathbf{x}}$, and the resulting regression variables $\mathbf{x}$, which will gener-

ally consist of the elements of $\tilde{\mathbf{x}}$ and possibly functions of the elements of $\tilde{\mathbf{x}}$, such as powers and cross-terms.

As before, the m joint observations of $\mathbf{x}$ and y are contained within the $m \times n$ matrix $\mathbf{X}$ and the $m \times 1$ column vector $\mathbf{y}$. As well as viewing $\mathbf{X}$ as a collection of the m row vectors $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \ldots, \mathbf{x}^{(m)}$, it can be represented in terms of the n column vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$, i.e.

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}^{(1)} \\ \mathbf{x}^{(2)} \\ \vdots \\ \mathbf{x}^{(m)} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \ldots & \mathbf{x}_n \end{bmatrix} \tag{4.24}$$

Here, $\mathbf{x}_i$ consists of the m observations of the single variable x_i.

As described in section 3.3, the coefficients $\mathbf{L}$ of the stochastic estimation model are found as the solution of:

$$\sum_{i=1}^{n} \langle F_i F_j \rangle L_i = \langle F_j G \rangle \tag{4.25}$$

for $j = 1, 2, \ldots n$. Defining $F_i^{(k)}$ as the kth observation of the ith element of $\mathbf{F}$, i.e. F_i, and $G^{(k)}$ as ith observation of G, (4.25) can be written as:

$$\sum_{i=1}^{n} \left(\frac{1}{m} \sum_{k=1}^{m} F_i^{(k)} F_j^{(k)} \right) L_i = \left(\frac{1}{m} \sum_{k=1}^{m} F_j^{(k)} G^{(k)} \right) \tag{4.26}$$

Using the equivalent linear regression variables, this can be written as:

$$\sum_{i=1}^{n} \left(\frac{1}{m} \mathbf{x}_i^T \mathbf{x}_j \right) L_i = \left(\frac{1}{m} \mathbf{x}_i \mathbf{y} \right) \tag{4.27}$$

for $j = 1, 2, \ldots n$. This system of equations can be expressed in matrix form

as:

$$\frac{1}{m}\begin{bmatrix} \mathbf{x}_1^T\mathbf{x}_1 & \mathbf{x}_1^T\mathbf{x}_2 & \dots & \mathbf{x}_1^T\mathbf{x}_n \\ \mathbf{x}_2^T\mathbf{x}_1 & \mathbf{x}_2^T\mathbf{x}_2 & \dots & \mathbf{x}_2^T\mathbf{x}_n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_n^Tx_1 & \mathbf{x}_n^T\mathbf{x}_2 & \dots & \mathbf{x}_n^T\mathbf{x}_n \end{bmatrix}\begin{bmatrix} L_1 \\ L_2 \\ \vdots \\ L_n \end{bmatrix} = \frac{1}{m}\begin{bmatrix} \mathbf{x}_1^T\mathbf{y} \\ \mathbf{x}_2^T\mathbf{y} \\ \vdots \\ \mathbf{x}_n^T\mathbf{y} \end{bmatrix} \tag{4.28}$$

which simplifies to:

$$\frac{1}{m}\mathbf{X}^T\mathbf{X}\mathbf{L} = \frac{1}{m}\mathbf{X}^T\mathbf{y} \tag{4.29}$$

and finally:

$$\mathbf{X}^T\mathbf{X}\mathbf{L} = \mathbf{X}^T\mathbf{y} \tag{4.30}$$

These are the normal equations as introduced in equation (4.13), which yield the OLS estimator $\hat{\mathbf{b}}_{OLS}$. Hence, the OLS coefficients and the stochastic estimation coefficients, $\mathbf{L}$ are identical, under assumption that the stochastic estimation model is built from joint observations of G and the *full* event vector $\mathbf{E}$.

4.1.2 Prediction Properties of Ordinary Least Squares Regression

This section concerns the finite-sample properties of OLS regression which, as a result of the equivalence demonstrated in section 4.1.1, also apply to stochastic estimation when the model is built from a finite number of joint observations of the conditional variable G and the full unconditional vector $\mathbf{E}$. OLS regression possesses a number of appealing properties, which are valid when the underlying data meets certain assumptions. The most famous of these properties are those which arise when the so-called *Gauss-Markov assumptions* are shown to hold, although there are a number of lesser quoted

properties and corresponding assumptions that will also be presented here. In this work, the OLS properties relevant to stochastic estimation are divided into a hierarchy of four distinct tiers, with each tier placing certain requirements on the nature of the underlying data. With each increasing tier, the resulting performance of the OLS estimator is improved, but the necessary requirements become stricter.

The treatment of OLS regression and its properties is commonplace throughout several disciplines. However, the presentation provided here differs somewhat from the prevalent approaches, which is necessitated by the differences in the ways linear regression and stochastic estimation are employed.

Perhaps the biggest discrepancy between the approaches, at least when considering linear regression in its "traditional" form, is the treatment of the independent variables $\mathbf{x} = [x_1 \; x_2 \; \ldots \; x_n]$. The convention in most statistics-orientated references is to require that the independent variables are non-random; that is, that the value of the variables is directly controllable by the practitioner. Clearly, this is inappropriate in stochastic estimation, where the model is invariably built from observations of random events. As such, the *conditional* form of the OLS properties are presented here, which is the approach commonly used in the field of econometrics.

Another difference between stochastic estimation and linear regression is that the linear regression model $y = \mathbf{x}\mathbf{b} + e$ explicitly models the instantaneous behaviour of the dependent variable y, whereas stochastic estimation models the conditional average, expressed as $\mathrm{E}[y|\mathbf{x}]$ in the current notation. In order to interpret the OLS properties in terms of stochastic estimation, it is therefore necessary to produce a model that directly describes the behaviour of y, but remains compatible with the approach taken by stochastic estimation. This is made difficult by the fact that the standard text book treatment of the

58

OLS properties requires that the regression model applied to the data is of an identical form to the hypothetical "true" model.

Of course, any attempt to define the "true" stochastic estimation model in terms of the governing equations of the flow is inappropriate. Hence, it is necessary to ignore the physical mechanisms that dictate the behaviour of the data, and consider the behaviour of y simply as if it is a function simply of the conditional average, and a random error term ϵ:

$$y = \mathrm{E}[y|\tilde{\mathbf{x}}] + \epsilon \tag{4.31}$$

where ϵ represents the part of y that is unexplained by $\tilde{\mathbf{x}}$. However, even with a suitably defined "true" model, there is no guarantee that the conditional average $\mathrm{E}[y|\tilde{\mathbf{x}}]$ can be modelled exactly using a linear model. Nor indeed should there be, as this requirement is fundamentally at odds with the purpose of stochastic estimation, which is defined merely as an approximation. In situations where $\mathrm{E}[y|\tilde{\mathbf{x}}] \not\equiv \mathbf{x}\mathbf{b}$, then the conventional OLS assumptions will not hold. Nonetheless, the OLS estimator still possesses some useful properties in these cases, although they are rarely considered in standard texts. The properties of OLS regression, when used to approximate unknown functions was detailed by White (1980), whose work provides the first tier of properties presented here, and also motivates the following definition of the "optimum" regression coefficients.

All the properties described in this section relate to how closely the finite-sample OLS coefficients $\hat{\mathbf{b}}_{OLS}$ can be expected to match the optimal model coefficients, denoted $\check{\mathbf{b}}$, which can be viewed as those which minimise the squared error between y and $\hat{y} = \mathbf{x}\mathbf{b}$, over the entire population of $\mathbf{x}$ and y. From equation (4.31), y is a function of the random variables ϵ and $\mathbf{x}$, which is

in turn a function of $\tilde{\mathbf{x}}$; hence, the behaviour $\mathbf{x}$, $\tilde{\mathbf{x}}$ and y are all determined by the joint distribution of $\tilde{\mathbf{x}}$ and ϵ, i.e. $F_{\tilde{\mathbf{x}},\epsilon}(\tilde{\mathbf{x}}, \epsilon)$. Following a similar approach to that described by White (1980), the optimum coefficients can then be defined as:

$$
\begin{aligned}
\check{\mathbf{b}} = \underset{b}{\operatorname{argmin}} &\int_\epsilon \int_{\tilde{\mathbf{x}}} \left(y - \hat{y}\right)^2 F_{\tilde{\mathbf{x}},\epsilon}(\tilde{\mathbf{x}}, \epsilon)\, d\tilde{\mathbf{x}}\, d\epsilon \\
= \underset{b}{\operatorname{argmin}} &\int_\epsilon \int_{\tilde{\mathbf{x}}} \left(\mathrm{E}[y|\tilde{\mathbf{x}}] - \mathbf{x}\mathbf{b} + \epsilon\right)^2 F_{\tilde{\mathbf{x}},\epsilon}(\tilde{\mathbf{x}}, \epsilon)\, d\tilde{\mathbf{x}}\, d\epsilon \\
= \underset{b}{\operatorname{argmin}} &\int_\epsilon \int_{\tilde{\mathbf{x}}} \left((\mathrm{E}[y|\tilde{\mathbf{x}}] - \mathbf{x}\mathbf{b}) + \epsilon\right)^2 F_{\tilde{\mathbf{x}},\epsilon}(\tilde{\mathbf{x}}, \epsilon)\, d\tilde{\mathbf{x}}\, d\epsilon \\
= \underset{b}{\operatorname{argmin}} &\int_\epsilon \int_{\tilde{\mathbf{x}}} \left((\mathrm{E}[y|\tilde{\mathbf{x}}] - \mathbf{x}\mathbf{b})^2 + 2\left(\mathrm{E}[y|\tilde{\mathbf{x}}] - \mathbf{x}\mathbf{b}\right)\epsilon + \epsilon^2\right) F_{\tilde{\mathbf{x}},\epsilon}(\tilde{\mathbf{x}}, \epsilon)\, d\tilde{\mathbf{x}}\, d\epsilon \\
= \underset{b}{\operatorname{argmin}} \Bigg(&\int_{\tilde{\mathbf{x}}} (\mathrm{E}[y|\tilde{\mathbf{x}}] - \mathbf{x}\mathbf{b})^2 \, F_{\tilde{\mathbf{x}}}(\tilde{\mathbf{x}})\, d\tilde{\mathbf{x}} \\
&+ 2 \int_\epsilon \int_{\tilde{\mathbf{x}}} \left(\mathrm{E}[y|\tilde{\mathbf{x}}]\epsilon\right) F_{\tilde{\mathbf{x}},\epsilon}(\tilde{\mathbf{x}}, \epsilon)\, d\tilde{\mathbf{x}}\, d\epsilon \\
&- 2 \int_\epsilon \int_{\tilde{\mathbf{x}}} \left(\mathbf{x}\mathbf{b}\epsilon\right) F_{\tilde{\mathbf{x}},\epsilon}(\tilde{\mathbf{x}}, \epsilon)\, d\tilde{\mathbf{x}}\, d\epsilon \\
&+ \int_\epsilon \epsilon^2 F_\epsilon(\epsilon)\, d\epsilon \Bigg)
\end{aligned}
$$

(4.32)

where $F_{\tilde{\mathbf{x}}}(\tilde{\mathbf{x}})$ and $F_\epsilon(\epsilon)$ are the *marginal distributions* of $\tilde{\mathbf{x}}$ and ϵ, i.e.:

$$
F_{\tilde{\mathbf{x}}}(\tilde{\mathbf{x}}) = \int_\epsilon F_{\tilde{\mathbf{x}},\epsilon}(\tilde{\mathbf{x}}, \epsilon)\, d\epsilon
$$

(4.33)

and equivalently for $F_\epsilon(\epsilon)$. The second and third term of (4.32) correspond to the expectations $2\mathrm{E}\left[\mathrm{E}[y|\tilde{\mathbf{x}}]\epsilon\right]$ and $2\mathrm{E}[\mathbf{x}\mathbf{b}\epsilon]$ respectively. As $\mathrm{E}[\epsilon] = 0$, and ϵ is independent of $\tilde{\mathbf{x}}$, both these terms are equal to zero. The fourth term is the expectation $\mathrm{E}[\epsilon^2]$, which is the variance of the unexplained error, i.e. σ_ϵ^2.

60

Therefore, equation (4.32) becomes:

$$\check{\mathbf{b}} = \underset{b}{\operatorname{argmin}} \left(\int_{\tilde{\mathbf{x}}} (\mathrm{E}[y|\tilde{\mathbf{x}}] - \mathbf{x}\mathbf{b})^2 \, F_{\tilde{\mathbf{x}}}(\tilde{\mathbf{x}}) \, d\tilde{\mathbf{x}} + \sigma_\epsilon^2 \right) \tag{4.34}$$

It can be seen that when the linear model is of the same functional form as the conditional average, $\mathbf{x}\mathbf{b} \equiv E[y|\tilde{\mathbf{x}}]$, then the optimum coefficients are defined so that the two are numerically identical. Conversely, when $\mathbf{x}\mathbf{b} \not\equiv E[y|\tilde{\mathbf{x}}]$, the distribution of $\tilde{\mathbf{x}}$ is instrumental in the definition of $\check{\mathbf{b}}$. This is demonstrated in figure 4.1. Each plot shows the result of fitting the misspecified model $y = \tilde{x}b + e$ to the true function $y = \tilde{x}^2 + e$ (denoted by the blue line) where the error term e is normally distributed. The red lines represent the optimum misspecified model, $\tilde{x}\check{b}$ that occurs when the distribution of $\tilde{x}$ is a) negatively skewed, and b) unskewed. In the presence of misspecification, it is clear that the accuracy of fit is compromised. Also, this loss in accuracy becomes more pronounced as the likelihood of the event $\tilde{x}$ decreases. This will be a particular problem when the emphasis is on the accurate prediction of rare events, and stresses the importance of choosing an appropriate linear model in such cases. However, when the interest is in the instantaneous prediction of y, the resulting optimum model will, on average, offer the best MSE performance.

With the necessary definitions in place, it is now possible to introduce the properties of the OLS estimator.

Tier 1

For the first tier to hold, a single assumption is needed:

- Assumption 1 - The matrix of observations $\mathbf{X}$ is full rank, and of size $m \times n$, where $m \geq n$.

(a) Positively skewed (b) Unskewed
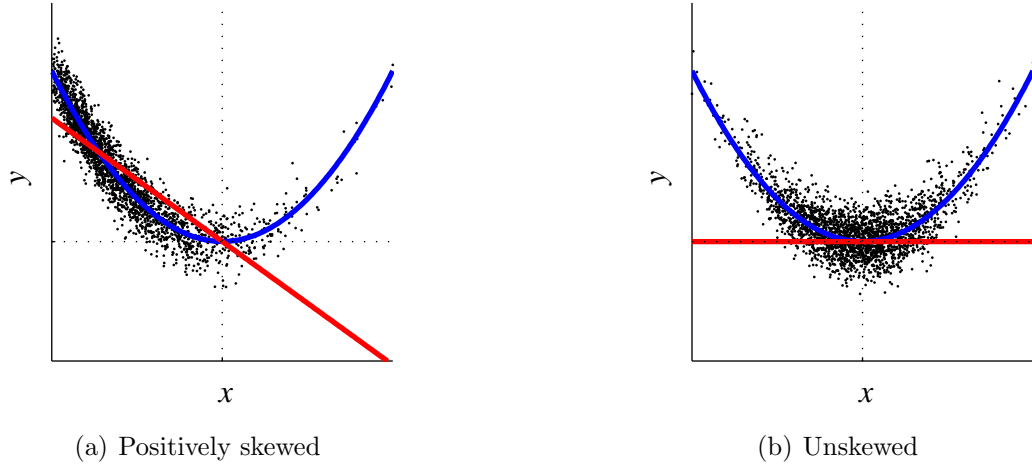
Figure 4.1: Comparison of optimum misspecified model (red) with true conditional average (blue) for different distributions of $\mathbf{x}$

The purpose of this assumption is simply to ensure that $\mathbf{X}^T\mathbf{X}$ is invertible, and so guarantee that the OLS estimator $\hat{\mathbf{b}}_{OLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ can be uniquely solved. If this is true, White (1980) demonstrated that as the number of observations n increases, the OLS estimator $\hat{\mathbf{b}}_{OLS}$ will tend to $\check{\mathbf{b}}$. This property, known as *consistency*, ensures that, asymptotically, $\hat{\mathbf{b}}$ provides the best possible prediction of y allowable by the specified model[1]. In practice, this means that as long as a sufficiently large number of observations are used, there will be good agreement between $\hat{\mathbf{b}}_{OLS}$ and $\check{\mathbf{b}}$. Of course, this property alone does not necessarily mean the resulting predictions will be "good". This requires, firstly, that $\mathrm{E}[y|\tilde{\mathbf{x}}]$ is a close approximation to y, and secondly, that the optimum linear model $\mathbf{x}\check{\mathbf{b}}$ is a close approximation to $\mathrm{E}[y|\mathbf{x}]$. However, this is still an important result, as the property holds regardless of the shape of $\mathrm{E}[y|\tilde{\mathbf{x}}]$, the type of linear model specified, and the distribution of the data involved. In other words, this is the minimum level of performance that is guaranteed

---

[1] the reader is referred to White (1980) for the full proof of this property, which is beyond the scope of this work

when using stochastic estimation.

Tier 2

It may be the case, either through virtue of the data, or through a judicious choice of linear model, that the functional form of the conditional average is equivalent to the chosen linear model, i.e. $\mathrm{E}[y|\tilde{\mathbf{x}}] \equiv \mathbf{x}\mathbf{b}$. In essence, this is the assumption that leads to the second tier of OLS properties. However, in line with the convention, it is usual to express this assumption in terms of e, which is the error term in the linear model $y = \mathbf{x}\mathbf{b} + e$. Taking the true underlying model to be $y = \mathrm{E}[y|\tilde{\mathbf{x}}] + \epsilon$, the error term is:

$$e = \mathrm{E}[y|\tilde{\mathbf{x}}] - \mathbf{x}\mathbf{b} + \epsilon \tag{4.35}$$

In the presence of a correctly specified model, however, the error e is simply ϵ, which is independent of $\mathbf{x}$. Hence the expectation of e, conditional on $\mathbf{x}$, is $\mathrm{E}[e|\mathbf{x}] = \mathrm{E}[\epsilon] = 0$. Therefore, the assumption can be expressed as:

- Assumption 2 - The expected error of the linear model, e conditional on $\mathbf{x}$, is zero, i.e. $\mathrm{E}[e|\mathbf{x}] = 0$.

The immediate implication is that, from the first assumption, the OLS estimator will not only tend to $\check{\mathbf{b}}$, but the resulting linear model $\mathbf{x}\check{\mathbf{b}}$ will be exactly the conditional average $\mathrm{E}[y|\tilde{\mathbf{x}}]$. Asymptotically, the OLS estimator will be the best possible predictor of y in terms of squared-error loss. Furthermore, it is now possible to establish the first finite-sample properties of the OLS estimator, which relate to the expected value of $\hat{\mathbf{b}}_{OLS}$, if it were to be repeatedly estimated using different observations of $\mathbf{y}$ and $\mathbf{X}$. In particular, it can be

shown (Appendix A.1) that:

$$E[\hat{\mathbf{b}}_{OLS}] = \check{\mathbf{b}} \qquad (4.36)$$

which shows that the average value of the OLS estimator is in fact the true underlying value, $\check{\mathbf{b}}$. In which case, the OLS estimator is said to be *unbiased*. Furthermore, the OLS estimator is also unbiased, conditional on $\mathbf{X}$:

$$\mathbf{E}\left[\hat{\mathbf{b}}_{OLS}|\mathbf{X}\right] = \check{\mathbf{b}} \qquad (4.37)$$

The concept of conditional unbiasedness can be interpreted by considering an infinite number of joint observations of $\mathbf{y}$ and $\mathbf{X}$; in this context, (4.37) would then describe the expected value of $\hat{\mathbf{b}}_{OLS}$, if it was repeatedly estimated from only the joint observations where $\mathbf{X}$ takes on a specific value. While unconditional unbiasedness is perhaps the more intuitive measure, unbiasedness conditional on $\mathbf{X}$ is the more desirable property. This is because unconditional unbiasedness only requires that the estimator is unbiased on average. This allows $\hat{\mathbf{b}}$ to be biased for particular values of $\mathbf{X}$, as long as the bias cancels out to zero over the entire distribution of $\mathbf{X}$. In this light, unbiasedness, conditional on $\mathbf{X}$, is the stricter property.

The unbiasedness of the OLS estimator ensures the resulting OLS predictor, $\hat{y} = \mathbf{x}\hat{\mathbf{b}}_{OLS}$, is also an unbiased predictor of y. In this case, the strictest form of unbiasedness is conditional on both $\mathbf{X}$, which is used to build the model, and $\mathbf{x}$, which is used to make the subsequent prediction. From Appendix A.2, it can be shown that:

$$E\left[y - \hat{y}|\mathbf{x}, \mathbf{X}\right] = 0 \qquad (4.38)$$
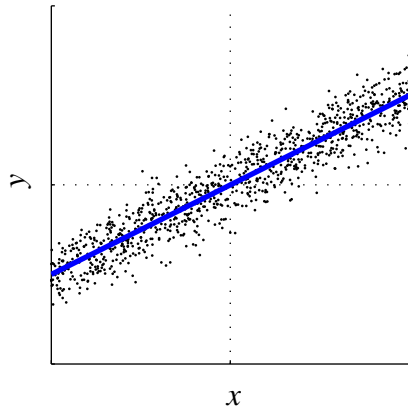
(a) Homoskedastic error, where the error variance is constant for all x.
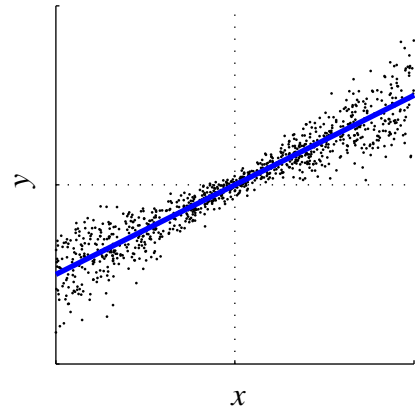
(b) Heteroskedastic error, where the error variance is roughly proportional to the magnitude of x.

Figure 4.2: Examples of homoskedastic and heteroskedastic error

Tier 3

The assumptions for the third tier again relate to the model error, e. The first assumption requires that the distribution of e is *homoskedastic*, which means that the variance of the error is constant for all $\mathbf{x}$. An example of homoskedastic error, and the converse, *heteroskedastic error*, is demonstrated in figure 4.2. The second assumption is that the error term for a given observation is uncorrelated with the error terms for all other observations.

- Assumption 3.a - The error term e is homoskedastic. i.e. $\text{var}(e|\mathbf{x}) = \sigma^2$.

- Assumption 3.b - $\text{Cov}(e_i, e_j) = 0$ for any two separate observations i and j.

The full set of assumptions 1 - 3 are known collectively as the *Gauss-Markov* assumptions. When met, the OLS estimator is said to be the most *efficient* amongst the class of linear unbiased estimators. The concept of efficiency will be introduced shortly, but first, the class of unbiased linear estimators will be

defined. An estimator $\tilde{\mathbf{b}}$ is *linear* if it can be expressed as a linear function of the matrix of observations, $\mathbf{y}$, in the form:

$$\tilde{\mathbf{b}} = \mathbf{A}\mathbf{y} \tag{4.39}$$

where the matrix $\mathbf{A}$ can depend on $\mathbf{X}$ and on non-random constants. Also, an estimator is *unbiased* if:

$$\mathrm{E}\left[\tilde{\mathbf{b}}|\mathbf{X}\right] = \check{\mathbf{b}} \tag{4.40}$$

It can be seen that the OLS estimator is both linear (where $\mathbf{A} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$) and unbiased (as established in equation (4.37)). The term *efficiency* relates to the amount of variation in an estimator over repeated estimations, with a low variance indicating a high efficiency. As demonstrated in appendix A.3, the variance of the OLS estimator can be shown to be lower than every other linear unbiased estimator, i.e:

$$\mathrm{Var}\left[\tilde{\mathbf{b}}|\mathbf{X}\right] \geq \mathrm{Var}\left[\hat{\mathbf{b}}_{OLS}|\mathbf{X}\right] \tag{4.41}$$

which will only hold with equality when $\tilde{\mathbf{b}} = \hat{\mathbf{b}}_{OLS}$. Hence, the OLS estimator is said to be the most efficient, or *best* linear unbiased estimator (BLUE). In practice, this means that for a given set of observations $(\mathbf{X}, \mathbf{y})$, the resulting OLS estimator is expected to be closer to the true value $\check{\mathbf{b}}$ than any other linear unbiased estimator. In this context, *efficiency* can be seen as the efficiency with which an estimator is able to extract information about $\check{\mathbf{b}}$ from a finite observation of data.

As expected, the optimal properties of the OLS estimator also apply to the subsequent predictions of the predictor $\hat{y} = \mathbf{x}\hat{\mathbf{b}}_{OLS}$. Defining the MSE of

66

the linear unbiased predictor $\tilde{y}$ as:

$$
\begin{aligned}
\text{MSE}\left[\tilde{y}|\mathbf{x}, \mathbf{X}\right] &= \text{E}\left[(y - \tilde{y})^2 |\mathbf{x}, \mathbf{X}\right] \\
&= \text{E}\left[\left(\mathbf{x}\check{\mathbf{b}} + e - \mathbf{x}\tilde{\mathbf{b}}\right)^2 |\mathbf{x}, \mathbf{X}\right]
\end{aligned}
\tag{4.42}
$$

it is shown in appendix A.4 that:

$$
\text{MSE}\left[\tilde{y}|\mathbf{x}, \mathbf{X}\right] \geq \text{MSE}\left[\hat{y}|\mathbf{x}, \mathbf{X}\right]
\tag{4.43}
$$

which means that the OLS predictor will perform either as well as, or better than, any other linear unbiased predictor.

This property provides obvious reassurances in situations when the Gauss-Markov assumptions are known to hold, but it also carries with it some negative connotations about the use of the OLS estimator. Firstly, it demonstrates that when the assumptions are not met, there will be a linear unbiased estimator that is more efficient than OLS, and secondly, even when the assumptions do hold, there is the possibility that there are *nonlinear* unbiased estimators that offer better performance. While both these statements are technically true, the appropriateness of using more efficient estimators in such cases is a complex issue, and is often only of theoretical, rather than practical, importance. A full discussion is saved for section 4.2.2, but for now, it will be shown that under a final assumption, the OLS estimator becomes the most efficient of all unbiased estimators.

Tier 4

- Assumption 4 - The error term, e, is normally distributed.

If this assumption holds, the OLS estimator is the *Uniformly Minimum Variance Unbiased Estimator* (UMVUE). This property is established by showing that the OLS estimator attains the *Cramér-Rao lower bound* (CRLB), which specifies the minimum possible variance attainable by any unbiased estimator.

Central to the task of establishing the CRLB is the concept of the *likelihood function*. To produce a likelihood function, it is assumed that the observed data $\mathbf{y}$ is sampled according to the probability density function (PDF) $F(\mathbf{y}|\theta)$, which is defined by the parameter vector θ. For example, consider the case where $\mathbf{y}$ is a vector of m independent observations of the normally distributed variable y. The PDF of y is defined by the mean μ and the variance σ^2, and so $\theta = [\mu \ \sigma^2]$. i.e.

$$F(y|\theta) = F(y|\mu, \sigma^2) = \left(2\pi\sigma^2\right)^{-\frac{1}{2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right) \qquad (4.44)$$

which is the well-known PDF of the normal distribution, As the observations of y are independent, the corresponding PDF of $\mathbf{y} = [y^{(1)} \ y^{(2)} \ \ldots \ y^{(m)}]^T$ can be found simply by multiplying the PDFs for each individual observation:

$$\begin{aligned} F(\mathbf{y}|\theta) =& F(y^{(1)}|\theta) \times F(y^{(2)}|\theta) \times \ldots \times F(y^{(m)}|\theta) \\ =& \left(2\pi\sigma^2\right)^{-\frac{m}{2}} \exp\left(-\frac{\sum_{i=1}^{m}\left(y^{(i)}-\mu\right)^2}{2\sigma^2}\right) \end{aligned} \qquad (4.45)$$

Once the PDF of the observed data $\mathbf{y}$ is established, the corresponding likelihood function, $\mathcal{L}(\theta|\mathbf{y})$, is simply equivalent to the PDF, i.e.:

$$\mathcal{L}(\theta|\mathbf{y}) = F(\mathbf{y}|\theta) \qquad (4.46)$$

Hence, the difference between $F(\mathbf{y}|\theta)$ and $\mathcal{L}(\theta|\mathbf{y})$ is simply one of interpretation. A PDF is considered a function of $\mathbf{y}$, which provides the probability density of $\mathbf{y}$ given a fixed set of parameters. Conversely, a likelihood function is considered a function of the parameter θ, with $\mathbf{y}$ held constant. In essence, a likelihood function provides a measure of the likelihood that the fixed observation $\mathbf{y}$ was observed from the population defined by θ. As well as its use in establishing the CRLB, the likelihood function also forms the basis of *maximum likelihood estimation* (MLE). This technique will be referred to in section 4.2.2, so the opportunity will be taken here to introduce the general concept of MLE.

The maximum likelihood estimate of θ is defined as:

$$\hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmax}} \, \mathcal{L}(\theta|\mathbf{y}) \qquad (4.47)$$

Therefore, $\hat{\theta}_{MLE}$ is the value of θ that maximises the likelihood of observing the finite sample $\mathbf{y}$. The approach to the maximisation will depend on the likelihood function in question. In some cases, an analytical solution is possible, by taking derivatives of either the likelihood function or the log likelihood function (the latter being appropriate when dealing with functions based on the exponential family of distributions), although it is often the case that an optimisation procedure must be applied.

Given the full set of assumptions 1-4, it is possible to establish a likelihood function that is relevant to the estimation of the linear regression coefficients $\mathbf{b}$, from the standard model $y = \mathbf{xb} + e$. Given that the error term is normally distributed, with variance σ^2 and a mean of zero, it follows that y, conditional on $\mathbf{x}$, is also normally distributed, this time with variance σ^2 and

a mean of $\mathbf{xb}$. From (4.44) the PDF of y conditional on $\mathbf{x}$ is simply:

$$F(y|\theta, \mathbf{x}) = \left(2\pi\sigma^2\right)^{-\frac{1}{2}} \exp\left(-\frac{(y - \mathbf{xb})^2}{2\sigma^2}\right) \tag{4.48}$$

where:

$$\theta = \begin{bmatrix} \mathbf{b} \\ \sigma^2 \end{bmatrix} \tag{4.49}$$

Similarly, for the set of m joint observations $(\mathbf{y}, \mathbf{X})$, it follows from (4.45) that the PDF of $\mathbf{y}$, conditional on $\mathbf{X}$, is:

$$
\begin{aligned}
F(\mathbf{y}|\theta, \mathbf{X}) &= \left(2\pi\sigma^2\right)^{-\frac{m}{2}} \exp\left(-\frac{\sum_{i=1}^{m}\left(y^{(i)} - \mathbf{x}^{(i)}\mathbf{b}\right)^2}{2\sigma^2}\right) \\
&= \left(2\pi\sigma^2\right)^{-\frac{m}{2}} \exp\left(-\frac{(\mathbf{y} - \mathbf{Xb})^T(\mathbf{y} - \mathbf{Xb})}{2\sigma^2}\right)
\end{aligned}
\tag{4.50}
$$

which is equivalent to the likelihood function $\mathcal{L}(\theta|\mathbf{y}, \mathbf{X})$.

As an aside, the MLE estimate $\hat{\mathbf{b}}_{MLE}$ can be found by differentiating the log-likelihood function, $\ln\mathcal{L}(\theta|\mathbf{y}, \mathbf{X})$, which is shown in appendix A.5 to produce:

$$
\begin{aligned}
\hat{\mathbf{b}}_{MLE} &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \\
&= \hat{\mathbf{b}}_{OLS}
\end{aligned}
\tag{4.51}
$$

Hence, under assumptions 1-4, the MLE and OLS estimate of $\mathbf{b}$ are identical.

The likelihood function defined by equation (4.50) can subsequently be used to obtain the CRLB for an estimate of $\mathbf{b}$. This is shown in appendix A.6 to produce:

$$\text{Var}(\tilde{\mathbf{b}}|\mathbf{X}) \geq \sigma^2\left(\mathbf{X}^T\mathbf{X}\right)^{-1} \tag{4.52}$$

70

From appendix A.3, the variance of the OLS estimator is:

$$\text{Var}(\hat{\mathbf{b}}_{OLS}|\mathbf{X}) = \sigma^2 \left(\mathbf{X}^T\mathbf{X}\right)^{-1} \tag{4.53}$$

which shows that the OLS estimator achieves the CRLB, and hence is the UMVUE.

4.2 The Properties of Ordinary Least Squares in the Context of Flow Prediction

This section considers the relevance of the OLS properties to the task of flow prediction, both in terms of their implications, and their potential uses. In the first section, the task of choosing a valid stochastic estimation model is discussed, and it will be shown how the properties of the OLS estimator can provide useful and alternative guidance to conventional practice.

In light of the finite sample efficiency properties of the OLS estimator, the validity of using the OLS estimator in cases where a more efficient estimator exists is also investigated. The theoretical properties of the OLS estimator and competing techniques will be considered alongside the practical issues of implementing them in a traditional stochastic estimation. It will be argued that the OLS estimator is often the most appropriate technique even in situations when a more efficient estimator exists.

4.2.1 Validity of the linear model

Although there are an infinite number of linear models of the form $y = \mathbf{x}\mathbf{b} + e$ that can be used describe the relationship between y and $\mathbf{x}$, the choice of

model for stochastic estimation is relatively straightforward. First of all, the model is generally limited to a truncated power series expansion (although Brereton (1992) proposed the use of a Laurent polynomial expansion, which contains both positive and negative powers), and it is widely accepted that only the first few orders of the expansion are needed to produce an accurate representation. Furthermore, stochastic estimation is considered a tool to *approximate* the conditional average, which permits the use of a simple model in situations that would otherwise require an overly complex model to capture the true conditional average exactly. Nonetheless, for the sake of accuracy, it is prudent to choose the stochastic estimation model that is best suited to the data in question.

It many cases, the correct choice of model may have already been established in past work of a similar nature. Often cited examples include the work of Tung and Adrian (1980), which demonstrated that only the linear terms are needed when both the unconditional and event data are velocities from isotropic turbulence, and Naguib *et al.* (2001), which showed the necessity of second order terms when estimating turbulent boundary layer velocities from wall pressure events. In some cases, it may be possible to infer the correct choice of model from an analysis of the underlying physical properties of the flow. For example Jordan *et al.* (2007) used linear stochastic estimation to model the relationship between the velocity field and far-field pressure of a turbulent jet, based on evidence that the sound generating mechanisms are predominantly linear. For further assurances, or in situations where the above approaches do not apply, the correct model can be inferred from the data itself. In these cases, a standard approach is to base the decision on the distribution of the data. It is often quoted that linear stochastic estimation is optimal when the distribution of the conditional and event data is jointly normal. As

such, any deviation from joint normality is often used as an indication that higher order terms are required in the model (e.g. (Brereton, 1992; Tinney *et al.*, 2006; Druault *et al.*, 2009)).

Of course, the problem of identifying an appropriate model is not unique to stochastic estimation; the detection and correction of model misspecification has received a great deal of attention in the field of regression. As a result, there are a wealth of techniques to aid this process, which can generally be divided into qualitative approaches, which involve plotting and visual inspection of some aspect of the data, and quantitative approaches, which provide an automatic approach to detecting and correcting the model fit. Of the two, visual inspection is by far the most common, partly due to its effectiveness and relative simplicity. The foundation of such approaches is commonly based on inspection of the residuals of the model, $\hat{e}$, where $\hat{e} = y - \hat{y}$. As demonstrated by assumption 2 of the OLS properties, the presence of a correctly specified model is indicated when the expectation of the residuals conditional on $\tilde{\mathbf{x}}$ is zero, i.e. $\mathrm{E}[e|\tilde{\mathbf{x}}] = 0$. Therefore, the presence of misspecification can be assessed by inspecting scatter plots of each $\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n$ versus the residuals that result from the given model. If the model is correctly specified, then the resulting distributions will be trendless and centred around $\hat{e} = 0$. Furthermore, when this is not the case, the distributions of the residuals can also provide clues as to both the amount and nature of the misspecification. For example, a symmetric trend in the residuals may indicate the need to include further even terms in the model, and asymmetric trends may indicate the need for odd terms. A problem arises when the size of $\tilde{\mathbf{x}}$ is large, which may make the inspection of each individual plot of $\tilde{x}_i$ vs. $\hat{e}$ prohibitively time consuming. In these cases, plotting the predicted values $\hat{y}$ versus the residuals can often yield the same, or at least similar, conclusions. Because $\hat{y}$ is a function of each

$x_1, x_2, \ldots, x_n$, the resulting plot will often convey the same information as the individual x_i vs. e plots. The effectiveness of this approach is compromised, however, if the model misspecification manifests itself differently amongst the $(\hat{e}, \tilde{x}_i)$ pairs, as the individual trends may no longer be obvious in the combined plot, and may even counteract each other, leading to the false conclusion that the model is correct.

As mentioned above, the process of inferring the stochastic model from the distribution of the underlying data has enjoyed a great deal of success within the field of stochastic estimation. Therefore, it is interesting to observe that the only property necessary for a correctly specified model, i.e. that $\text{E}[e|\tilde{\mathbf{x}}] = 0$, makes no reference to the distribution of the data at all. Furthermore, although a linear model can be proven to be optimal in the case of joint normality (Papoulis, 1985), it should be noted that theoretically at least, the presence of non-normality does not require that the underlying conditional average is non-linear. Therefore, the inclusion of higher order terms in the presence of non-normality can be considered more of a well-established guideline, rather than a strict requirement.

A final approach is to consider the accuracy of the predictions provided by competing models, and to simply choose the one that yields the lowest mean-square error between the true values of y, and the predicted values $\hat{y} = \mathbf{x}\hat{\mathbf{b}}_{OLS}$. Although this approach may seem reasonable when model selection is considered a linear regression problem, the same is not necessarily true if the prevailing view of stochastic estimation is taken. This is due to the fact a stochastic estimation model is viewed as an estimate of a conditional average, rather than as a model of the behaviour of y. As such, model error is viewed in the same light; for example, it is stated in Tropea *et al.* (2007) that "the accuracy of an estimate must rest in direct comparison with the

[true] conditional average". This can be a problem in practice, because the direct calculation of a conditional average can be difficult to obtain. Indeed, as described in section 3.3, the original motivation for stochastic estimation was to avoid the problems of doing so in the first place.

If the absolute accuracy between a stochastic estimation model and the true conditional average is sought, then the above statement is certainly true. However, for the purpose of model selection, it is the relative accuracy of the models, rather than absolute accuracy, which is required. In which case, a great deal can be learnt by considering the MSE of prediction. As established in section 4.1.2, the behaviour of y is assumed to be dictated by the true model:

$$y = \mathrm{E}[y|\tilde{\mathbf{x}}] + \epsilon \qquad (4.54)$$

The prediction error is therefore given by $\mathrm{E}[y|\tilde{\mathbf{x}}] - \hat{y} + \epsilon$. This can be viewed as consisting of two components; the unexplained error, ϵ, and the difference between the true conditional average and the estimated linear model. As the contribution of ϵ to the MSE is a constant, regardless of the chosen model, any improvement in accuracy can be attributed to a reduction in the error between $\mathrm{E}[y|\tilde{\mathbf{x}}]$ and the estimated model $\mathbf{x}\hat{\mathbf{b}}_{OLS}$. As such, the model which provides the lowest MSE can be considered to be the best specified, from the point of view of both linear regression and stochastic estimation.

When using this approach, care must be taken to avoid over-fitting. If the calculation of the MSE is performed on the same data used to build the estimated model, then increasing the complexity will invariably reduce the error, but the resulting model may be too specific to the sample data and could actually be a worse model for the population in general. There are a variety of procedures to avoid over-fitting, such as cross validation, which is

employed in chapter 5 for this purpose.

The final part of this section will highlight a potentially unrecognised form of model misspecification in the standard stochastic estimation procedure, which occurs when including higher order terms in the model. The stochastic estimation model, in its standard form, omits the zeroth order, or constant term, b_c from the series expansion. This decision may have been motivated by the fact that b_c will be zero when a linear stochastic model is applied to the fluctuating components of the data in question. Unfortunately, this is not necessarily the case when higher order terms are included in the model. To illustrate this, a linear stochastic estimation model and a higher order stochastic estimation will be considered, this time with the zeroth order term included. It will be shown that in the latter case, the value of $\check{b}_c$ in the optimum model is likely to be non-zero when second order terms and higher are included, which suggests that its omission is not justified in these situations.

For simplicity, the following work will consider the case where both x and y are univariate. In line with the standard stochastic estimation procedure, the model is applied to the fluctuating components of the data, $x' = x - \bar{x}$ and $y' = y - \bar{y}$, where $\bar{x} = \mathrm{E}[x]$ and $\bar{y} = \mathrm{E}[y]$. In this situation, the true relationship between x' and y' is given by:

$$y' = \mathrm{E}[y'|x'] + \epsilon \tag{4.55}$$

First, consider the case where (4.55) is approximated with a model of the form $y' = b_c + x'b_1 + e$, which leads to the optimum model $y' = \check{b}_c + x'\check{b}_1 + e$. Taking

the expectation of the optimum model gives:

$$E[y'] = E[\check{b}_c + x'\check{b}_1 + e]$$
$$= \check{b}_c + E[x']\check{b}_1 + E[e] \tag{4.56}$$

which is equivalent to:

$$E[y - \bar{y}] = \check{b}_c + E[x - \bar{x}]\check{b}_1 + E[e] \tag{4.57}$$

and so the value of $\check{b}_c$ is given by:

$$\check{b}_c = E[y - \bar{y}] - E[x - \bar{x}]\check{b}_1 - E[e] \tag{4.58}$$

A property of the optimum model is that the unconditional expectation of the error term is zero, $E[e] = 0$, assuming that a constant term is included in the model (Appendix A.7). Also, $E[y - \bar{y}] = 0$ and $E[x - \bar{x}] = 0$. Hence:

$$\check{b}_c = 0 \tag{4.59}$$

which means that the omission of b_c from the linear stochastic estimation model is valid. Now consider the optimum linear model consisting of terms up to the jth order, i.e.:

$$y' = \check{b}_c + x'\check{b}_1 + x'^2\check{b}_2 + \ldots + x'^j\check{b}_j + e \tag{4.60}$$

As before:

$$E[y - \bar{y}] = \check{b}_c + E[x - \bar{x}]\check{b}_1 + E\left[(x - \bar{x})^2\right]\check{b}_2 + \ldots + E\left[(x - \bar{x})^j\right]\check{b}_j + E[e] \tag{4.61}$$

77

and:

$$\check{b}_c = \mathrm{E}[y - \bar{y}] - \mathrm{E}[x - \bar{x}]\check{b}_1 - \mathrm{E}\left[(x - \bar{x})^2\right]\check{b}_2 - \ldots - \mathrm{E}\left[(x - \bar{x})^j\right]\check{b}_j - \mathrm{E}[e]$$

$$= -\mathrm{E}\left[(x - \bar{x})^2\right]\check{b}_2 - \mathrm{E}\left[(x - \bar{x})^3\right]\check{b}_3 - \ldots - \mathrm{E}\left[(x - \bar{x})^j\right]\check{b}_j$$

$$(4.62)$$

here, the term $\mathrm{E}\left[(x - \bar{x})^i\right]$ corresponds to the ith moment of the independent variable x. As such, $\check{b}_c$ cannot be assumed to be zero. Of course, this requires a contribution from at least one of the terms $\mathrm{E}\left[(x - \bar{x})^i\right]\check{b}_i$. This in turn is dependent on two factors; firstly, that the inclusion of the ith order term of the model actually contributes to the estimation of the conditional average, which ensures that $\check{b}_i$ is non-zero, and secondly, that the ith moment of the distribution of x exists (all the odd moments of a symmetric distribution are zero, for example).

As it is commonly only the second order term that is included in practice (in the form of *quadratic* stochastic estimation), it will often only be the term $\mathrm{E}\left[(x - \bar{x})^2\right]\check{b}_2$ that will be of concern. Unfortunately, this is the very term which is likely to cause the most problems. The decision to include the second order term in the first place is likely to be based on the belief that it will make some contribution to the overall prediction of y, in which case $\check{b}_2$ will be non-zero, and the second moment of x, which is the variance, will always exist and will generally be non-negligible. The resulting omission of $\check{b}_c$ can therefore amount to a serious misspecification of the model. An example of this misspecification is shown in figure 4.3. Both x and y have a mean of zero. The red line shows the resulting fit when the zeroth order term is included, and the blue line shows the fit with the zeroth order term omitted. It can be seen that the omission of the zeroth order term forces the curve to pass though the origin, which may not coincide with the correct intercept of the y-axis.
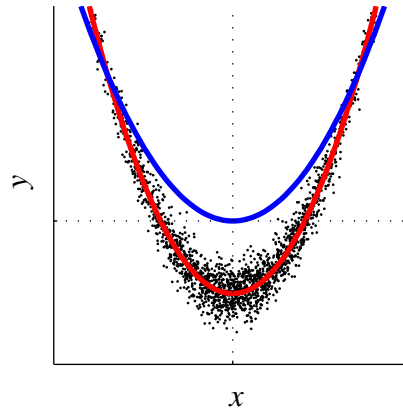
Figure 4.3: Model misspecification due to omitted zeroth order term

4.2.2 Validity of the Stochastic Estimation Procedure

The discussion of the OLS properties introduced two forms of optimality for an estimator; optimality within the class of linear unbiased estimators (BLUE) and the more preferable, optimality within the entire class of unbiased estimators (UMVUE). When the OLS estimator does not meet the requirements to possess these properties, it is possible that there are alternative estimators that do. For example, it is well-known that the *generalised least squares* (GLS) estimator is the BLUE in cases when the 3rd set of assumptions are not met (Rao *et al.*, 2008) (in fact, the OLS estimator can be viewed as a specific case of the GLS estimator). The UMVUE is slightly more complicated, as its precise form, as well its existence, depends on the distribution of the data in question. However, when such an estimator does exist, the technique of maximum likelihood estimation (introduced in section 4.2) can generally be used to find it ((Panik, 2005), for example).

With this in mind, it is reasonable to question the validity of choosing the OLS estimator in situations when a potentially more efficient estimator exists. As the finite sample properties of stochastic estimation seem not to

have been considered previously, such issues have rarely been addressed before. Perhaps the only work to question the validity of the stochastic estimation procedure, relative to other techniques, was by Brereton (1992). In this paper, Brereton related stochastic estimation to the technique of maximum likelihood estimation (MLE). He concluded that the OLS procedure, as implemented by stochastic estimation, was optimal when the model error, e, was normally distributed. When this was not the case, he suggested that other techniques may be more appropriate, and proposed MLE for this purpose.

Viewing these conclusions in the context of the OLS properties, it is clear that Brereton's conditions for optimality of stochastic estimation correspond with the fourth tier of properties. This reveals the exact nature of Brereton's optimality in this situation; namely, that it corresponds to the Uniformly Minimum Variance Unbiased Estimator, or UMVUE. However, as Brereton points out, the property of normality is rarely an accurate model in turbulence. As such, Brereton's conclusions cast doubt on the validity of stochastic estimation in the very situations where the technique is commonly applied.

While these concerns about the optimality, or otherwise, of the OLS estimator are theoretically valid, they do not account for the complexities of adopting an alternative technique in the setting of stochastic estimation.

A fundamental issue with the aforementioned techniques is that they require a specific knowledge of the underlying error term. For example, the application of MLE requires that the underlying distribution of the error term can be modelled by a parametric model, which must also be specified beforehand. In applications of stochastic estimation, neither of these requirements seems likely to be fulfilled exactly, although an approximate model of the empirical distribution of the error term may be available. Unfortunately, any

error between the true and estimated distribution will not only reduce efficiency (i.e. the estimator will no longer be UMVUE, if it would have been otherwise), but may also remove the property of consistency as well. A similar problem occurs with the use of GLS estimation, which requires exact knowledge of both the conditional variance of the error term, $E[e|\mathbf{x}]$, and the error covariance, $Cov[e_i, e_j]$, in order to be considered the BLUE. This is less restrictive than the requirements of MLE, but again, this information will generally have to be approximated from empirical data (leading to the technique of *Feasible Generalised Least Squares* (Cameron and Trivedi, 2005)), leading to a loss of efficiency and consistency.

Ultimately, this puts the benefits of the OLS estimator into context. As an estimator, it is unique in the fact that so many of its properties are guaranteed with minimal requirements on the underlying data. Although the OLS estimator may rarely be able to claim the property of either the BLUE or UMVUE when applied to turbulence, the task of finding these optimum estimators requires a knowledge of the underlying distribution that may simply not be available. Therefore, it seems unlikely that the pursuit of such optimal estimators will be fruitful in the field of stochastic estimation.

This established, it is still beneficial to be aware of the situations when the OLS estimator is not optimal, as alternative techniques may still be able to produce better, even if not the best, estimators when the error distribution deviates sufficiently far from the ideal of homoskedastic, normally distributed error. For example, feasible GLS can outperform OLS in the case of heteroskedastic error (Cameron and Trivedi, 2005), and MLE based approaches can be more reliable when the error distribution exhibits fat or thin tails, or skewness (Tiki and Akkaya, 2004). In these cited examples, the improvement in accuracy was achieved even though the underlying error distribution was

unknown, and had to be estimated from the data. As explained earlier, this tends to invalidate the properties of such estimators, which means that their ability to outperform OLS cannot be inferred from the properties alone, and instead requires a careful analysis of the resulting prediction accuracy. This highlights a further disadvantage to the use of these techniques, which is the considerable effort required to implement and validate their use, coupled with the fact that this must be carried out for each specific application. Given that the OLS estimator will always possess the property of consistency, it must be assessed whether it is worth the time and effort of introducing a different technique, when the act of simply acquiring more data to produce the OLS estimate may offer a similar improvement in performance.

Finally, assuming the practitioner is prepared to adopt an alternative to OLS estimation in order to improve accuracy, it should be noted that there are techniques besides those discussed here that may be better suited to the task. Both GLS and MLE are limited by the Cramér-Rao lower bound (CRLB), which places a restriction on the maximum accuracy of all unbiased estimators. However, in many circumstances relevant in flow prediction, there are certain biased techniques which, no longer restricted by the CRLB, can provide an improvement in MSE that far exceeds the performance of even the UMVUE. These techniques form the basis of the next section.

4.3 Biased Regression Techniques

This section introduces the concept of biased regression techniques, and considers their prediction performance in the context of stochastic estimation. When first considered, it may not be immediately obvious why such estimators would be of use. Purposefully introducing a systematic error seems contrary to the

goal of minimising the error of the estimation, and naturally, it requires giving up the properties of consistency and small-sample unbiasedness, and the assurances of validity that accompany them. In many situations, however, a phenomenon known as the *bias-variance tradeoff* means that the introduction of bias can greatly improve the stability of the estimation, to the point where the accuracy of subsequent predictions is vastly superior to that of the OLS estimation. While this still sacrifices the property of unbiasedness, it is often the case that the amount of bias required is negligible. Furthermore, the careful introduction of bias, followed by rigorous validation of the model, provides a high level of confidence in the estimate, and removes any need to rely on unbiasedness as a guarantee of validity.

The study of biased estimation is particularly relevant to the field of stochastic estimation, as the properties of fluid flow are often conducive to the use of biased techniques. Furthermore, it transpires that certain variations of the stochastic estimation technique can in fact be viewed as biased estimation techniques themselves. In the remainder of this chapter, the relationship between bias and accuracy is considered for a number of techniques, both from within the field of stochastic estimation and from the more general field of linear regression. This leads on to the work in the next chapter, where the performance of the competing techniques is compared in a series of simulated flow prediction tasks.

Much of the discussion in this section will focus on the *bias-variance decomposition* of prediction error, which shall be derived for each of the techniques under investigation. These expressions will subsequently be used to investigate the mechanisms by which bias is able to reduce the overall error in prediction. Each of the techniques has a unique bias-variance decomposition,

but the general form is:

$$\text{MSE}[\hat{y}] = \text{unexplained variance} + \text{model bias} + \text{model variance} \qquad (4.63)$$

The decomposition contains three elements. The *unexplained variance* is the variance of the error term, e, that is present in the optimum linear model $y = \mathbf{x}\check{\mathbf{b}} + e$. Hence, its presence in the MSE is unavoidable, even if the true value, $\check{\mathbf{b}}$ is known, and for this reason it is sometimes referred to as the *irreducible* variance. The remaining terms constitute the model error, which is the error between the estimated model and the true underlying value. This, in turn, consists of the *model bias*, which is a systematic deviation from the true value, and the *model variance*, which is the degree of variation in the model when repeatedly estimated from independent observations of the data. As such, the model variance is synonymous with the model's *stability*.

Once again, the techniques here are presented in the context of instantaneous flow prediction, although the findings are just as relevant when the goal of stochastic estimation is considered in terms of estimating the conditional average. Here, the error in the estimate is the difference between the true conditional average, $\mathbf{xb}$, and the finite sample estimate $\mathbf{x}\hat{\mathbf{b}}$. Therefore, the MSE does not contain a contribution from the error term e, and the bias-variance decomposition becomes:

$$\text{MSE}[\mathbf{x}\hat{\mathbf{b}}] = \text{model bias} + \text{model variance} \qquad (4.64)$$

In the following work, the bias-variance decompositions of $\text{MSE}[\hat{y}]$ can equally be considered as decompositions of $\text{MSE}[\mathbf{x}\hat{\mathbf{b}}]$, simply by ignoring the unexplained variance term.

84

All the biased techniques in this section can be thought of as variations of the standard OLS estimator, so it makes sense to begin with a decomposition of the MSE of the OLS estimator itself. The approach employed here and in subsequent derivations is based loosely on the work of Næs and Mevik (2001). What follows is a brief summary of the derivation; the full version is given in appendix B.1.

The MSE of prediction is defined as:

$$
\begin{aligned}
\mathrm{MSE}\left[y_{\hat{O}LS}\right] &= \mathrm{E}\left[(y-\hat{y})^2\right] \\
&= \mathrm{E}\left[\left(y - \mathbf{x}\hat{\mathbf{b}}_{OLS}\right)^2\right] \\
&= \mathrm{E}\left[\left(\mathbf{x}\check{\mathbf{b}} + e - \mathbf{x}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}\right)^2\right]
\end{aligned}
\tag{4.65}
$$

If assumptions 1-3 are assumed to be valid this expression can be reduced to the form:

$$
\mathrm{MSE}\left[y_{\hat{O}LS}\right] = \sigma^2 + \sigma^2\mathrm{E}\left[\mathbf{x}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{x}^T\right]
\tag{4.66}
$$

The covariance matrix of $\mathbf{X}$ is defined as $\mathbf{S} = \frac{1}{m}\mathbf{X}\mathbf{X}^T$, which allows $(\mathbf{X}^T\mathbf{X})^{-1}$ to be expressed as $(m\mathbf{S})^{-1} = \frac{1}{m}(\mathbf{S})^{-1}$. Therefore:

$$
\mathrm{MSE}\left[\hat{y}\right] = \sigma^2 + \frac{\sigma^2}{m}\mathrm{E}\left[\mathbf{x}(\mathbf{S})^{-1}\mathbf{x}^T\right]
\tag{4.67}
$$

Representing $\mathbf{S}$ as a function of its eigenvectors (i.e. the principal components[2]

---

[2]Strictly speaking, the definition of principal component analysis given in section 3.4 requires that the observations of $\mathbf{x}$ have a mean of zero, which is not assumed here. However, the general principle remains the same

of $\mathbf{X}$) yields $\mathbf{S} = \mathbf{ULU}^T = \sum_{k=1}^{n} l_k \mathbf{u}_k \mathbf{u}_k^T$:

$$
\begin{aligned}
\mathrm{MSE}\left[\hat{y}\right] &= \sigma^2 + \frac{\sigma^2}{m}\mathrm{E}\left[\mathbf{x}(\sum_{k=1}^{n} l_k \mathbf{u}_k \mathbf{u}_k^T)^{-1}\mathbf{x}^T\right] \\
&= \sigma^2 + \frac{\sigma^2}{m}\mathrm{E}\left[\sum_{k=1}^{n} \frac{\mathbf{x}\mathbf{u}_k \mathbf{u}_k^T \mathbf{x}^T}{l_k}\right] \\
&= \sigma^2 + \frac{\sigma^2}{m}\mathrm{E}\left[\sum_{k=1}^{n} \frac{(\mathbf{x}\mathbf{u}_k)^2}{l_k}\right]
\end{aligned}
\tag{4.68}
$$

At this point, an alternative approach to that of Næs and Mevik (2001) is taken. By making the simplifying assumption that the sample principal components of $\mathbf{X}$ are identical to those of the underlying population, i.e. $\mathbf{u}_i = \boldsymbol{\gamma}_i$, then $\mathrm{E}\left[(\mathbf{x}\mathbf{u_k})^2\right] = \mathrm{E}\left[(\mathbf{x}\boldsymbol{\gamma}_k)^2\right] = \lambda_k$, which is the variance accounted for by the kth principal component:

$$
\mathrm{MSE}\left[\hat{y}\right] = \sigma^2 + \frac{\sigma^2}{m}\mathrm{E}\left[\sum_{k=1}^{n} \frac{\lambda_k}{l_k}\right]
\tag{4.69}
$$

Also, assuming the sample and population eigenvalues to be equal, $l_k = \lambda_k$:

$$
\mathrm{MSE}\left[\hat{y}\right] = \sigma^2 + \frac{\sigma^2}{m}\mathrm{E}\left[\sum_{k=1}^{n} \frac{\lambda_k}{\lambda_k}\right]
\tag{4.70}
$$

This leads to the ultimate result:

$$
\mathrm{MSE}\left[\hat{y}\right] = \sigma^2 + \frac{n}{m}\sigma^2
\tag{4.71}
$$

The MSE decomposition consists of two elements; the unexplained variance, σ^2, and the model error, $\frac{n}{m}\sigma^2$. As the OLS estimator has the property of unbiasedness, then the model error in this case consists solely of the model

86

variance term. The expression for the model error provides a theoretical basis for some of the observations that have been made previously; that the model error is directly proportional to the model complexity, i.e. n, and inversely proportional to the number of observations used to construct the model, m. Furthermore, the model error is also proportional to the unexplained variance. Hence, even when the aim of stochastic estimation is the estimation of the conditional average, the unexplained variance will affect the accuracy of the resulting estimate.

With equation (4.71) as a guide, it is clear that for any application of the OLS estimator, the ratio $\frac{n}{m}$ should be as small as possible. However, achieving a sufficiently small ratio may not always be possible in many applications of stochastic estimation. Common sources of data, such as PIV or numerical simulations, are capable of providing a huge (and ever increasing) number of velocity vectors which may be used to generate event data, and the inclusion of higher orders into the estimation will increase the size of n even further. Conversely, time, cost or storage constraints may place an upper limit on the size of m. In situations where $\frac{n}{m}$ is unacceptably large even after the maximum number of observations has been reached, it may still be possible to improve the accuracy by reducing n.

Rather than reduce the size of $\mathbf{x}$ on an element-wise basis, a more prudent approach may be to instead replace $\mathbf{x}$ with a reduced-order model, which effectively reduces n, while (hopefully) retaining the important features of the original data. This approach provides the motivation for a number of biased techniques. Two of which, namely *Principal Component Regression* and *Partial Least Squares*, have been chosen for discussion due to their suitability and relevance to the field of stochastic estimation.

4.3.1 Principal Component Regression

Conceptually, principal component regression (PCR) can be thought of as OLS estimation, except where the original data, $\mathbf{x}$, is first transformed to the scores $\mathbf{z}$ using the first k principal components of the matrix $\mathbf{X}$, i.e.:

$$\mathbf{z} = \mathbf{x}\tilde{\mathbf{U}} \tag{4.72}$$

where $\tilde{\mathbf{U}}$ is an $n \times k$ matrix of the first k principal components of $\mathbf{X}$. This leads to the regression model:

$$y = \mathbf{z}\mathbf{c} + e \tag{4.73}$$

where the coefficients $\mathbf{c}$ are estimated by OLS:

$$\hat{\mathbf{c}} = (\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T\mathbf{y} \tag{4.74}$$

It is possible to express this regression in the standard form of the linear regression model, $y = \mathbf{x}\mathbf{b} + e$. Given that $\mathbf{z} = \mathbf{x}\tilde{\mathbf{U}}$, it is possible to express y as:

$$\begin{aligned} y &= \mathbf{x}\tilde{\mathbf{U}}\mathbf{c} + e \\ &= \mathbf{x}\mathbf{b}_{PCR} + e \end{aligned} \tag{4.75}$$

with $\hat{\mathbf{b}}_{PCR}$ given by:

$$\begin{aligned} \hat{\mathbf{b}}_{PCR} &= \tilde{\mathbf{U}}\left((\mathbf{X}\tilde{\mathbf{U}})^T\mathbf{X}\tilde{\mathbf{U}}\right)^{-1}(\mathbf{X}\tilde{\mathbf{U}})^T\mathbf{y} \\ &= \tilde{\mathbf{U}}\left(\tilde{\mathbf{U}}^T\mathbf{X}^T\mathbf{X}\tilde{\mathbf{U}}\right)^{-1}\tilde{\mathbf{U}}^T\mathbf{X}^T\mathbf{y} \end{aligned} \tag{4.76}$$

From appendix B.2, this can be further reduced to:

$$
\begin{aligned}
\hat{\mathbf{b}}_{PCR} &= \tilde{\mathbf{U}}\tilde{\mathbf{U}}^T \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y} \\
&= \tilde{\mathbf{U}}\tilde{\mathbf{U}}^T\hat{\mathbf{b}}_{OLS} \\
&= \underrightarrow{\mathbf{P}}_{\tilde{V}}\hat{\mathbf{b}}_{OLS}
\end{aligned}
\tag{4.77}
$$

where $\underrightarrow{\mathbf{P}}_{\tilde{V}}$ is a projection matrix that defines the projection onto the subspace $\tilde{V}$, as determined by the first k principal components of $\mathbf{X}$. Therefore, the regression model is:

$$
y = \mathbf{x}\underrightarrow{\mathbf{P}}_{\tilde{V}}\hat{\mathbf{b}}_{OLS} + e
\tag{4.78}
$$

Hence, PCR operates by restricting the independent variable $\mathbf{x}$ to a reduced subspace. In order to understand how this can improve the performance of standard OLS regression, a decomposition of the MSE of $\hat{\mathbf{y}}$ is once again performed. Following a similar approach to before (see appendix B.3), the MSE can be decomposed into:

$$
\begin{aligned}
\mathrm{MSE}\left[\hat{y}\right] &= \sigma^2 + \frac{k}{m}\sigma^2 + \mathrm{E}\left[\left(\mathbf{x}\underrightarrow{\mathbf{P}}_{\tilde{V}_\perp}\check{\mathbf{b}}\right)^2\right] \\
&= \text{unexplained variance} + \text{model variance} + \text{model bias}
\end{aligned}
\tag{4.79}
$$

where the matrix $\underrightarrow{\mathbf{P}}_{\mathbf{V}_\perp}$ is the projection onto the null space, $\tilde{V}_\perp$, i.e. the subspace defined by the discarded principal components. Compared with the MSE of the OLS prediction, it can be seen that the model variance has been reduced from $\frac{n}{m}\sigma^2$ to $\frac{k}{m}\sigma^2$, at the expense of the introduction of potential model bias. Any actual inclusion of bias is dependent on two factors; first, the expected value of $\mathbf{x}$ must contain components that encroach into the null space, $\tilde{V}_\perp$ (leading to a non-zero value of $\mathbf{x}\underrightarrow{\mathbf{P}}_{\tilde{V}_\perp}$), and second, that these components would otherwise have contributed to the estimation of y (leading to a non-zero

value of $\mathbf{x}\underrightarrow{\mathbf{P}}_{\tilde{V}_\perp}\check{\mathbf{b}}$). By removing the principal components of $\mathbf{x}$ that contain no, or little, variance, PCR can be seen to address the first of these factors. However, no regard is paid toward the second, which may limit its effectiveness when the subspace of $\mathbf{x}$ that is relevant to y only comprises a small part of the total subspace spanned by x. As such, the use of PCR will generally include more dimensions of $\mathbf{x}$ than are necessary for the prediction of y. In applications of stochastic estimation, this may be a concern when $\mathbf{x}$ and y are taken from domains with large spatial or temporal separations, and/or when $\mathbf{x}$ and y are observations of two distinct properties of the flow. As an example, this phenomena was observed by Jordan *et al.* (2007), who investigated the sound production of a simulated turbulent jet using a novel technique called the *Most Observable Decomposition* (MOD). They found that while the jet's velocity field resided in a large subspace (more than 350 principal components were required to captured 50% of the kinetic energy), only a very small part of this was related to radiated sound energy.

In situations where the effectiveness of PCR is compromised in this manner, the technique of partial least squares may offer a more suitable alternative.

4.3.2   Partial Least Squares

To some extent, partial least squares (PLS) is similar in approach to PCR, in that it seeks to decompose the n-dimensional prediction data $\mathbf{x}$ into a lower dimensional set of scores, $\mathbf{t}$, using the "weight" matrix $\mathbf{R}$:

$$\mathbf{t} = \mathbf{x}\mathbf{R} \tag{4.80}$$

and then carry out a regression of the form:

$$\mathbf{y} = \mathbf{t}\mathbf{Q} + \mathbf{e} \tag{4.81}$$

where $\mathbf{Q}$ is estimated using OLS, i.e. $\hat{\mathbf{Q}} = (\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T\mathbf{Y}$. Again, PLS can be expressed in terms of the linear model:

$$\begin{aligned}
\mathbf{y} &= \mathbf{x}\mathbf{R}\mathbf{Q}^T + \mathbf{e} \\
&= \mathbf{x}\mathbf{B}_{PLS} + \mathbf{e}
\end{aligned} \tag{4.82}$$

The major conceptual difference between PCR and PLS is the manner in which $\mathbf{x}$ is decomposed. Whereas PCR provides a decomposition that is optimal only in terms of $\mathbf{x}$, PLS explicitly aims to extract components of $\mathbf{x}$ that are relevant to $\mathbf{y}$, so that the prediction of $\mathbf{y}$ from $\mathbf{t}$ can be carried out with as small a value of k as possible. Also, note that the single element y has been replaced by a p dimensional vector, $\mathbf{y}$. Although both OLS and PCR can be applied to a multivariate y (see section 4.1), the resulting regression yields the same result as p independent regressions on each element $y_1, y_2, \ldots, y_p$. The same is not true for PLS, however, and is recommended that PLS is applied to the full vector $\mathbf{y}$ whenever there is correlation between elements (Wold *et al.*, 2001). This is likely to occur in most applications of stochastic estimation, so the multivariate case will be considered here.

PLS can be considered a general framework to regression, open to a degree of interpretation, rather than an explicitly defined technique. As such, since its introduction by Wold (1975), it has undergone several revisions and adaptations, to the point now where there are many competing algorithms available. Central to all these implementations is the task of calculating the

matrices $\mathbf{R}$ and $\mathbf{Q}$, necessary for the construction of $\mathbf{B}_{PLS}$ (although there are generally several intermediate and non-essential variables available, for diagnostics and interpretation of the results). The discussion here will focus specifically on the SIMPLS algorithm (de Jong, 1993), which is regarded for its computational efficiency and ease of interpretation (ter Vraak and de Jong, 1998; Boulesteix and Strimmer, 2005). Given a joint observation of $(\mathbf{Y}, \mathbf{X})$, the SIMPLS algorithm decomposes $\mathbf{X}$ into the score matrix $\mathbf{T}$:

$$\mathbf{T} = \mathbf{XR} \tag{4.83}$$

where $\mathbf{R}$ is found by iteratively estimating its columns, $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_k$, according to the following criteria:

1. The first vector $\mathbf{r}_1$ is a unit vector that maximises the covariance between $\mathbf{Y}^T$ and the resulting score vector, $\mathbf{t}_1$, i.e:

$$\begin{aligned} \mathbf{r}_1 &= \mathrm{argmax}\left(\left|\mathbf{Y}^T \mathbf{t}_1\right|\right) \\ &= \mathrm{argmax}\left(\left|\mathbf{Y}^T \mathbf{X} \mathbf{r}_1\right|\right) \end{aligned} \tag{4.84}$$

2. Subsequent vectors $\mathbf{r}_i$ where $i = 2, 3, \ldots, k$, are unit vectors that maximise the covariance between $\mathbf{Y}^T$ and the resulting $\mathbf{t}_i$, subject to the constraint that $\mathbf{t}_i$ is orthogonal to all preceding vectors $\mathbf{t}_1, \ldots, \mathbf{t}_{i-1}$.

It is important to note that SIMPLS does not place any requirement on the orthogonality of the vectors $\mathbf{r}_i$. This is of relevance if it is necessary to reconstruct $\mathbf{x}$ from $\mathbf{t}$; although $\mathbf{t} = \mathbf{xR}$, the quantity $\mathbf{tR}^T$ is not a suitable approximation to $\mathbf{x}$, as would be the case if the columns of $\mathbf{R}$ were orthogonal (such as if $\mathbf{R}$ consisted of principal components of $\mathbf{X}$, for example). Instead,

$\mathbf{x}$ is reconstructed using OLS regression, in the same way as $\mathbf{y}$:

$$\hat{\mathbf{x}} = \mathbf{t}\hat{\mathbf{P}} \qquad (4.85)$$

where $\hat{\mathbf{P}} = (\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T\mathbf{X}$. Calculation of $\mathbf{r}_i$ is achieved by means of the singular value decomposition. For the first iteration, the choice of $\mathbf{r}_1$ that maximises $\left(\left|\mathbf{Y}^T\mathbf{X}\mathbf{r}_1\right|\right)$ is simply the largest right singular vector of the matrix $\mathbf{Y}^T\mathbf{X}$. To ensure that subsequent score vectors $\mathbf{t}_i$ are orthogonal, the covariance matrix $\mathbf{Y}^T\mathbf{X}$ is "deflated" after every iteration, to ensure that the contribution to $\mathbf{X}$ made by the previous values of $\mathbf{t}$ is not present in the covariance matrix for future calculations. Details of this procedure, along with a summary of the entire algorithm, are presented in appendix B.4.

Although the concept of PLS may seem to be a significant departure from OLS regression, and hence from stochastic estimation, it has a number of attractive features that may be relevant for the study and prediction of flow phenomena. Firstly, it addresses PCR's potential efficiency problem, which occurs when the $\mathbf{y}$-relevant components of $\mathbf{x}$ reside in a small part of the total subspace of $\mathbf{x}$. As before, the bias-variance decomposition of the MSE is a useful guide to the performance of PLS. This time, the prediction yields the multivariate vector, $\hat{\mathbf{y}}_{PLS}$, so the MSE is redefined as:

$$\begin{aligned} \mathrm{MSE}[\hat{\mathbf{y}}_{PLS}] &= \frac{1}{p}\mathrm{E}\left[(\mathbf{y} - \hat{\mathbf{y}}_{PLS})(\mathbf{y} - \hat{\mathbf{y}}_{PLS})^T\right] \\ &= \frac{1}{p}\mathrm{E}\left[\|(\mathbf{y} - \hat{\mathbf{y}}_{PLS})\|^2\right] \end{aligned} \qquad (4.86)$$

which describes the average MSE of the p elements of $\hat{\mathbf{y}}_{PLS}$. Due to the similarities between PCR and PLS (as described above), the resulting decomposition

is of the same form as equation (4.79):

$$\text{MSE}\left[\hat{\mathbf{y}}\right] = \frac{\sigma^2}{p} + \frac{k}{pm}\sigma^2 + \frac{1}{p}\text{E}\left[\left\|\mathbf{x}\underline{\mathbf{P}}_{\rightarrow V_{\mathbf{R}_\perp}}\mathbf{b}\right\|^2\right]$$

$$=\text{unexplained variance} + \text{model variance} + \text{model bias}$$

(4.87)

where σ^2 is the total unexplained variance contained within $\mathbf{y}$, and the matrix $\underline{\mathbf{P}}_{\rightarrow V_{\mathbf{R}_\perp}}$ defines the projection onto the null space of the columns of $\mathbf{R}$. As with PCR, prediction performance is maximised by limiting this null space to the regions of the space spanned by $\mathbf{x}$ that are irrelevant to the prediction of $\mathbf{y}$, while keeping k, the number of factors, as low as possible. Ultimately, the success of PLS rests on how closely the matrix $\mathbf{R}$ is able to meet this criteria. Unfortunately, because of the iterative nature of the technique, and the relative infancy of the SIMPLS algorithm, little is known about the exact properties of $\mathbf{R}$, so it is difficult to draw any strong conclusions about the predictive ability of PLS, at least from a theoretical point of view. Nonetheless, given how closely the above criteria coincide with the intended goals of PLS, it is worth investigating how well it is able to achieve them in practice.

A second reason for the inclusion of PLS is that it has potential uses as an analytical tool in the study of coherent structures in turbulence. A recent development in this field has seen the use of POD, and POD-like techniques, used as a means of simultaneously extracting correlated structures from separate, but related, domains. These approaches include *extended* proper orthogonal decomposition (Borée, 2003), and the *Most Observable Decomposition* (Jordan *et al.*, 2007). PLS can also be viewed within this context, as a means of performing a joint decomposition of the observed quantities $\mathbf{X}$ and $\mathbf{Y}$ in the

form:

$$\hat{\mathbf{X}} = \mathbf{TP}$$
$$\hat{\mathbf{Y}} = \mathbf{TQ}$$

(4.88)

where $\hat{\mathbf{X}}$ and $\hat{\mathbf{Y}}$ are the parts of $\mathbf{X}$ and $\mathbf{Y}$ that are explained by the scores $\mathbf{T}$. By considering the ith joint observation $(\mathbf{x}(i), \mathbf{y}(i))$ (where $i = 1, 2, \ldots, m$), the PLS decomposition can be written in the style of conventional POD:

$$\hat{\mathbf{x}}(i) = \sum_{j=1}^{k} t_j(i)\mathbf{p}_j$$
$$\hat{\mathbf{y}}(i) = \sum_{j=1}^{k} t_j(i)\mathbf{q}_j$$

(4.89)

Here, the scores t_j can be thought of as a type of expansion coefficient, which describe the temporal evolution of the spatial "modes" $\mathbf{p}_j$ and $\mathbf{q}_j$. Further work would be required to assess whether PLS is capable of providing useful and reliable information in this setting. However, the motivation behind the technique, at least, is well suited to the task, as is the intuitive nature of the resulting model.

4.3.3 Low Dimensional Modelling of Multivariate Dependent Variables

The biased techniques considered so far have used a decomposition of the independent variable $\mathbf{x}$ in order to improve prediction performance. In cases where the response variable is multivariate, performance benefits may also be possible by carrying out the regression using a decomposition of $\mathbf{y}$ instead. This section will focus on a specific approach to this concept, which involves partially decomposing $\mathbf{y}$ into the k-element score vector $\mathbf{t} = \mathbf{y}\tilde{\mathbf{U}}^T$, where the

k orthonormal columns, or *modes*, of the $p \times k$ matrix $\tilde{\mathbf{U}}$ form the basis of a reduced subspace of $\mathbf{y}$. This results in a linear model of the form:

$$\mathbf{t} = \mathbf{xC} + \mathbf{e} \tag{4.90}$$

which is then solved using the standard OLS estimator, $\hat{\mathbf{C}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{T}$. For a prediction of $\hat{\mathbf{t}}$, the subsequent value of $\mathbf{y}$ is then given by:

$$\hat{\mathbf{y}} = \hat{\mathbf{t}}\tilde{\mathbf{U}} \tag{4.91}$$

This approach is recognisable as the *complimentary technique* (herein referred to as complementary stochastic estimation (CSE)), which is described in 3.4. Here, the columns of the matrix $\tilde{\mathbf{U}}$ are chosen to be the first k principal components of the observed response matrix $\mathbf{Y}$. Outside the field of stochastic estimation, the technique of *reduced rank regression* (RRR) can be represented in this form, only this time using the principal components of the predicted response $\hat{\mathbf{Y}} = \mathbf{X}\hat{\mathbf{B}}_{OLS}$ instead. In both cases, the approach is normally applied because a low dimensional model is explicitly sought for analysis purposes, whereas in the context of this work, the low-dimensional nature of the approach is considered solely as a means of improving the prediction of $\mathbf{y}$.

Once again, the predictive ability of these techniques can be understood by investigating the bias-variance decomposition of the MSE. In order to express the decomposition in a manner which is applicable to both CSE and RRR, $\tilde{\mathbf{U}}$ will be left undefined. For simplicity, this general approach will be tentatively titled *low-dimensional response* (LDR) regression.

First, it is necessary to express LDR regression in terms of the standard

linear model:

$$\mathbf{y} = \mathbf{x}\mathbf{C}\tilde{\mathbf{U}} + \mathbf{e}$$

$$= \mathbf{x}\mathbf{B}_{LDR} + \mathbf{e}$$

(4.92)

where an estimate of $\mathbf{B}_{LDR}$ is provided by:

$$\hat{\mathbf{B}}_{LDR} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{T}\tilde{\mathbf{U}}$$

$$= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}\tilde{\mathbf{U}}^T\tilde{\mathbf{U}}$$

$$= \hat{\mathbf{B}}_{OLS}\tilde{\mathbf{U}}^T\tilde{\mathbf{U}}$$

$$= \hat{\mathbf{B}}_{OLS}\underline{\mathbf{P}}_{\rightarrow\tilde{V}}$$

(4.93)

where $\underline{\mathbf{P}}_{\rightarrow\tilde{V}}$ is the projection on to the subspace $\tilde{V}$, defined by the matrix $\tilde{\mathbf{U}}$. This leads to:

$$\hat{\mathbf{y}}_{LDR} = \mathbf{x}\hat{\mathbf{B}}_{OLS}\underline{\mathbf{P}}_{\rightarrow\tilde{V}}$$

$$= \hat{\mathbf{y}}_{OLS}\underline{\mathbf{P}}_{\rightarrow\tilde{V}}$$

(4.94)

This result shows any performance improvement is due to the projection of the OLS estimate of $\mathbf{y}$ onto the subspace $\tilde{V}$. It can be observed that although the technique can be carried out by a decomposition of the *original* data $\mathbf{y}$, the technique is ultimately acting on the *predicted* data $\hat{\mathbf{y}}_{OLS}$. Naturally, $\mathbf{y}$ and $\hat{\mathbf{y}}_{OLS}$ differ due to the inclusion of prediction error in the latter, but there can also be a fundamental difference in terms of subspaces that these vectors inhabit. Both vectors are of length p, and so can be considered as points within p-dimensional space $\mathbb{R}^p$. However, whereas $\mathbf{y}$ can theoretically inhabit the full p-dimensional space of $\mathbb{R}^p$, $\hat{\mathbf{y}}_{OLS}$ is potentially limited to a smaller subspace, the size of which is dependent on the rank of $\hat{\mathbf{B}}_{OLS}$. At most, $\hat{\mathbf{B}}_{OLS}$ has a rank of $r = \min\left(\text{rank}(\mathbf{X}), \text{rank}(\mathbf{Y})\right) = \min(n, m, p)$. Therefore, in cases where the number of independent variables n, or observations m, is less than the length of $\mathbf{y}$, then $r < p$, and $\hat{\mathbf{y}}_{OLS}$ will be constrained to an r dimensional subspace

within $\mathbb{R}^p$. This subspace shall be denoted V_{OLS}.

The differences between the true and predicted values of $\mathbf{y}$ have certain implications on the use of CSE, both as a means of improving prediction accuracy, and for the technique in general. This can be illustrated with a simple example, which involves the prediction of a three dimensional dependent variable $\mathbf{y} = [y_1 \ y_2 \ y_3]$. The behaviour of each element of $\mathbf{y}$ is determined by the underlying model of the form:

$$y_1 = x_1 + x_2 + x_3 + e_1 \tag{4.95a}$$

$$y_2 = x_1 + x_2 + x_3 + e_2 \tag{4.95b}$$

$$y_3 = 2x_3 + e_3 \tag{4.95c}$$

For simplicity, x_1, x_2, x_3, e_1, e_2 and e_3 are independent, normally distributed random variables, all with zero mean and variance σ^2. The resulting distribution of $\mathbf{y}$ is shown in figure 4.4. Also shown on the plots is the direction of the largest principal component, $\mathbf{u}_1 = \left[\frac{\sqrt{3}}{3} \ \frac{\sqrt{3}}{3} \ \frac{\sqrt{3}}{3} \right]^T$. Consider a linear regression model, $\mathbf{y} = \mathbf{x}\mathbf{B} + \mathbf{e}$, based on the independent variables x_1 and x_2 *only*, i.e.:

$$y_i = b_1 x_1 + b_2 x_2 + e \tag{4.96}$$

for $i = 1, 2, 3$. An OLS estimate of $\mathbf{B}$ is made using the observed matrices $\mathbf{Y}$ and $\mathbf{X}$. Assuming that the number of observations $m \geq 3$, the resulting OLS predictions $\mathbf{x}\hat{\mathbf{B}}$ will have a rank of 2, which is limited by the number of independent variables, $n = 2$. As such, the subspace V_{OLS} is limited to a 2-dimensional region of $\mathbb{R}^3$, as shown in figure 4.5 (a), along with some example predictions of $\hat{\mathbf{y}} = \mathbf{x}\hat{\mathbf{B}}_{OLS}$, which are constrained to the plane V_{OLS}. From equation (4.95c), it can be seen that the regression model has no predictive

power over y_3, so it follows that V_{OLS} resides in the $y_1 - y_2$ plane at $y_3 = 0$.

The differing approaches of RRR and CSE will now be considered. From equation (4.94), both techniques involve the projection of $\mathbf{x}\hat{\mathbf{B}}$ onto a subspace $\tilde{V}$, defined by the user-specified matrix $\tilde{\mathbf{U}}$. In situations where the data already resides in a subspace V_{OLS}, the projection from V_{OLS} to $\tilde{V}$ can produce some unexpected results, which depend on the relationship between the two subspaces.

In the case of RRR, the predicted values $\hat{y}$ are projected onto a subspace $\tilde{V}_{RRR}$, defined by principal components of the predicted data $\mathbf{Y} = \mathbf{X}\hat{\mathbf{B}}$. This means that $\tilde{V}_{RRR}$ is necessarily contained within V_{OLS}, and so the projection of $\mathbf{x}\hat{\mathbf{B}}$ on to $\tilde{V}_{RRR}$ comprises simply of a reduction of subspace V_{OLS} (figure 4.5 (b)). However, for CSE, the matrix $\tilde{\mathbf{U}}\mathbf{y}$ is obtained from the principal components of the original data $\mathbf{Y}$ rather than $\hat{\mathbf{Y}}$, which means that the subspace $\tilde{V}_{CSE}$ and V_{OLS} need not be related. This is the case in the example, where the subspace defined by the largest principal component of $\mathbf{Y}$ intersects V_{OLS} at a large angle. This is illustrated in figure 4.5 (c), and more clearly in 4.6.

The projection of $\mathbf{x}\hat{\mathbf{B}}$ on to an unrelated subspace can be problematic for three reasons. Firstly, it can be seen from the figures that the principal components of $\mathbf{y}$, which are optimal for the data they are built from, need not provide an optimal decomposition of $\mathbf{x}\hat{\mathbf{B}}$. Hence, when a compact low dimensional representation of the predicted data $\hat{\mathbf{y}}$ is sought, the CSE approach is not necessarily the most appropriate approach. In the context of biased regression, this is likely to reduce the performance improvement provided by CSE, relative to RRR, as will be discussed shortly. Another concern relates to the interpretation of the resulting CSE predictions. This is most apparent in figure 4.6, which shows that the CSE predictions have effectively been rotated
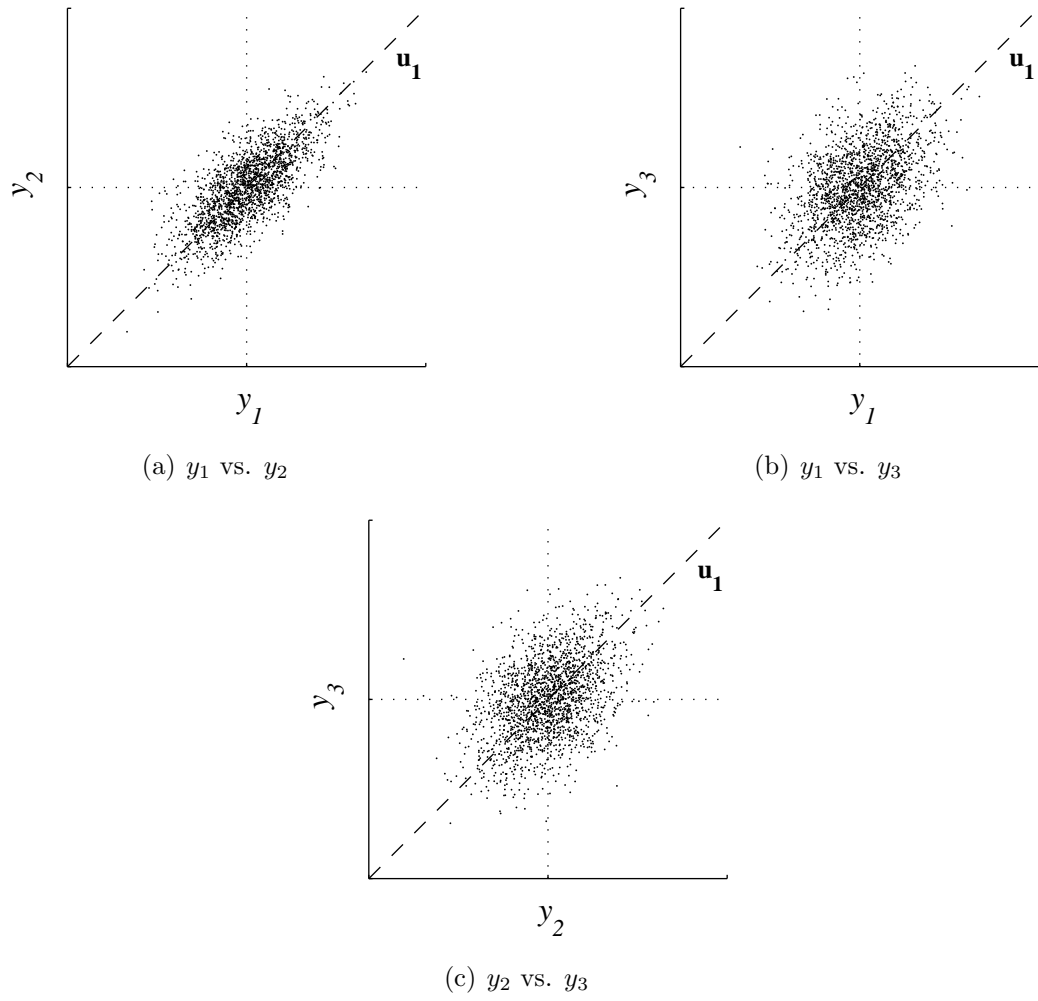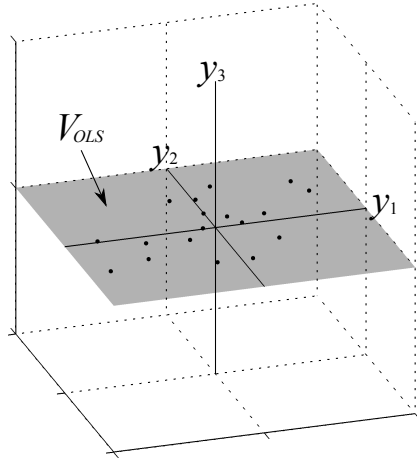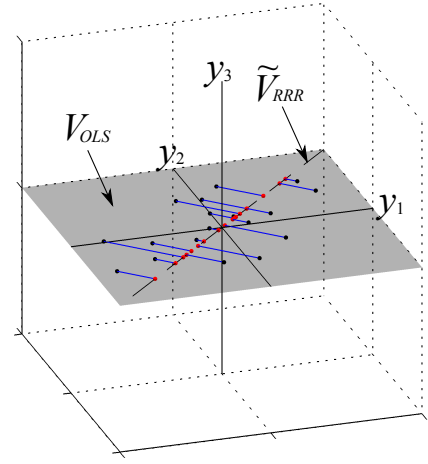
(a) y_1 vs. y_2

(b) y_1 vs. y_3

(c) y_2 vs. y_3

Figure 4.4: Distribution of data in **y**. The dashed line indicates the direction of the first principal component, $\mathbf{u}_1$

(a) Example OLS predictions of **y**, denoted by the black points. The predictions are contained within the 2-dimensional subspace V_{OLS}, illustrated by the grey plane

(b) The RRR predictions of **y** (red dots), resulting from the projection of the OLS predictions on to the subspace $\tilde{V}_{RRR}$ as defined by the largest principal component of $\hat{\mathbf{Y}}_{OLS}$. Note that $\tilde{V}_{RRR}$ is contained within V_{OLS}.
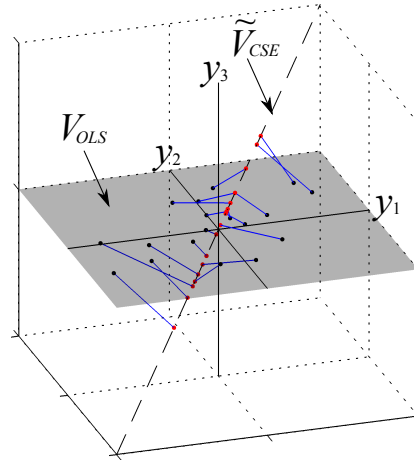
(c) The CSE predictions of **y** (red dots). The subspace $\tilde{V}_{CSE}$, defined by largest principal component of $\mathbf{Y}$ is not contained within V_{OLS}.

Figure 4.5: Comparison of the OLS, RRR and CSE predictions of **y**

101
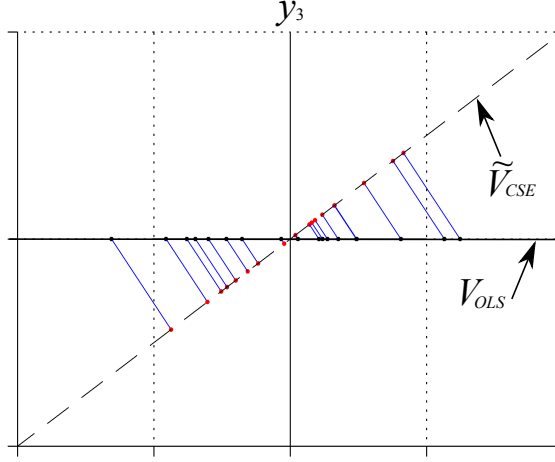
Figure 4.6: CSE predictions of $\mathbf{y}$, viewed in a plane defined by $\hat{y}_3$ and the line $\tilde{V}_{CSE}$.

from their original positions, and the data now possesses a $\hat{y}_3$ component. This is at odds with the underlying regression model (equation (4.96)), which has no predictive power over y_3. This leads to the erroneous conclusion that the behaviour of y_3 is related to the behaviour of x_1 and/or x_2. Of course, this situation has been manufactured to illustrate this problem, and is unlikely to occur to such an extent in practical situations. Nonetheless, this does show the risks involved in applying CSE. The most pressing concern, however, is that the projection from V_{OLS} to a potentially unrelated subspace $\tilde{V}$ results in behaviour that is too complex to account for in the derivation of the bias-variance decomposition. As such, for the purposes of the following work, it is necessary to assume that $\tilde{V}$ *must* lie within V_{OLS}. Obviously, this limits the applicability of the bias-variance decomposition to CSE when $r < p$, although it is still valid when $r = p$, in which case V_{OLS} spans the full $\mathbb{R}^p$, which will contain all possible subspaces $\tilde{V}$. Also, the decomposition is valid for all cases of RRR. While the limited applicability of the decomposition of CSE is

unfortunate, it is noted that the expression is not intended to be used to make exact calculations of the MSE, merely to provide an insight into the nature of the techniques' predictive performance. The assumptions made here will hopefully not detract from this goal.

As the error between $\mathbf{y}$ and $\hat{\mathbf{y}}_{RRR}$ is a $1 \times p$ vector, the MSE of $\hat{\mathbf{y}}_{LDR}$ shall be defined as:

$$
\begin{aligned}
\mathrm{MSE}[\hat{\mathbf{y}}_{LDR}] &= \frac{1}{p}\mathrm{E}\left[(\mathbf{y} - \hat{\mathbf{y}}_{LDR})(\mathbf{y} - \hat{\mathbf{y}}_{LDR})^T\right] \\
&= \frac{1}{p}\mathrm{E}\left[\|(\mathbf{y} - \hat{\mathbf{y}}_{LDR})\|^2\right]
\end{aligned}
\tag{4.97}
$$

The resulting bias-variance decomposition (as described in Appendix B.5) is:

$$
\begin{aligned}
\mathrm{MSE}[\hat{\mathbf{y}}_{LDR}] &= \frac{\sigma^2}{p} + \hat{\sigma}_{\perp}^2\frac{n}{pm} \\
&\quad + \tilde{\sigma}^2\frac{n}{pm} + \frac{1}{p}\mathrm{E}\left[\left\|\mathbf{x}\check{\mathbf{B}}\left(\underset{\rightarrow}{\mathbf{P}}_{V_{OLS}} - \underset{\rightarrow}{\mathbf{P}}_{\tilde{V}}\right)\right\|^2\right] \\
&= \text{unexplained variance} + \text{unremovable model error} \\
&\quad + \text{removable model variance} + \text{model bias}
\end{aligned}
\tag{4.98}
$$

where:

- σ^2 is the total unexplained variance in $\mathbf{y}$.

- $\hat{\sigma}_{\perp}^2$ is the component of unexplained variance that resides in the null space of $\hat{\mathbf{Y}}_{OLS}$, i.e. the region of $\mathbb{R}^p$ that excludes the subspace V_{OLS}. This term is only present when the rank r of $\mathbf{B}_{OLS}$ is less than the size p of $\mathbf{y}$.

- $\tilde{\sigma}^2$ is the component of unexplained variance contained in the subspace $\tilde{V}$, which is defined by the choice of $\tilde{\mathbf{U}}$.

- $\underline{\mathbf{P}}_{\rightarrow V_{OLS}}$ is the projection onto the subspace V_{OLS}.

- $\underline{\mathbf{P}}_{\rightarrow \tilde{V}}$ is the projection onto the subspace $\tilde{V}$.

- $\underline{\mathbf{P}}_{\rightarrow V_{OLS}} - \underline{\mathbf{P}}_{\rightarrow \tilde{V}}$ defines the projection on to the resulting subspace when $\tilde{V}$ is removed from V_{OLS}.

Equation (4.98) differs from previous bias-variance decompositions due to the inclusion of a fourth term, which can be interpreted as *unremovable model error*. This term is present when $r < p$, and corresponds to a component of model error that cannot be removed from the prediction error, regardless of the choice of $\tilde{\mathbf{U}}$. The remaining three terms in the expression have similar meanings to before.

The bias-variance decomposition demonstrates the improvement that LDR regression can bring to the stability of the estimated model. In the case of OLS estimation, the full sum of unexplained variance in $\mathbf{y}$ contributes to the model variance, whereas in LDR regression this contribution is restricted to the unexplained variance $\tilde{\sigma}^2$ that lies within a reduced subspace of $\mathbb{R}^p$, defined by the matrix $\tilde{\mathbf{U}}$. As the number of modes in $\tilde{\mathbf{U}}$ is decreased, the size of this subspace is reduced, and it can be expected that the model variance will fall. However, it is noted that the exact nature of the reduction of $\tilde{\sigma}^2$ is not defined; contrast this with the cases of PCR and PLS, where the bias-variance expressions reveal that the model variance reduces proportionally to the number of modes removed. The choice of $\tilde{\mathbf{U}}$ also limits the subspace in which resulting predictions can reside, leading to the introduction of bias of the form $\mathrm{E}\left[\left\|\mathbf{x}\check{\mathbf{B}}\left(\underline{\mathbf{P}}_{\rightarrow \hat{U}} - \underline{\mathbf{P}}_{\rightarrow \tilde{U}}\right)\right\|^2\right]$. Hence, the aim is to limit $\hat{\mathbf{y}}_{OLS}$ to as small a subspace as possible, while only removing dimensions that are unimportant to the true prediction quantity $\mathbf{x}\check{\mathbf{B}}$, under the assumption that the discarded part of the estimated prediction contributes more towards the model variance

than it does towards the correct prediction. Of course, the true value $\mathbf{x}\check{\mathbf{B}}$ is not known, and so a sensible option is to remove the unimportant dimensions of the estimated quantity $\mathbf{X}\hat{\mathbf{B}}_{OLS}$ instead, which is the approach taken by RRR. CSE on the other hand, bases the choice of $\tilde{\mathbf{U}}$ on the principal components of $\mathbf{Y}$. As shown previously, such modes may not be ably to efficiently capture the behaviour of $\mathbf{x}\check{\mathbf{B}}_{OLS}$. In terms of prediction accuracy, at least, this would suggest that CSE may not be the most appropriate technique.

Regardless of the relative benefits of RRR over CSE, both these techniques are ultimately dependent on how well $\mathbf{x}\hat{\mathbf{B}}_{OLS}$ can be represented in a reduced subspace. $\mathbf{x}\hat{\mathbf{B}}_{OLS}$ resides in the subspace V_{OLS}, which, as explained earlier, is of size $r = \min\left(\text{rank}(\mathbf{X}), \text{rank}(\mathbf{Y})\right) = \min(n, m, p)$. The success of both these techniques therefore requires that $\mathbf{x}\hat{\mathbf{B}}_{OLS}$ can be accurately represented in a subspace of size k, where $k < r$. The mechanisms that allow for a reduction of V_{OLS} can be identified by considering the situation where the information in $\mathbf{x}$ is entirely contained within a subspace of size $k_{\mathbf{x}}$, and the information in $\mathbf{y}$ is similarly contained within a subspace of size $k_{\mathbf{y}}$. Assuming $k_{\mathbf{x}}, k_{\mathbf{y}} < m$, then $\mathbf{x}\hat{\mathbf{B}}_{OLS}$ will be contained within a subspace of size $k = \min(k_{\mathbf{x}}, k_{\mathbf{y}})$. Hence, if $y_{\mathbf{x}}$ and/or $y_{\mathbf{y}}$ is less than r, then the subspace of V_{OLS} can be reduced, providing CSE and RRR with the potential to outperform OLS. This does not represent the complete picture though, as not all information in $\mathbf{x}$ may be relevant to the prediction of $\mathbf{y}$. For example, in the extreme case where the quantities $\mathbf{x}$ and $\mathbf{y}$ are completely unrelated, the sizes of $k_{\mathbf{x}}$ and $k_{\mathbf{y}}$ are rendered irrelevant, as the size of V_{OLS} in this case will be zero. This can be accounted for by redefining $k_{\mathbf{x}}$ as the size of the subspace of $\mathbf{x}$ *relevant to the prediction of* $\mathbf{y}$. In most situations, it is highly unlikely that either $\mathbf{x}$ or $\mathbf{y}$ will reside *exactly* within reduced subspaces, although this analysis is nonetheless a useful guide to identifying potential situations where CSE

and RRR may be successful. In practice, it is more appropriate to consider the following rule-of-thumb:

$\mathbf{x}\hat{\mathbf{B}}_{OLS}$ can be accurately represented in a reduced subspace if:

1. the subspace of $\mathbf{x}$ that contributes to the prediction of y can be accurately represented in a subspace $k < r$; and/or

2. $\mathbf{y}$ can be accurately represented in a subspace $k < r$.

One final aspect of performance to be considered is the inclusion of the unremovable model error term, $\hat{\sigma}^2_\perp \frac{n}{pm}$, which is present whenever $r < p$. This term cannot be removed by either CSE or RRR, and so is likely to place a limit on the amount of model variance that can be removed. Also, it can be seen that the size of the term is proportional to the unexplained variance in the null space of $\mathbf{x}\hat{\mathbf{B}}$, and so is likely to grow as the ratio between r and p increases. Commonly, stochastic estimation is employed to predict a flow field at a large number of points p, given knowledge of the flow at only a few points, n. Because $r = \min(n, m, p)$, this is likely to produce an unfavourable ratio between p and r, and may limit the effectiveness of this approach in this setting.

4.4 Conclusion

The work in this section comprises a comprehensive study of the finite-sample performance of stochastic estimation. This aspect of the technique has rarely been explored previously, and so this work contributes further to the understanding of the field. Furthermore, it is hoped that many of the concepts introduced here will be of significant practical use, which will lead to improvements in the mean-square prediction performance of stochastic estimation in

many situations. The first section concerns the performance of the OLS estimation, which has been shown to be equivalent to stochastic estimation when the prediction model is built from joint observations of the conditional variable and the full set of unconditional variables. The finite sample properties of OLS estimation were presented in a form both relevant and compatible with stochastic estimation, and their implications have been considered. In light of the fact that in many cases, alternative techniques are theoretically able to outperform OLS, it was argued that their use in practice is inappropriate, and that the use of OLS is generally valid in most conceivable applications of stochastic estimation. It was also shown how some of the OLS assumptions can be used to determine whether the chosen prediction model is correctly specified, which may be useful to practitioners when choosing which terms to include in the stochastic estimation model.

The second section concerns the use of biased regression techniques, and their ability to offer potential improvements over the unbiased technique of OLS estimation. Each of the techniques considered involves the projection of some part of the standard OLS model to a reduced subspace, where the subspace is defined by a set of user defined *modes*. Some of the techniques discussed are already established in the field of stochastic estimation, and are the result of combining the technique with principal component analysis. However, such approaches are conventionally used for the analysis of coherent structure, where PCA is used to restrict the prediction model to particular scales within the flow. However, the focus in this chapter is somewhat different, and it has been shown that by selectively discarding parts of the original data, it is possible to produce a more accurate model than one built from the full data set.

Analysis of the biased technique has focused on the trade-off between

model bias and variance; a bias-variance decomposition has been derived for each technique under investigation, which demonstrate the mechanisms by which prediction performance can be improved. Also, the requirements placed on the underlying data have been considered in each case, and have been used to infer how suitable the techniques will be in typical applications of stochastic estimation.

For PCR and PLS, the bias-variance decompositions reveal that the improvement in accuracy is proportional to the number of dimensions removed from the independent variables $\mathbf{x}$. The number of modes that can be removed is dependent on the data; for PCR to be effective, the information in $\mathbf{x}$ must be contained in a small number of dimensions, whereas for PLS, the information relevant to the prediction of $\mathbf{y}$ must be contained in a small number of modes. As the latter case permits a larger number of dimensions to be removed, it would appear that PLS is the more suitable technique, although it should be noted that this conclusion is based on the stated goal of PLS, rather than any theoretical evidence. Unfortunately, the complex, iterative nature of the technique makes it difficult to assess its actual performance.

CSE and RRR act by projecting the OLS estimator $\hat{\mathbf{y}} = \mathbf{x}\hat{\mathbf{B}}_{OLS}$ onto a smaller subspace. Unlike PLS and PCR, the bias-variance decomposition for CSE and RRR does not give a clear relationship between the number of dimensions removed and the level of improvement obtained, merely indicating that an improvement is theoretically possible. It has been shown that any such improvement requires one of two criteria to be met, namely that:

- the data in $\mathbf{x}$ relevant to $\mathbf{y}$ can be accurately contained in a subspace of size k.

- the data in y can be accurately contained in a subspace of size k.

In each case, k must be smaller than the subspace of the original OLS estimator, which has a size $r = \min(n, m, p)$. If either of these criteria is true, then the resulting prediction $\mathbf{x}\hat{\mathbf{B}}_{OLS}$ can accurately be reduced to a k dimensional subspace, and the unexplained variance in the discarded dimensions will no longer contribute to the model error.

Assuming that a reduction of $\mathbf{x}\hat{\mathbf{B}}_{OLS}$ is possible, it appears that RRR may be better suited to doing so than CSE, due to the different choice of subspace used by the techniques. For CSE, the largest principal components of the observed matrix $\mathbf{Y}$ are chosen to define the reduced subspace, whereas RRR uses the modes of the predicted data $\hat{\mathbf{Y}} = \mathbf{X}\hat{\mathbf{B}}_{OLS}$. The performance of both techniques appear to be limited when the number of dependent variables, p is larger than r, which may be an issue in many applications of stochastic estimation, where it is likely that $p >> n$, and hence $p >> r$.

Given the limitations of both CSE and RRR, it would appear that PCR and PLS offer a more effective means of improving the predictions in stochastic estimation applications. Therefore, the evidence suggests that the best performance can be obtained by PLS, followed by PCR, RRR then CSE. Of course, this conclusion is based purely on the theoretical work of this chapter. As already mentioned, this is by no means a complete description of the techniques' performance, and involves some simplifying assumptions. In order to confirm these conclusions, and address any omissions, an empirical study of the practical performance of the techniques has been performed, which is detailed in the following chapter.

Chapter 5

Quantitative Accuracy Analysis of Regression Techniques

5.1 Introduction

In the following work, the quantitative performance of the regression techniques from the previous chapter are assessed in a series of simulated experiments. This will hopefully bring further understanding to the techniques under consideration, provide evidence for the bias-variance decompositions derived in section 4.3, and serve to identify situations when a particular technique may be effective.

This chapter also introduces a rigorous methodology for assessing the practical performance of a regression technique, which can be used to obtain the optimum number of modes k to build a biased regression model, and ultimately select the most appropriate regression technique for the application in question.

All the following tests were performed on data obtained from a direct numerical simulation of a channel flow, which is described section 5.5. Using this data, a number of simulated stochastic estimation experiments were created, each of which is designed to investigate a particular aspect of the regression procedure that dictates the accuracy of resulting predictions.

The process of obtaining a reliable, quantitative measure of a regression model's performance is commonly referred to as *model validation*. The chapter begins with a discussion on this topic, before going on to discuss the practical aspects of implementing the validation procedure for the current work. The chapter ends with the presentation and discussion of the validation results from the various test scenarios.

5.2   Model Validation

The requirements for successful model validation are two-fold: first, it is important to choose a suitable metric with which to judge performance, and secondly, the validation must be carried out in a manner that ensures the resulting metric is representative of the regression model's behaviour in general, not just for the data it was derived from.

As the focus of this work is prediction, the obvious measure of a regression model's performance is the mean square error (MSE) of prediction. However, the value of the MSE cannot be interpreted without knowledge of the underlying data that is being predicted, which makes it difficult to compare performance between different data sets. Therefore, it is useful to consider the error in the more interpretable and generalisable form of *unexplained* variance. The unexplained variance of prediction is defined as:

$$\frac{\text{MSE}[\hat{y}]}{\text{Var}[y]} \times 100\% \tag{5.1}$$

Although this metric should technically range from 0% (when there is no prediction error), to 100% (when the regression model has no predictive power over y), in practice, the regression model error can be so severe that the resulting MSE is actually higher than the variance of y, resulting in an unexplained variance of $> 100\%$.

A closely related statistic is the *explained* variance, which is simply:

$$\left(1 - \frac{\text{MSE}[\hat{y}]}{\text{Var}[y]}\right) \times 100\% \tag{5.2}$$

The percentage explained variance is equivalent to the *coefficient of determination*, R^2, which is a commonly used measure in linear regression.

Again, it is possible to produce a percentage explained variance of less than 0%.

In the following sections, particular emphasis is placed on the performance of the biased techniques relative to that of OLS. Therefore, a final measure will be introduced, which demonstrates the improvement in prediction error of a biased regression technique over that of standard OLS regression. This is given as a percentage, defined as:

$$\frac{\text{MSE}[\hat{y}_{OLS}] - \text{MSE}[\hat{y}_{biased}]}{\text{MSE}[\hat{y}_{OLS}]} \times 100\% \tag{5.3}$$

In order to compare the relative performance of the regression techniques, it is first necessary to determine the optimum number of modes, k, for each biased regression technique. From the previous chapter, it was demonstrated that all the biased approaches are subject to the bias-variance tradeoff, where any improvement in model variance is accompanied by a subsequent increase in bias. This is shown schematically in figure 5.1. For prediction purposes, the optimum choice of k is that which yields the minimum prediction error, which can be obtained in practice by building biased regression models for every possible k, and then choosing the model with the lowest prediction error. An example result from this procedure is shown in figure 5.2. The real example has similarities to the theoretical case, with the exception that the theoretical case considers only the model variance and bias, while the example shows the full MSE (which consists of the model variance, model bias and the irreducible variance). Although not apparent from figure 5.2, the unexplained variance reaches 100% when $k = 0$, which is the case for all the biased techniques under consideration here. Conversely, when the full set of modes is used, all the biased techniques are equivalent to OLS regression.

113
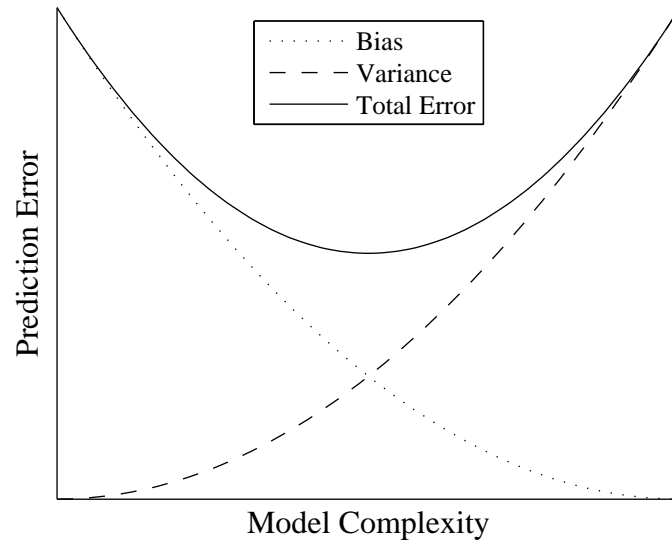
Figure 5.1: Schematic representation of the bias-variance tradeoff (taken from Faber (1999))
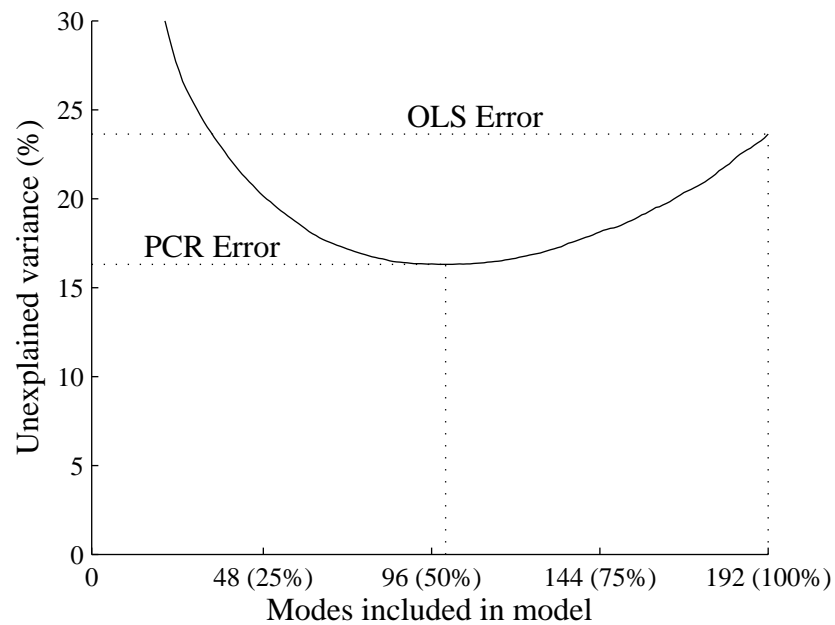
Figure 5.2: Example plot of prediction error v.s number of modes, indicating the potential improvement over OLS regression
Data acquired from the validation of PCR regression, test scenario 1, sample size = 350 (see section 5.6.2)

The process of model validation is a complex subject, which has received a great deal of attention in literature. It should be noted there are many alternative, and potentially superior, approaches to the one outlined here, although their complexity places them beyond the scope of this work. A more detailed discussion of the subject is given in Harrell Jr. (2001).

As explained previously, it is important that the chosen approach to validation produces a *representative* performance estimate. That is, it must accurately reflect the performance of future predictions made by the model. This rules out the option of validating a regression model on the same data set used to create the model, as it yields considerably biased results that will invariably underestimate the prediction error. Also, such an approach is particularly inappropriate in the current context because the OLS estimator, by design, will produce the estimate that minimises the error of prediction for the data set it is built from. Furthermore, for a biased technique, the value of k that minimises the prediction error will automatically default to the maximum number of modes, which corresponds to the OLS estimator. Therefore, it is vital that the validation is carried out on data that is independent from the sample used to build the model.

For the purpose of this work, a simple form of *k-fold cross-validation* is proposed. This is an iterative approach that involves partitioning the available data into k sections. At each iteration, a different partition is removed from the full data set. The removed partition becomes the *test set*, which is used to assess the prediction performance of the model built from the remaining $k-1$ partitions (the *training set*). This is carried out for a total of k iterations, each time using a different partition for the test set. The resulting k validations are then averaged to produce the final error measure.

The most appropriate choice of k is, again, a complex matter. In prac-

115

tice however, *10-fold* cross validation is often used as standard, and is adopted for this work, as well for subsequent validations in later chapters.

5.3 Practical Implementation of the Regression and Cross-Validation Procedures

The process of 10-fold cross-validation requires each regression model to be estimated ten times. Furthermore, for each biased regression technique, a model must be built for every possible number of modes, k at each iteration of the cross-validation procedure. The maximum value of k, and hence the total number of regression models, varies with technique. For PCR, PLS and RRR, the maximum number of modes is determined by the rank of $\mathbf{X}$, which is at most $\min(n, m)$, and for CSE, it is determined by the rank of $\mathbf{Y}$, which is $\min(p, m)$. In practice then, the total number of individual regression models required to validate a technique can be very high; in the current work, this number can reach $5{,}000$. Fortunately, with the use of optimised software and a number of algorithmic improvements, a daunting number of regressions does not necessarily translate into a time consuming or resource intensive computation.

In the current work, the cross-validation procedure was performed on a standard 32-bit Windows XP desktop computer, equipped with a 3GHz Pentium processor and 1GB RAM, and the task of fully validating all five of the regression techniques was accomplished in the order of minutes for all the scenarios considered in this chapter.

The validation software was implemented using MATLAB 7.4. The majority of the code consists of matrix operations, which makes the use of

MATLAB particularly suited to this task, as it provides access to efficient and optimised linear algebra libraries (in the form of the LAPACK library and an architecture optimised implementation of the BLAS library). The actual implementation of the regression techniques differs slightly from their descriptions given in the previous chapters. In all the following test scenarios, the dependent variable $\mathbf{y}$ is multivariate, and so the multivariate forms of OLS and PCR are used (see section 4.1). Using the multivariate forms, much of the computational workload involved is reduced to a series of matrix multiplications, which are able to make use of the optimised linear algebra libraries. A more fundamental change relates to the calculation of $\hat{\mathbf{B}}_{OLS}$, which has been described thus far as $\hat{\mathbf{B}}_{OLS} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$. While theoretically correct, the direct implementation of this expression can be numerically unstable (Trefethen and D Bau, 1997), and so $\hat{\mathbf{B}}_{OLS}$ is instead solved using the pseudoinverse of $\mathbf{X}$:

$$\hat{\mathbf{B}}_{OLS} = \mathbf{X}^+\mathbf{Y} \tag{5.4}$$

where the pseudoinverse is calculated by means of the SVD of $\mathbf{X}$:

$$\mathbf{X}^+ = \mathbf{V}_R\mathbf{S}^{-1}\mathbf{U}_L^T \tag{5.5}$$

Here, $\mathbf{S}$ is the matrix of singular values of $\mathbf{X}$, and $\mathbf{V}_R$ and $\mathbf{U}_L$ are the corresponding left and right singular vectors. A further advantage of this approach is that the pseudoinverse exists for any rectangular matrix, which permits the calculation of $\hat{\mathbf{B}}_{OLS}$ even when there are fewer observations than variables, $m < n$. It should be noted however, that this violates the first OLS assumption (section 4.1.2), which means that the corresponding properties do not hold. Also, the bias-variance decomposition for the OLS predictor, which

117

was derived using this assumption, is invalid when $m < n$. An alternative derivation for this situation is provided in section 5.4.

Finally, the implementation of PCA (required in the calculation of the PCR, RRR and CSE models) is performed using the SVD method (see section 3.4) wherever possible[1]. Again, this is due to potential instabilities in the direct implementation either the "standard" or snapshot method of PCA (Trefethen and D Bau, 1997).

The computational complexity of the regression techniques (excluding PLS, which will be discussed separately) is dominated by two operations; the calculation of $\hat{\mathbf{B}}_{OLS}$, and the calculation of the principal components. However, there are some simple modifications that allow these operations to be kept to a minimum. Firstly, in the process of validating a biased technique for every possible value of k, the same set of principal components are used each time. Therefore, for a given biased technique the PCA need only be performed once for each iteration of the cross-validation. Also, once $\hat{\mathbf{B}}_{OLS}$ has been calculated, it can be reused by each biased regression technique, and only has to be calculated once for each iteration of the cross-validation.

The case of PLS differs slightly from the other techniques, due to the iterative nature in which it produces the regression model. If a particular k-mode estimate is sought, then the PLS algorithm must go though k iterations to find it. This in itself is not a problem, and can be exploited for the purposes of cross-validation simply by setting the algorithm to iterate through the full set of modes, and storing the current estimate of $\mathbf{B}$ at the end of every iter-

---

[1] The computational cost of the SVD is $\min[\mathrm{O}(mn^2), \mathrm{O}(nm^2)]$ floating point operations, or *flops*, as opposed to $\mathrm{O}(n^3)$ required for the traditional PCR approach (based on the cost of computing the EVD of $\mathrm{X}^T\mathrm{X}$), and $\mathrm{O}(m^3)$ for the method of snapshots (based on the EVD of XX^T). Hence, when $n >> m$ or $m >> n$, the use of SVD becomes inefficient, and the most appropriate EVD-based approach is used instead

ation. Therefore, at the end of the procedure, PLS models for all values of k are available, and so the algorithm only needs to be run once for each iteration of the cross-validation. Nonetheless, the computational cost of validating PLS is still much higher than for the other techniques, due to the fact that the SVD of either the covariance matrix $\mathbf{Y}^T\mathbf{X}$, or the deflated covariance matrix $\mathbf{Y}^T\mathbf{X}\underset{\rightarrow}{\mathbf{P}}_{V_{\hat{\mathbf{P}}_\perp}}$, must be calculated for each iteration. Moreover, the computational cost of computing the SVD of these matrices can often be substantial. For an $m \times n$ matrix $\mathbf{X}$ and a $m \times p$ matrix $\mathbf{Y}$, the size of the covariance matrix is $p \times n$, which may be difficult, or even impossible to solve in some cases. It is shown in appendix D.1 how the SVD of a $p \times n$ covariance matrix of the form $\mathbf{Y}^T\mathbf{X}$ can be reduced to a series of EVDs of an $m \times m$ matrix, which can be significantly more efficient than the direct approach when m is smaller than n and p. This process is then generalised to account for the case of the deflated covariance matrix $\mathbf{Y}^T\mathbf{X}\underset{\rightarrow}{\mathbf{P}}_{V_{\hat{\mathbf{P}}_\perp}}$.

5.4 MSE of OLS behaviour for $m < n$

In several of the conditions considered in this chapter, the sample size m, is smaller than the number of independent variables n. A prediction can in this case still be made using OLS by using the pseudo-inverse approach described above, although the theoretical expression for the MSE of the OLS predictor must be refined:

The MSE for $m \geq n$ is:

$$\text{MSE}\left[\hat{y}\right] = \sigma^2 + \frac{n}{m}\sigma^2 \tag{5.6}$$

Recall from the derivation of this expression (esp. equation (4.70)) that the

119

value n corresponds to the number of principal components present in the matrix $\mathbf{X}$, which is determined by the number of independent variables, m when $m \geq n$. In cases where $m < n$, the number of principal components is limited by the number of samples, m. Hence, when $m < n$, the model variance term becomes:

$$\frac{m}{m}\sigma^2 = 2\sigma^2 \tag{5.7}$$

However, for future predictions of $\mathbf{y}$ from $\mathbf{x}$, the omitted principal components act to remove a subspace from $\mathbf{x}$ that can no longer contribute towards the prediction of $\mathbf{y}$. This introduces a bias term into the MSE expression. Essentially, OLS is behaving in an identical manner to PCR in this situation, so the bias term can be taken from the bias-variance decomposition for PCR (4.79), leading to:

$$\mathrm{MSE}\left[\hat{y}\right] = 2\sigma^2 + \mathrm{E}\left[\left(\mathbf{x}\underrightarrow{\mathbf{P}}_{\tilde{V}_\perp}\check{\mathbf{b}}\right)^2\right] \tag{5.8}$$

In order to generalise this expression for any m, the quantity $q = \min(m, n)$ is introduced, giving:

$$\mathrm{MSE}\left[\hat{y}\right] = \sigma^2 + \frac{q}{m}\sigma^2 + \mathrm{E}\left[\left(\mathbf{x}\underrightarrow{\mathbf{P}}_{\tilde{V}_\perp}\check{\mathbf{b}}\right)^2\right] \tag{5.9}$$

where $\mathbf{P}_{\tilde{V}_\perp}$ is the projection onto the subspace defined by the omitted principal components.

5.5   Overview of Data Set

The data used in this chapter, and subsequently in chapter 6, is taken from a direct numerical simulation (DNS) of a turbulent channel flow subject to wall suction and blowing, which was provided courtesy of Dr Yongmann Chung
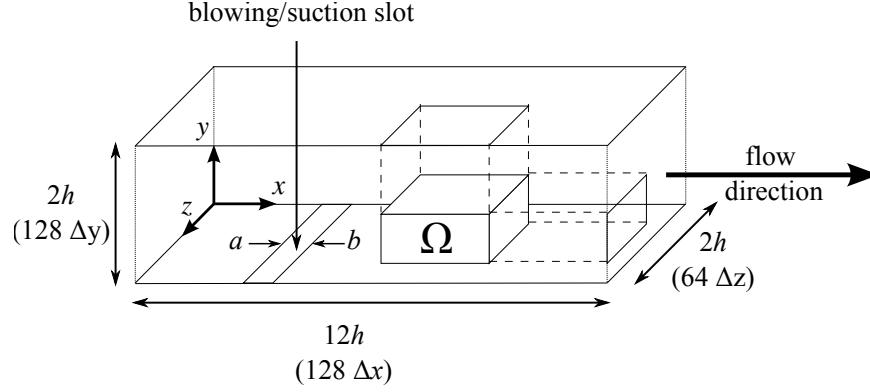
Figure 5.3: Flow configuration, showing the approximate position of the domain Ω, which provides the data used in this work.

and Tariq Talha of the University Warwick. The flow configuration consists of a 3-dimensional domain, as illustrated in figure 5.3. The domain is bounded at the top and bottom with impermeable wall conditions, except for a slot on the lower wall that runs the entire spanwise direction z, which provides the blowing and suction. Periodic boundary conditions are applied in the spanwise direction, and the inflow boundary employs instantaneous data from a separate, identically sized simulation of a fully developed turbulent channel flow. A continuous, unvarying wall blowing and suction is applied through the spanwise slot by imposing a sinusoidal wall-normal velocity profile along the streamwise direction of the slot, as demonstrated in figure 5.4. The remaining velocity components are set to zero. The flow variables are nondimensionalised using the streamwise mean velocity U_m and the channel half-height h. The Reynolds number of the simulation is 5600, which is defined by $Re = U_m h/\nu$, where ν is the kinematic viscosity. The flow is homogeneous in the spanwise dimension z, inhomogeneous in y, and the presence of the blowing/suction introduces inhomogeneity into the streamwise direction x. The $x \times y \times z$ dimensions of the flow domain in terms of the channel half height h are $12 \times$
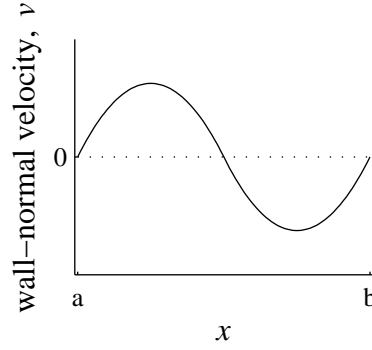
Figure 5.4: Blowing/Suction profile imposed along the x direction of slot.

2×2. The grid resolution is $128 \times 128 \times 64$ The grid spacing is uniform in x and y, with a spacing of $\Delta x = 0.1h$ and $\Delta z = 0.031h$ The grid in y is generated using a hyperbolic tangent function, which produces a dense grid close to the walls of the domain, with a coarser grid in the centre of the channel.

The data used in this thesis is taken from a $32 \times 32 \times 32$ subregion Ω, which is located downstream from the blowing/suction slot. The position and size has been chosen to capture the lower boundary layer of the flow. A total of 250 time-uncorrelated instances of Ω were available, although for the majority of the work, only 2-dimensional $x - y$ slices of the data were required. Due to the homogeneity in z, it was possible to extract multiple planes from each instance of omega; analysis of the data revealed that up to 3 planes could be taken from each instance, while still ensuring enough spatial separation to ensure that the individual planes were uncorrelated. For the work in this chapter, 600 slices were made available for the quantitative testing. The details of the data used in later work is provided in chapter 6.

5.5.1 Note Regarding Errors due to the Misrepresentation of the CFD Grid Spacing

Unfortunately, all the processing and analysis of the simulated data in this thesis was performed under the erroneous assumption that the grid was uniform and equal in all dimensions, i.e. $\Delta x = \Delta y = \Delta z$. As such, the data has been represented on a grid with incorrect spacing, as illustrated in figure 5.5. The depiction in 5.5 (a) is only an approximation to the actual grid, however; currently the exact parameters of the inverse hyperbolic function are unknown, and so the true grid cannot be replicated.

This issue was only discovered at a late stage in the work, and there was not sufficient time to rectify the problem in either the calculations or in their subsequent presentation and analysis. It should be stressed that for the majority of the work, which involves the quantitative comparison of the true data with predicted data, this is of little concern. Such calculations are independent of the spatial distribution of the data, and so are valid regardless of whether the true grid is known. However, this issue should be borne in mind in the following chapters, where the spatial locations of the velocity vectors used in the calculations are presented on the incorrect grid, as demonstrated in 5.5.

The only major concern arises in chapter 6, where a qualitative analysis of the MP-VSE procedure is made using the simulated data. In this section, MP-VSE is used to predict instances of the full volume Ω, and the predicted vorticity fields are compared with the true values. The calculated vorticity data, which is a function of grid spacing, is therefore incorrect. Nonetheless, the emphasis of the section is on the relative behaviour of the true and predicted flow field, and even though the absolute calculations are wrong, the

(a) Representation of true grid

(b) The same volume, represented on the grid used in the following work
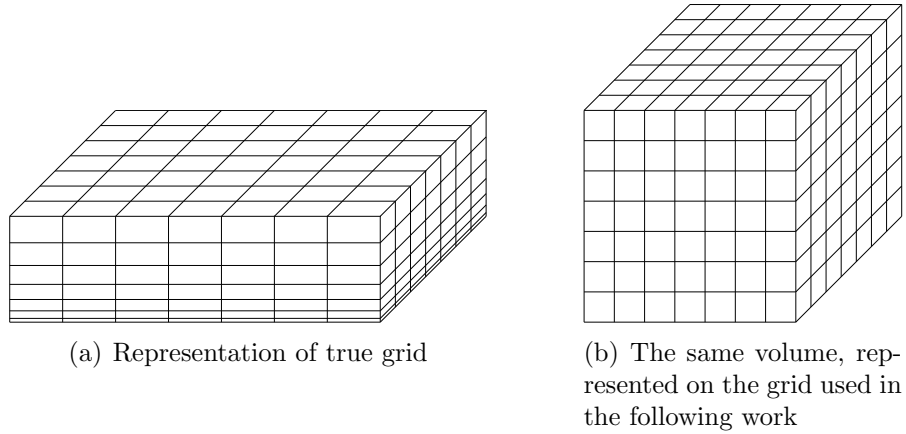
Figure 5.5: Comparison of true grid and erroneous grid spacing

problem manifests itself identically for both the true and predicted data. As such, the incorrect vorticity data is still useful as a means of assessing the ability of MP-VSE to predict derivatives of the velocity field.

Although the use of the incorrect grid should not affect the conclusions of any of the work presented in the thesis, the correction of the issues discussed here is currently being pursued as a matter of priority.

5.6 Performance of Regression Techniques: Results

5.6.1 Overview

The following sections describe the performance of the five regression techniques under investigation for each of the test scenarios. In each scenario, one test parameter is varied, with the remaining held fixed, yielding a set of individual *test conditions* for each scenario. The accuracy of each technique was determined for each test condition through the process of cross-validation.

The four test scenarios are:

- **Test Scenario 1** - Number of observations, m, used to produce the regression model.

- **Test Scenario 2** - Number of independent variables, n.

- **Test Scenario 3** - Multicollinearity in the independent variables.

- **Test Scenario 4** - Number of dependent variables, p.

Scenarios 1, 2 and 4 are self-explanatory, although scenario 3 requires further elaboration. Collinearity refers to the presence of correlation between two variables, and multicollinearity is simply the generalisation of this concept, to account for cases with more than 2 variables. The presence, and degree of, multicollinearity in a data set plays an important role in the performance of all the biased techniques considered here, as it dictates the intrinsic dimensionality of a data set; for example, if the variables are all highly correlated, then it is likely that the data can be accurately modelled with only a small number of modes. As described in chapter 4, this is a necessary requirement for the success of each of the biased techniques. In test scenario 3, the multicollinearity is controlled by varying the distance between the locations of the independent variables. This is motivated by the assumption that as the spacing becomes smaller, the behaviour of the independent variables will become more similar, hence increasing the multicollinearity.

For each test condition, a stochastic estimation problem was constructed by defining a set of vectors within the 32×32 flow field to act as the conditional and unconditional data (as shown in figure 5.6, for example). To the extent possible, these conditional and unconditional vectors were chosen to provide a realistic stochastic estimation procedure in each test condition.
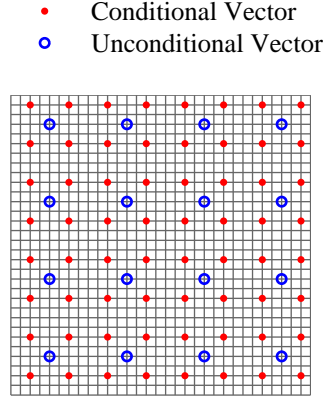
Figure 5.6: Example vector placement within flow domain to provide conditional and unconditional data

The unconditional and conditional vectors were used to create the independent variables $\mathbf{x}$ and dependent variables $\mathbf{y}$, of the linear model $\mathbf{y} = \mathbf{x}\mathbf{B} + \mathbf{e}$. In the data set used, each vector consists of 3 velocity components, so the resulting sizes of $\mathbf{x}$ and $\mathbf{y}$ (i.e. n and p) are 3 times the number of vectors used to build them.

For the purposes of cross-validation, the 600 frames were split into 10 partitions of equal size. Thus, at each iteration of the cross-validation, 60 samples were available for the test set and a total of 540 samples could be used for the training set. However, the actual size of the training set was usually limited to 500 for simplicity, except in test scenario 1, where the number of samples was varied from 50 to 500. For a given sample size, care was taken to ensure that the samples were selected as equally as possible from each of the nine partitions that made up the available data.

In many of the test scenarios, the position of the conditional and unconditional vectors overlapped, which meant that the resulting regressions were predicting vectors whose values were already known. This in itself is not a

126
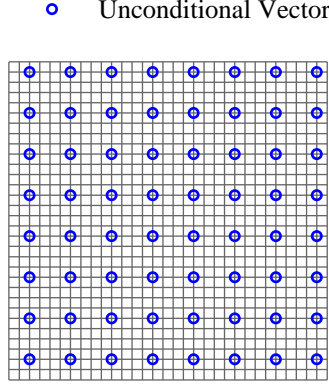
Figure 5.7: Unconditional vector placement within flow domain

problem, although it has the effect of artificially decreasing the overall MSE. In these situations, the prediction was performed with the overlapping vectors included in the data, but they were subsequently masked out before the MSE was calculated, in order to provide a more representative error measure.

In all the scenarios, a standard linear stochastic estimation model was employed, i.e. only the first order terms of the independent variable were included. The use of higher-order terms was investigated (see appendix D.2), but it was concluded that the linear terms were sufficient to correctly specify the model, and no benefit was to be had from including further terms.

5.6.2   Test Scenario 1: Effect of Sample Size, m

For the first scenario, the sample size, m, used to build the regression models was varied from 50 to 500, in steps of 50. Here, 64 uniformly spaced unconditional vectors were used to predict the full 32×32 flow field (see figure 5.7). A summary of the parameter values is given in table 5.1, the optimum number of modes is listed in table 5.2 and the results are illustrated in figures 5.8 and 5.9.

Parameter	Value
Sample size, m	$50, 100, 150, \ldots, 500$
Unconditional Vectors	64
Independent Variables, n	192
Conditional Vectors	1024
Dependent Variables, p	3072

Table 5.1: Summary of parameter values for Scenario 1
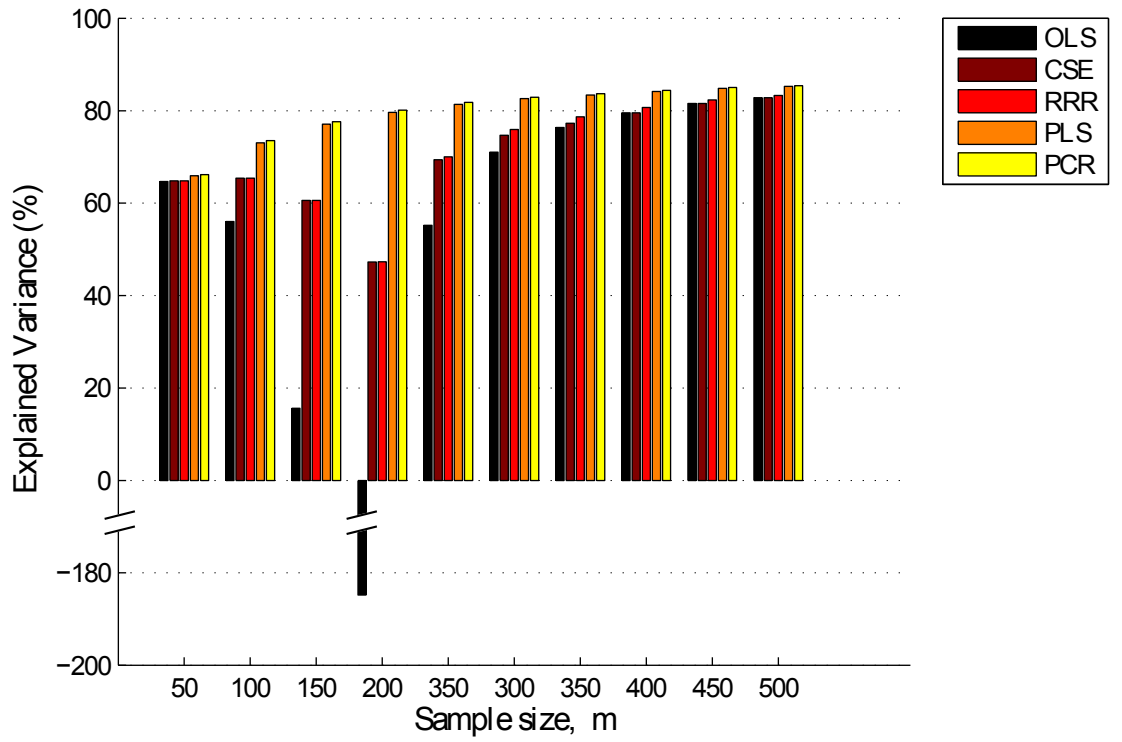
Figure 5.8: Average explained variance vs. sample size
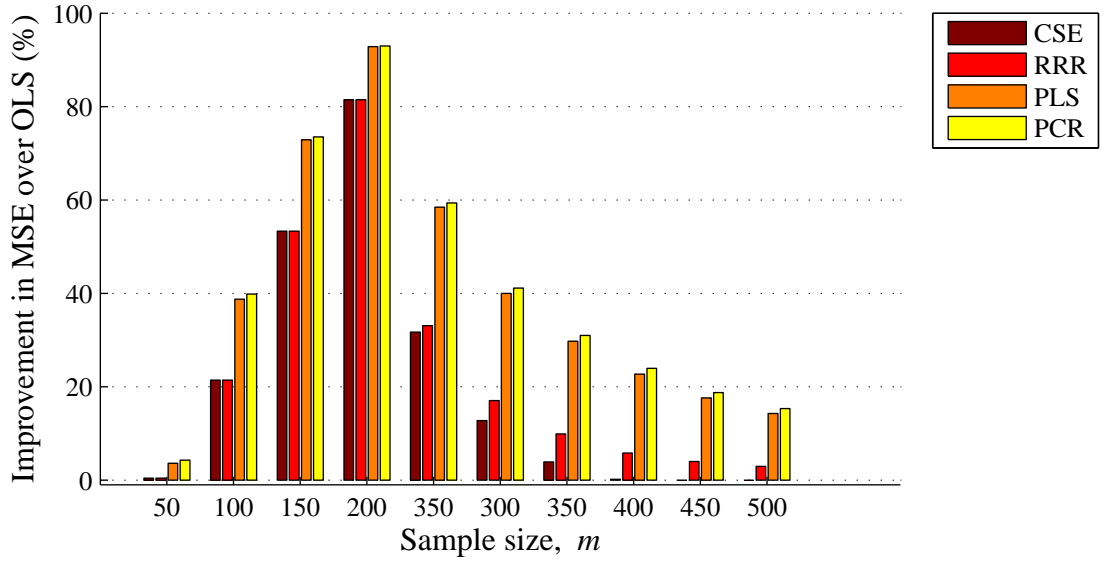
Figure 5.9: Percentage improvement over OLS prediction vs. sample size

Samples	CSE	RRR	PLS	PCR
50	45 (*50*)	45 (*50*)	38 (*50*)	39 (*50*)
100	37 (*100*)	37 (*100*)	54 (*100*)	54 (*100*)
150	22 (*150*)	22 (*150*)	66 (*150*)	66 (*150*)
200	10 (*200*)	10 (*192*)	75 (*192*)	78 (*192*)
250	36 (*250*)	36 (*192*)	86 (*192*)	93 (*192*)
300	56 (*300*)	65 (*192*)	92 (*192*)	92 (*192*)
350	83 (*350*)	80 (*192*)	98 (*192*)	101 (*192*)
400	126 (*400*)	97 (*192*)	102 (*192*)	107 (*192*)
450	500 (*450*)	122 (*192*)	104 (*192*)	108 (*192*)
500	500 (*500*)	117 (*192*)	103 (*192*)	113 (*192*)

Table 5.2: Optimum number of modes for differing sample sizes (*the quantity in brackets denotes the maximum permissible number of modes for the given case*).

Discussion

The results show a number of interesting, and unexpected, phenomena, which are best explained by considering the particular case of the OLS predictions. From section 5.4, the MSE of OLS has been derived as:

$$\text{MSE}\left[\hat{y}\right] = \sigma^2 + \frac{q}{m}\sigma^2 + \text{E}\left[\left(\mathbf{x}\underrightarrow{\mathbf{P}}_{\tilde{V}_\perp}\check{\mathbf{b}}\right)^2\right] \tag{5.10}$$

where $q = \min(n, m)$. This suggests that when $m \geq n$, the prediction accuracy will decrease with m, due to the increase in model variance. For $m < n$, the model variance will remain constant, but the model bias is likely to increase with decreasing m, and so the accuracy will continue to fall. This theoretical behaviour is demonstrated in figure 5.10, which has been based on equation (5.10). An estimate of the irreducible variance, σ^2 was produced from the convergence of the results in figure 5.8. The bias, which cannot easily be estimated, has been chosen under the assumption that the largest principal components of $\mathbf{x}$ are more important to the prediction of $\mathbf{y}$ than the smallest. Also shown in figure 5.10 are the results of the OLS prediction. Clearly, there is a large discrepancy between the theoretical and actual behaviour. Starting at $m = 50$, the results show that with increasing m, the accuracy initially falls, to the point where $m = 200$ (where $m \approx n$), and then begins to rise again, with the rate of growth slowing as m increases. Also, the theoretical value consistently underestimates the true error.

The discrepancy between the theoretical and true behaviour is likely to arise from assumptions used in the construction of the MSE decomposition. The problem appears to lie with the assumption that the empirical roots and principal components are identical to those of the underlying population. This results in a situation where each principal component contributes an

130
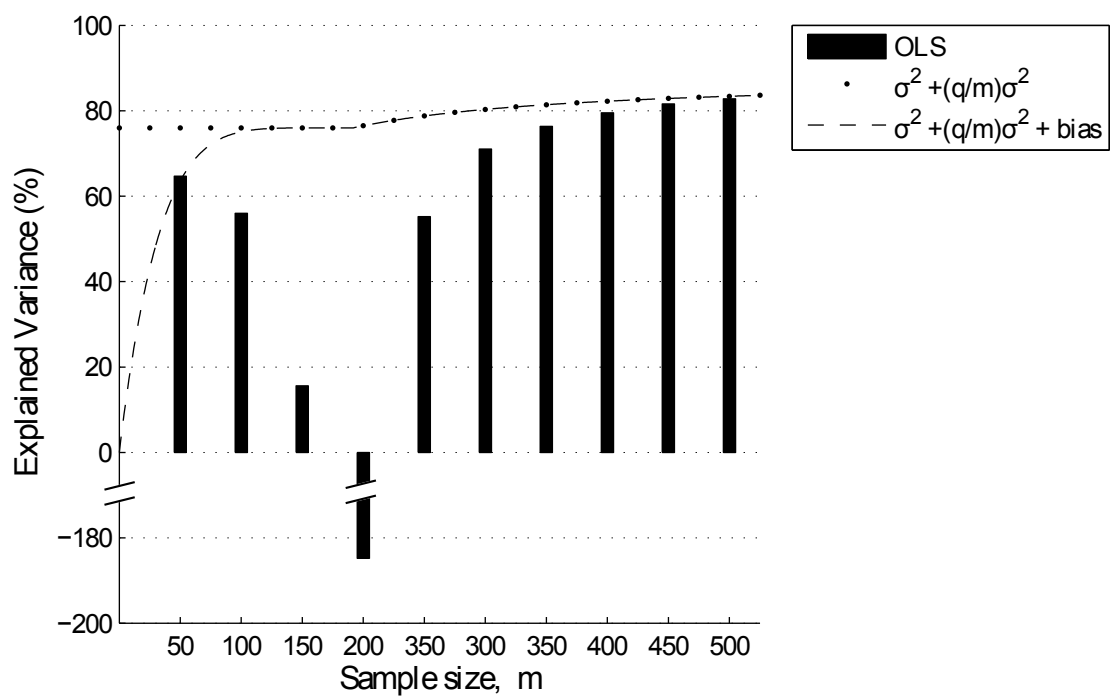
Figure 5.10: Percentage improvement over OLS prediction vs. sample size

equal amount to the overall MSE (equations (4.68) to (4.71), section 4.3). In reality, the empirically calculated roots and principal components are subject to a degree of estimation error, which becomes larger as the sample size, m, decreases. As discussed in appendix D.3), it appears that this error causes the smallest principal components of $\mathbf{X}$ to contribute a disproportionate amount to the MSE. This contribution can be extremely large when $n \approx m$, to the point where a few small principal components act to vastly inflate the MSE, far beyond what the theoretical expression suggests. This is observed in the case where $m = 200$. At this point, the resulting MSE is larger than the total variance of $\mathbf{y}$, resulting in a negative explained variance.

However, when $m < n$, the rank deficiency of $\mathbf{X}^T\mathbf{X}$ effectively removes the smallest principal components from the model, which would otherwise cause the inflation of the MSE. As m decreases, more principal components are removed, and the model variance reduces. Again, further explanation is provided in appendix D.3, along with empirical evidence to support this theory.

This line of reasoning is helpful to explain the performance of the remaining techniques. PCR acts by removing the smallest principal components from $\mathbf{x}$, and so directly combats the source of the problem. PLS also removes dimensions from $\mathbf{x}$, and although the selection criteria is more complex than PCR, the fact that the performance of PLS and PCR is roughly equal suggests that it is just as capable of avoiding the problems caused by the smallest principal components. As a result, these two techniques do not appear to be affected by this phenomena at all, at least in this scenario. CSE and RRR, however, perform a decomposition on $\mathbf{x}\hat{\mathbf{B}}_{OLS}$, rather than $\mathbf{x}$, and so are unable to act directly on the underlying cause of the MSE inflation. This can be observed in the fact that their performance, while often superior to OLS, still follows the same pattern shown by the OLS predictions. On the whole RRR is

more accurate than CSE, although their performance is roughly equal in the first four cases. This is likely due to the fact that RRR performs an optimal decomposition of $\mathbf{x}\hat{\mathbf{B}}$, while CSE does not (CSE instead uses the principal components of $\mathbf{Y}$ to decompose $\mathbf{x}\hat{\mathbf{B}}$).

Care must be taken when attempting to generalise the results here. On the one hand, the increase in estimation error in the principal components due to decreasing m appears to be independent of the underlying data, and so the poor performance of OLS seems certain whenever $m \approx n$. What is more debatable is whether biased techniques will always offer a suitable solution, as is the case here. In this scenario, it is probable that the multicollinearity in $\mathbf{x}$ ensures that the small principal components are unimportant to the prediction, and can be removed without incurring a large bias penalty. However, there may be situations where the small dimensions of $\mathbf{x}$ are important for the prediction. Here, the removal of these components will prevent the aforementioned catastrophic error due to a reduction in model error, but the resulting prediction will still be poor due to the large bias error. In these cases, the only real solution will be to ensure that enough data is acquired to prevent the problem in the first place.

A final observation concerns the number of modes used to build the optimum biased models, which can be seen to vary considerably for each technique over the range of test cases, (see table 5.2). This highlights an aspect of biased techniques' performance that has yet to be considered in detail. Firstly, it is noted that all the biased techniques correspond to the OLS model when the full set of modes are used; essentially, the biased techniques act by removing modes from the OLS model in order to improve its stability. In the theoretical study of the biased techniques of section 4.3, the ability of all the techniques was attributed to the ability to accurately represent the behaviour

of a multivariate quantity within a reduced subspace, i.e. the reduction of $\mathbf{x}$ in the case of PCR and PLS, and $\mathbf{x}\hat{\mathbf{B}}_{OLS}$ for CSE and RRR. This is certainly an important criteria, as it allows modes to be removed from the OLS regression model without incurring a large bias penalty. However any improvement over OLS ultimately requires that any increase in bias is smaller than the reduction in model variance. As such, the level of model variance in the OLS predictions is also of importance to the performance of the biased techniques. This effect can be observed in the results of this scenario. Each condition considered here uses the same $\mathbf{x}$ and $\mathbf{y}$, which suggests that the distribution of data in the relevant subspaces will remain roughly the same for all conditions. Hence, for a given technique, the amount of bias introduced by removing modes from the OLS model should also be similar across the conditions. Furthermore, the unexplained variance also remains constant, which implies that the difference in the performance of the biased techniques is purely due to changes in model variance of the OLS estimator. Inspection of the optimum modes used by the biased techniques shows that, in general, the number of modes removed by the optimum model is correlated with the model variance of the OLS estimator. Furthermore, the reduction in the number of modes is also directly linked to the relative improvement of all the biased techniques over OLS. This shows, all else being equal, as the model variance of OLS increases, the benefit of discarding a given mode from a biased regression becomes larger. This is certainly the case here when $m \geq n$, although the situation is slightly more complicated for $m < n$, where the effect manifests itself differently for the regression techniques considered here. The number of modes used by PLS and PCR appears to be roughly proportional to m, while for CSE and RRR, whose behaviour is closely coupled with the performance of OLS, the number of modes is proportional to the accuracy of the OLS prediction. Nonetheless,

134

Parameter	Value
Sample size, m	500
Unconditional Vectors	4,16,64, 256
Independent Variables, n	12,48,192,768
Conditional Vectors	1024
Dependent Variables, p	3072

Table 5.3: Summary of parameter values for Scenario 2

the results still serve to show that a large model variance in the OLS model is just as important to the performance of the biased techniques as the loosely defined concept of "accurately representing the behaviour of a multivariate quantity within a reduced subspace".

5.6.3 Test Scenario 2: Effect of the number of independent variables, n

The second scenario investigates the effect that the number of independent variables, n has on the performance of the regression techniques. The number of independent variables was controlled by varying the number of unconditional vectors, as illustrated in figure 5.11. For each test condition, the regression models were built from a sample size, m of 500, and the entire 32×32 flow field was predicted each time. A summary of the parameters for this scenario is given in table 5.3, while the optimum number of modes chosen to build the biased models is given in table 5.4. The resulting accuracy of the regression techniques, and the biased technique's improvement over OLS is shown in figures 5.12 and 5.13, respectively.
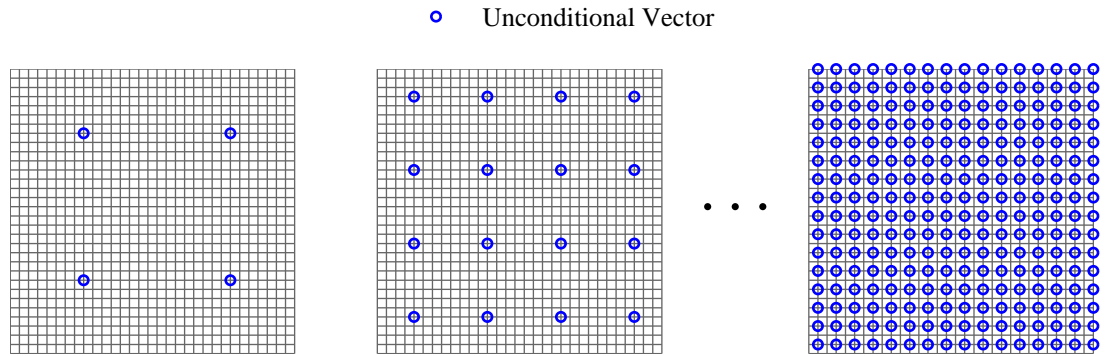
Figure 5.11: Unconditional vector placement within flow domain, for $4, 16, \ldots, 254$ vectors
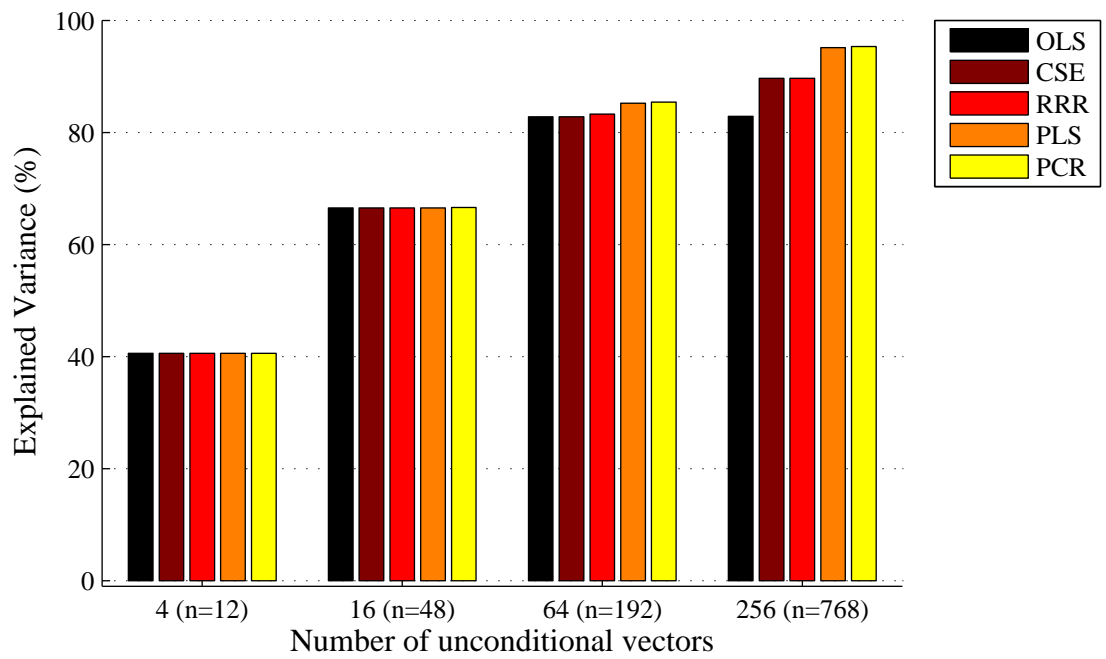
Figure 5.12: Average explained variance vs. number of unconditional vectors
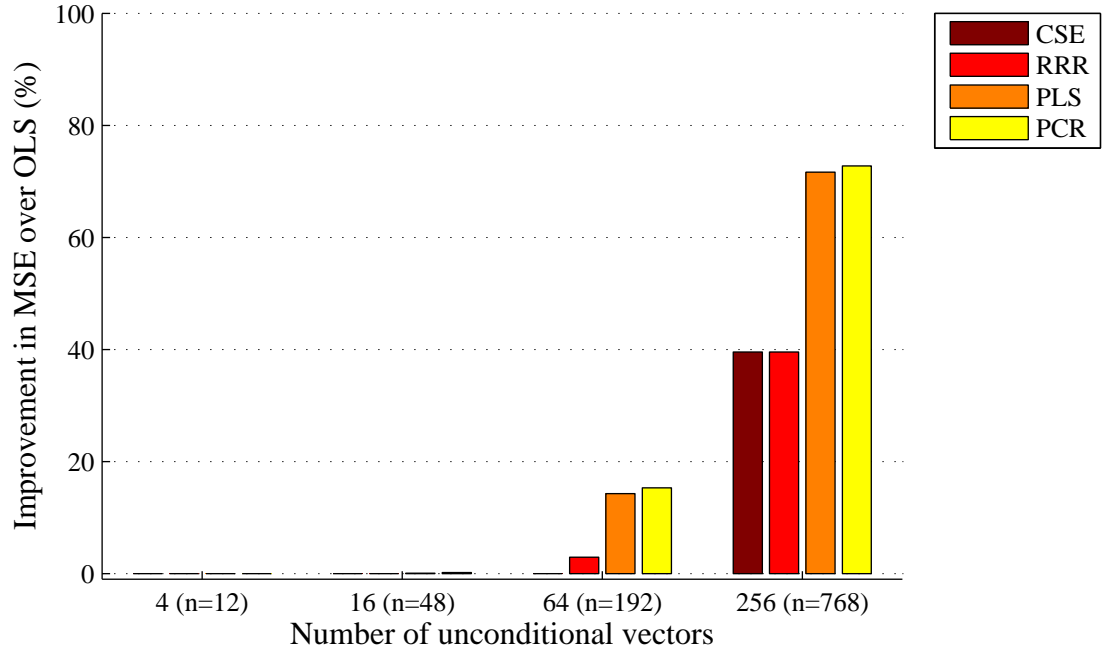
Figure 5.13: Percentage improvement of different biased techniques over OLS prediction vs. number of unconditional vectors

Vectors	CSE	RRR	PLS	PCR
4	500 (*500*)	12 (*12*)	12 (*12*)	12 (*12*)
16	500 (*500*)	48 (*48*)	46 (*48*)	40 (*46*)
64	500 (*500*)	117 (*192*)	103 (*192*)	113 (*192*)
256	139 (*500*)	139 (*500*)	248 (*500*)	263 (*500*)

Table 5.4: Optimum number of modes for differing numbers of unconditional vectors.

Discussion

To begin, the accuracy of the OLS predictions will be discussed. From the decomposition of the MSE for the OLS predictor (equation 4.71), the accuracy in this scenario is ultimately dictated by two conflicting phenomena: firstly, as the number of unconditional vectors is increased, the knowledge of $\mathbf{y}$ contained within $\mathbf{x}$ increases, and so the unexplained variance, σ^2, becomes smaller. At the same time, n becomes larger, which means that the term $\frac{n}{m}$ in the model variance $\frac{n}{m}\sigma^2$ increases. Figure 5.12 shows that the accuracy of the OLS prediction increases with n, up until $n = 192$ which has approximately the same accuracy as $n = 768$, presumably because the model variance begins to outweigh the reduction in σ^2 at this point.

As discussed previously, although each of the biased techniques differ in their approach, their success is greatly influenced by their ability to accurately represent some multivariate quantity in a much reduced subspace. Also, from the first scenario, the importance of a large model variance is also important. In this scenario, changing n ultimately effects both of these aspects, although the results indicate that the model variance is of far less importance here, as the performance of the biased techniques can be seen to improve with increasing n, even though the model variance is actually reducing. Interpreting the results of the PCR and PLS regressions is fairly straightforward, where the subspace in question is that of $\mathbf{x}$. In the present study, as the number of unconditional vectors is increased, the spacing between these vectors is reduced, leading to an increase in correlation between the elements of $\mathbf{x}$ (i.e. the level of multicollinearity increases). This, in turn, has the effect of concentrating the variance of $\mathbf{x}$ into a reduced subspace. Once the level of multicollinearity is high enough, the reduction in model variance due to the removal of small

and/or unimportant dimensions from $\mathbf{x}$, will be larger than the resulting introduction of bias. In the current scenario, this phenomena appears to occur once n reaches 192, at which point both PLS and PCR are able to outperform OLS. As expected, the improvement increases further for $n = 768$. In both cases, the performance of PLS and PCR is roughly equivalent. The situation for CSE and RRR is slightly more complicated, as these techniques act by reducing the subspace spanned by the OLS predictor $\mathbf{x}\hat{\mathbf{B}}_{OLS}$. From section 4.3.3, the size of this subspace is at most $r = \min\left(\mathrm{rank}(\mathbf{X}), \mathrm{rank}(\mathbf{Y})\right) = \min(n, m, p)$. The success of both these techniques therefore requires that $\mathbf{x}\hat{\mathbf{B}}_{OLS}$ can be accurately represented in a subspace of size k, where $k < r$. As discussed in section 4.3.3, the situations that allow for a reduction of V_{OLS} to a subspace of size k arise when:

1. the subspace of $\mathbf{x}$ that contributes to the prediction of $\mathbf{y}$ can be accurately represented in a subspace $k < r$, and/or

2. $\mathbf{y}$ can be accurately represented in a subspace $k < r$

In the current scenario, $m = 500$, $p = 3000$, and $n = 12, 48, 192, 768$. Therefore, $r = n$ for all but the latter case, where $r = m = 500$. Considering the 2 requirements above, it seems unlikely that the second requirement will be fulfilled, as the accurate reduction of the p-dimensional quantity $\mathbf{y}$ to an r-dimensional subspace will be difficult, given that $p >> r$ in all but the last test case. The first requirement seems more plausible, as this is the very requirement needed for the successful use of PLS and PCR. This is confirmed by the results, which demonstrate that when PCR and PLS are able to outperform OLS, then RRR also outperforms OLS. CSE, however, only produces an improvement in the final case.

Parameter	Value
Sample size, m	500
Unconditional Vectors	64
Independent Variables, n	191
Unconditional Vector Spacing	$4, 2, 1$
Conditional Vectors	1024
Dependent Variables, p	3072

Table 5.5: Summary of parameter values for Scenario 3
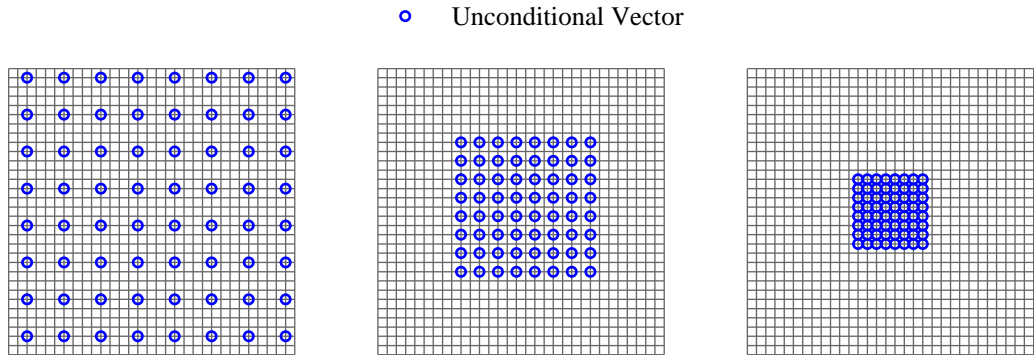
Figure 5.14: Unconditional vector placement within flow domain, for a separation of 4, 2 and 1 vectors

5.6.4 Scenario 3: Multicollinearity

Scenario 3 has been devised to show the effect of multicollinearity in $\mathbf{x}$ on the performance of the regression techniques. Although this has already been observed in scenario 2, the effect was an indirect result of varying n. Here n is held fixed, and the degree of multicollinearity is controlled directly by adjusting the spacing between the unconditional vectors, as shown in figure 5.14. As the spacing is reduced, the degree of collinearity is expected to increase due to the closer proximity of the vectors. The results are shown in figures 5.15 and 5.16, and a summary of parameters is in table 5.5.
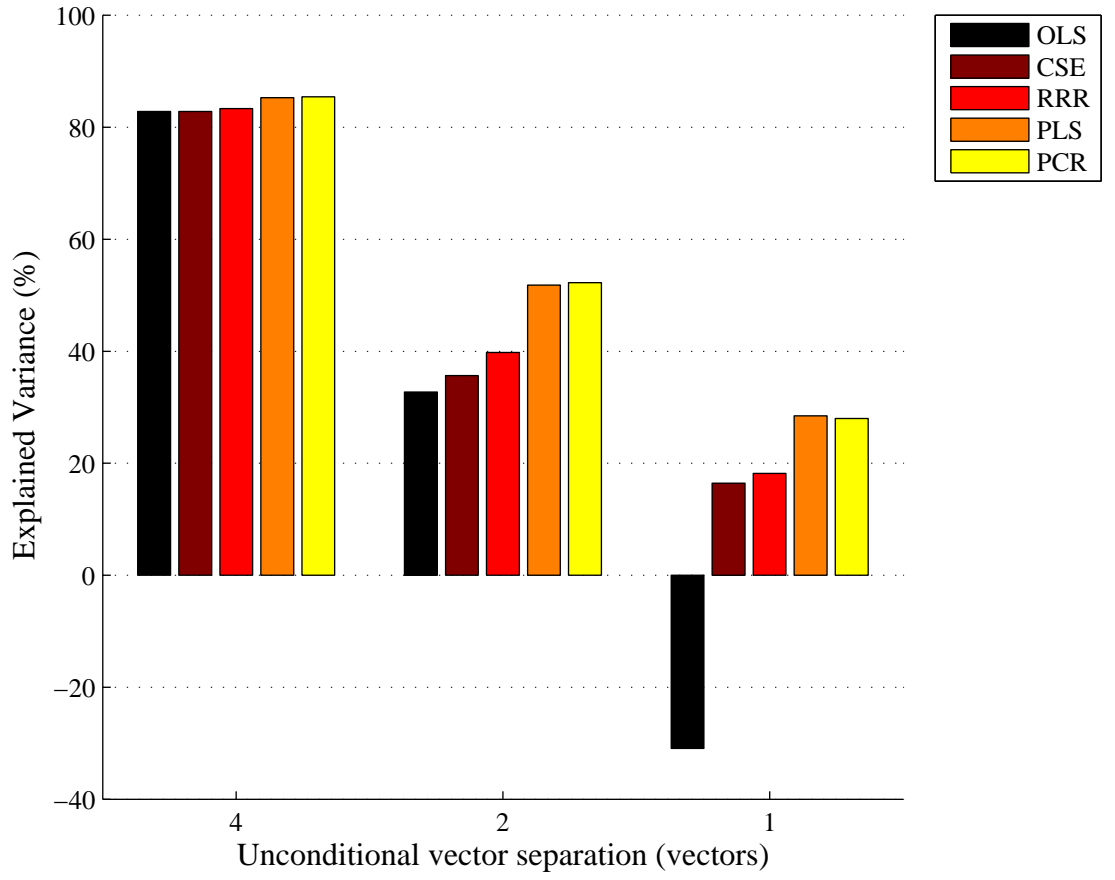
Figure 5.15: Average explained variance vs. unconditional vector separation

Separation	CSE	RRR	PLS	PCR
4	500 (*500*)	117 (*192*)	103 (*192*)	113 (*192*)
2	23 (*500*)	16 (*192*)	48 (*192*)	61 (*192*)
1	1 (*500*)	1 (*192*)	37 (*192*)	33 (*192*)

Table 5.6: Optimum number of modes for differing separation between unconditional event vectors
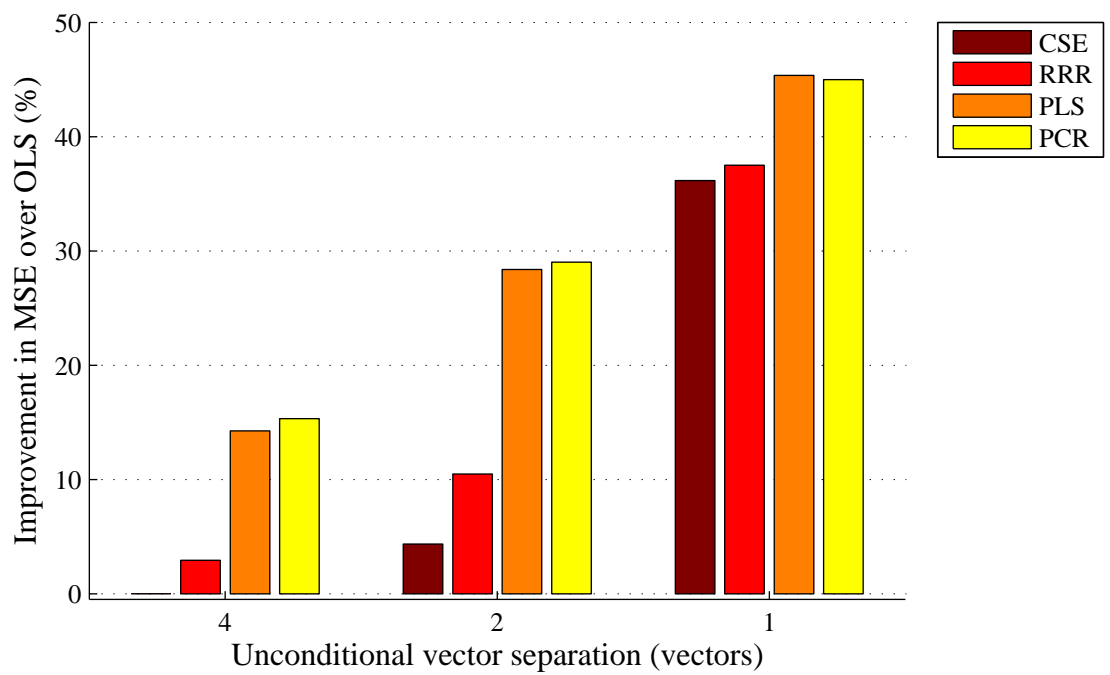
Figure 5.16: Percentage improvement over OLS prediction vs. unconditional vector separation

5.6.5 Discussion

Although this scenario has been devised to study the effect of multicollinearity in $\mathbf{x}$, a result of decreasing the spacing between unconditional vectors is that the ability to predict $\mathbf{y}$ from $\mathbf{x}$ is reduced. This can be observed in absolute performance of all the results (figure 5.15), which falls as the vector spacing decreases, due to a loss of predictive power over the vectors on the periphery of the flow field (see figure 5.14). In the last condition, the performance of OLS is particularly bad, where the combination of unexplained variance and model error is larger than the variance of $\mathbf{y}$, leading to a negative unexplained variance. As such, the reduction in vector spacing is ideal for the use of biased techniques for two reasons. Firstly, it was shown in scenario 2 that as the multicollinearity increases, the variance in both $\mathbf{x}$ (needed for PLS and PCR) and $\mathbf{x}\hat{\mathbf{B}}_{OLS}$ (for RRR and CSE) will be concentrated in a reduced subspace, which reduces the amount of bias introduced into the predictions when modes are removed. Secondly, as the predictive power reduces, the unexplained variance σ^2 increases, which means that the OLS model error $\frac{n}{m}\sigma^2$ will be large. The joint effect of these phenomena can clearly be observed in the relative improvement of the biased techniques with decreasing vector spacing, and also in the reduction in the number of modes used by the optimum biased models (table 5.6).

5.6.6 Scenario 4: Number of Dependent Variables

The fourth scenario investigates how the prediction accuracy is affected by the number of dependent variables included in the model. So far, the scenarios have all investigated parameters that directly relate to the independent variables, $\mathbf{x}$, while the dependent variables $\mathbf{y}$ have consistently been defined

Parameter	Value
Sample size, m	500
Unconditional Vectors	64
Independent Variables, n	192
Conditional Vectors	$4, 16, 64, 256, 1024$
Dependent Variables, p	$12, 48, 192, 868, 3072$

Table 5.7: Summary of parameter values for Scenario 4

as the full 32×32 flow field. The subsequent results have shown that PCR has offered the best performance overall (although the improvement over PLS is marginal). However, unlike PCR, the biased techniques of PLS, CSE and RRR, all take the dependent variables into account when producing the prediction model. If any of these techniques are able to outperform PCR, then it is likely to be for this reason, and it may be the case that until now, the choice of **y** has not been favourable for their use. This certainly implied by the bias-variance decomposition of CSE and RRR, which showed that their ability to reduce the model error is limited when the rank of the OLS predictor, $r = \min(n, m, p)$, is lower than than the number of dependent variables p. This has been true of all the previous scenarios, where the full 32×32 flow field has been used as the dependent variables, resulting in a size of p which has been considerably larger than r in most cases.

By varying the number of, and position of, the conditional vectors, as illustrated in figure 5.17, this scenario aims to uncover situations which may be favourable for the techniques of PLS, CSE and RRR. A summary of the parameters for this scenario are given in table 5.7 and the optimum number of modes used to build the biased models in table 5.8. Results are shown in figures 5.18 and 5.19.
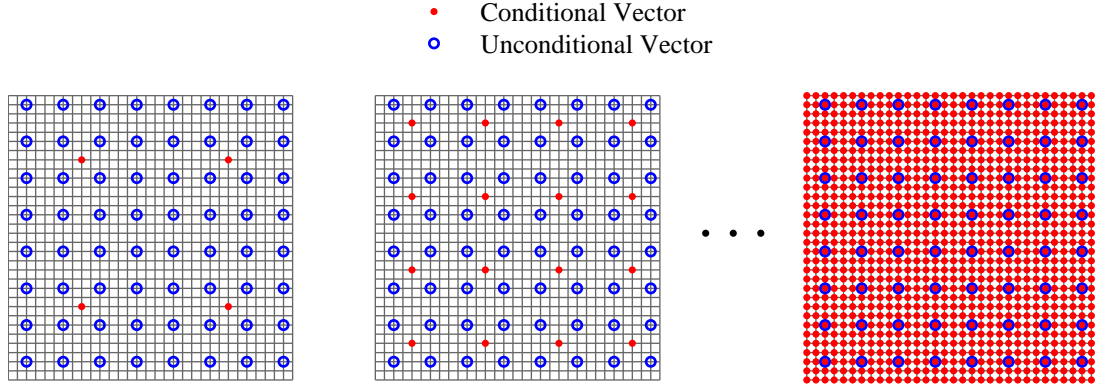
Figure 5.17: Conditional vector placement within flow domain, for $4, 16, \ldots, 1024$ vectors
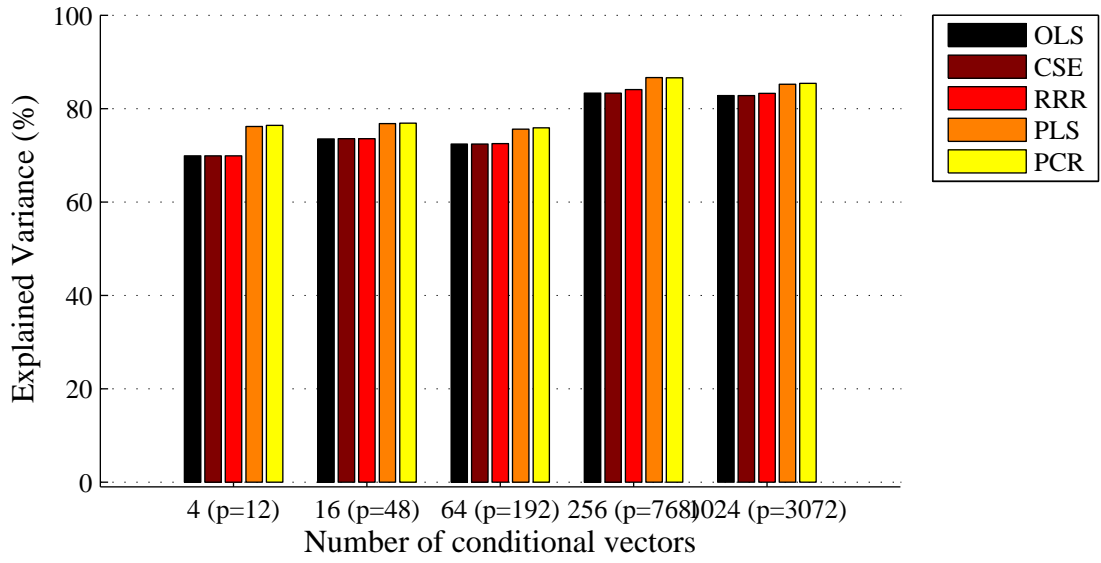
Figure 5.18: Average explained variance vs. number of conditional vectors
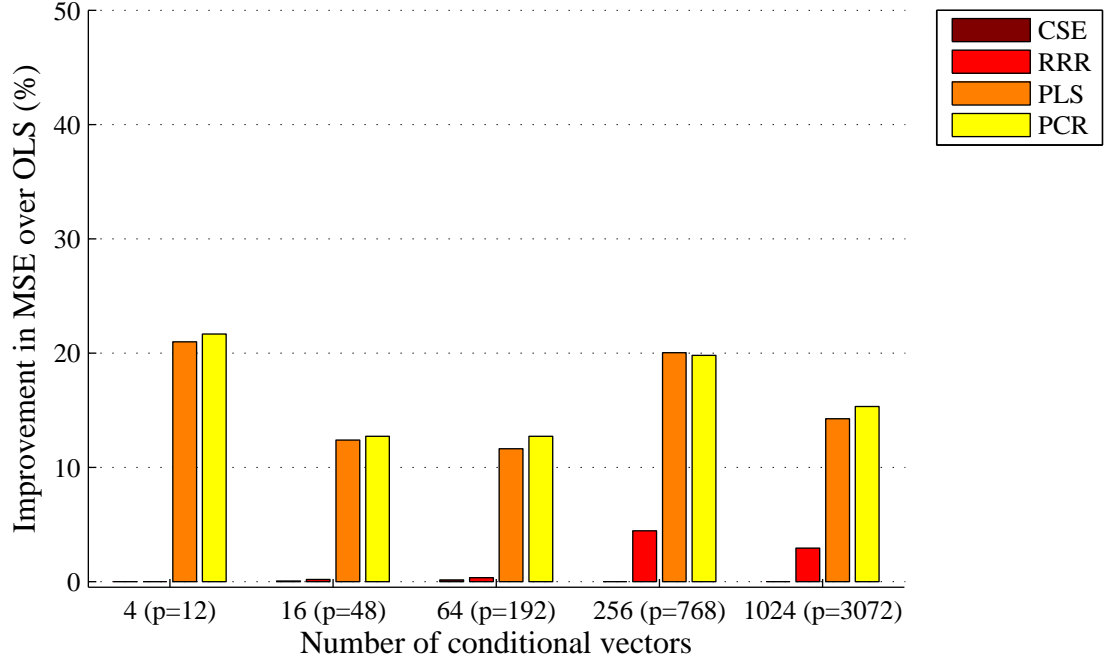
Figure 5.19: Percentage improvement over OLS prediction vs. number of conditional vectors

Vectors	CSE	RRR	PLS	PCR
4	12 (*12*)	12 (*12*)	32 (*192*)	78 (*192*)
16	44 (*48*)	43 (*48*)	81 (*192*)	122 (*192*)
64	157 (*192*)	135 (*192*)	107 (*192*)	122 (*192*)
256	500 (*500*)	127 (*192*)	76 (*192*)	75 (*192*)
1024	500 (*500*)	117 (*192*)	103 (*192*)	113 (*192*)

Table 5.8: Optimum number of modes for differing numbers of conditional vectors.

Discussion

First of all, it can be seen from figure 5.15 that the performance of the techniques varies somewhat arbitrarily over the range of conditions considered here. Unlike previous scenarios, the data being predicted differs for each condition under consideration, which makes interpretation of the results slightly more complex. However, it is shown that regardless of varying the parameter p, the relative performance of the techniques is comparable to that observed in all the previous scenarios. In particular, the performance of PLS and PCR here is approximately equal, and consistently better than the other techniques. Also, RRR is clearly better than CSE in the latter two conditions, with CSE barely outperforming OLS across the entire scenario.

In hindsight, the similarity of the results between this and previous scenarios is unsurprising; regardless of the intentions of this scenario, the effect of varying p in the manner prescribed here is unlikely to be conducive to the performance of either PLS, CSE or RRR. As observed in scenario 2, where the number of independent variables was varied, the effect of reducing the number of dependent variables increases the spacing between vectors, leading to a reduction in the multicollinearity in $\mathbf{y}$. For CSE and RRR, the theoretical analysis of the techniques implies that this is actually detrimental to their performance. Any improvement requires the accurate representation of $\mathbf{x}\hat{\mathbf{B}}_{OLS}$ in a reduced subspace, requiring that:

1. the subspace of $\mathbf{x}$ that contributes to the prediction of y can be accurately represented in a subspace $k < r$, and/or

2. $\mathbf{y}$ can be accurately represented in a subspace $k < r$

Because $\mathbf{x}$ is fixed for all conditions, the subspace that contributes to the

prediction of $\mathbf{y}$ is unlikely to change considerably, and as p decreases, the decrease in multicollinearity will reduce the chances of the second condition being met. This alone shows that in this scenario, the reduction in p offers no benefits for the performance of CSE and RRR. What is worse, is that in the first three conditions, the subspace r is now dictated by the size of p, which makes the fulfilment of either of the two conditions even less likely, as the relevant subspace must be reduced further before any benefit will occur. This can be seen in the results, where the performance of the two techniques for the first three cases (where $p = r$) is barely distinguishable from that of OLS.

Although not immediately apparent from bias-variance decomposition of PLS, it is nonetheless cited that multicollinearity in y is favourable for the use of PLS (Wold *et al.*, 2001). As such, it is understandable that the conditions in this scenario do not produce any notable improvement over PCR. Interestingly, it should be noted that the decrease in multicollinearity with p does not appear to adversely affect its relative performance either.

In conclusion, rather than identifying situations where the behaviour of $\mathbf{y}$ can be exploited for the benefit of PLS, CSE and RRR, the results of this scenario simply serve to highlight further conditions where PLS and PCR offer very similar performance, and further consolidate the fact that RRR, and particularly CSE, are ill-suited to the purpose of improving the prediction performance of stochastic estimation.

5.7 Conclusions

The results and discussion presented in this chapter form a comprehensive study into the behaviour of the prediction techniques under consideration. Often, the results have served to corroborate the theoretical analysis provided

in chapter 4.3. In other cases, the results have found limitations in the theory, and identified the need for further study. Ultimately though, the main accomplishment of this work is the wealth of practical information it provides towards the goal of accurate flow prediction.

One of the most important observations here is that the use of OLS, which is the standard stochastic estimation method, can perform badly, and may even fail catastrophically. What is most surprising is that the situations that led to this behaviour, namely when the number of samples, m, is slightly larger than the number of independent variables, n, do not invalidate any of the OLS assumptions. Furthermore, this behaviour is not predicted by the theoretical expression of the MSE for OLS predictions derived in section 4.3. This shows the importance of validating a regression model before use, as the presence and extent of this phenomenon may not be apparent from theoretical considerations alone. And once again, this demonstrates the need to acquire as many observations as possible, especially if OLS is to be used.

The issues with the theoretical expression of the MSE for OLS have been attributed to the presence of estimation error in the principal components and eigenvalues of $\mathbf{X}$, which was not accounted for due to assumptions in the derivation. These assumptions also appear to produce a theoretical MSE that is an underestimate of the true value. Unfortunately, the assumptions have been used in the derivation of the model variance terms in all subsequent bias-variance decompositions for the biased techniques, and so similar issues will be present in these expressions as well. However, this need not detract from the fact that, in general, the expressions have served as useful indicators of the technique's performance, and have been in good agreement with the quantitative results. In particular, even though the theoretical form of the OLS model variance, $\frac{n}{m}\sigma^2$ may not be numerically exact, its implication that model

149

variance increases with n and σ^2, and decreases with m, has been demonstrated extensively. Similarly, the bias-variance decompositions for the biased models has often been useful in explaining the results observed in this chapter.

The performance of the biased techniques conclusively shows their suitability to flow prediction, as the techniques under consideration have been able to outperform OLS in many situations. This relative improvement has been shown to increase with both multicollinearity in $\mathbf{x}$, and with the model variance of the OLS estimator. The most impressive of the these techniques have undoubtedly been PLS and PCR, which have consistently outperformed or equalled all the other techniques. Furthermore, unlike the remaining techniques, PLS and PCR have provided relatively accurate predictions, regardless of sample size. As such, they have not suffered from the catastrophic errors that affect OLS, and they have even provided accurate predictions when the number of observations is far less than the number of independent variables. In all these cases, the superiority of PLS and PCR has been attributed to the presence of multicollinearity in $\mathbf{x}$, which will be a common occurrence in many applications of stochastic estimation.

For this data set, the relative performance of PLS and PCR is very close, but over the entire set of scenarios, it appears that PCR offers slightly higher accuracy. Also, considering that the computational effort of PLS is far higher, PCR would appear to be the best choice for prediction. However, on paper, PLS does have some potential benefits over PCR, in that it theoretically only selects modes from $\mathbf{x}$ that are relevant for the prediction of $\mathbf{y}$, whereas PCR chooses the largest principal components of $\mathbf{x}$, regardless of their importance to the prediction. It is quite plausible that the data set used here was not able to make full use of PLS's potential; in these scenarios, $\mathbf{x}$ and $\mathbf{y}$ are both velocities from the same flow domain, and it is quite likely that the dominant

150

dimensions of $\mathbf{x}$ are the ones that best predict the dominant dimensions of $\mathbf{y}$. However, in cases where $\mathbf{x}$ and $\mathbf{y}$ are taken from considerably different measurements (either due to large spatial/temporal separation, or when two different forms of measurement are made), PLS may be able to offer superior performance. Another potential use of PLS, not investigated here, is its use as a means of producing a low dimensional dynamical relationship between two related flow quantities, as described in 4.3.2. For both these reasons, PLS may still prove to be a useful tool in experimental fluids applications.

The performance of CSE and RRR, while often better than OLS, has rarely rivalled that of PCR and PLS. A particular problem is that the accuracy of CSE and RRR is closely linked with that of the OLS predictions, and while they do not suffer to the same extent, their performance may still be too poor to be of any practical use. From the results here, as well as the theoretical discussion in section 4.3.3, it appears that the requirements for any improvement over OLS are quite specific, and have been difficult to obtain even in this artificial setting. Furthermore, even in favourable conditions, these techniques have not always produced a noticeable improvement over OLS. Clearly, the processes that dictate their performance are more complex than the present study can account for. Nonetheless, even with an incomplete understanding of these techniques, the available evidence strongly suggests that their use for instantaneous flow-prediction is limited.

While the results in the chapter appear to rule out the use of CSE as a tool to improve the prediction accuracy of OLS, it is perhaps unfair to be too critical of the technique, as it is not conventionally used for this purpose. Instead, it is employed as means of producing a low-order model of the flow (i.e. the variable $\mathbf{y}$ in the current terminology), using a reduced set of principal components of $\mathbf{Y}$, which is usually considered to represent the

coherent structure of the flow. However, the results of this chapter have shown that RRR is generally better at predicting $\mathbf{y}$ than CSE, and it follows that a low-order model of $\mathbf{y}$ built from the principal components of $\mathbf{X}\hat{\mathbf{B}}_{OLS}$ should be more realistic that one built from the principal components of $\mathbf{Y}$. This may be of particular use in applications of real-time flow sensing and control. Here, the use of CSE is of benefit as it allows the prediction of the flow at an indefinitely large number of points to be reduced to the prediction of a small number, k of principal component scores, which intrinsically describe the behaviour of the full flow field without the large computational overhead. However, instead of predicting the scores of the principal components of $\mathbf{Y}$, predicting the scores of the principal components of $\hat{\mathbf{Y}}$ may be more appropriate, as it should provide a better description of the flow for a given k.

Furthermore, in situations where a low-dimensional representation of $\mathbf{y}$ using the principal components of $\mathbf{Y}$ is explicitly sought, the biased techniques considered here may still be of use as a means of improving the accuracy of the model. It has been established that because CSE acts on the OLS prediction, $\mathbf{x}\hat{\mathbf{B}}_{OLS}$, it suffers when the OLS prediction is poor. A more accurate approach may be to instead predict $\mathbf{y}$ using either PLS or PCR, and then produce the low order model from this quantity instead.

Chapter 6

Prediction of a Flow Volume using MP-VSE

6.1 Introduction

This chapter introduces the proposed technique of multiple plane volumetric stochastic estimation (MP-VSE), which extends the use of stochastic estimation to provide a volumetric prediction of a flow that possesses a homogeneous dimension. The chapter begins with a step-by-step outline of the technique, followed by a description of a synthetic MP-VSE experiment, performed to demonstrate the feasibility of the approach. Finally, the issues regarding the practical implementation MP-VSE are considered.

6.2 Theory

The basic approach of MP-VSE is illustrated in this section, using the example of a flow that possesses a Cartesian homogeneous dimension. Here, the goal is to predict a uniformly spaced $i \times j \times k$ grid of velocity vectors within the volume Ω, as illustrated in figure 6.1. The grid spacing is defined by Δx, Δy and Δz, so the size of Ω is $i\Delta x \times j\Delta y \times k\Delta z$. The flow is homogeneous in the z dimension, which means that the statistical properties of the flow are independent of the z location. It is assumed that all measurements are carried out using stereoscopic particle image velocimetry (stereo PIV), which provides 3-component velocity vectors on a uniformly spaced planar grid.

6.2.1 Required measurements

The proposed technique involves two stages, each requiring a separate experimental configuration. The first stage involves the measurement of m stereo PIV measurements at a rectangular region, P, in a plane normal to the homogeneous dimension. Each PIV measurement consists of a set of velocity
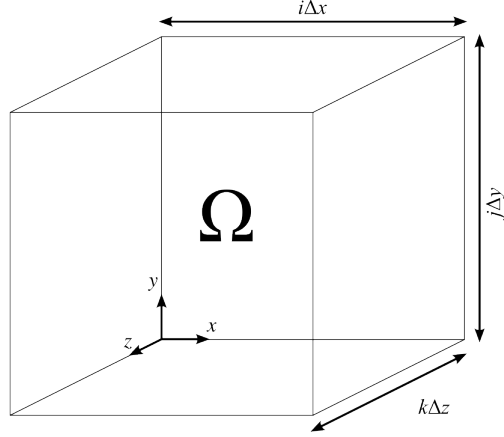
Figure 6.1: The prediction volume Ω. It is assumed that the flow in question is homogeneous in the z dimension

vectors arranged on an $i \times j$ uniformly space grid. In the current example, this is the $x - y$ plane, as shown in figure 6.2 (a). This configuration is used to acquire the data for the prediction model. The second stage consists of l separate measurement planes, $Q_1, \ldots, Q_l$, which must be parallel to the homogeneous dimension (illustrated in figure 6.2 (b) for 3 measurement planes). At each plane, a grid of $i \times k$ vectors is measured. This configuration provides the instantaneous data that is used to predict the full volume Ω.

The data acquired at P is used to build a stochastic estimation model, where the unconditional vectors are defined by the intersections of $Q_1, \ldots, Q_l$ with P, as indicated in figure 6.3. The full $i \times j$ flow field is used as the conditional vector set.

MP-VSE is appropriate for both stationary as well as non-stationary flows, although the approach is slightly different in each case. For stationary flows, the acquisition of data at a plane P can be carried out without regard to the time at which the observations are made, other than to ensure that the
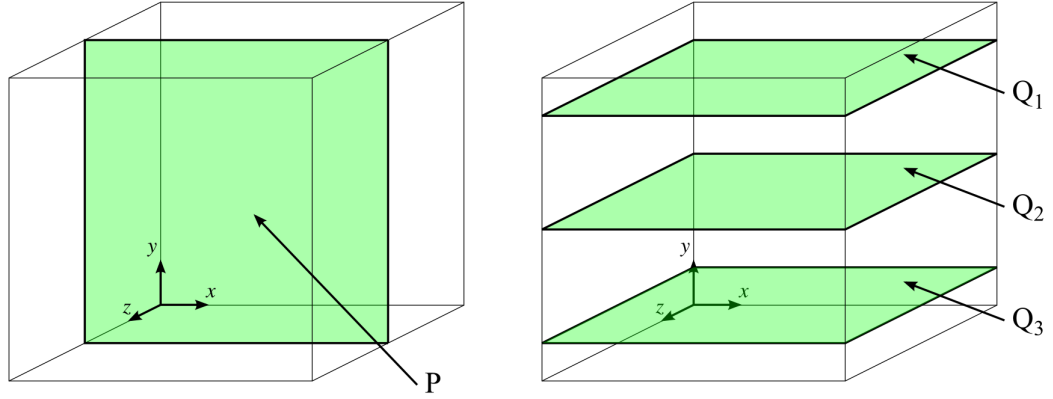
155

(a) Configuration 1: Position of plane P, which is used to build the prediction model

(b) Configuration 2: Position of planes Q_1, Q_2 and Q_3, which provide the instantaneous measurements for the prediction

Figure 6.2: Measurement plane configuration for the prediction of Ω

set of m observations are statistically independent. For non-stationary flows, the acquisition of the data used to build the prediction model must be made at the same point in time, over multiple runs of the same experiment.

Once the prediction model has been created, volumetric prediction is possible. Once again, the temporal behaviour of the flow must be taken into account; for a stationary flow, the prediction can be made at any point in time, while for a non-stationary flow, the model is only valid for predicting at the specific point in time that it was created from. As such, prediction of the flow at multiple points in time requires a separate model to be built for each occasion.
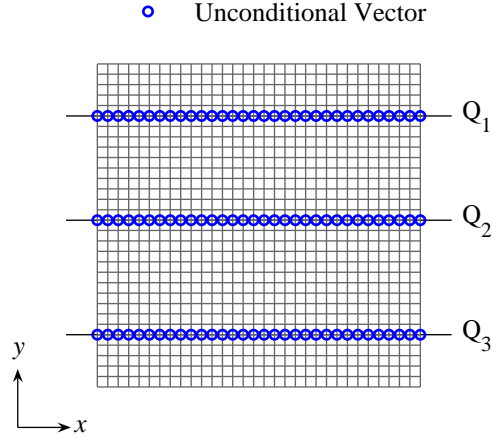
Figure 6.3: Position of unconditional vectors for the prediction model, using the data acquired from plane P. All the vectors that lie on the intersection between Q_1, Q_2, Q_3 and P are used as unconditional vectors

6.2.2 Prediction

The prediction of the flow at time t is carried out using instantaneous measurements from the planes $Q_1, \ldots, Q_l$, also acquired at time t. In the current example, each Q consists of an $i \times k$ grid of vectors aligned in the $y - z$ plane. At each z position, the corresponding vectors in $Q_1, \ldots, Q_l$ form the unconditional vectors for the prediction model illustrated in figure 6.3, which allows the full $x - y$ plane to be predicted for the current z position. Performing the prediction at $z = 0, \Delta z, 2\Delta z, \ldots, k\Delta z$, allows the full $i \times j \times k$ volume to be reconstructed. This process is illustrated in figure 6.4.
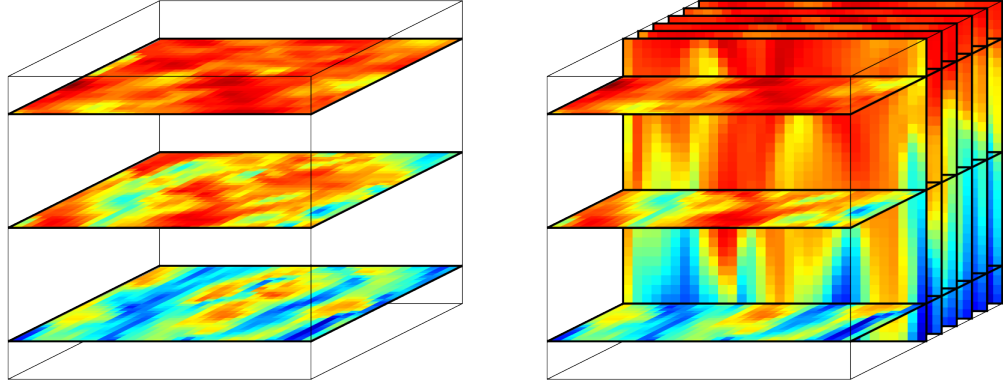
(a) The process begins with a the instantaneous measurements of $Q_1, \ldots, Q_l$ at time, t

(b) The values of vectors in $Q_1, \ldots, Q_l$ at each z are used to predict the corresponding $x - y$ plane
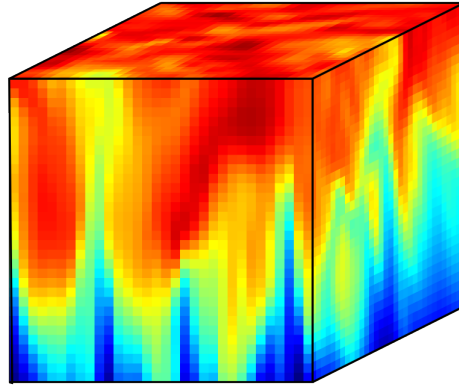
(c) The full set of reconstructed $x - y$ planes provides the volumetric prediction of Ω at time, t

Figure 6.4: Overview of the volumetric prediction process

6.3 Proof of Concept - Volumetric Prediction of Simulated Channel Flow

In order to demonstrate the feasibility of MP-VSE, the technique is applied to a simulated experiment based on volumetric CFD data. The benefit of carrying out the technique using simulated data is that the true velocity is known throughout the full volume, and can thus be compared with the prediction. In this section, the comparison is carried out both quantitatively and qualitatively. A further purpose of this test is to assess the suitability of the biased regression techniques for MP-VSE applications. The data used for this work was taken from the channel flow simulation, described in section 5.5. A $32 \times 32 \times 32$ volume, located in the position indicated in figure 5.3, was selected as the prediction volume Ω. The flow is homogeneous in the z dimension.

A total of 250 independent observations of the volume Ω were available. From this data set, an artificial MP-VSE experiment was created by selecting slices of the volume to act as planar stereo PIV measurements, as illustrated in figure 6.2, with the position of P placed at $z = 16\Delta z$ in the $x-y$ plane and the three Q planes positioned at $y = 5, 16, 26\Delta y$. The resulting position of the unconditional vectors in the prediction model is determined by the intersection of the P and Q planes, as shown in figure 6.5.

6.3.1 Quantitative comparison

10-fold cross validation was employed to assess the accuracy the five regression techniques for the prediction of the P plane data. As such, the 250 volumes were partitioned into 10 sets. However, to build the prediction model, two $x-y$ planes were extracted from each of the 250 volumes, with sufficient separation
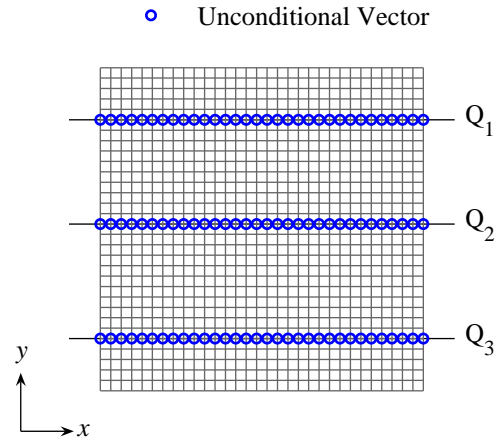
Figure 6.5: Position of unconditional vectors in the prediction model

(a) Position of Q planes

(b) Position of R planes
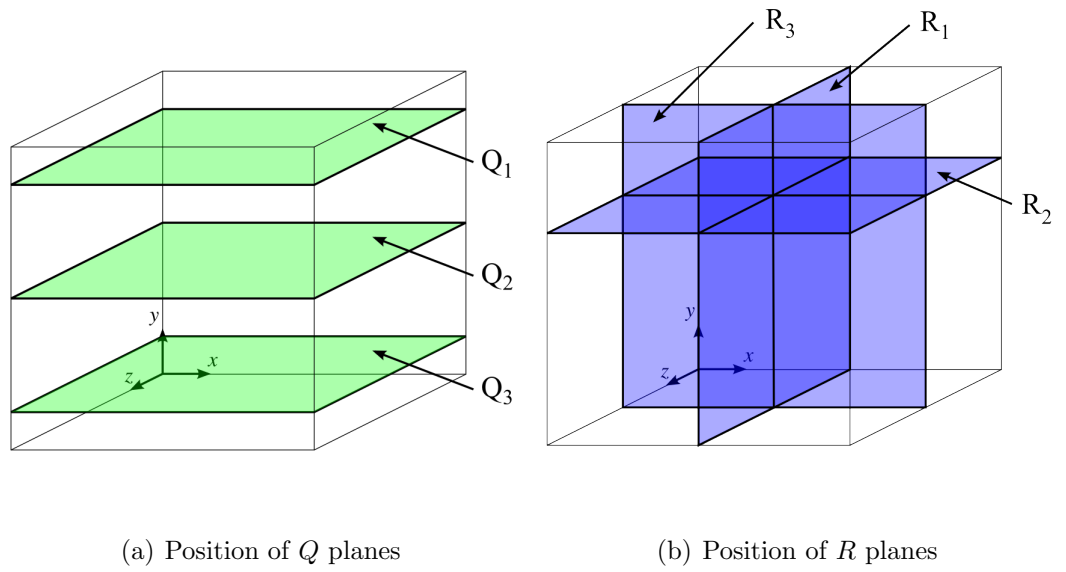
Figure 6.6: Positions of the planes used to compare the true and predicted flow

Parameter	Value
Sample size, m	450
Unconditional Vectors	96
Independent Variables, n	288
Conditional Vectors	1024
Dependent Variables, p	3072

Table 6.1: Summary of parameters

CSE	RRR	PLS	PCR
85 (*450*)	85 (*288*)	137 (*288*)	141 (*288*)

Table 6.2: Optimum number of modes used for biased regression techniques

in z to ensure statistical independence, This allowed a total of 500 planes to be used for the cross-validation procedure. At each stage of the cross-validation, 450 observations of P were therefore available to build the model, and a test set of 50 planes were used to measure the MSE of prediction. At the end of the process, the average MSE was calculated for each model, and the optimum model was identified for each biased technique.

The overall accuracy results are presented in figures 6.7 and 6.8. The number of modes used in the biased models is shown in table 6.2, and the regression parameters are summarised in table 6.1.

The results here clearly exhibit the same trend identified in the results of chapter 5. All the biased techniques outperform the standard OLS prediction and once again, the best performance is obtained using PCR regression, beating PLS by a small margin. This behaviour is unsurprising, as the close spacing of the unconditional vectors along the intersection of the Q planes (see figure 6.5) is likely to provide the high levels of multicollinearity conductive to
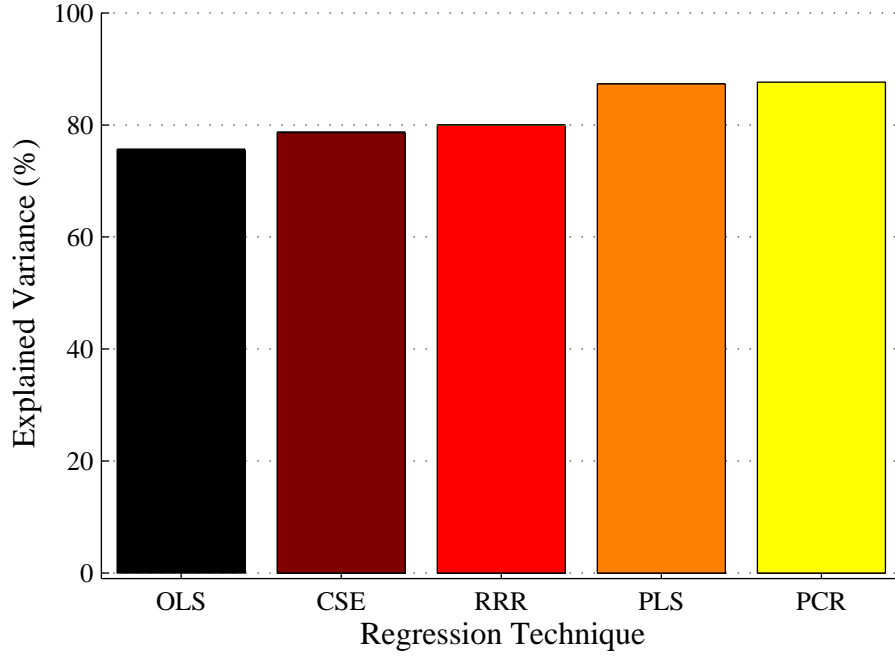
Figure 6.7: Average variance explained vs. regression technique

the performance of the biased techniques, particularly PLS and PCR.

To some extent, this artificial experiment creates an unrealistically favourable setting for the MP-VSE procedure, as the data is free from the numerous sources of error that will be present in any practical implementation. Nonetheless, the results are promising. In this example, the unconditional vectors make up approximately 9% of total vectors in the volume. Therefore, the fact that the best performing regression model (141-mode PCR) accounts for approximately 88% of the total variance of the flow is impressive, especially as this figure does not include the variance contribution of the unconditional vectors themselves (as described in section 5.3, this component of the variance is removed in order to provide a more representative error measurement).
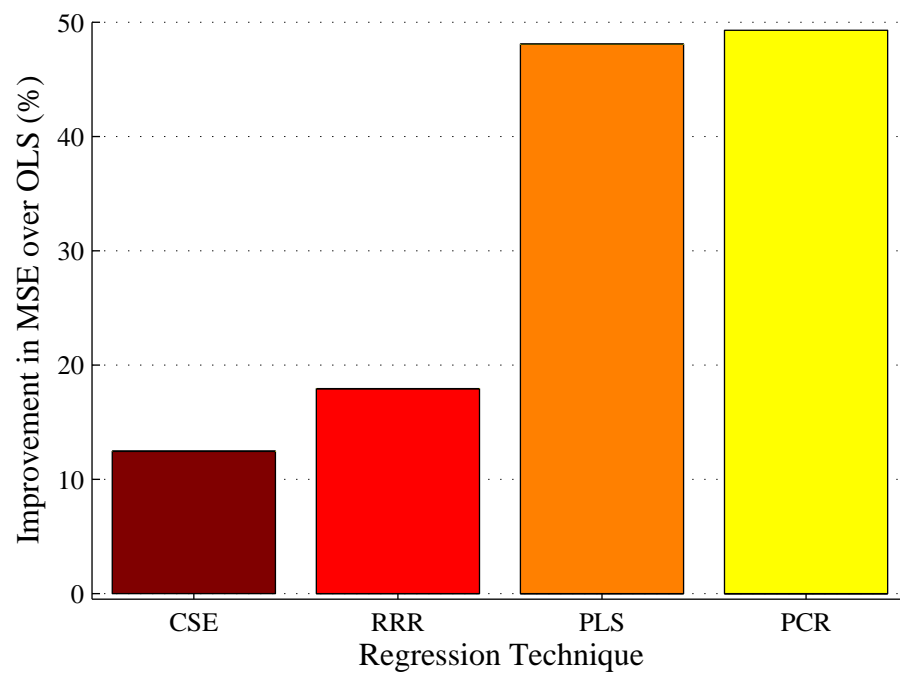
Figure 6.8: Percentage improvement over OLS prediction vs. biased regression techniques

6.3.2 Qualitative comparison

Further understanding of the performance of MP-VSE can be obtained from a visual comparison of the true and predicted data. For this purpose, a set of predicted volumes were obtained by performing the 10-fold cross-validation procedure again. This time, only the optimum 141-mode PCR model was built, and the resulting model was used to predict the full volume of the 25 observations of Ω in the test set, which was then stored for future analysis. Hence, 250 predicted volumes were available at the end of the procedure. Comparisons of the true and predicted flow field were made at selected planes, R_1, R_2 and R_3, shown in figure 6.6. At each plane, predicted and true flow quantities were compared visually. The results for a representative volume are presented here. Figures 6.9, 6.10 and 6.11 compare the u, v and w velocity components at planes R_1, R_2 and R_3, respectively, and figure 6.12 compares the plane-normal vorticity for each R.

On the whole, the location, size and shape of the large-scale structure in both the velocity and vorticity fields is represented well in the predicted data. The ability to accurately capture the vortical structure is particularly remarkable, as vorticity, which is based on the derivative of the quantity being predicted, will be particularly sensitive to error in the prediction. Also of interest is the fact that the planes R_1 and R_2 do not show any noticeable artifacts from the prediction process. Unlike R_3, which is the result of a single prediction, R_1 and R_2 are made up of slices from 32 individual planar predictions. An initial concern was that the errors between neighbouring planar predictions would result in sharp discontinuities along the homogeneous dimension. However, inspection along the z axis of the relevant figures shows that this does not appear to be a problem.

(a) Contours of true u

(b) Contours of predicted u

(c) Contours of true v

(d) Contours of predicted v
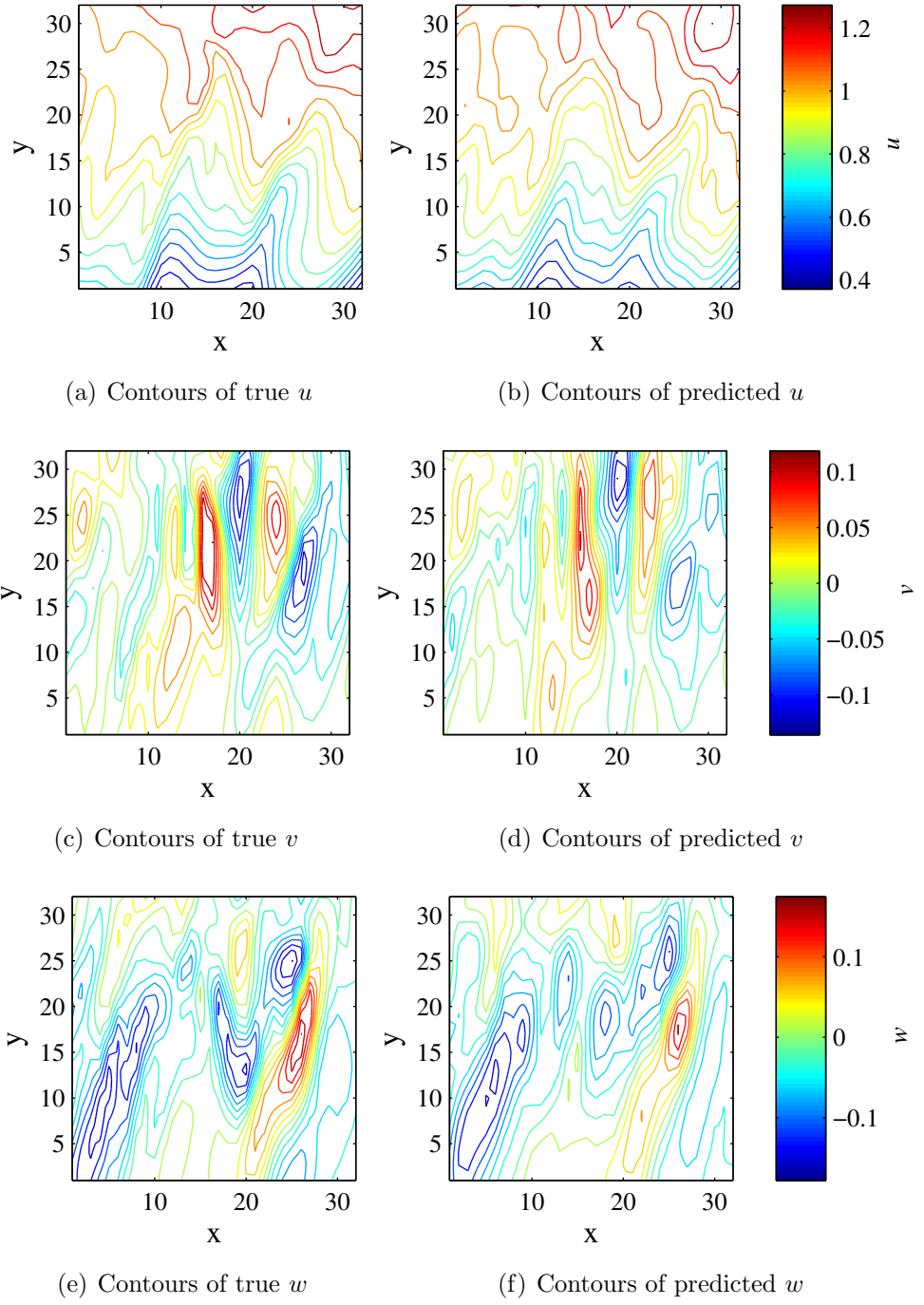
(e) Contours of true w

(f) Contours of predicted w
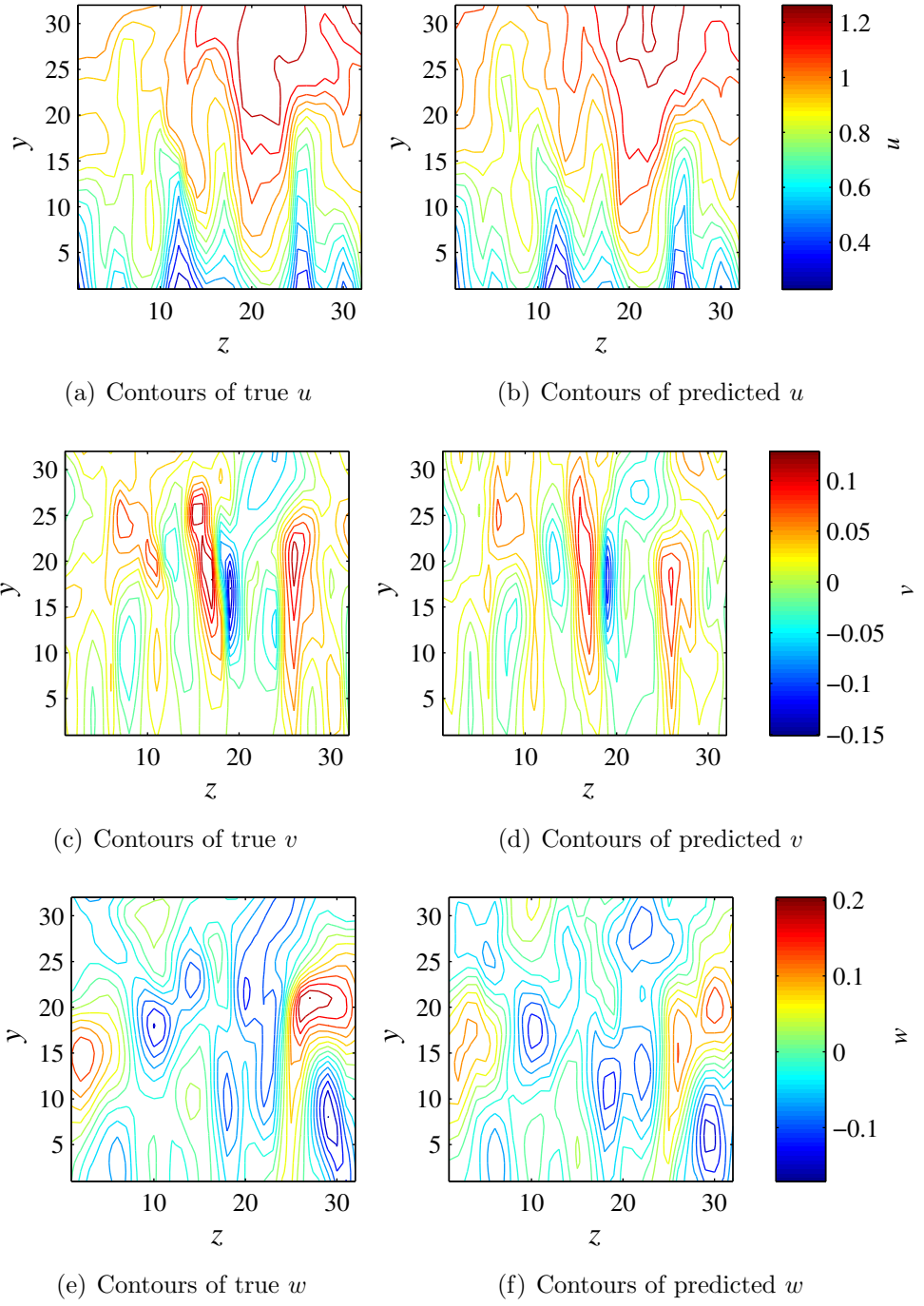
Figure 6.9: Comparison of true and predicted instantaneous velocity field in $y - z$ plane, $x = 16\Delta x$

(a) Contours of true u

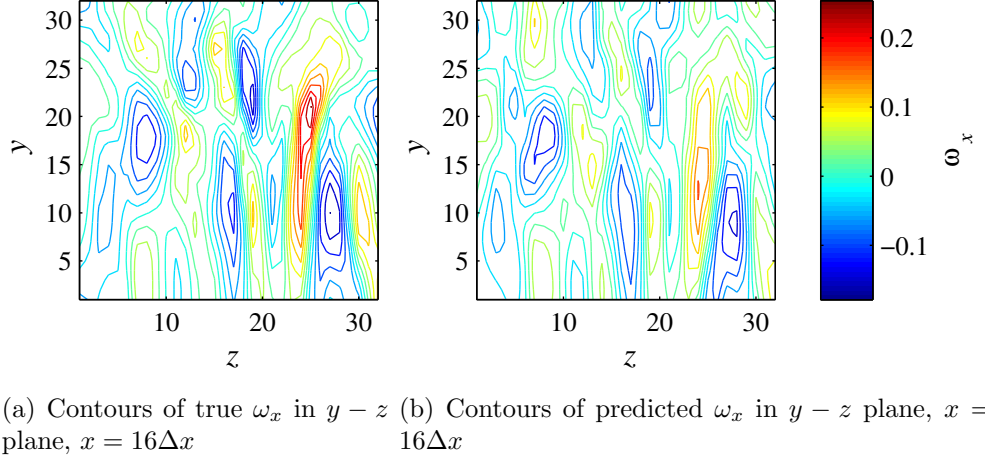(b) Contours of predicted u

(c) Contours of true v

(d) Contours of predicted v

(e) Contours of true w

(f) Contours of predicted w
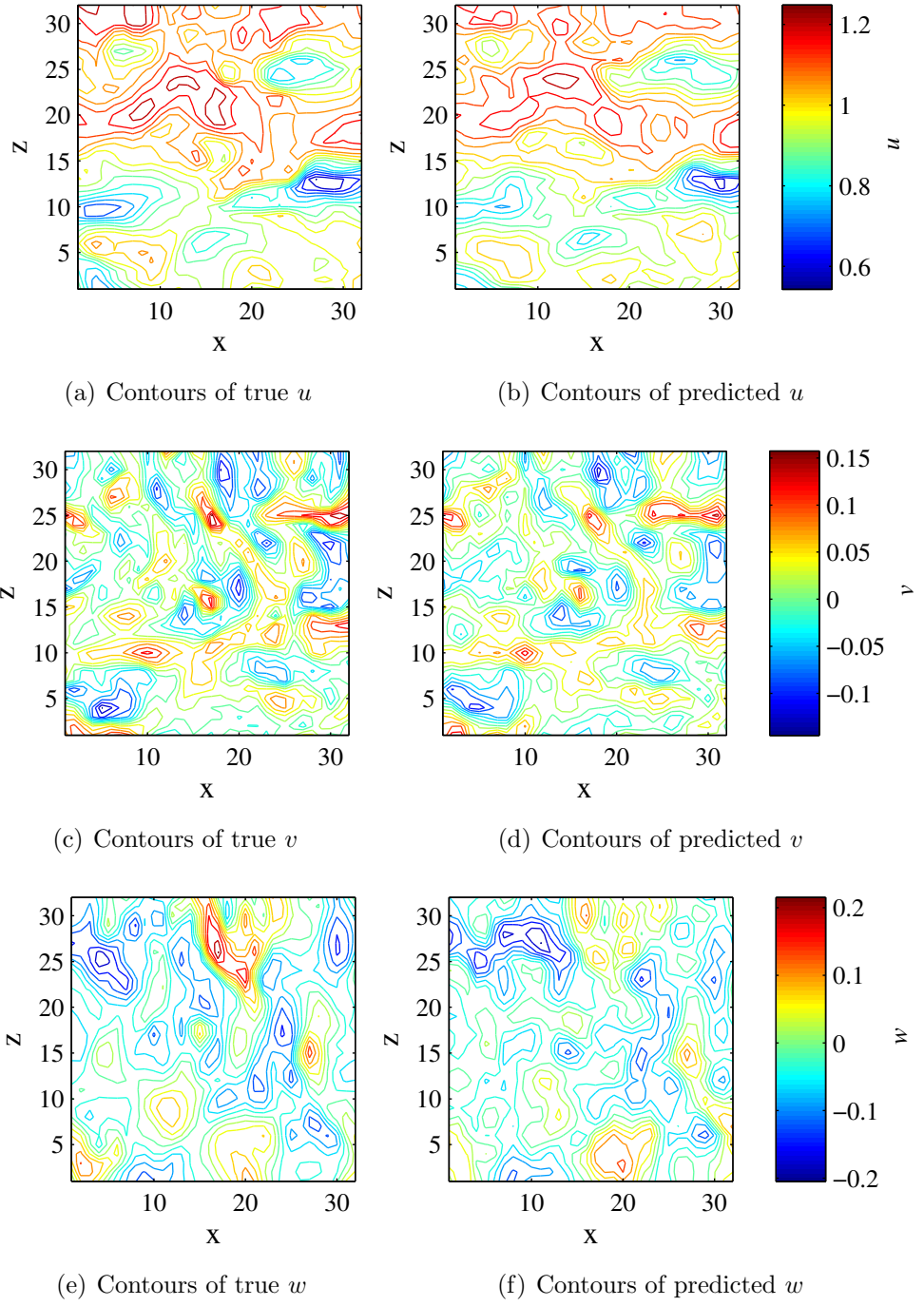
Figure 6.10: Comparison of true and predicted instantaneous velocity field in $x - z$ plane, $y = 22\Delta y$

166

(a) Contours of true u (b) Contours of predicted u

(c) Contours of true v (d) Contours of predicted v

(e) Contours of true w (f) Contours of predicted w

Figure 6.11: Comparison of true and predicted instantaneous velocity field in $x - y$ plane, $z = 16\Delta z$

(a) Contours of true ω_x in $y - z$ plane, $x = 16\Delta x$

(b) Contours of predicted ω_x in $y - z$ plane, $x = 16\Delta x$

(c) Contours of true ω_y in $x - z$ plane, $y = 22\Delta y$

(d) Contours of predicted ω_y in $x - z$ plane, $y = 22\Delta y$

(e) Contours of true ω_z in $x - y$ plane, $z = 16\Delta z$

(f) Contours of predicted ω_z in $x - y$ plane, $z = 16\Delta z$
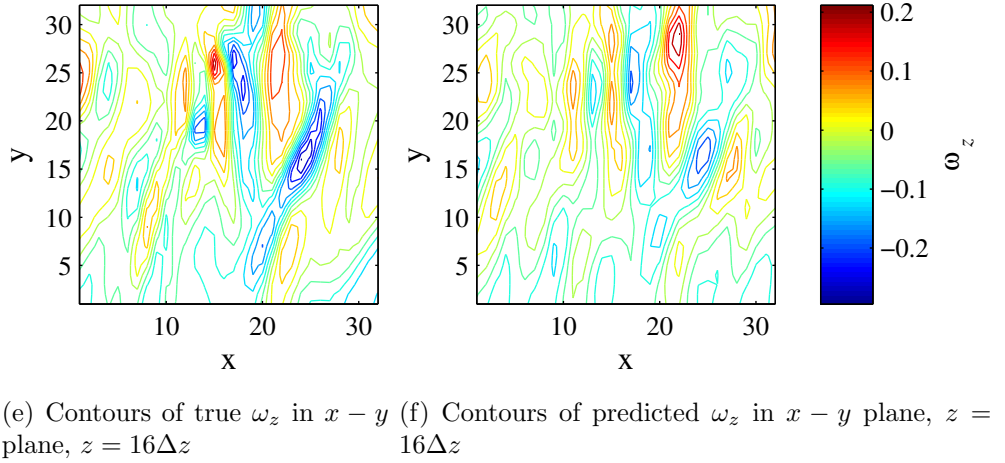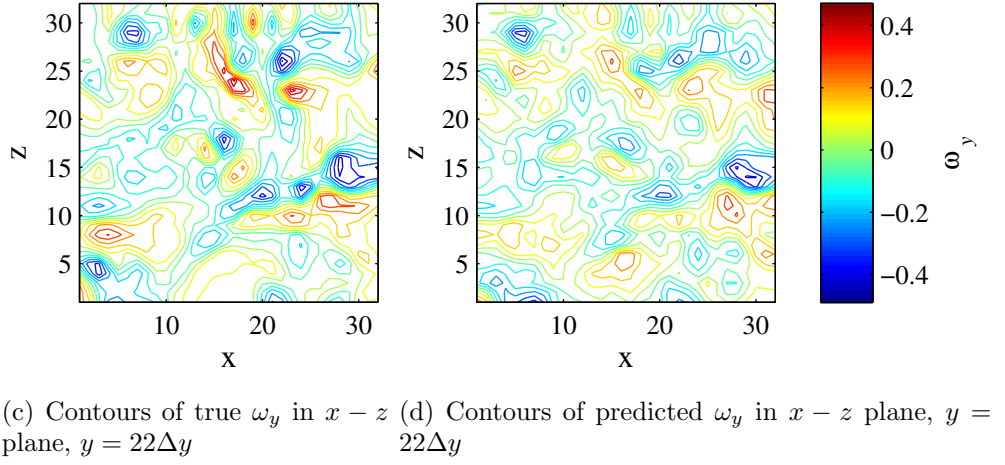
Figure 6.12: Comparison of true and predicted instantaneous vorticity fields

168

6.4   Practical Implementation of Technique

The basic principle of MP-VSE allows a great deal of flexibility in configuration; by tailoring the position, quantity, and spacing of the measurement planes, the technique can be put to a great many uses, over a wide range of flow phenomena. This section presents an overview of potential applications, along with other practicalities that must be considered in any experimental setup.

6.4.1   Potential uses of MP-VSE

Clearly, the ability to produce instantaneous 3-dimensional, 3-component velocity data lends itself to many applications, although it must be remembered that the results are predictions, rather than a measurement of the flow. Unfortunately this precludes use of MP-VSE for a number of tasks that would be possible with true volumetric measurements, particularly when the accurate measurement of statistical quantities is sought. Instead, the strength of MP-VSE is likely to be found in the study of coherent structures. The results from the previous section confirm its potential to accurately characterise the dominant flow structures in both the predicted velocity and vorticity fields. Also, as a form of stochastic estimation, MP-VSE lends itself naturally to many of the applications commonly found in this field, with the obvious benefit of providing a full 3D-3C reconstruction. In particular, the reconstructed velocity fields could potentially provide the basis for subsequent low-dimensional analysis and modelling using POD.

6.4.2 Temporal response

One aspect of MP-VSE not yet considered is the temporal response. For time-stationary flows, the rate at which predictions can be made is limited by the speed at which successive Q plane measurements can be acquired. Hence, with high-speed cameras and/or slow moving flows, the ability to make time-resolved volumetric predictions may be possible. As explained in section 6.3, there is an inherent difficulty producing predictions at multiple points in time for non-stationary flows, as a separate prediction model must be made for each time of interest. As such, MP-VSE is poorly suited for time-resolved predictions of non-stationary flows; even with the availability of high-speed cameras, the data and processing requirements to produce the prediction models would quickly become prohibitive.

6.4.3 Practical limitations to volume

Although it is unable to rival the accuracy of existing 3D-3C measurement techniques, MP-VSE does offer some unique advantages. Perhaps most importantly, neither the size nor resolution of the reconstructed volume are subject to the same limitations as other techniques (as discussed in section 2.2). The ability to predict large volumes is due to the extensibility of the technique, whereby the size of the volume can be increased with the addition of further Q planes. A wide variety of multiple plane PIV approaches exist (see section 2.2), which all have the potential to be used for MP-VSE, but each configuration appears to have an upper limit of the number of planes that can be measured in practice. Furthermore, the increase in the number of planes is likely to correspond to a increase in the cost of the equipment required. As such, there will always be an upper limit to the size of volume that can be

predicted.

Choice of number and positions of Q planes

A second, no less important limitation to the size of the prediction volume is due to the flow itself. Accurate prediction of the flow requires that the conditional vectors are well correlated with at least some of the unconditional vectors. Regardless of the flow in question, it can be expected that the correlation between two points will ultimately decrease as their separation becomes larger. Generally speaking, given a finite number of Q planes, it is likely that the accuracy of prediction will drop as the volume of interest is increased, and the spacing between the Q planes becomes larger. Conversely, a reduction in volume should be accompanied by an increase in accuracy.

Even though this effect serves to place a limit on both the volume and accuracy of MP-VSE, the ability to easily trade one for the other, and hence adapt the setup for the particular application, highlights the flexibility of the approach. For any given application, however, an appropriate choice for the number and position of Q planes will have to be made with regard to the behaviour of the flow. As a minimum requirement, some knowledge of structure size or correlation length would be necessary to make an informed decision on the spacing of the planes, and to provide an indication of the resulting predictive power. However, a more rigorous approach to Q plane positioning can be achieved if preliminary PIV measurements of the P plane are available first. As shown in section 6.2, the MP-VSE prediction model is derived from the P plane measurements, in a progress that begins by specifying the vectors in the P plane that correspond to the intersection of the Q planes. With this in mind, prediction models can be build from the preliminary P

171

plane measurements with Q planes in varying positions. Then, in the same way that cross validation is used to assess the relative performance of competing regression models, it can be applied to compare the prediction accuracy of the different Q plane configurations. Viewed in this light, the task of Q plane positioning is a straightforward mathematical optimisation problem where the aim is to maximise prediction accuracy, which is a function of the Q plane coordinates. In practice, this maximisation could achieved though any number of approaches, ranging from simple trial-and-improvement, through to a fully automated numerical optimisation.

6.5   Error sources

A requirement of any regression-based prediction is that the data used to build the regression model is taken from the same population as the data used in the subsequent predictions. However, the nature of the MP-VSE approach means that the measurement setup used to build the model is different to the setup used to provide the prediction data, which allows the potential for variations between the model data and prediction data. Furthermore, the majority of subsequent predictions are not made in the same position that the model was created, which may introduce further error if the flow is not perfectly homogeneous.

Thus, in the ideal case, the following criteria must be fulfilled:

1. the flow in question must have a homogeneous dimension, at least within the bounds of the measurement volume

2. the P plane must be normal to the homogeneous dimension of the flow

3. the Q planes must be parallel to the homogeneous dimension

4. the intersection of $Q1, \ldots, Q_l$ and P must be known exactly, so that the unconditional vectors in the prediction model are correctly defined

5. the distribution of the measurement error in the P and Q planes must be identical.

Any deviation from these criteria will introduce error into the predictions, beyond what is indicated by the process of cross-validation. Of course, the likelihood of meeting any of these criteria in practice is unlikely, must it is nonetheless important to ensure that these requirements are fulfilled to the extent possible. Then, once all reasonable steps have been taken, the experimental setup and flow properties must be assessed in order ensure that any error is within tolerable limits. Section 7.4 provides details of this process.

It can be seen that three of these criteria relate to the accurate positioning of the PIV light sheet planes. Given the importance of this aspect of the experiment, a simple and accurate light sheet measurement system has been developed, which provides the means of detecting the alignment of a light sheet, relative to an absolute reference, as well as to the position of another light sheet. This approach is discussed in Appendix C.

6.6   Conclusion

The technique of MP-VSE has been introduced. A proof of concept has been performed on simulated channel flow data, which has shown the potential of the technique to predict a high percentage of the flow variance, as well as the ability to accurately capture the large scale structure in both the velocity and vorticity fields. The results also indicate that the biased techniques, especially PCR and PLS, are ideally suited for use in MP-VSE applications, due to the

173

close spacing, and hence high multicollinearity, in the unconditional vectors. Potential applications of MP-VSE, and the practicalities of implementing the procedure have been considered.

Chapter 7

Experimental Prediction of Vortex Rings

<div align="center">(a) Laminar vortex ring   (b) Unstable vortex ring with an
$n = 8$ azimuthal mode</div>

Figure 7.1: Visualisation of vortex rings in water. Dye has been injected into the rings at formation, which reveals the position and shape of the ring core. *(Courtesy of M. Brend, University of Warwick)*

7.1 Introduction

This chapter describes the practical application of the MP-VSE technique, which was used to investigate the three dimensional behaviour of unstable vortex rings. The nature of a vortex ring's development is determined by the conditions of its creation (Lim and Nickels, 1995). Commonly, vortex rings are formed by the ejection of fluid through an orifice. A low ejection velocity will produce a laminar vortex ring, as shown in figure 7.1 (a), that will propagate smoothly and decay naturally due to viscosity. At moderate velocities, a laminar vortex ring will form, but small perturbations on the ring are amplified, leading to the production and growth of wavy instabilities on the core (figure 7.1 (b)). These instabilities continue to grow until a point where they quickly break down, culminating in a turbulent ring. At sufficiently high ejection velocities, the ring will be turbulent from the outset.

The following work is concerned with moderate velocity rings, specif-

ically during their period of instability. This stage of development is itself divided into two distinct stages, known as the linear and non-linear phases. The first, which is characterised by the growth of stationary azimuthal waves around the circumference of the ring, is well understood thanks to a wealth of theoretical, numerical and experimental investigation (which is reviewed in Lim and Nickels (1995)). The second stage, however, has received less attention (Dazin *et al.*, 2006b). The most commonly observed phenomenon in this latter stage is the presence of a net "swirling" flow around the circumference of the ring, which has been observed experimentally (Maxworthy, 1977; Naitoh *et al.*, 2002; Dazin *et al.*, 2006b) and in numerical simulations (Shariff *et al.*, 1994; Archer *et al.*, 2008). The advancement of CFD and flow measurement technology has recently lead to the observation of further phenomena, which include the development of submodes and harmonics of the dominant mode from the linear instability stage, and the formation of a 3-dimensional vortical structure that wraps around the vortex ring core (Shariff *et al.*, 1994; Dazin *et al.*, 2006b; Archer *et al.*, 2008). So far, the latter has only been observed in numerical studies (Shariff *et al.*, 1994; Archer *et al.*, 2008) and in qualitative planar laser-induced phosphorescence (PLIF) measurements (Dazin *et al.*, 2006b).

As such, there is considerable worth in the use of MP-VSE to investigate these phenomena, which provides a means of observing unstable vortex rings experimentally, quantitatively and in three dimensions.

While an investigation of this nature is a worthwhile pursuit in its own right, it also provides an ideal yet challenging flow to test the performance of MP-VSE. Statistically, the vortex rings are axisymmetric, but individually, their structure exhibits a combination of both randomness and periodicity of arbitrary phase, and their trajectory and orientation is subject to a large

degree of deviation from the average. The ability to capture such behaviour will be a major step in establishing the merit of the MP-VSE concept.

The following chapter is organised as follows. To begin, section 7.2 provides a general overview of the experimental facility used in this work, and describes the specific form of the MP-VSE procedure applied in this setting. A detailed discussion of the experimental setup follows in section 7.3.

Prior to carrying out the volumetric predictions, the statistical properties of the stereo PIV data were analysed to assess the accuracy of the light sheet alignment, and confirm the mean-axisymmetry of the vortex rings. These results are given in section 7.4. Section 7.5 provides the results of the cross-validation procedure, which is used to select the most appropriate regression technique for the current application. In section 7.6, a selection of reconstructed vortex rings is presented and their behaviour is compared with findings from previous literature. Conclusions and further work are given in section 7.7.

7.2 Overview of MP-VSE Procedure

The MP-VSE technique was applied to produce volumetric predictions of water based vortex rings. The investigation was limited to a single point in the lifetime of the vortex rings, which corresponds to the non-linear phase of development. The rings were formed by pumping a finite volume of water into a submerged box with a circular orifice. The fluid ejected from the orifice forms a laminar vortex ring, which propagates forwards, and eventually becomes unstable.

A preliminary dye visualisation experiment was carried out to determine an experimental configuration capable of producing rings of this nature, and
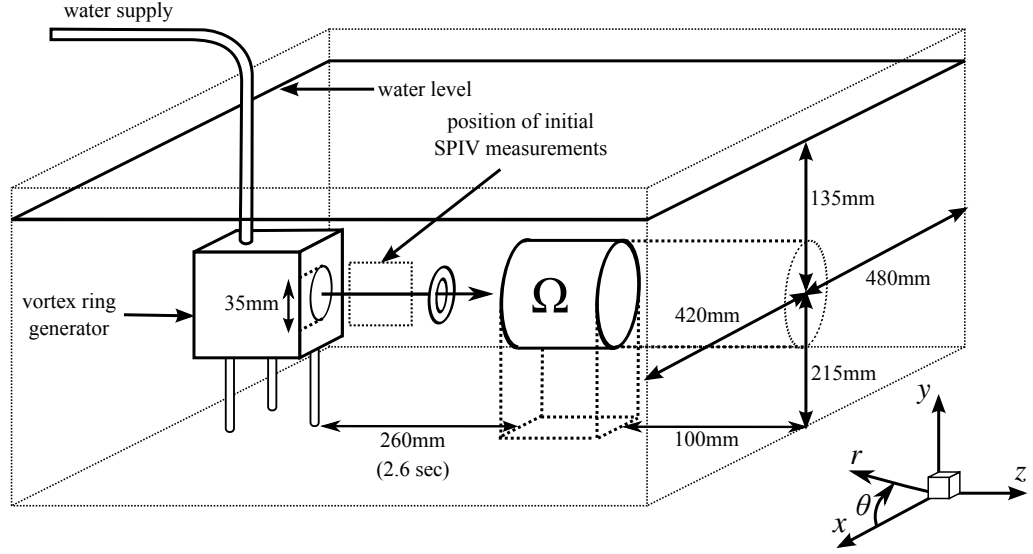
Figure 7.2: Experimental configuration

subsequently identify the average position and time at which the non-linear phase occurs. This was used to determine the location of the prediction volume Ω for the subsequent MP-VSE experiment, as shown in figure 7.2.

The characteristic properties of the vortex rings were calculated from stereo PIV measurements of the rings just after generation, at the position indicated in 7.2. One hundred vortex rings were measured, from which the average initial ring radius R_0 , core radius a_0, ring translational speed V_0, and circulation Γ_0 were calculated. These properties are depicted in figure 7.3. The circulation was obtained by calculating the velocity line integral along a rectangular path around the outside of the ring core (Fouras and Soria, 1998). The average Reynolds number (as defined by Glezer (1988)) was then computed as $Re_0 = \Gamma_0/\nu$, where ν is the kinematic viscosity. The resulting parameters are provided in table 7.1.

Figure 7.4 shows the geometry of the prediction volume Ω, along with the location and orientation of the coordinate systems used in this work. The
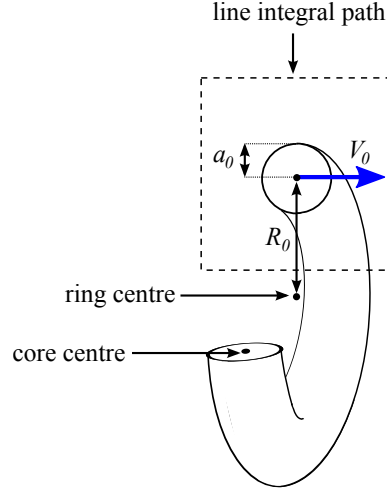
Figure 7.3: Overview of parameters obtained from the initial stereo PIV experiment. The rectangular box indicates the approximate position of the circuit used to calculate the circulation.

Parameter	Value
Circulation, Γ_0	$0.0045 \text{ m}^2/\text{s}$
Kinematic viscosity, ν (water @ 15°C)	$1.14 \times 10^{-6} \text{m}^2/\text{s}$
Reynolds number, $Re_0 = \Gamma_0/\nu$	≈ 4000
Ring radius, R_0	0.02 m
Core radius, a_0	0.007 m
Ring translational speed, V_0	0.05 m/s

Table 7.1: Summary of average vortex ring parameters upon generation. All quantities are averaged from 100 individual rings, and were obtained from stereo PIV measurements at the position indicated in figure 7.2.
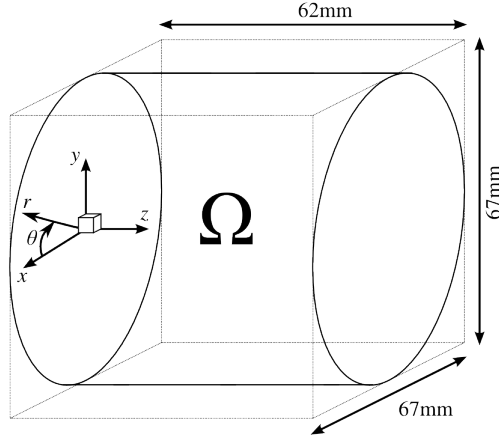
Figure 7.4: The prediction volume Ω. It is assumed that the flow in question is homogeneous in θ.

position of Ω was chosen so that the mean path of the vortex rings passed directly through the centre of the volume and along the z dimension. In this configuration, the homogeneous dimension of the vortex ring flow corresponds to the azimuthal dimension, θ, of the cylindrical coordinate system (r, θ, z). For convenience, both the Cartesian (x, y, z) and cylindrical coordinate systems share the same origin and a common axis, z.

As described in the MP-VSE theory chapter (chapter 6), the procedure requires two separate experimental configurations. The first configuration involves the acquisition of data at the plane P, to provide the data to build the prediction model. Plane P was positioned in the $y - z$ plane at $x = 0$, normal to the homogeneous dimension of the flow, θ (figure 7.5 (a)) For the second configuration, the position of the Q planes must then be parallel to θ, which corresponds to the $x - y$ plane. Three Q planes were used, positioned according to the configuration shown in figure 7.5 (b).

The unconditional vectors of the prediction model were determined by
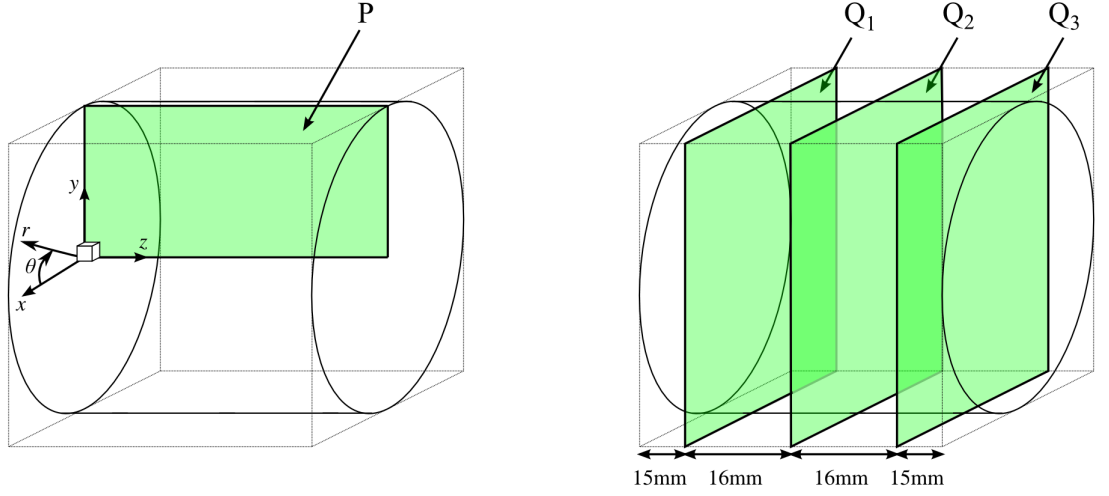
(a) Configuration 1: Position of plane P, which is used to build the prediction model

(b) Configuration 2: Position of planes Q_1, Q_2 and Q_3, which provide the instantaneous measurements for the prediction

Figure 7.5: Measurement plane configuration for the prediction of Ω

the intersection of the P and Q planes (figure 7.6) and the full 33×62 vector field was used as the conditional vectors.

Given an instantaneous set of observations of the Q planes (figure 7.7 (a)) , the volume is created by predicting planes of vectors at successive positions along the θ axis. At each θ, the corresponding line of vectors is extracted from the Q planes to use as the unconditional vectors in the prediction model (figure 7.7 (b)). However, the vectors in the Q planes, which are obtained from stereo PIV measurements, are arranged on a uniformly-spaced grid and must first be interpolated to match the required positions for use in the prediction model, as shown in figure 7.7 (c). In the present work, the planes were reconstructed at 88 equally spaced positions along θ. Finally, the resulting volume was re-interpolated back onto a Cartesian grid, resulting in a $67 \times 67 \times 62$ grid, with $\Delta x = \Delta y = \Delta z = 1$mm.
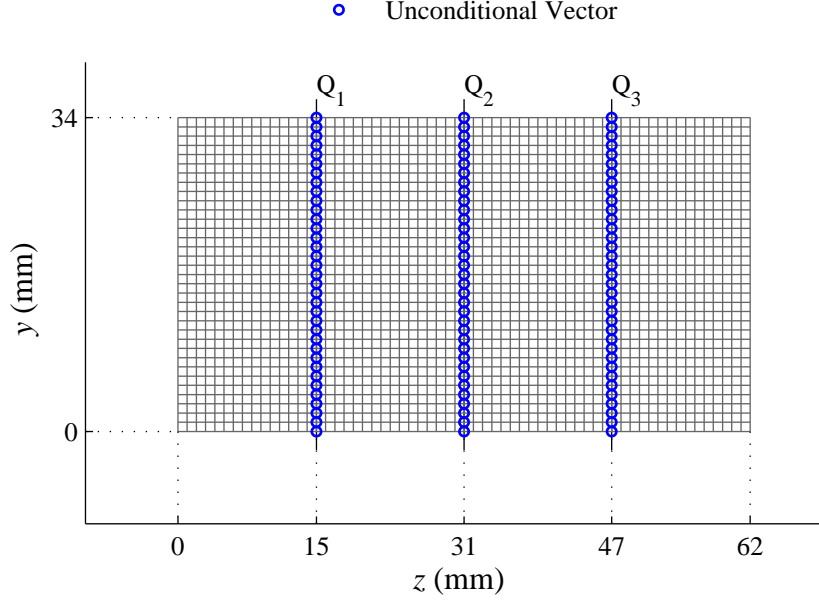
Figure 7.6: Location of the unconditional vectors in the prediction model

7.2.1 Note regarding the number and position of Q planes

The decisions regarding the number and position of the Q planes in the experimental setup was determined, to a large part, by practical considerations. The choice of three planes, for example, was dictated by the number of lasers available for the experiment. The use of more Q planes would almost certainly have improved the predictions, but was simply not possible. In order to maximise accuracy using the available Q planes, the size of the volume Ω was then kept as small as possible, while still ensuring that the significant majority of the generated vortex rings were within the volume when the PIV measurements were triggered. Other decisions were made for convenience; it was decided that separation between Q planes would be constant, and that the Q planes would intersect exactly with the vector grid of the P plane, in order to simplify the alignment and prediction stages of the MP-VSE pro-

183
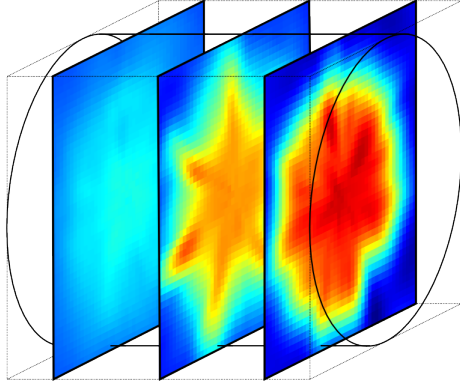
(a) The process begins with a set of instantaneous measurements of $Q_1, \ldots, Q_l$ at time t
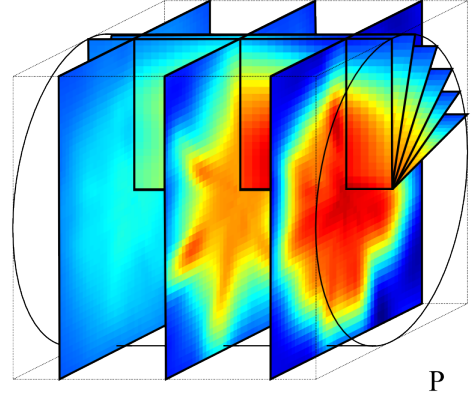
(b) The vector values along each θ in Q_1, Q_2 and Q_3 are used to predict the corresponding plane
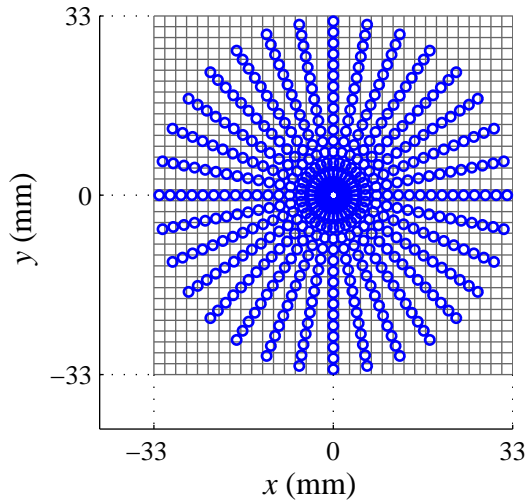
(c) In order to extract unconditional vectors from the Q planes, the vectors must be interpolated on to a polar grid (only 32 azimuthal angles are shown here; 88 are used in practice
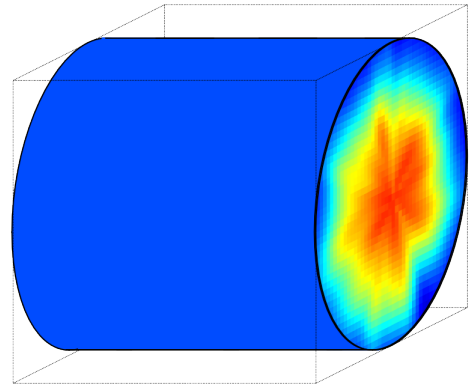
(d) The full set of reconstructed planes provides the volumetric prediction of Ω at time, t

Figure 7.7: Overview of the volumetric prediction process

cedure. However, within these constraints there were a number of potential configurations that had to be chosen from. For this purpose, a simple cross-validation procedure was employed to assess the relative prediction accuracy of the competing configurations (as described in section 6.4.3). This required the P plane PIV system to be installed prior to the Q planes, so that P plane measurements could be used for the cross validation procedure. The Q plane configuration shown in figure 7.5(b), was found to offer the highest accuracy out of a broad, albeit incomplete, range of possible options (a more rigorous comparison was deemed unnecessary as the accuracy was relatively insensitive to small changes in Q plane position).

7.3 Experimental Configuration

The experimental set-up is shown in figures 7.8 and 7.9. The vortex rings were produced in a transparent Perspex tank, which provided optical access for both the light sheets and camera systems. Submerged in the tank was a vortex ring generator (provided courtesy of Prof. P. Thomas, University of Warwick), which comprised a box of dimensions of approximately $300 \times 300 \times 300$ mm, with five sides made out of Perspex, and the 6th being a metal plate with a 35mm diameter precision-machined circular hole in the centre. Water was injected via an inlet pipe on the top of the box, expelling a slug of water out of the orifice in the metal plate, resulting in the formation of a vortex ring. The inlet pipe was connected to the mains water supply, via a solenoid valve and a flow control valve. Injection of water into the generator was obtained by briefly opening the solenoid valve. The flow control valve ensured a constant flow velocity into the generator, regardless of fluctuations in the mains water pressure. The generator was mounted on three height-adjustable legs, the front of which was fixed to a position in the tank but able to rotate. One of the rear legs was attached to a micrometer traverse, which allowed the horizontal adjustment of the leg's position. This caused the tank to rotate around the front leg, and allowed precise adjustment of the vortex ring's trajectory.

Measurements were obtained using an individual stereo PIV system for each plane. For the measurement of the Q planes, discrimination of the individual sheets was achieved by introducing a small temporal separation between each sheet. The full measurement system can be categorised into three parts, consisting of the camera systems, the light sheet generation system and the control system. Each part is described in detail in the following sections.

186
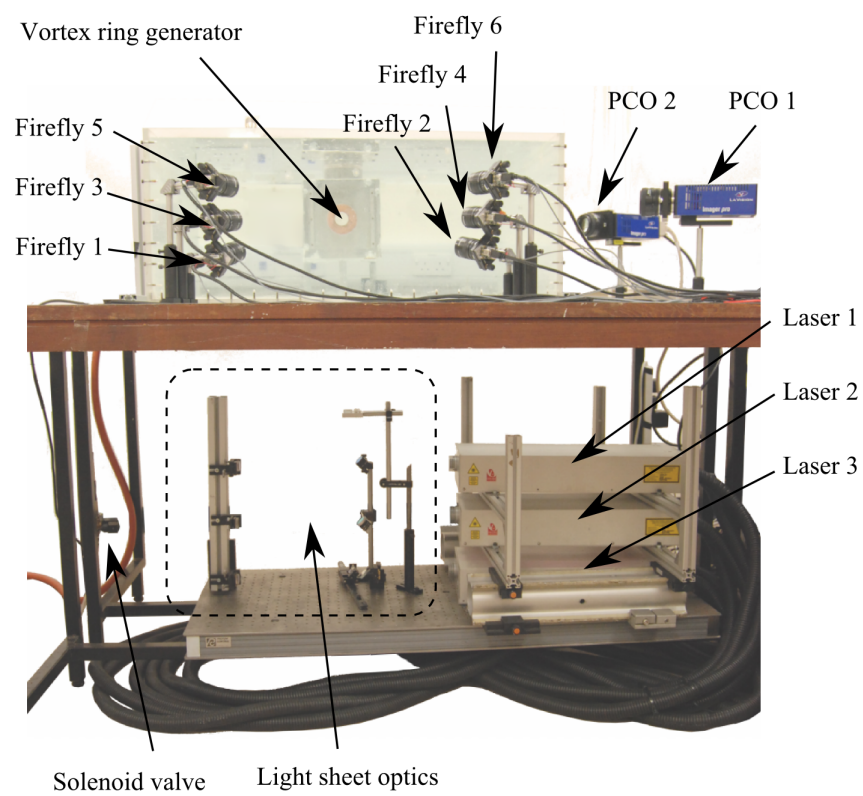
Figure 7.8: Side-view of experiment

187
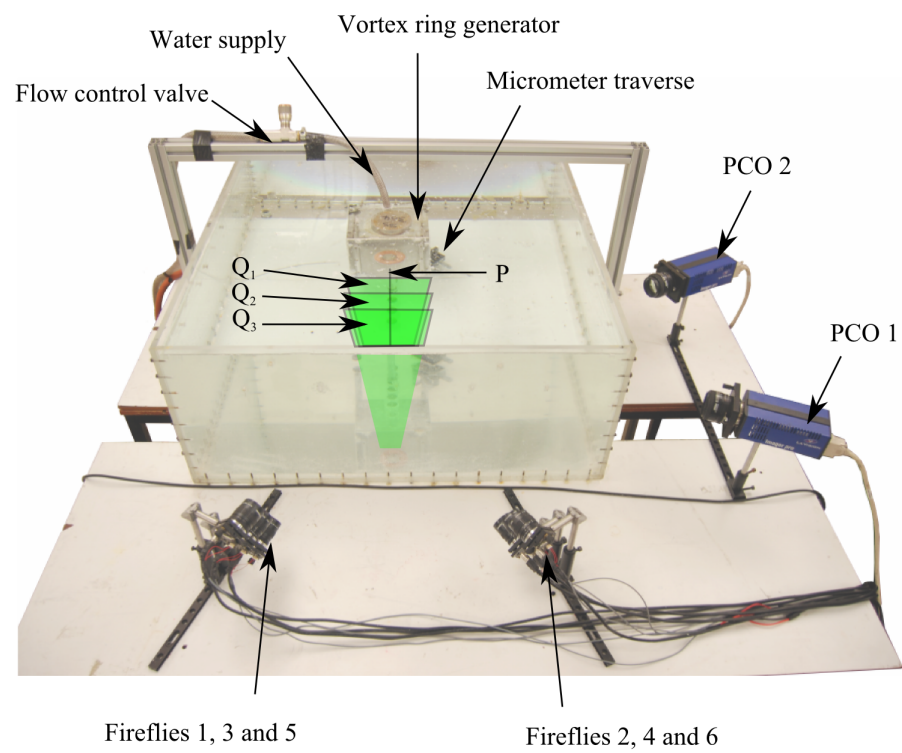
Figure 7.9: Top-view of experiment, overlaid with the position of the P and Q planes.

7.3.1 Camera System

The four separate stereo PIV systems required the use of eight cameras in total. The measurements of the P plane were obtained using two PCO Gmbh. PCO.2000 cameras with Nikon 50mm f1.4 lenses (figure 7.10 (a)). For the three Q planes six Point Gray Research Inc. Firefly MV board cameras with Cosmicar (Pentax) 25mm 1.4 lenses were used (figure 7.10 (b)). The full specification of both types of camera is provided in appendix D.4. Although the Firefly cameras have a considerably lower specification than the PCO.2000 cameras, they were nonetheless suitable for the low-speed flow in the present application, and offered excellent value for money (all six cameras were purchased for around \$1300). They also have the major benefit of hardware triggering, which allowed for the precise synchronisation required for the temporal separation method of light sheet discrimination.

Each of the stereo PIV systems employed the *angular displacement* method of arrangement (Prasad, 2000), which meant that the cameras viewed the light sheet at an oblique angle. In a standard camera system the image plane, lens plane and the plane of focus are all parallel to each other, which means that the plane of focus would not coincide with the light sheet plane in this configuration. As such, achieving accurate focus across the entire field of view can be difficult; a small lens aperture may produce a sufficiently large depth of field to focus the full light sheet, but this comes at the expense of a reduced light intensity. This problem is avoided by employing a *Scheimpflug* configuration to the cameras, which allows a camera to focus on an off-axis plane by rotating the lens and/or sensor so that the image plane, lens plane and object plane (i.e. the light sheet) intersect along a common line (Walker, 2002). This principle is illustrated in figure 7.11.

189
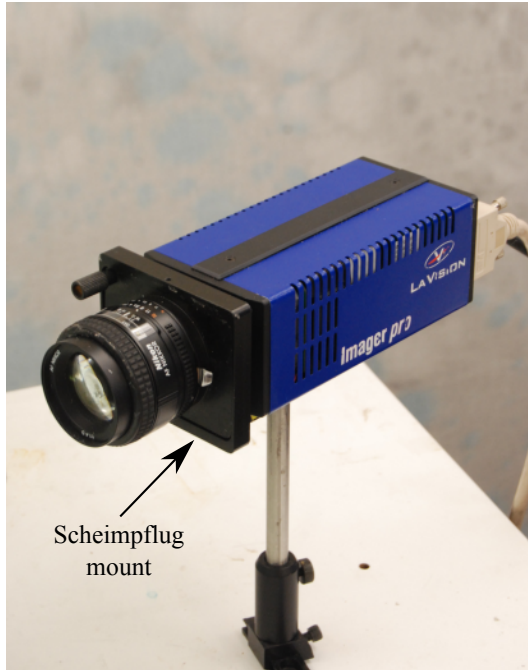
(a) PCO camera 1 with Nikon 50mm lens and Scheimpflug adaptor, used for P plane measurement.
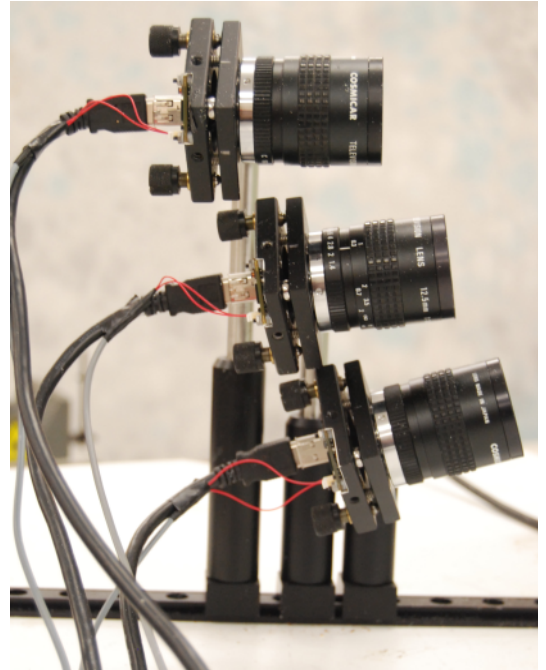
(b) Firefly cameras 1, 3 and 5 (bottom, middle and top, respectively) fitted with Cosmicar 25mm lenses and custom Scheimpflug adaptors, used for Q plane measurements.
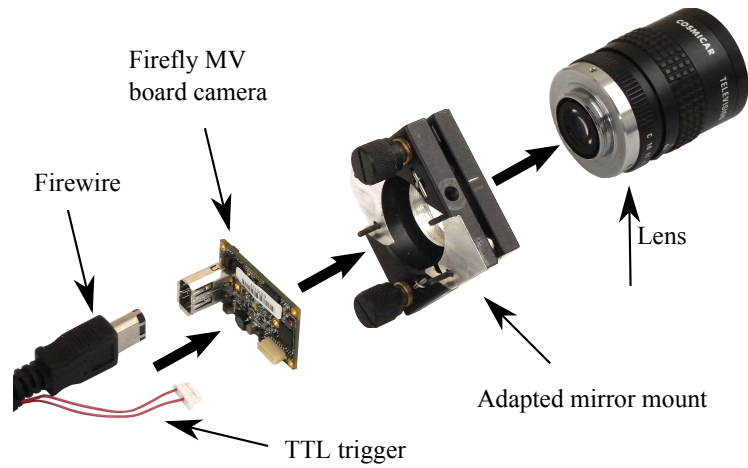
(c) Individual components of a Firefly camera system

Figure 7.10: Configuration of camera systems used for stereo PIV measurements.

At the time of the experiment, Scheimpflug adaptors were available for the PCO cameras, but not for the Firefly cameras. However, the need to configure the Firefly cameras in the Scheimpflug arrangement was particularly important, due to a combination of the poor light sensitivity of the cameras which necessitated a large aperture, and the large angle between the cameras and the light sheet planes (as shown in figure 7.10 (b)). A further complication arose from the fact that the bottom two sets of Firefly cameras were also tilted about a horizontal axis (figure 7.9). As such, the lower cameras necessitated a two-axis Scheimpflug configuration, as shown in figure 7.11 (b). A solution was obtained by modifying a set of Thor Labs kinematic mirror mounts to act as Scheimpflug adaptors. This involved the addition of a C-mount lens thread, and removing part of the back of the mount to house the Firefly board at the correct distance from the lens. The resulting camera configuration is shown in figure 7.10 (c).

After positioning and focusing the cameras, the camera systems were calibrated using LaVision Davis 7.2 software. A camera pinhole model was employed, which was fitted using an image of a 3-dimensional target (shown in figure C.14 of appendix C.4), placed in the light sheet plane. The calibration was then refined using the built in self-calibration procedure described by Wieneke (2005).

7.3.2 Light Sheet Generation

Four separate light sheets are required for the experiment, each produced by the expansion of a collimated laser beam with a cylindrical lens. As only three light sheets were ever required at one time, only three lasers were needed, with one laser used for the measurement of both P and Q_3. Three Litron

191
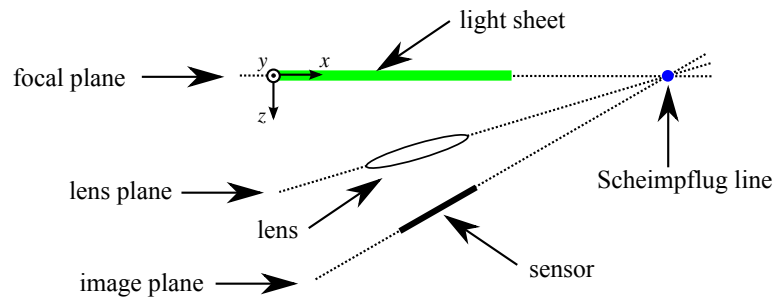
(a) 2-dimensional view of a one-axis Scheimpflug condition, which is met by rotating the lens plane around a single axis of the coordinate system.
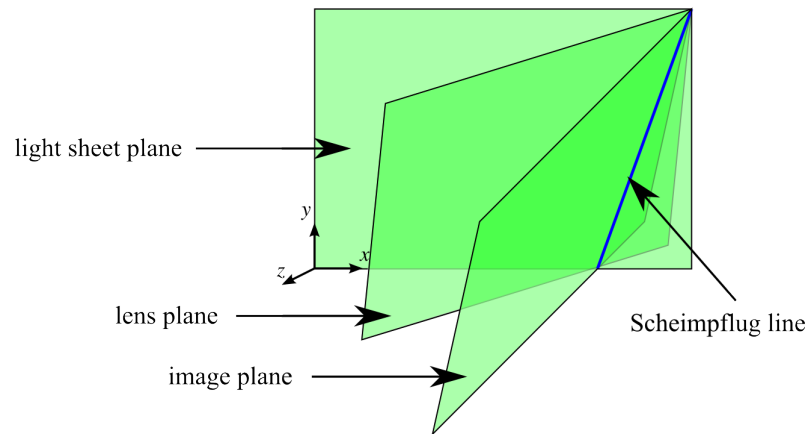
(b) A two axis Scheimpflug condition. The lens plane must be rotated around two axes of the coordinate system to meet the Scheimpflug criterion.

Figure 7.11: Two examples of the Scheimpflug principle

Nano S PIV Nd:YAG lasers are utilised for this purpose. Each laser contains two separately controlled 1064nm heads, each with a pulse duration of 4ns. The output of both heads passes through a single frequency doubling crystal to produce a double-pulsed 532nm visible beam, rated at 140mJ. The ability to control the individual pulses allows the accurate control of pulse separation Δt. For this work, the choice of Δt was dictated by the minimum inter-frame time of the Firefly cameras, which was 17ms.

The choice of this parameter, along with the light sheet thickness, should allow for sufficient particles in the first PIV image to be present in the second, which is determined by the maximum speed of the flow in the out-of-plane direction, w_0. As a guide it is suggested that $w_0 \Delta t / \Delta z_0 < 0.25$, where Δz_0 is the light sheet thickness (Keane and Adrian, 1990). For the Q plane measurements, which contain predominantly out-of-plane motion, the maximum out-of-plane velocity was determined to be approximately $w_0 = 0.08$ m/s. With the minimum possible Δt of 17 ms, this suggests that the light sheet thickness should be around 5mm. However, for convenience, a sheet thickness of 4mm was chosen, which corresponded to the diameter of laser beams and allowed for a simpler optical arrangement for the light sheet generation. The P plane measurements used identical values of Δt and Δz_0 for consistency.

Light sheet discrimination was achieved by introducing a slight separation, $\delta t = 160 \mu s$, between the pulses of the light sheets for Q_1, Q_2 and Q_3, to allow each camera system to acquire an image of the relevant light sheet without interference from the others. The sequence of laser pulses is shown in figure 7.12.

Theoretically, MP-VSE requires the Q plane measurements to be acquired instantaneously, although it can be seen from the diagram that the measurements were actually acquired over a period of $2\delta t = 320 \mu s$. For this
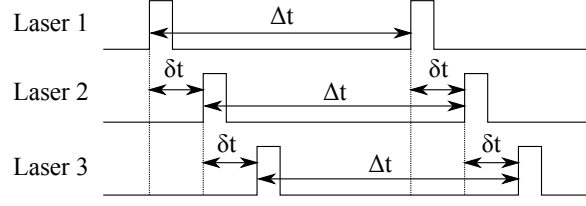
Figure 7.12: Laser pulse sequence for Q plane measurements. Pulse separation time $\Delta t = 17$ms, time between light sheets Q $\delta t = 160$ μs.

approach to be valid, this period must be sufficiently small compared to the timescales of the phenomena under investigation, so that the flow is effectively stationary over the period of acquisition. An indication of the timescales involved can be inferred for the numerical study of (Shariff *et al.*, 1994), which modelled the growth of vortex rings up to the early non-linear phase. In this work the time variable was non-dimensionalised using the characteristic timescale R_0^2/Γ_0. The subsequent simulations used a minimum time step of 0.05 time units, which was sufficient to capture the behaviour of the ring. Applying the same procedure here, the acquisition period of 320 μs corresponds to a non-dimensionalised time-step of 0.02 time units. This would suggest that a 320 μs period is sufficient to ensure the stationarity of the ring.

As described in section 6.4, the successful application of MP-VSE requires the fulfilment of a number of criteria, three of which directly relate to the alignment of the light sheets:

1. the P plane must be normal to the homogeneous dimension of the flow

2. the Q planes must be parallel to the homogeneous dimension

3. the intersections of Q_1, Q_2 and , Q_3 with P must be known, so that the unconditional vectors in the prediction model are correctly defined

For the experiment considered here, the task of meeting these require-

194

ments was aided with the light sheet measurement technique proposed in appendix C, and was further simplified with the use of a custom-built cylindrical lens mount shown in figure 7.13.

The mount consists of two rails, positioned at exactly 90° to each other, which provide a physical reference for the cylindrical lens position; each of the lenses was placed with its side in contact with the rail which ensured that the optical axes of the three cylindrical lenses for the Q planes were exactly parallel, and that the axis of the P plane cylindrical lens was exactly normal to the Q plane lenses. 4mm apertures were then placed on the base of the mount (see figure 7.13 (b)), to ensure the three Q planes had the correct separation (16mm between the centre of each sheet).

An unconstrained plane has three degrees of freedom (DOF), each of which must be set during the process of alignment. However, using the lens mount, only one of the planes needs to be positioned using the full three DOF. Once this plane is aligned, the remaining planes only have a single DOF left to adjust. Assuming that the first plane is positioned accurately, the constraints imposed by the mount guarantee the accuracy of the remaining planes' positions, at least up to the adjustment of the final DOF. Also, this approach has the considerable benefit of accelerating the alignment process.

The complete light sheet forming system is shown in figure 7.14 (a). The cylindrical lens mount was placed directly under the Perspex tank. The beams were directed to the lens mount using a system of mirrors on kinematic mounts, which provided fine adjustment of the light sheet positions. The process of moving laser 3 from the P plane configuration to the Q plane configuration was simplified with the use of a mirror mounted on a micrometer traverse. This could be moved into the path of the beam from laser 3, which diverted it from the Q cylindrical lens (figure 7.14 (b)) to the P lens (figure 7.14 (c)).
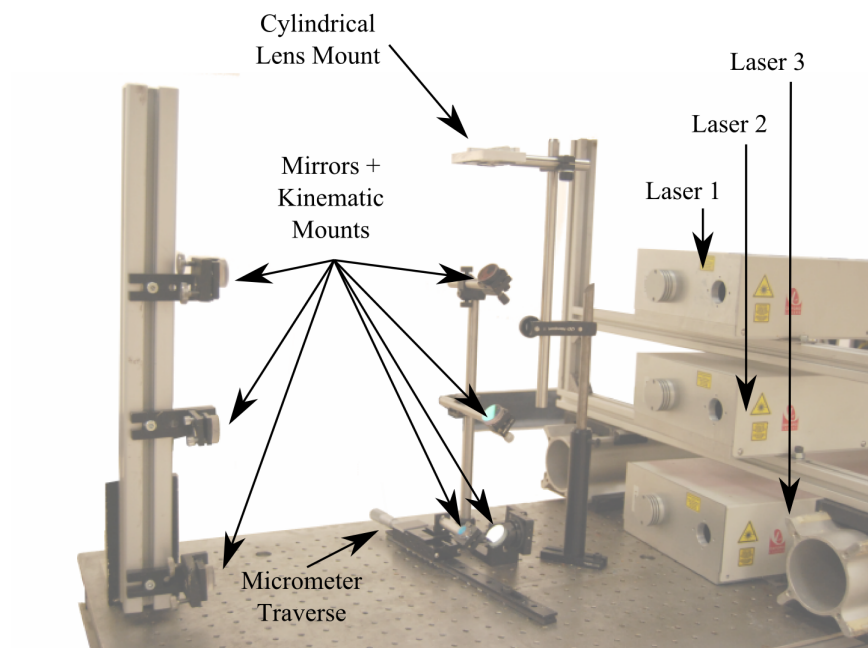
195
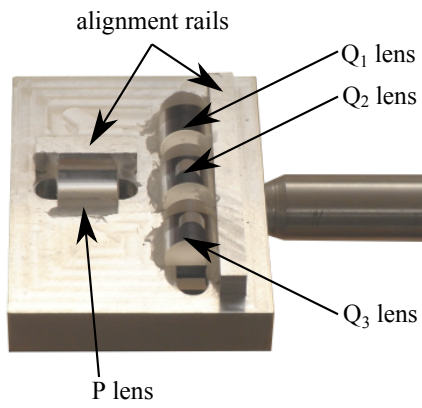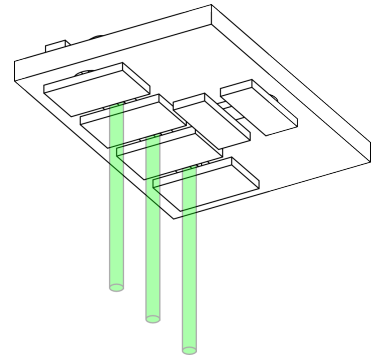
(a) Top view

(b) Bottom view of the mount, showing the position of apertures and path of the Q plane laser beams
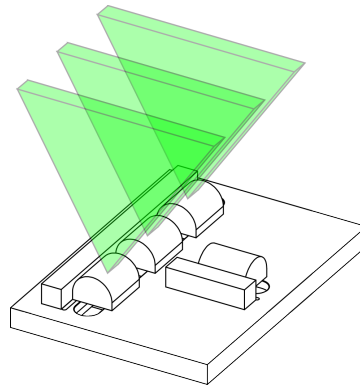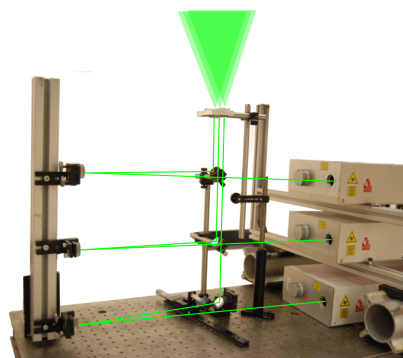
(c) Top view of the mount, showing the generation of the Q plane light sheets
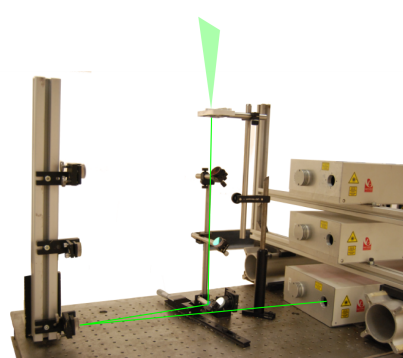
Figure 7.13: Mounting platform for cylindrical lenses

(a) Light sheet forming system

(b) Q plane configuration

(c) P plane configuration

Figure 7.14: The light sheet generation process

The alignment process was carried out in the stages shown in figure 7.15. The first stage involves the alignment of the P plane, which was facilitated by attaching a long, straight target to the top of vortex generator, directly above, and perpendicular to, the orifice. The light sheet and vortex ring generator were then adjusted so that the sheet directly intersected the target (figure 7.15 (a)). The correct alignment was confirmed by inspecting the ensemble average of measurements from the P plane stereo PIV system, to confirm that the average vortex ring passed directly through the sheet. The next step was to align the Q_1 plane (figure 7.15 (b)). Although this plane only had a single DOF to adjust, this stage was the most time consuming; unlike before, there is no physical reference with which to position the sheet, and so the position of the sheet had to be assessed solely from measurements with the Q_1 stereo PIV system, which were used to check the symmetry of vortex rings passing through the sheet. Once this plane was set, however, it was then possible to employ the light sheet measurement technique (described in appendix C) to position the remaining Q planes (figure 7.15 (c)). First, the light sheet measurement frame was used to align Q_2 relative to Q_1, and then the process was then repeated for the position of Q_3 relative to Q_2. Finally, in order to calculate the positions of the intersections of the Q and P planes, a ground glass screen was placed directly in the path of the P plane (figure 7.15 (d)). The three Q plane light sheets were then generated, which projected lines onto the screen (figure 7.15 (e)). An image of the projected lines was acquired using the P plane stereo PIV system, which provided a means of identifying the exact position of the Q plane intersections within the P plane measurements.

Upon completion of this process, a final validation was performed using stereo PIV measurements from each of the planes. This was to ensure that the three criteria mentioned above, as well as the remaining criteria from section

198
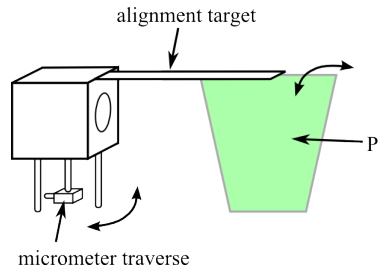
(a) A thin alignment target, perpendicular to the vortex generator plate, is used to align the P plane with the direction of the vortex ring flow
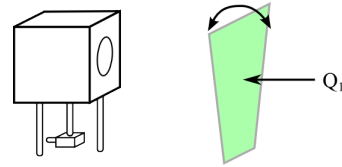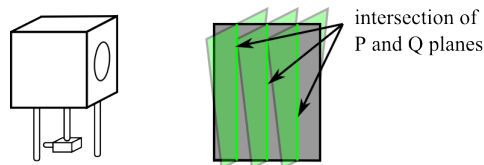

(b) Plane Q_1 is positioned to be exactly normal to the direction of flow, which is determined from PIV measurements of vortex rings passing through the Q_1 plane.
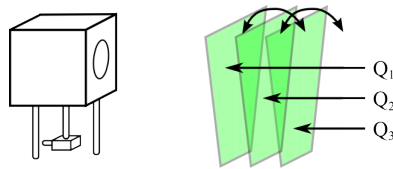
(c) Planes Q_2 and Q_3 are aligned parallel to Q_1, using the proposed calibration technique
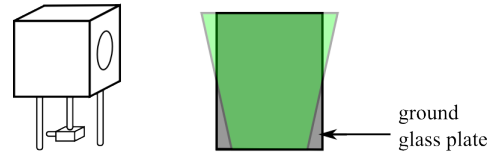
(d) A thin ground glass screen is placed directly in the P plane

(e) The projection of the Q planes on the screen is used to obtain the intersection of the P and Q planes

Figure 7.15: Light sheet alignment process

6.4, were adequately met. The results from the validation are presented in section 7.4.

7.3.3   Control System

Almost all aspects of the experiment were automated and operated remotely by computer. The control system is detailed in figure 7.16. At the centre of the system was a purpose-built timing generator (figure 7.17), which provided the triggering signals to the cameras and lasers, as well as the 24V dc power supply for the solenoid valve which initiated the formation of the vortex rings.

The timing box was built around a Microchip dsPIC30 microcontroller, which was programmed in C. Accurate generation of timing signals is achieved via the on-board 32 bit timer, which generates an interrupt signal whenever the state of the output channels needs to be changed. A serial port connection allows communication with a PC, in order to download the timing sequence programme to the timing box, which is stored to non-volatile memory. The serial connection is also used to initiate the timing sequence, although the box could be operated without connection to a PC if necessary. All cameras and lasers were triggered using TTL-level signals. As described in section 7.3.2, each laser consists of two separately triggered heads. Each head in turn requires two separate TTL-level signals, one for the flash lamp, and one, 160 μs later, to initiate the pulse by triggering the Q-switch. Therefore, firing each laser requires four trigger signals, and so 12 TTL channels were needed in total. The timing generator has a total of eight TTL-level channels, so a demultiplexer system (courtesy of Dr Andrew Skeen, University of Warwick) was incorporated into the system. This allowed both the flashlamp and Q-switch of each laser head to be controlled with a single channel. As such, six

Figure 7.16: Overview of control system.

(a) Front view   (b) Rear view

Figure 7.17: Timing generator.

channels were used to control the lasers, with a final channel used to control the demultiplexer.

Due to the shortage of channels, camera triggering was obtained by combining the signals from the two corresponding flash lamps using a NOR-gate (the negation was required because the lasers are triggered from the rising edge of the signal, and the cameras from the falling edge).

The images from the cameras were downloaded to either the master PC or one of the slave PCs. Images on the slave PCs were then transferred, via a 1 gigabit Ethernet connection, to a raid array on the master PC. Control of, and acquisition from, the Firefly cameras was achieved using software written by the author, which was based on the Point Gray Research Inc, FlyCapture software development kit. Acquisition from the PCO cameras was handled with the provided PCO CamWare software.

Control of the master PC was possible with two terminals; terminal one was placed behind a safety screen, and was used for control and monitoring purposes while the experiment was running. Terminal 2 was placed in close proximity to the experiment, which allowed convenient access to the live feeds

from the cameras during camera alignment and focusing, and similarly allowed access to the software for the light sheet position measurement system.

7.3.4 Overview of Stereo PIV procedure

The MP-VSE experiment consisted of stereo PIV measurements of 300 vortex rings in the P plane configuration, and 80 rings in the Q plane configuration. In both configurations, the measurements were made 2.6 seconds after generation of the vortex ring and at one minute intervals, to allow the water in the tank to settle. Prior to the procedure, the Perspex tank was seeded with Plascoat Talisman 30, a powdered coating for metal with a particle size of $100-200$ μm (Yoshimoto, 2009). A high concentration of seeding was also placed directly in the vortex generator to ensure that there was sufficient seeding within the vortex core. After each run of 50 vortex rings, the experiment was paused, so that further seeding could be added to the vortex ring generator, and to drain water from the tank to keep the water level roughly constant.

Calculation of the stereo PIV vector maps was performed using LaVision DaVis 7.2, using iterative multi-pass correlation with a reducing window size. For all correlations, the final window size was chosen to be as close to 4 mm as possible (12×12 pixels for Firefly MV cameras, 32×32 pixels the PCO.2000 cameras), in order to match the light sheet thickness. This way, all vector measurements were derived from an approximately cubic volume of size $\approx$ $4 \times 4 \times 4$ mm, which ensured that the degree of spatial averaging inherent in the PIV processing was the same for the in and out-of-plane velocities, and consistent between the Q and P plane measurements.

A simple estimate of the measurement error was obtained by assuming a cross-correlation uncertainty of 0.1 pixels, a figure which has been shown to

Parameter	Value
Total number of measurements	300
Valid measurements	288
Pulse separation, Δt	17ms
Particle image size	2048×2048 pixels
Final correlation window size	32×32 pixels
Correlation window overlap	87%

Table 7.2: Summary of P plane stereo PIV parameters

be valid in idealised conditions (Tropea *et al.* (2007)). Based on the resolution, field of view and Δt of the camera systems, this equates to an uncertainty of 2×10^{-3} m/s for the Q plane measurements and 7×10^{-4} m/s for the P plane measurements. Expressed as a percentage of the RMS velocity magnitude of the measurements, this translates as $< 3\%$ and $< 0.5\%$ for the Q and P planes respectively. As the Q plane measurements would go on to be interpolated during the MP-VSE procedure, the cross-correlation was performed with a high window overlap (87%), to ensure that a dense vector grid was available for the interpolation process.

After the experiment, inspection of the stereo PIV vector maps revealed that a small number of vortex rings had failed to form correctly and/or had trajectories that were outside of the measurement planes. These measurements were discarded, leaving 288 P plane measurements, and 66 Q plane measurements.

The parameters of the stereo PIV procedure for the P and Q planes are summarised in tables 7.2 and 7.3 respectively.

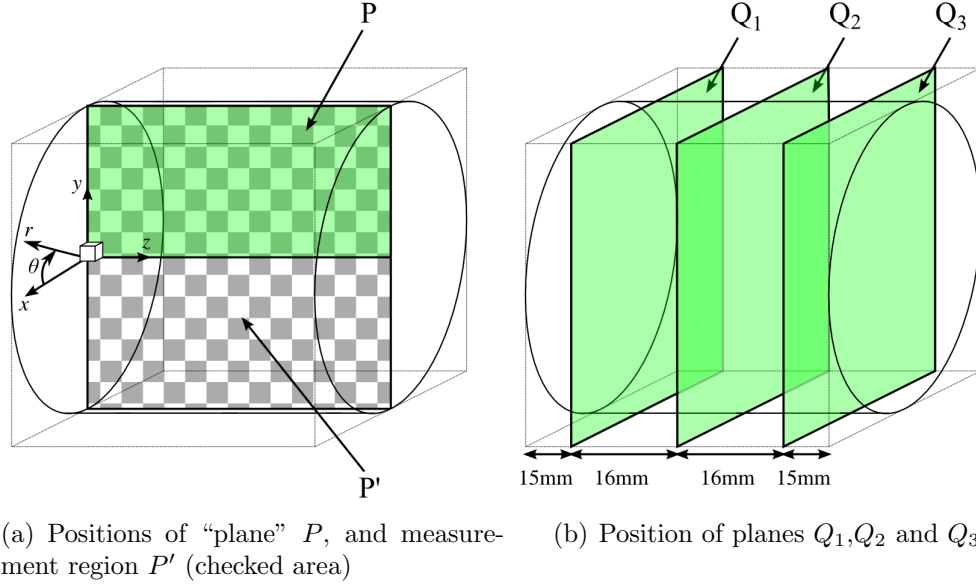| Parameter	Value
Total number of measurements	80
Valid measurements	66
Pulse separation, Δt	17ms
Particle image size	640×480 pixels
Final correlation window size	12×12 pixels
Correlation window overlap	50%

Table 7.3: Summary of Q plane stereo PIV parameters

(a) Positions of "plane" P, and measurement region P' (checked area)

(b) Position of planes Q_1, Q_2 and Q_3

Figure 7.18: Position of measurement regions

7.4 Validation of MP-VSE procedure

Before beginning the vortex ring prediction process, the statistical properties of the stereo PIV measurements were analysed to assess the accuracy of the light sheet alignment, confirm the axisymmetry of the vortex ring data, and ensure that the P and Q plane measurements were consistent with each other. A summary of the acquired measurements is provided in figure 7.18. Note that the P "plane", as it was referred to in previous sections, is not strictly a plane, but a 62×34 mm rectangular area. In reality, the P plane measurements consist not only of this region (i.e. the vectors above $y = 0$), but also of an identically sized region directly below. This full 62×68 mm measurement area, denoted by the checked area in 7.18 (a), is referred to as the region P'.

In the presence of correct light sheet alignment and perfect axisymmetry, a number of properties should be evident in the measurements. Under these conditions, the vortex rings should, on average, travel directly along the

z axis at $(x, y) = (0, 0)$. This should manifest itself in the ensemble-averaged measurements of P' in the form of a rotational symmetry. i.e. the lower region of P' should be identical to the top region, rotated about the z axis. This property was later exploited to extract two separate measurements of P from a single measurement of P'. Previous studies have shown that unstable vortex rings commonly exhibit a degree of azimuthal rotation (e.g. (Naitoh *et al.*, 2002; Dazin *et al.*, 2006b)). Under ideal settings, it is expected that the occurrence of this rotation should occur in a clockwise and anti-clockwise direction with equal probability, and so the mean vortex ring behaviour should not show any rotational behaviour. Hence, the out-of-plane velocity component should be zero across the entire field.

For the Q plane measurements, the ensemble averages should ideally show a perfectly circular velocity distribution, centred about the origin.

Figure 7.19 shows that the P and Q measurements intersect along three separate lines, at $z = 15, 31, 47$ mm. Hence, along these lines, the distribution of the measured velocity components should be identical for both P and Q measurements. This indicates that the relative alignment of the P and Q sheets is correct, and also shows that any error in the measurements is consistent in both data sets.

The aim of the following sections is to investigate the extent to which these properties hold. The nature of the measurements means that throughout this work, the velocity fields are presented in terms of both the cylindrical and Cartesian coordinate systems, as defined in figure 7.18 (a). The velocity components in Cartesian coordinates are given by u, v and w, , corr which correspond to the x, y and z directions. In cylindrical coordinates the radial, v_r, azimuthal, v_θ, and axial, w, velocity components correspond to r, θ and z respectively. The w component is the same in both systems.
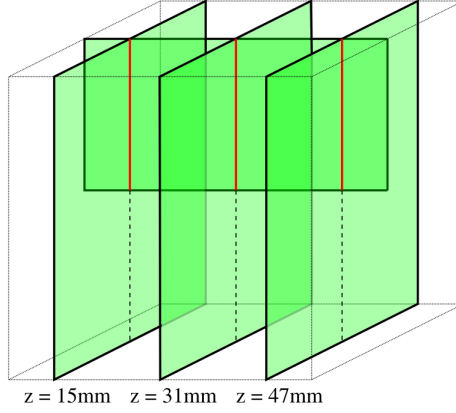
207

z = 15mm   z = 31mm   z = 47mm

Figure 7.19: Intersection of P and Q planes, as indicated by the red lines

7.4.1 Analysis of P' measurements

The mean in-plane velocity components (v and w) of the P' measurements are displayed in vector form in figure 7.20. The individual components are shown in figure 7.21.

Both the v and w components demonstrate rotational symmetry about the z axis at $y = 0$. The mean u component however shows a distinct, albeit small, non-symmetric and non-zero pattern, which would seem to indicate a degree of misalignment. It is difficult to determine the exact nature of the misalignment from these plots alone; the light sheet may be displaced from the path of the mean-vortex ring, it may be at an angle to the path of the mean vortex ring, or a combination of the two. There is also the possibility that the pattern in the mean w component could, in part, be due to errors in the stereo PIV calibration, which could lead to errors in the out-of-plane velocity field. In any case, the average magnitude of this component is considerably lower than the average in-plane velocity, (by a factor of about 30:1), and should not
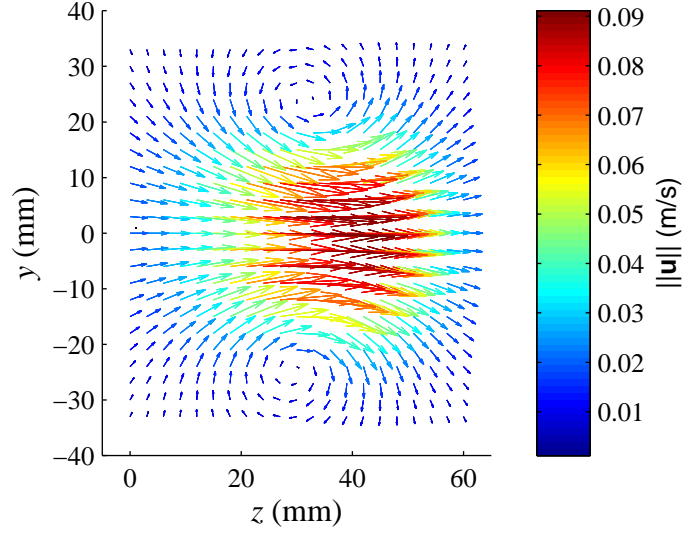
Figure 7.20: Vector map of average in-plane velocity of the P' measurements. Vector colour corresponds to the velocity magnitude

be a major concern.

7.4.2 Analysis of the Q Measurements

The mean velocity magnitude of the Q plane measurements is shown in figure 7.22. These plots serve to demonstrate the alignment of the light sheets and provide an indication of the mean axisymmetry of the flow, as indicated by a perfectly circular velocity distribution about the origin. To all practical purposes, this is shown in the contour plots of all three planes. The small sample set (66 measurements) used to produce the mean vector maps is likely to account for the unevenness of the contours, which is particularly pronounced in the centre of the plots.
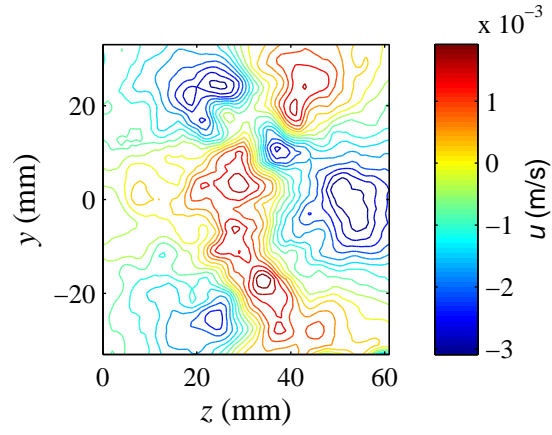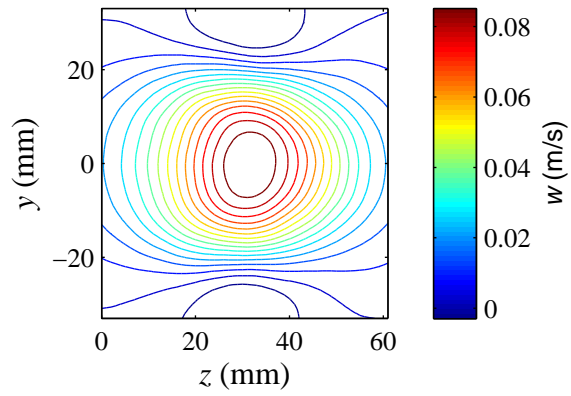
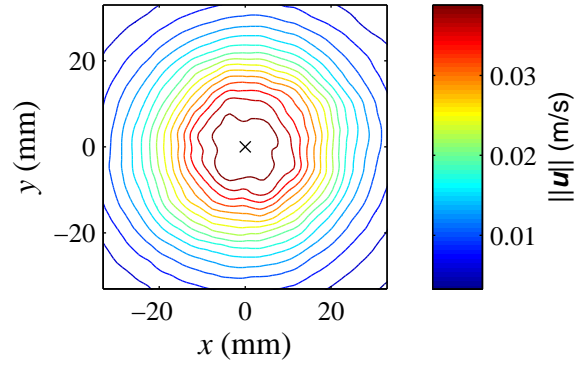(a) Contours of mean u velocity
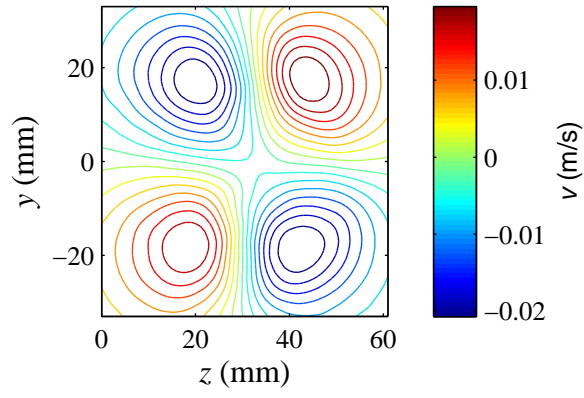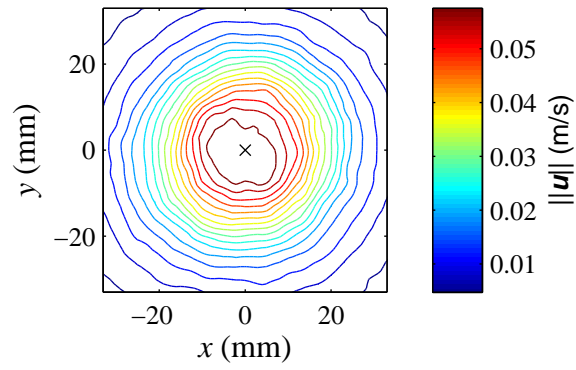

(b) Contours of mean v velocity
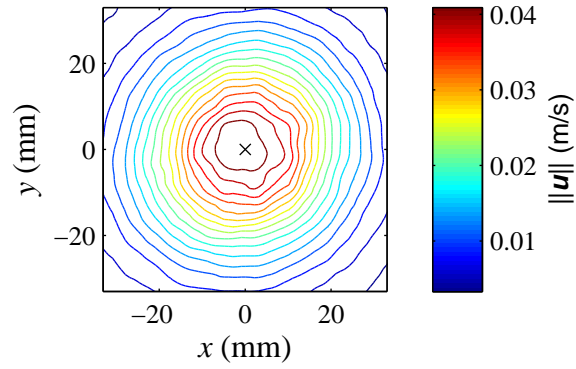

(c) Contours of mean w velocity

Figure 7.21: Mean velocity field of P plane measurements

(a) Contours of mean velocity magnitude for Q_1

(b) Contours of mean velocity magnitude for Q_2

(c) Contours of mean velocity magnitude for Q_{3*}

Figure 7.22: Mean velocity magnitude for Q plane measurements. Crosses indicate position of $(x, y) = (0, 0)$.

7.4.3   Comparison of P and Q Statistics

The P and Q planes intersect along three lines along the r axis at $z = 15, 31, 47$ mm and $\theta = 90°$. The statistical properties of the P and Q measurements along these lines should be identical; any deviation will be due to misalignment between the sheets, differences in the errors associated with stereo PIV measurement, and/or deviations from axisymmetry of the flow. As discussed previously, only 66 measurements of the Q planes were available, but under the assumption that rings are mean-axisymmetric, the statistics of the ring should be independent of θ, which means that the sample size may be increased by extracting multiple profiles at different θ from each Q plane measurement. To achieve this, 88 lines were extracted from each vector map, resulting in 5808 samples for each Q plane to base the statistics on. By extracting two measurements of P from each instance of P', 576 samples were available for the P plane. The resulting mean velocity profiles are shown in figure 7.23, and the velocity variances are shown in 7.24.

In general, the statistics of the radial v_r and axial w components appear to match quite well. The relative shape of the profiles are similar, which suggests that the relative light sheet positioning was accurate. The largest deviations occur in the azimuthal velocity statistics. In the case of the mean profiles, the corresponding P and Q measurements are largely dissimilar (figure 7.23 (b)). However given that the magnitudes of the mean azimuthal velocities are relatively small compared to the other components, this does not necessarily indicate a problem with alignment, or with the assumption of axisymmetry of the flow. Interestingly, the azimuthal velocity variances only differ significantly in the case of the $z = 31$ mm profile, with the profiles at $z = 15$ mm and z = 47 mm showing reasonable agreement. It is not immedi-

ately apparent why this should be the case, given that all the mean azimuthal profiles are dissimilar.

As a final indication of light sheet alignment, the 2-point correlation function $R_{ij}(r, r')$ was calculated at each of the three z positions. The correlation function is defined as:

$$R_{ij}(r, r') = \langle \mathbf{u}_i(r)\mathbf{u}_j(r') \rangle \tag{7.1}$$

where the subscripts i, j denote the velocity component, i.e. v_r, v_θ or w. Because two-point correlations, or covariances, form the heart of the stochastic estimation procedure, the similarity between the two-point correlation functions for P and Q measurements gives an important indication of how successful the reconstruction process will be. Contour plots of the two-point correlations for each velocity component are shown in figures 7.25, 7.26 and 7.27.

In each map, the diagonal line defined by $r = r'$ corresponds to the variance along r, and so is identical to the corresponding profile in figure 7.24. These maps reiterate some of the previous conclusions; generally, the shape of the correlation maps is closely matched, with the only significant deviation occurring in the case of the azimuthal velocity variance at $z = 31$ mm (figures 7.26 (c) and (d)).

Overall, the results in these sections confirm that the calibration of the experiment was largely successful, and indicate that the assumption of mean-axisymmetry is valid. A slight degree of misalignment of the P plane is suggested, but its effect on the measurements is marginal. The comparison of means, variances and two-point correlations for the P and Q measurements shows that the alignment of the P and Q light sheet planes is good. The similarity of these statistics also indicates that the use of data from the Q

213
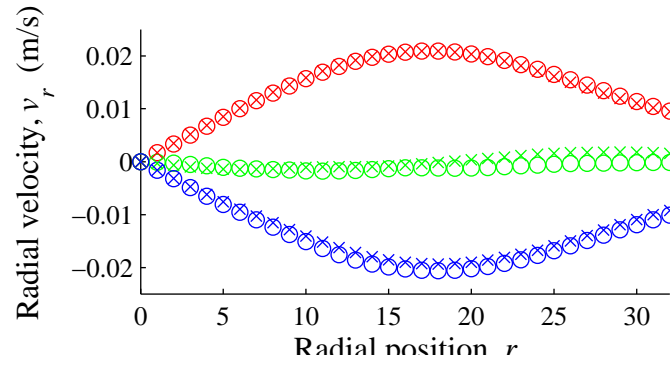
(a) Radial velocity profile, v_r
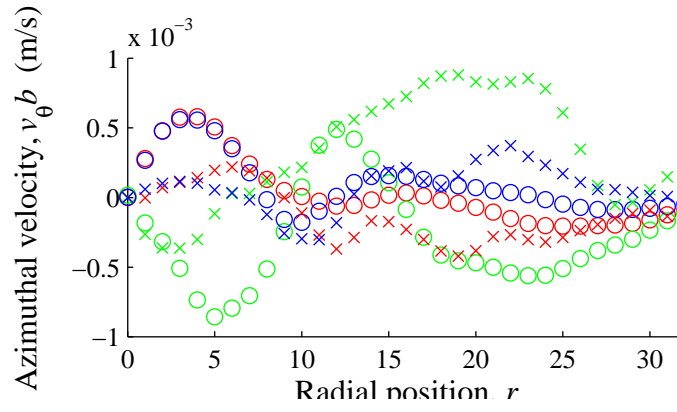
(b) Azimuthal velocity, v_θ
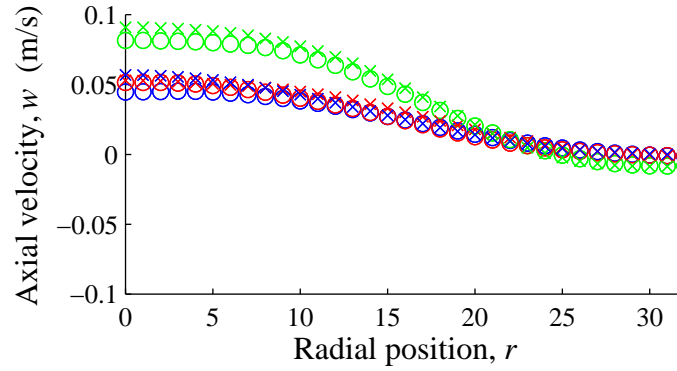
(c) Axial velocity profile, w

Figure 7.23: Mean velocity profiles. Crosses indicate the P plane measurements, circles indicate Q plane measurements. Red: $z = 15$mm, green: $z = 31$mm, blue: $z = 47$mm
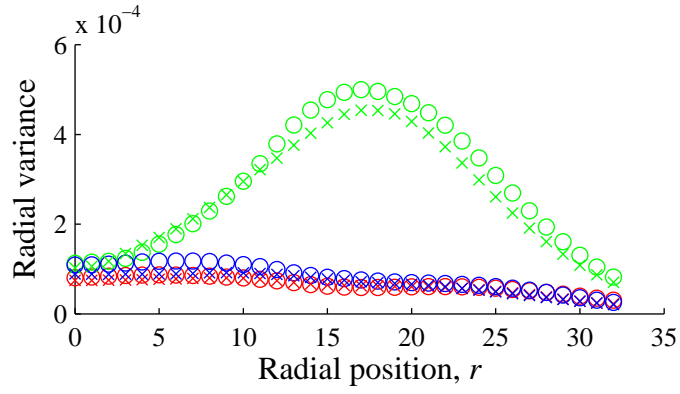
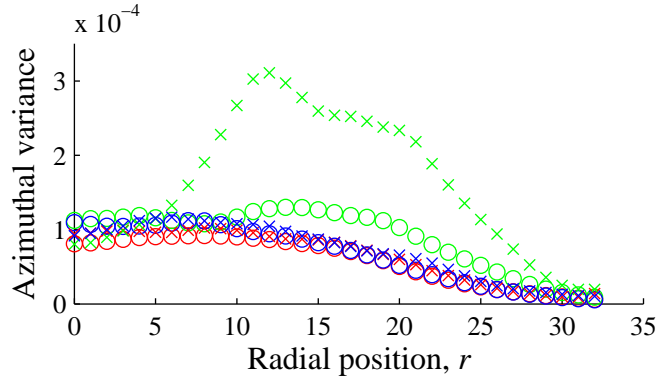(a) Radial velocity variance

(b) Azimuthal velocity variance
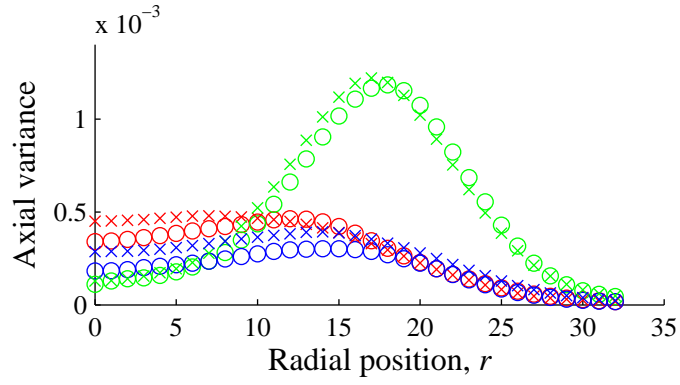
(c) Axial velocity variance

Figure 7.24: Velocity variances. Crosses indicate the P plane measurements, circles indicate Q plane measurements. Red: $z = 15$mm, green: $z = 31$mm, blue: $z = 47$mm

(a) P plane, radial velocity (R_{11})

(b) Q_1 plane, radial velocity (R_{11})

(c) P plane, azimuthal velocity (R_{22})

(d) Q_1 plane, azimuthal velocity (R_{22})

(e) P plane, axial velocity (R_{33})

(f) Q_1 plane, axial velocity (R_{33})

Figure 7.25: two-point correlations at $z = 15$mm

216

(a) P plane, radial velocity (R_{11})

(b) Q_2 plane, radial velocity (R_{11})
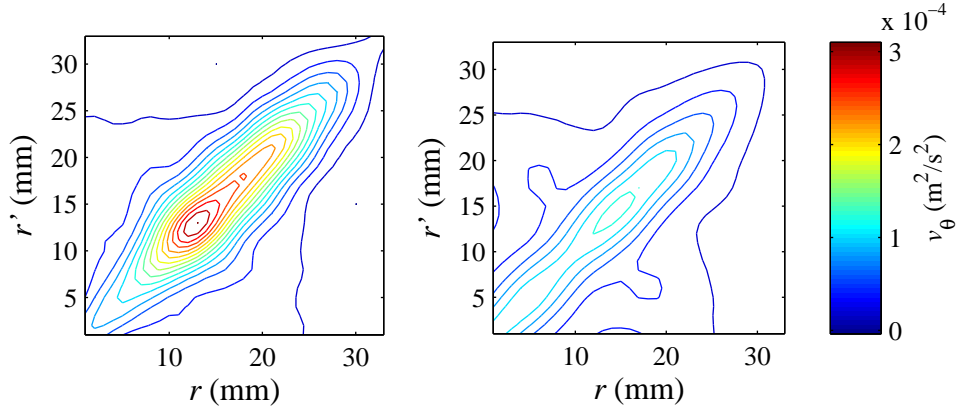
(c) P plane, azimuthal velocity (R_{22})

(d) Q_2 plane, azimuthal velocity (R_{22})
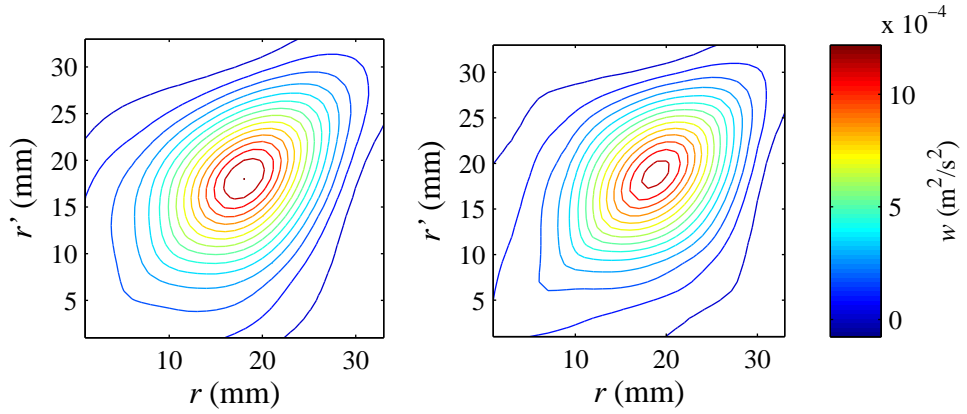
(e) P plane, axial velocity (R_{33})

(f) Q_2 plane, axial velocity (R_{33})

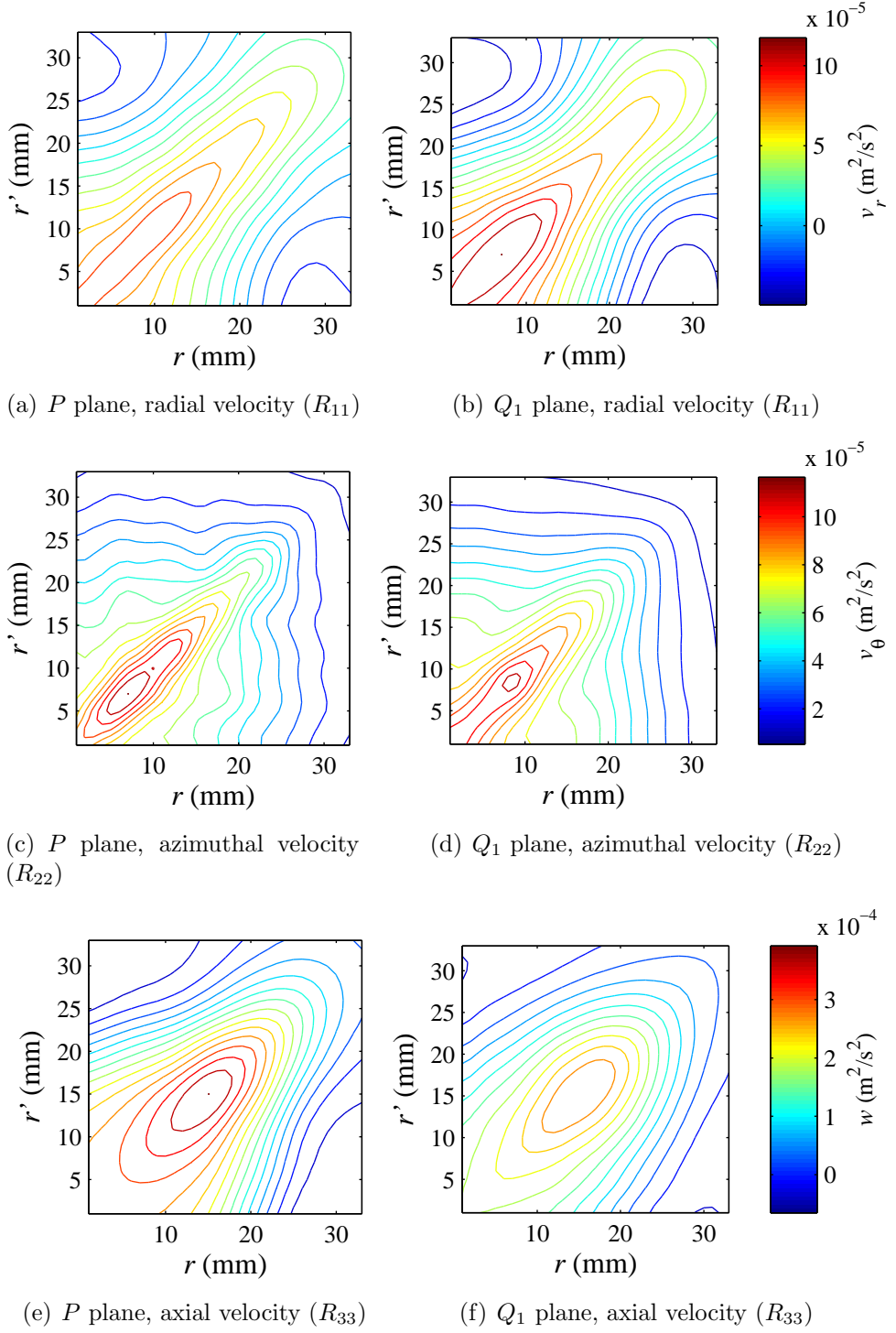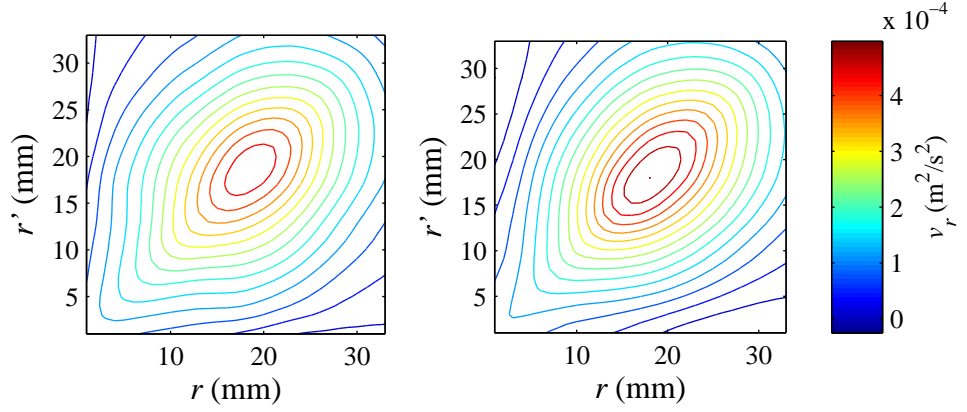Figure 7.26: two-point correlations at $z = 31$mm
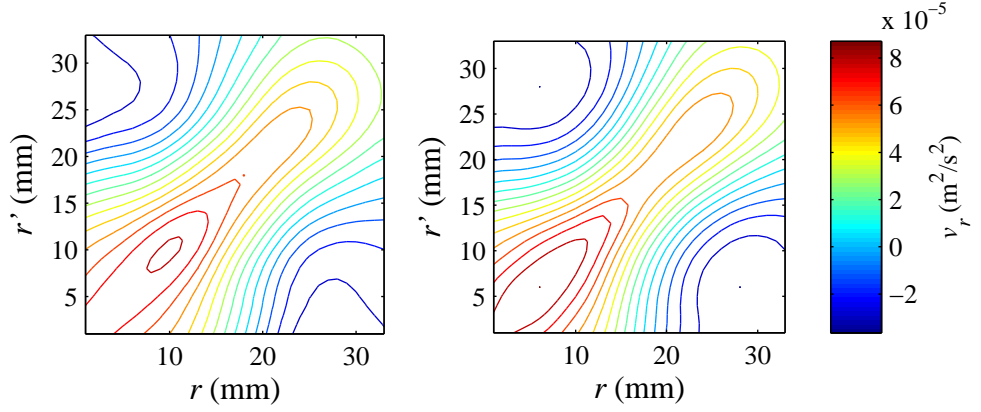
(a) P plane, radial velocity (R_{11})

(b) Q_3 plane, radial velocity (R_{11})
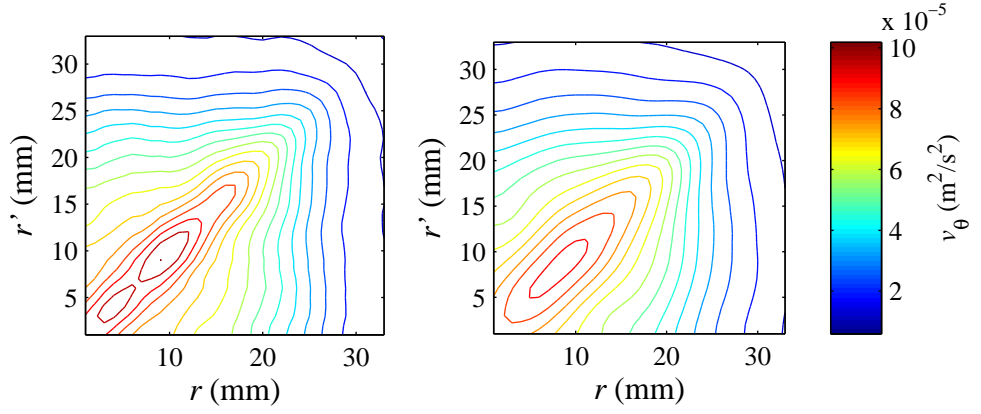
(c) P plane, azimuthal velocity (R_{22})

(d) Q_3 plane, azimuthal velocity (R_{22})
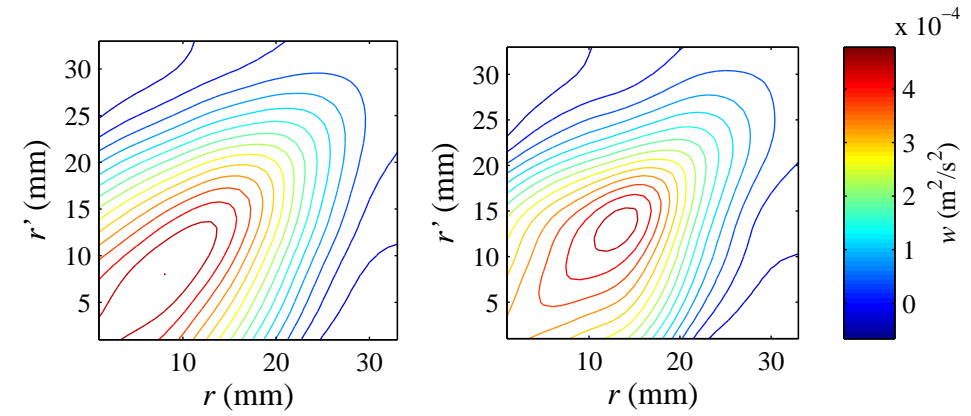
(e) P plane, axial velocity (R_{33})

(f) Q_3 plane, axial velocity (R_{33})

Figure 7.27: two-point correlations at $z = 47$mm

measurements to make predictions with a model built from the P measurements is valid. Although deviations between the statistics derived from the azimuthal velocities, particularly at $z = 31$ mm, is evident, it is assumed that the reconstructed vortex rings will not be adversely affected, due to the relative insignificance of this velocity component.

Parameter	Value
Sample size, m	518
Unconditional Vectors	102
Independent Variables, n	306
Conditional Vectors	2278
Dependent Variables, p	4556

Table 7.4: Summary of parameters for the cross-validation procedure

7.5 Model Construction

Using the P measurements, 10-fold cross-validation was used to obtain the prediction performance of the five regression techniques discussed in chapters 4 and 5. Only the standard linear stochastic estimation model was considered for this work. Given that the prediction model was to be built with separate data from that used for the subsequent predictions, it was felt that the use of higher order terms would decrease the stability of the model, which may amplify the prediction error due to deviations between the P and Q data.

The prediction model was built using the vectors shown in figure 7.6, with the 102 unconditional vectors used to predict the full 62×33 grid as indicated. As explained in section 7.4, due to the rotational symmetry present in the full P' region measurements, a total of 576 observations of P could be extracted from the 288 measurements of P'.

For each iteration of the cross-validation procedure, 512 samples were used to build the model, with the remaining 64 used for the validation. The regression parameters are summarised in table 7.4. Results of the cross-validation are presented in figures 7.28, 7.29, and table 7.5.

All the regression models provide excellent performance with this data set, with each explaining over 90% of the variance. The relative performance
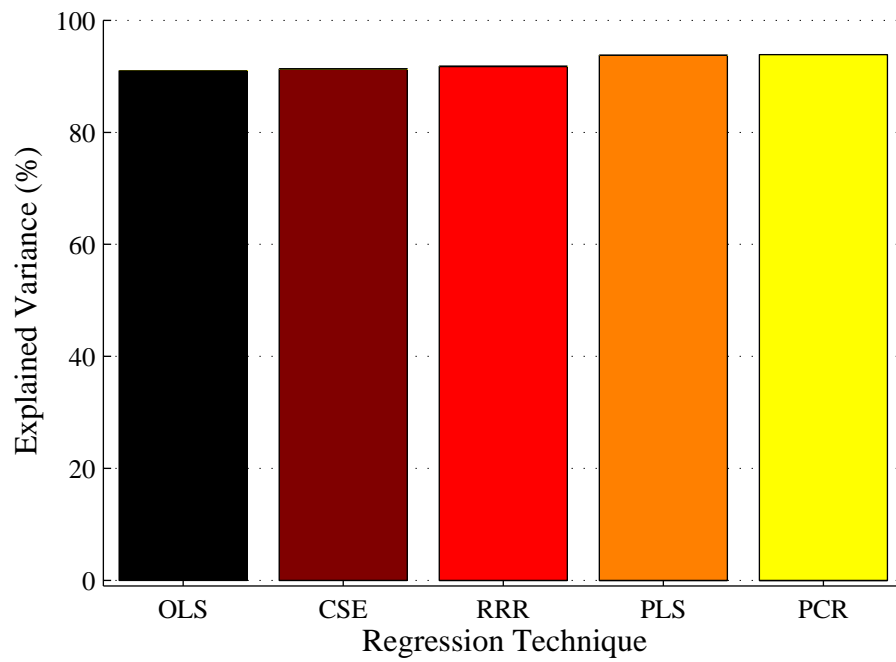
Figure 7.28: Percentage explained variance vs. regression technique
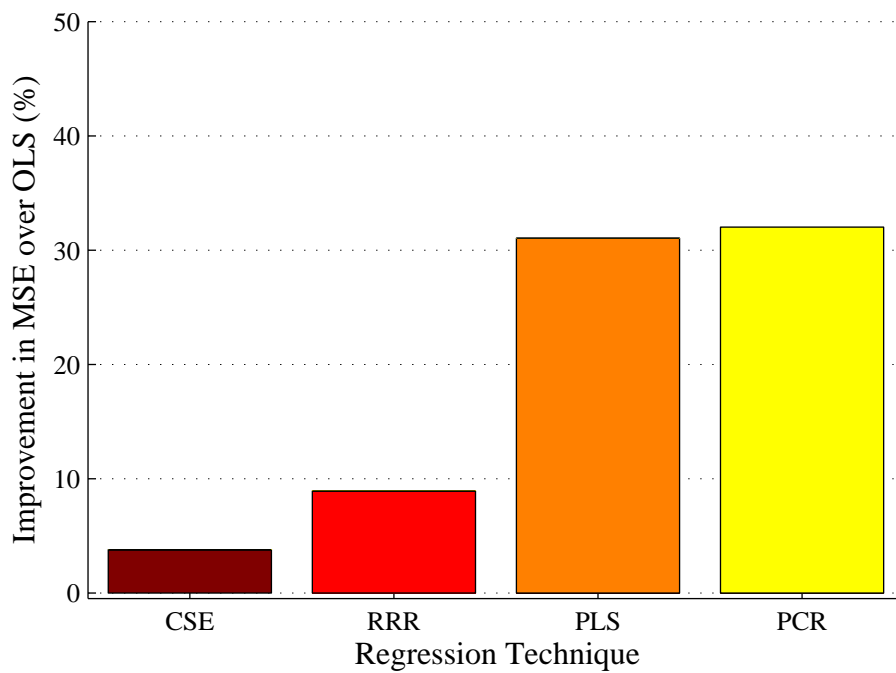
Figure 7.29: Percentage improvement over OLS vs. biased regression technique

CSE	RRR	PLS	PCR
20 (*518*)	21 (*518*)	24 (*518*)	27 (*518*)

Table 7.5: Optimum number of modes used for biased regression techniques

of the biased techniques over OLS is consistent with that seen previously in chapter 5 and section 6.3, although the high accuracy of OLS means that in absolute terms, the difference between the competing techniques is small. Nonetheless, the 27-mode PCR model provides the best performance (94%), and was ultimately used for the volumetric reconstructions. To maximise the accuracy of the predictions, the model was rebuilt using the full set of 576 measurements before carrying out the prediction process.

7.6 Results

7.6.1 Introduction

This section presents the results from the MP-VSE procedure. The work begins with the visualisation of some example vortex rings, followed by an investigation into some specific phenomena that have been observed in previous numerical and experimental studies. Many of the results in this section are expressed in terms of a coordinate system which is aligned with the instantaneous position of the vortex ring in question, as shown in figure 7.30. The reference plane of the cylindrical coordinate system bisects the vortex ring core, and the normal vector z' corresponds to the direction of travel of the ring. The $v_{r'}$, $v_{\theta'}$ and w velocity components correspond to r', θ' and z' respectively. Figure 7.30 also indicates the two dominant velocity components of the vortex ring's movement; namely, the translational velocity of the ring, V, and the core tangential velocity, v_T.

7.6.2 Visualisation of Vortex Ring Core

Visualisation of the vortex core is commonly based on vorticity magnitude, which for experimental rings should be "strongly peaked inside an inner core and then decay algebraically until it drops to zero at the ill-defined edge of the core" (Saffman, 1978). Unfortunately, the use of vorticity magnitude as a visualisation tool was ill-suited for use with the predicted data, as the distribution of vorticity does not coincide directly with the core position. This is illustrated in figure 7.31, which shows the distribution of vorticity in one of the predicted P planes.

Identical behaviour is exhibited in the raw measurements in the P plane,

223
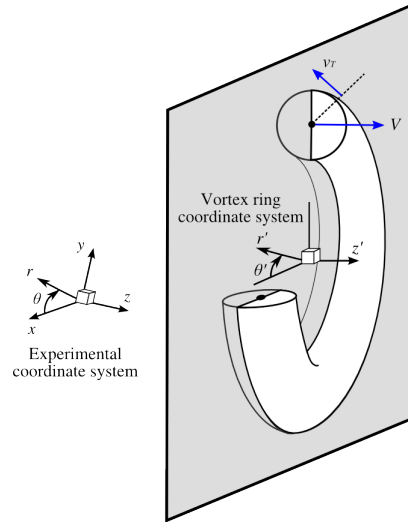
Figure 7.30: Definition of the vortex ring coordinate system. The blue arrows indicate the main velocity components of the vortex ring; V is the core translational velocity, and v_T is the core tangential velocity.
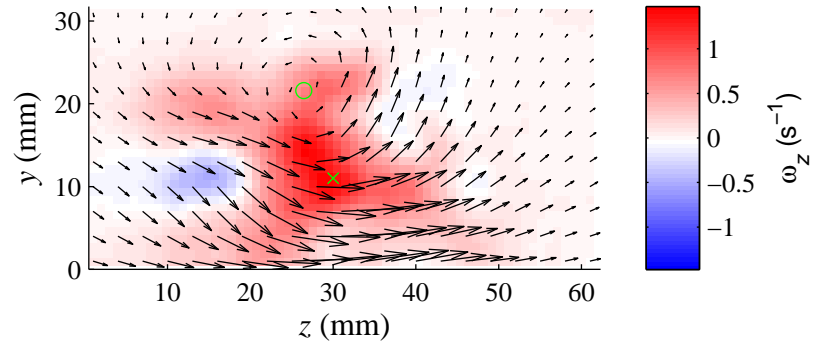
Figure 7.31: Vector map from an example predicted P plane. The scalar background shows the out-of-plane vorticity (ω_θ). The green cross shows the peak vorticity magnitude, and the green circle is the result of the vortex core detection algorithm proposed by Jiang *et al.* (1990)

which rules out the prediction process as the cause of the unusual vorticity distribution. Instead, it may be due to the large correlation window size employed in the PIV calculations, which was ultimately necessitated by the large interframe time of the Firefly cameras (see the discussion in section 7.3.2). It is possible that the window size was too large to sufficiently capture the velocity gradient within the core, causing a reduction in the vorticity in this region. Regardless of the cause, the result is that the peak vorticity, indicated by the green cross in figure 7.31, rarely corresponds to the centre of the vortex ring core, which lies at the point of zero tangential velocity (i.e. the centre of the swirling region of the vector field). As an alternative approach to core detection, the algorithm proposed by Jiang *et al.* (1990) was implemented, which uses concepts from combinatorial topology to detect vortex core regions using only the angle of the vectors within the velocity field. The green circle in figure 7.31 shows the result of this procedure, which can be seen to coincide with the centre of the swirling region of the flow. Applying this approach to each of the predicted planes in the volume provides the 3D coordinates at 88 points along the vortex core (figure 7.32 (a)). However, to improve both the resolution and accuracy of the detection process, a further stage was introduced to the visualisation process. The w velocity component was found at each detected point, and averaged to provide an estimate of the translational velocity of the ring. This was then subtracted from the velocity field, leaving only the tangential velocity as the dominant velocity component in the flow. Using the fact that the core centre has zero tangential velocity, and that the velocity increases monotonically to the edge of the core, a threshold of the velocity magnitude provides an accurate indication of the core position. An isosurface of velocity magnitude at $\|\mathbf{u}\| = 0.01$ m/s is shown in figure 7.32 (b). The thresholded velocity isosurface shows excellent agreement with

225

(a) Core detection process proposed (b) Velocity magnitude thresholding
by Jiang *et al.* (1990)
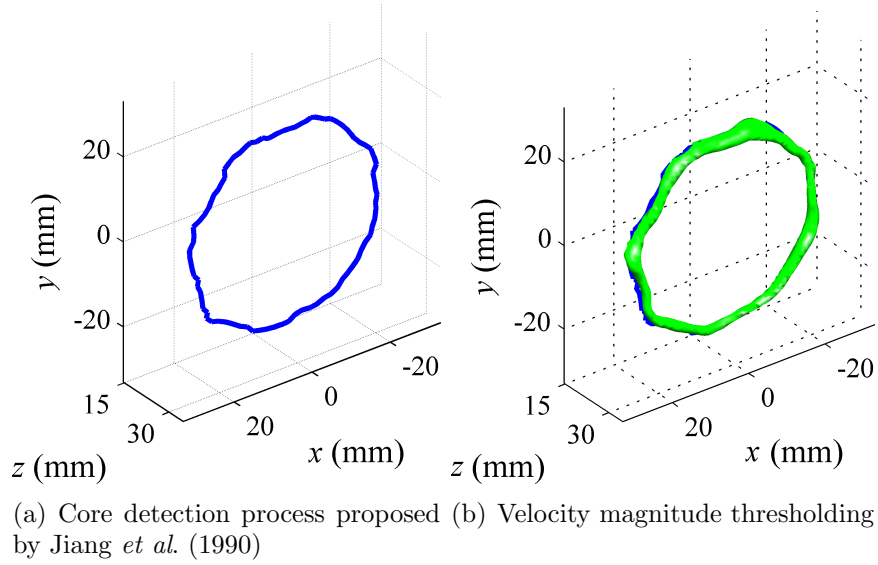
Figure 7.32: Approaches to visualising the vortex core centre

the core centre as detected by the method of Jiang *et al.* (1990), and has
the further benefit of indicating how the relative core size and shape varies
along the ring circumference. The choice of threshold value used here and
in the following examples has been selected; firstly to aid the visualisation of
the core instability and secondly, because the behaviour of the ring in this
region is markedly different to that of the outer region of the core, as will
be demonstrated in section 7.6.4. It should be noted that the threshold has
not been chosen to coincide with any theoretical definition of the ring core (for
example, the velocity threshold is lower than the maximum tangential velocity,
which from Saffman (1978), defines the theoretical "inner core" boundary).
Instead, it is helpful to consider the visualisations as defining a "filament" of
finite thickness that encompasses the core centre.

Four individual rings at nominally the same stage of evolution have
been chosen for further study, which are shown in figures 7.33 to 7.36.
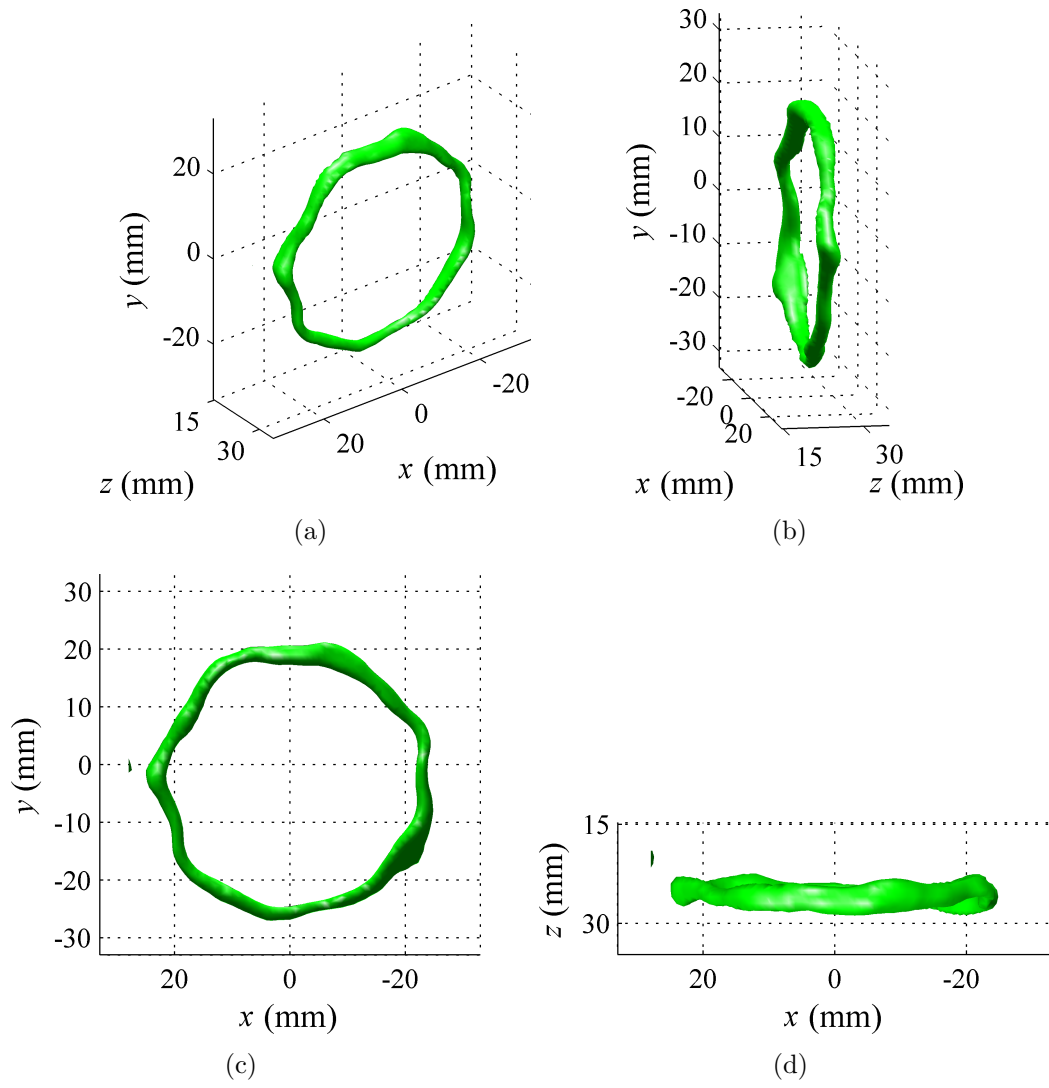
Figure 7.33: Views of ring 1 core (isosurface of $\|\mathbf{u}\| = 0.01$ m/s)
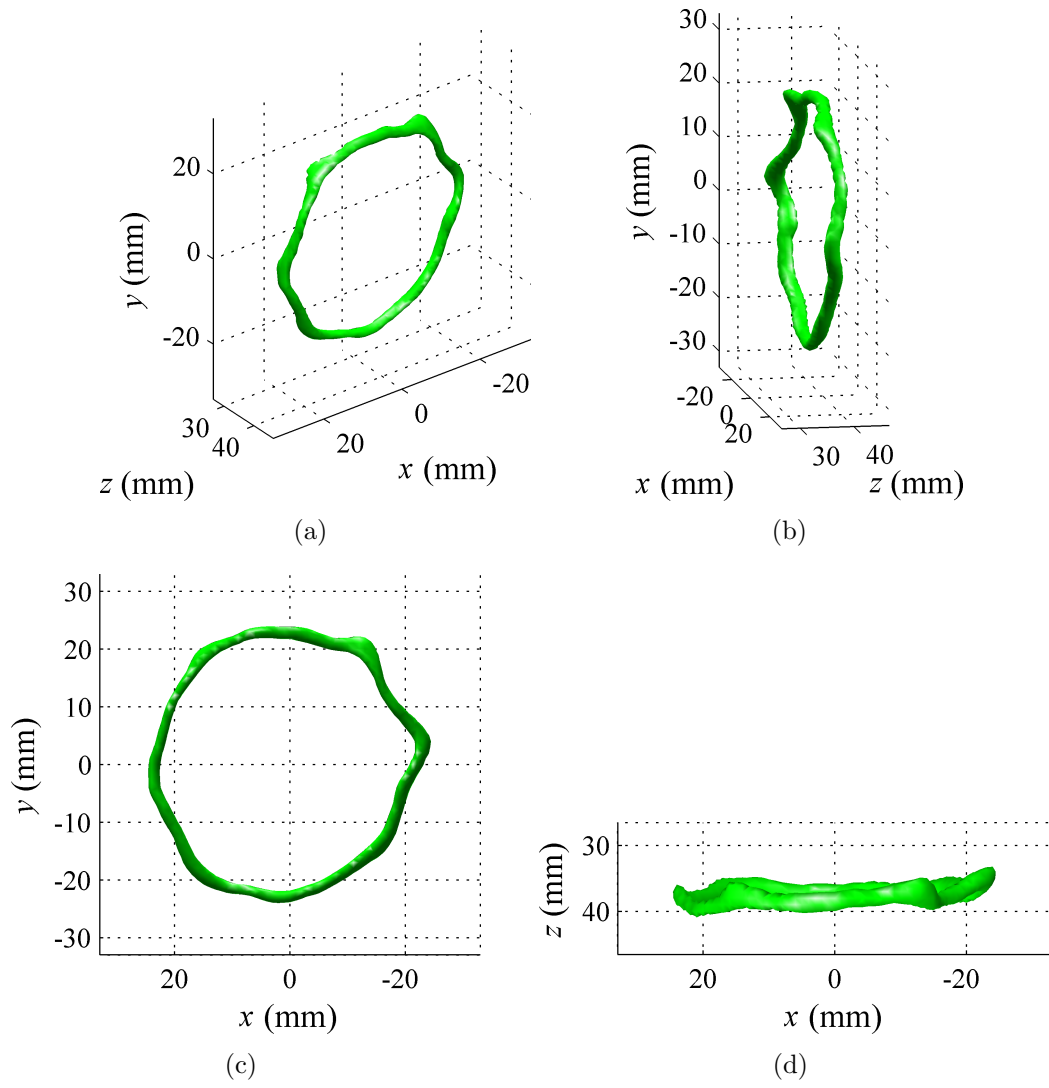
Figure 7.34: Views of ring 2 core (isosurface of $\|\mathbf{u}\| = 0.01\text{m/s}$)
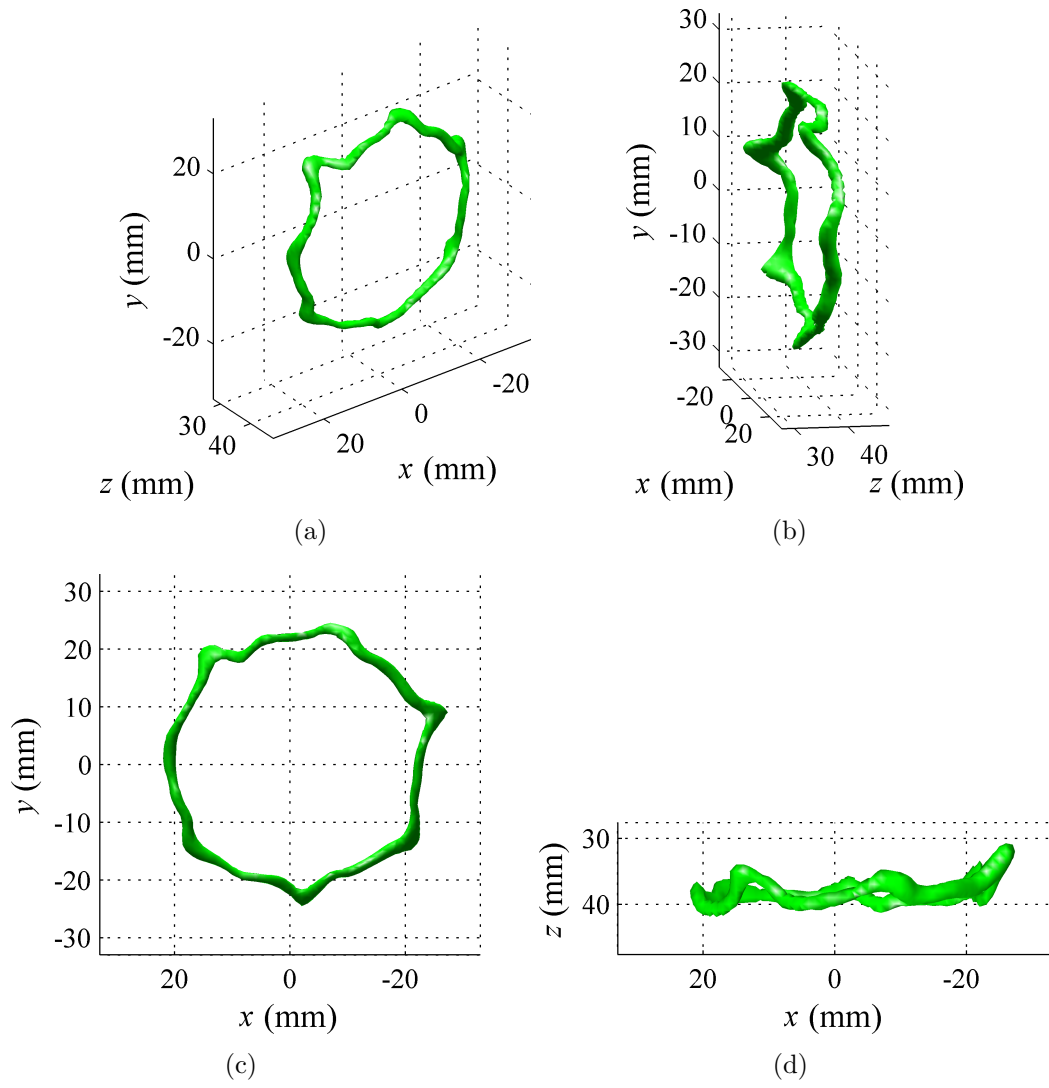
Figure 7.35: Views of ring 3 core (isosurface of $\|\mathbf{u}\| = 0.01\text{m/s}$)
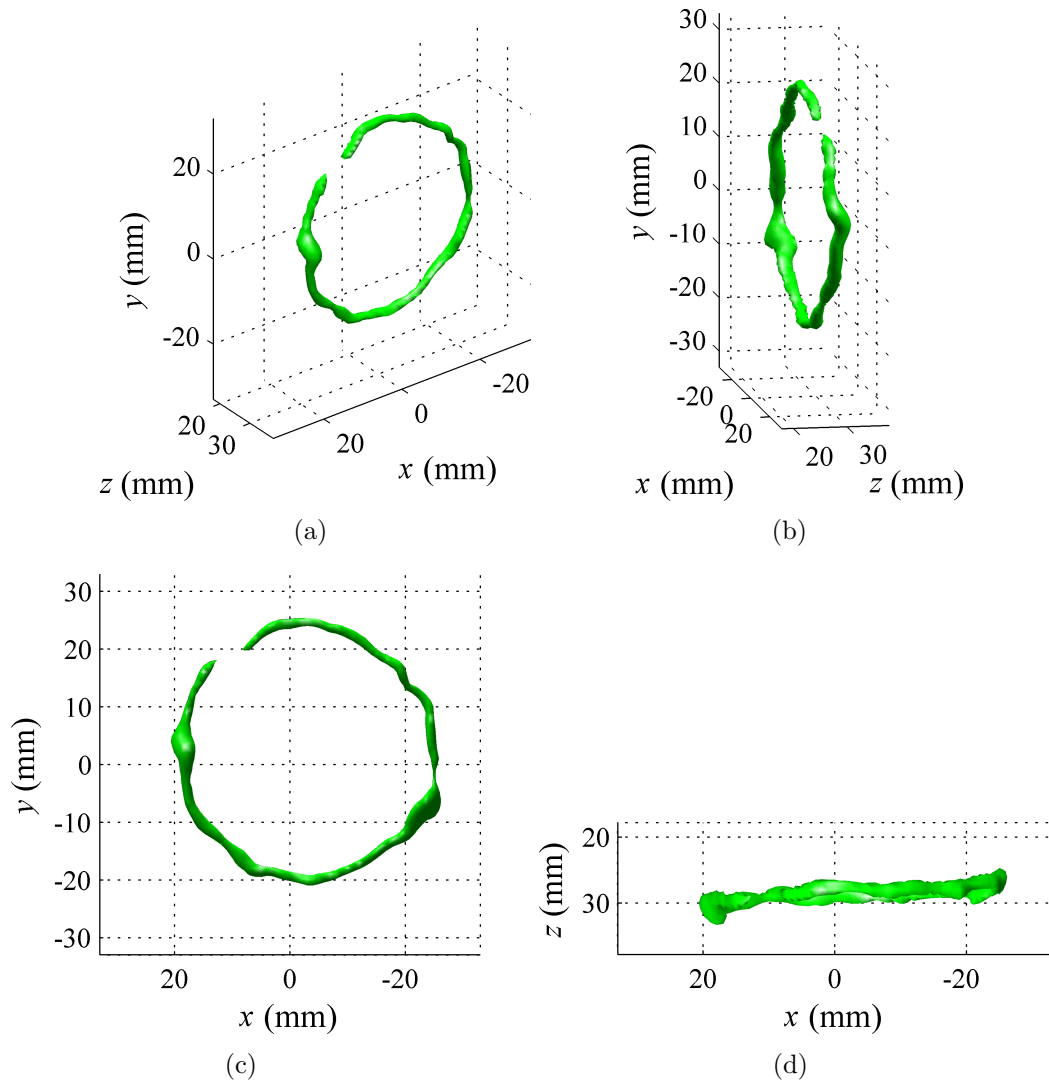
Figure 7.36: Views of ring 4 core (isosurface of $\|\mathbf{u}\| = 0.01\mathrm{m/s}$)

7.6.3 Investigation of Core Instability

The development of the instability on a vortex ring begins with the *linear phase*, which is characterised by the growth of one or more bands of waves around the ring core, orientated at 45° to the direction of travel. The nature of the instability was the focus of a great deal of experimental and theoretical work in the 1970s, which led to a detailed understanding of the physical process by which the instability is formed and gave rise to models that can accurately predict many aspects of its behaviour. However, it is only recently that this theory has been truly validated. This was made possible by the quantitative data provided by numerical simulations (Shariff *et al.*, 1994; Bergdorf *et al.*, 2007; Archer *et al.*, 2008) and stereo PIV measurements (Dazin *et al.*, 2006a,b), which has allowed the Fourier spectra of the instability to be accurately characterised. Some of the spectra from the work of Dazin *et al.* (2006b) are reproduced in figure 7.37. The spectra were calculated from the radial velocity of the ring, sampled at equally spaced points along the vortex ring's radius. Figure 7.37 (a) shows the spectrum of a vortex ring during the linear phase, revealing a band of waves centred around $n = 8$, where n is the number of waves in the radial direction.

These recent investigations have also brought about a greater understanding of the vortex ring behaviour during the *non-linear* stage, which is the focus of this study. It is now understood that the non-linear stage begins with the growth of harmonics of the dominant modes from the linear phase (figure 7.37 (b)), which subsequently decay, and give way to the development of low order modes (figure 7.37 (c)).

The volumetric data produced by the MP-VSE is ideally suited for such Fourier-based analysis. As demonstrated in (Dazin *et al.*, 2006a) and (Dazin

(a) Spectrum during linear phase

(b) Spectrum during early non-linear phase

(c) Spectrum during late non-linear phase
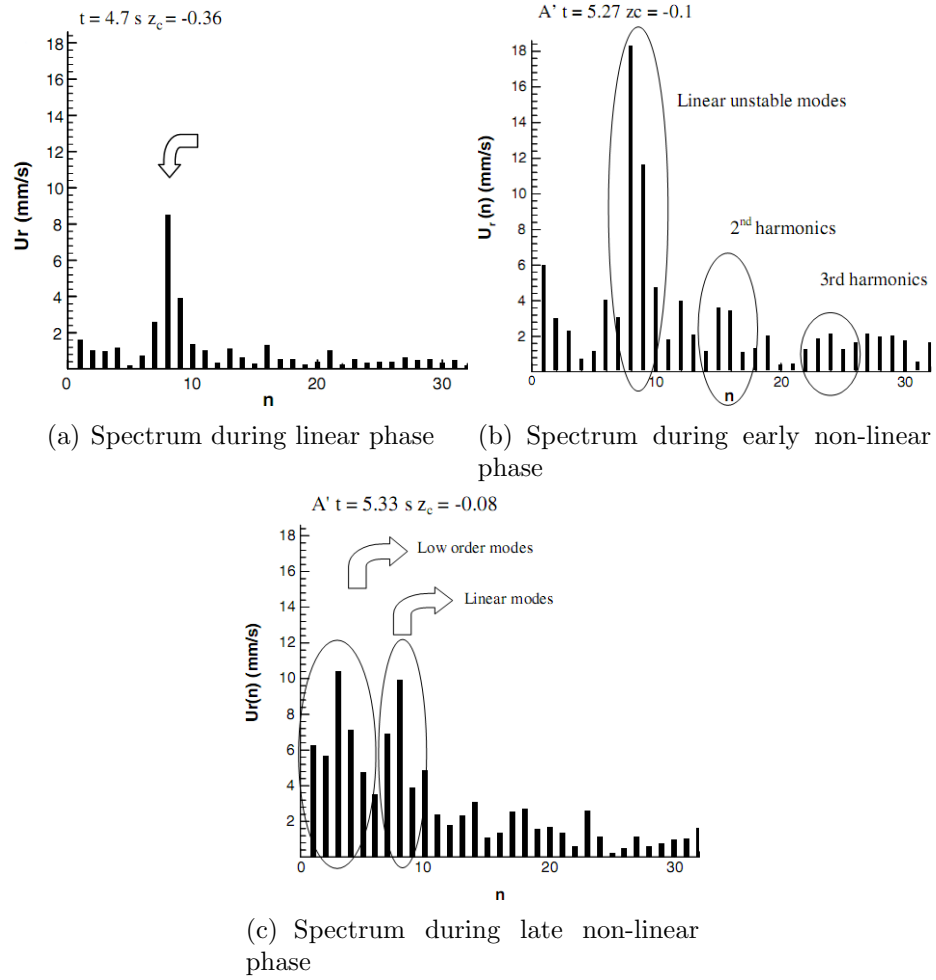
Figure 7.37: Radial velocity spectra of a vortex ring at different points in its development. (Reproduced from (Dazin *et al.*, 2006b))
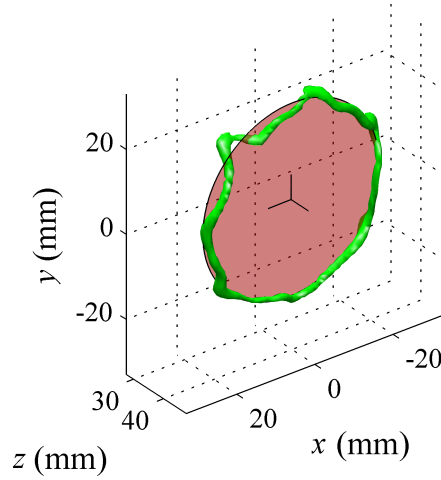
Figure 7.38: Position and radius of the vortex ring, as detected by the circle fitting algorithm.

et al., 2006b), this requires planar measurements of the ring velocity, but the accuracy is dependent on the ring passing directly normal to the measurement plane. If this is not the case, then the misalignment will cause an artificial increase in the $n = 1$ mode in the Fourier spectrum, as discussed in Dazin *et al.* (2006b). However, with access to the full volumetric velocity field, it is possible to position the plane directly through the vortex ring, thus avoiding any misalignment.

In practice, this was achieved using a non-linear least squares circle fitting algorithm[1], which was used to fit a circle through the core coordinates from the method of Jiang *et al.* (1990). An example is shown in figure 7.38. The axial velocity at 88 equally-spaced points was then extracted along the circumference of the fitted circle, and was used for the subsequent Fourier analysis.

---

[1] `ls3dcircle.m`, from the Least Squares Geometric Elements Library, developed by the National Physics Laboratory and available from the EUROMETROS software repository (`http://www.eurometros.org`)

233
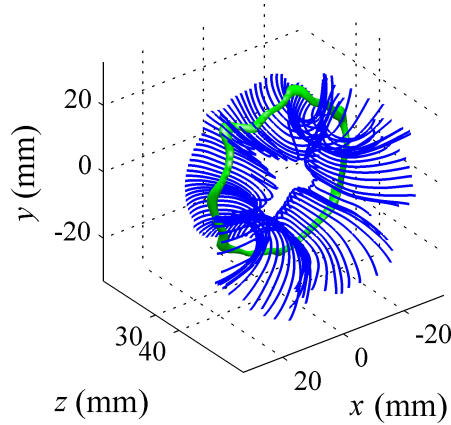
Figure 7.39: Vortex ring streamlines, which are seeded at equal increments of θ, 10mm upstream from the core centre

The resulting spectra are presented in this section alongside visualisations of the ring core. To further emphasise the nature of the instability, streamlines have been included in the visualisations, which are seeded directly 10mm upstream from the vortex core, as shown in figure 7.39.

As the predictions were made at a single point in time (5 seconds after closing the solenoid valve), the behaviour of the rings in their earlier stages of development is not known exactly. However, using the average initial vortex ring parameters, (shown in table 7.4), it is possible to predict which waves will appear on the ring during the linear phase. For this purpose, the model proposed by Saffman (1978) is employed (detailed in appendix D.5), which predicts the growth of bands of waves centred around $n = 7.8$, $n = 13.5$ and $n = 18.2$. The relative sizes of the peaks has not been calculated, although it is usually the case that the lowest peak will dominate (Saffman, 1978). This coincides well with the spectra of the predicted rings, which commonly show the presence of a dominant, (or at least, a large) peak at $n = 6$ and $n = 7$, which is likely to be the dominant mode of the linear phase of development. Of

course, this prediction is based on the average vortex ring data, and individual rings will show a degree of variation from the average. Given the simple method used to generate vortex rings, it is possible that the variation between rings is quite large. This lack of repeatability is not a problem in itself; in fact, it means that a wide range of vortex rings, at a variety of stages in their development, are available for study. These differences are clearly evident in the four rings selected for study here, whose spectra are shown in figures 7.40 to 7.43.

Of the rings considered here, it appears that ring 3 is in the earliest stage of development, perhaps only just in the non-linear phase. Two clear bands of peaks are visible, centred around $n = 7$ and $n = 13$, which is in excellent agreement with the prediction from the Saffman model. Also evident is the $45°$ angle of the instability, relative to the axis of travel. The presence of these phenomena is typical of the ring during the linear stage of development, although the ring shows evidence of the secondary vortical structure that occurs during the non-linear stage, which is shown in the next section.

Ring 4 appears to have progressed slightly further though its development, evidenced by clear harmonics of the dominant $n = 5$ mode at $n = 10, 15, 20$, and possibly also $n = 25$.

The spectra of rings 1 and 2 are dominated by low order modes ($n = 2$ for ring 1 and $n = 3$ for ring 2), which suggests that the rings are in the latter stages of the non-linear phase. This is further supported by the visualisations, which no longer exhibit the organised periodic behaviour of the linear phase.

(a) Visualisation of ring core

(b) Spectrum of radial velocity component
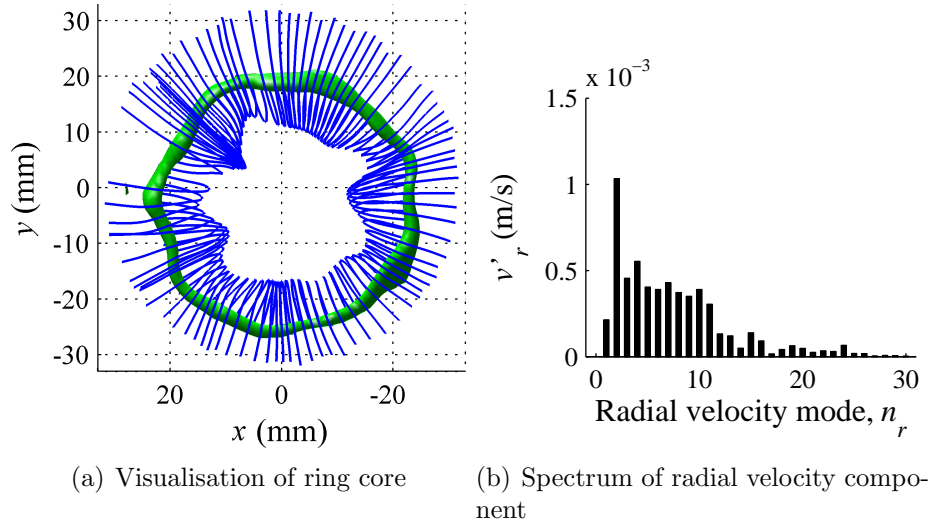
Figure 7.40: Ring 1 instability

(a) Visualisation of ring core

(b) Spectrum of radial velocity component
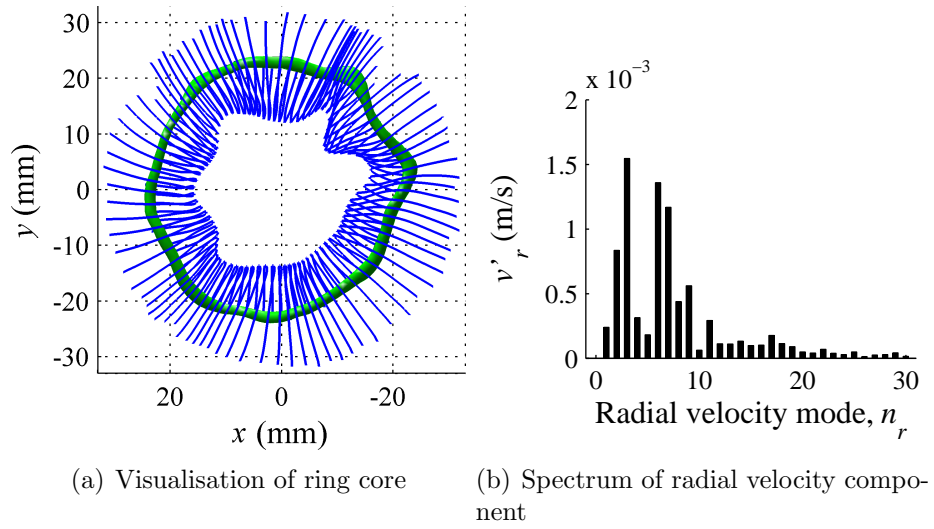
Figure 7.41: Ring 2 instability

(a) Visualisation of ring core

(b) Spectrum of radial velocity component
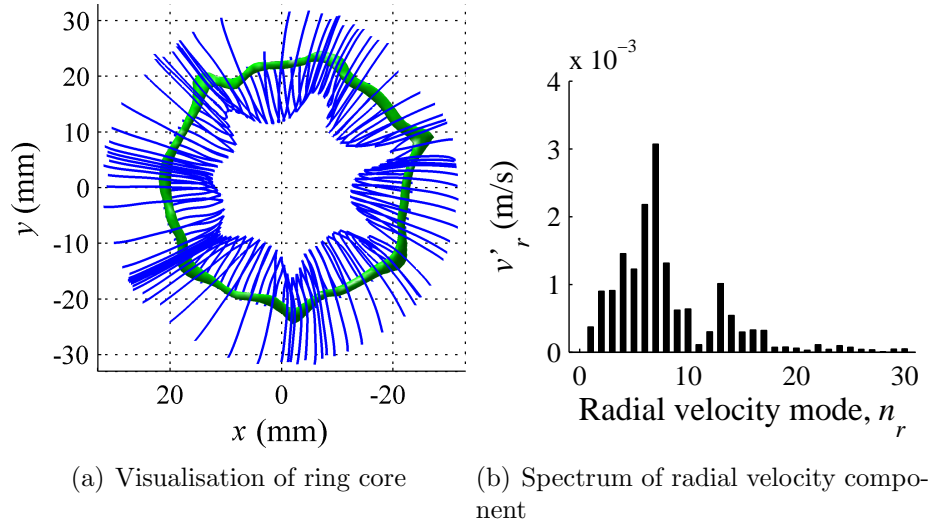
Figure 7.42: Ring 3 instability

(a) Visualisation of ring core

(b) Spectrum of radial velocity component
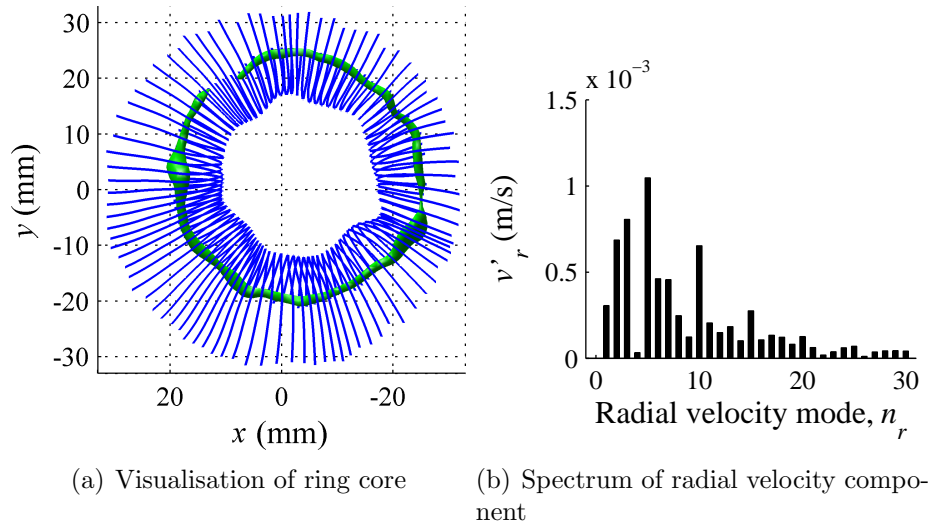
Figure 7.43: Ring 4 instability

7.6.4 Visualisation of Secondary Vortical Structure

This section investigates the presence of the secondary vortical structure that has been observed in the numerical simulations of Bergdorf *et al.* (2007) and Archer *et al.* (2008), and in the PLIF experiments of Dazin *et al.* (2006b). The secondary structure, or "halo vorticity" (Archer *et al.*, 2008), consists of pairs of vortical "loops" of alternating sign, that wrap around the inner core. The neighbouring loops join together in the peaks and troughs of the waves on the inner core, forming a lattice structure around the ring. Examples of this structure from previous literature are reproduced in figure 7.44.

Due to the issues described in section 7.6.2, the use of vorticity to visualise the secondary structures was largely unsuccessful, as the major contribution to the vorticity arises in the centre of the rings, apparently from regions of high shear. A number of alternative visualisation approaches were trialled, and it was found that the *curl angular velocity* (CAV) was most suitable. The CAV is defined as $\omega \cdot \frac{\mathbf{u}}{\|\mathbf{u}\|}$[Evorticity]$\omega$vorticity, which is the projection of the vorticity onto the normalised velocity vector. The properties of the CAV make it particularly appropriate for the visualisation of the secondary structure. In regions of high shear (shown in the centre of the ring in figure 7.45), the resulting vorticity vector will be approximately orthogonal to the velocity vector, and the CAV will be small. However, it is apparent from the images in figure 7.44 that the secondary vortical structure arises from the localised rotation of the flow around an axis approximately tangential to the inner core. As the velocity in this region is dominated by the tangential velocity around the ring core, the axes' vorticity and velocity should be closely aligned (also shown in figure 7.45), producing a large CAV.

Isosurfaces of the CAV at $2\mathrm{s}^{-1}$ (blue) and $-2s^{-1}$ (red) are plotted for the
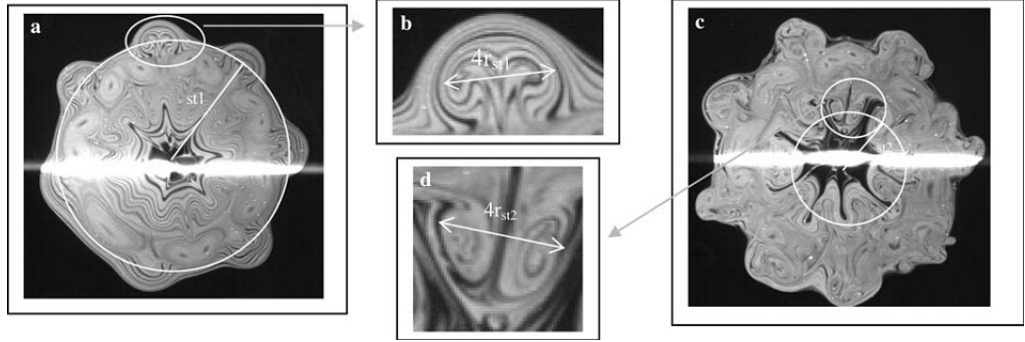
(a) PLIF visualisation of the vortical structure, from (Dazin *et al.*, 2006b). The image shows the pairs of counter-rotating structures, uniformly aligned on the inside and outside of the ring.
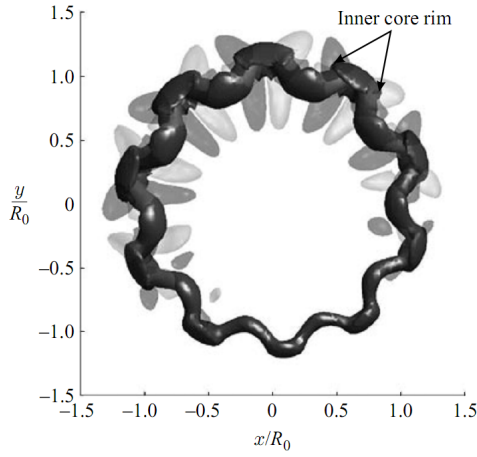
(b) Isosurface of positive and negative z component vorticity, ω_z, which shows the presence of alternating loops, aligned with the instability on the inner core (from (Archer *et al.*, 2008))
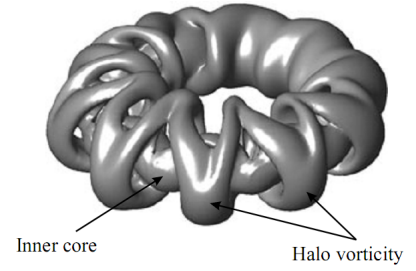
(c) Isosurface of the second invariant of the velocity gradient tensor (a common vortex identification method, also known as the "Q" criterion). The vortical structure can be seen to form a mesh around the inner core region (from (Archer *et al.*, 2008))

Figure 7.44: Examples of the secondary vortical structure

239
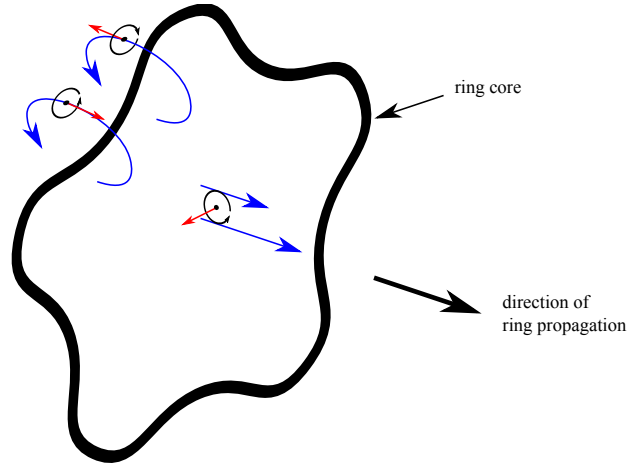
Figure 7.45: Orientation of velocity vectors (blue arrows) and vorticity vectors (red arrows) in a hypothetical vortex ring.

four rings in figures 7.46 to 7.49. Evidence of secondary structure is present to some extent in all rings, but perhaps clearest in rings 1 and 3, which clearly exhibit the alternating regions of positive and negative rotation, demonstrated in the PLIF and numerical visualisations in figures 7.44 (a) and (b). It is apparent in rings 2 and 4 that there is some difficulty in distinguishing between the inner core and secondary structure. However, for ring 2, the large region of negative (red) CAV to the left of ring, and positive (blue) CAV to the right of ring, appear to match the shape and orientation of the loop structure indicated in figure 7.44 (c).
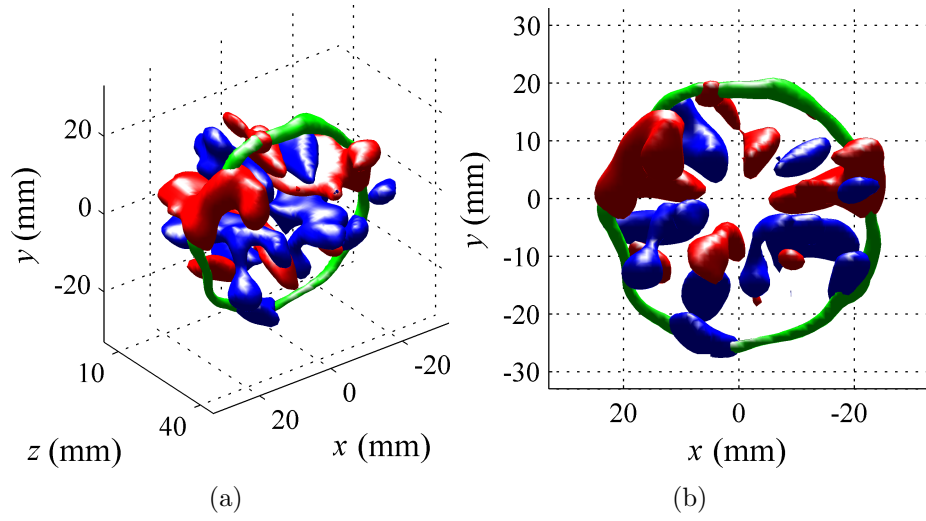
Figure 7.46: Isosurfaces of CAV $= \pm 2\text{s}^{-1}$ for ring 1
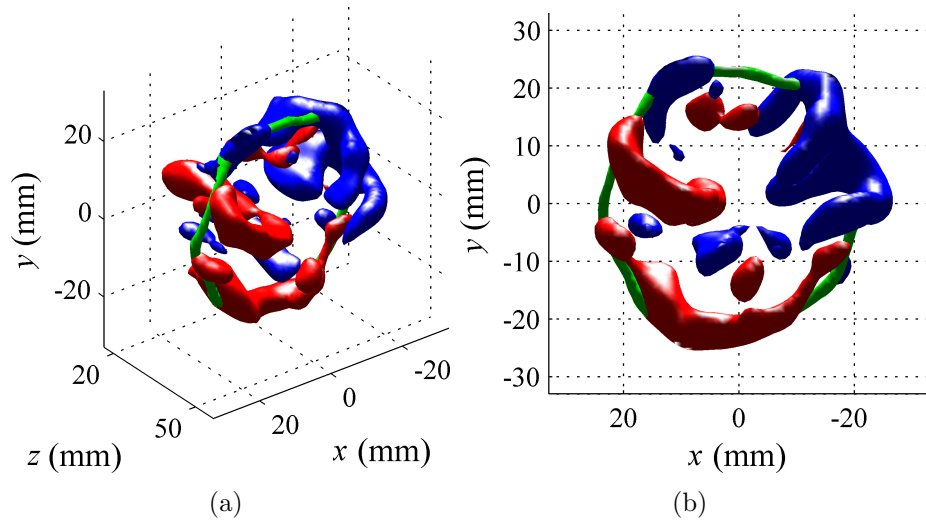
Figure 7.47: Isosurfaces of CAV $= \pm 2\text{s}^{-1}$ for ring 2

(a)

(b)
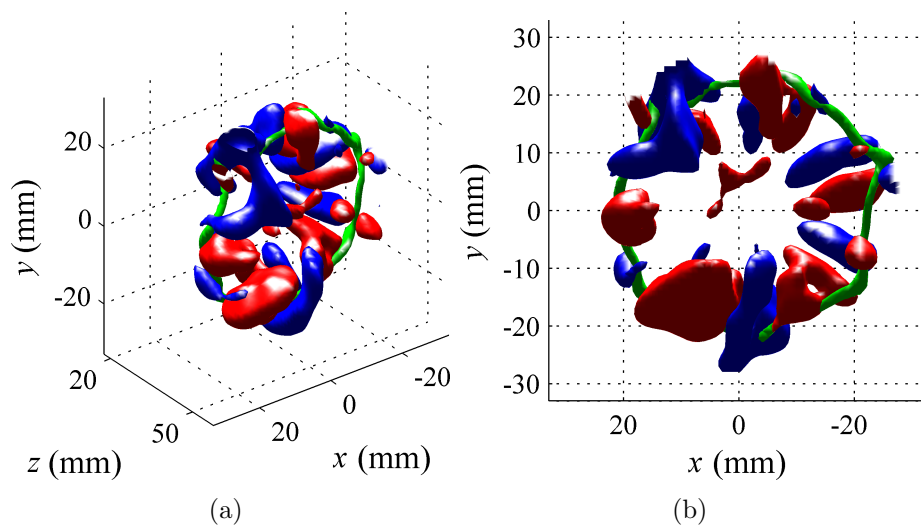
Figure 7.48: Isosurfaces of CAV = $\pm 2\text{s}^{-1}$ for ring 3

(a)

(b)
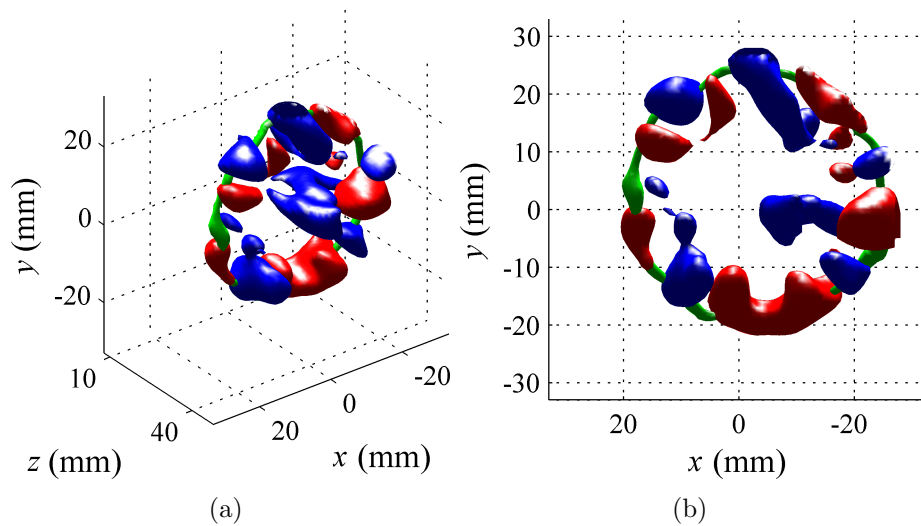
Figure 7.49: Isosurfaces of CAV = $\pm 2\text{s}^{-1}$ for ring 4

7.7 Conclusions

This chapter describes an experimental implementation of the MP-VSE procedure, which has been employed to investigate the behaviour of unstable vortex rings. The experiment has yielded a series of unique volumetric visualisations of vortex rings during an important stage of their development. In addition, it has introduced a version of the MP-VSE procedure that is applicable for the prediction of statistically-axisymmetric flows.

The rings selected for analysis here corroborate observations and theory from previous literature; Fourier analysis of the core instability shows many of the traits indicated in the work of Dazin *et al.* (2006a,b), and the number of waves on the dominant modes is in good agreement with the corresponding theoretical model of Saffman (1978). Plots of curl angular velocity isosurfaces show the presence of a secondary vortical structure arranged periodically along the core, which resembles those previously visualised from simulated data and PLIF experiments.

The cross-validation procedure described in section 7.5 shows that the prediction model used for the reconstructions is able to predict the P plane data with remarkable accuracy, yielding an explained variance of 93%. However, this is unlikely to correspond to the accuracy of the final volumetric predictions due to the introduction of further error, which arises from misalignment of both the P and Q planes, PIV error and any deviation from statistical-axisymmetry of the vortex ring flow. The extent of many of these problems was investigated in section 7.4, and it was found that in general, their contribution was minimal. Further assurances of prediction accuracy can be had from the fact that the behaviour of predicted rings coincides so well with previous observations in literature.

Nonetheless, the conclusions drawn were only qualitative in nature, and a quantitative measure of the actual accuracy of the final predictions has not been calculated. However, a preliminary investigation into this task is described in appendix D.6, where it is shown how the various sources of error propagate through the procedure, and influence the ultimate MSE of prediction. However, calculation of the MSE would require the variance of the many potential error sources to estimated, which, due to time constraints, is beyond the scope of this work.

The work in this chapter has served to demonstrate the feasibility of the MP-VSE procedure in practice. In particular, it has shown that the task of designing, implementing and calibrating an MP-VSE experiment can be achieved without undue difficulty or excessive cost. Given that the experiment was intended, first and foremost, as a proof of concept, it can be deemed a considerable success. Nonetheless, due to equipment and budget constraints, there are a number of improvements that could be made in future iterations of the experiment. For example, the availability of only three lasers has served to limit the number of Q planes that can be measured simultaneously, which ultimately limited both the accuracy and the size of the predicted volume. Another important constraint was the minimum interframe time of the Firefly cameras; as explained in section 7.3.4, the inability to reduce the Δ_t of the PIV measurements below 17 ms meant that the light sheet thickness had to be set to 4 mm in order to ensure that sufficient particle pairs were present in the PIV images. For consistency, the cross correlation window sizes in the subsequent PIV processing were then chosen to match the light sheet thickness, which meant that all vector measurements were derived from approximately 4 mm^3 regions of space. The size of this volume determines the level of spatial averaging in the PIV measurements, and places a limit on the smallest de-

tectable structure size. In the present experiment the 4mm^3 volume may have been too large to resolve some of the smallest structure. Any future MP-VSE investigations into unstable vortex rings, especially where the secondary vortex structure is of interest, would likely benefit from the use of faster-framing cameras, allowing for a reduction in the size of this volume.

Chapter 8

Conclusion

The work in this thesis falls into two distinct but related categories. The first comprises of a study into the prediction accuracy of stochastic estimation, and the second concerns a novel application of stochastic estimation, which provides a means of using multiple instantaneous stereo PIV measurements to predict a volumetric flow region. The conclusions, contributions, and avenues of further work in both these areas are provided in the following two sections.

8.1   Stochastic Estimation Theory

A major contribution of this work has been the investigation into the finite-sample behaviour of the stochastic estimation procedure. The main motivation for this work is the equivalence of linear regression and stochastic estimation, which has allowed a wealth of existing knowledge, and a number of alternative techniques, to be applied in the context of flow prediction. Of particular importance is the fact that stochastic estimation is identical to ordinary least squares (OLS) regression. The OLS properties have been presented in a manner appropriate to flow prediction, which will hopefully bring further rigour and confidence to the procedure. In practical terms, the properties help to clarify the relevance of the underlying distribution of the data, and provide guidance when choosing the most appropriate stochastic estimation model to apply to the given data set.

It is noted that OLS properties apply to stochastic estimation only in applications where the model is built from simultaneous observations of the complete set of variables under investigation. However, an alternative is to use a "piecewise" approach, where the model is built from multiple independent observations of subsets of the variables, allowing stochastic estimation to be applied using a small number of point measurement devices. This approach

has fallen out of favour in recent years due to the availability of full-field measurement techniques, but for completeness, it would be beneficial to obtain a set of small sample properties for the piecewise stochastic estimation model. Although not presented here, preliminary quantitative and analytical studies into this task indicate that the piecewise model appears to suffer from a systematic bias, which manifests itself in an overestimate of the magnitude of the model coefficients. With further work, it may be possible to identify the exact nature of this bias, which may subsequently allow for its correction in future applications of the approach.

A second contribution concerns the use of biased regression techniques, and their applicability to flow prediction problems. Biased regression involves introducing an amount of bias into the regression model, which can improve the model stability and led to improved accuracy over the unbiased OLS regression. The exact mechanisms that cause this phenomena have been investigated by deriving theoretical expressions for the mean-square error (MSE) of prediction for both OLS, and four competing biased techniques. The ability of the biased techniques to outperform OLS in practical situations was then investigated in a comprehensive series of simulated stochastic estimation problems. This procedure lead to several notable conclusions. Firstly, it was observed that OLS, which corresponds to the standard implementation of stochastic estimation, can fail catastrophically in situations where the number of observations, m used to build the model is similar to the number of independent variables, n. Also, the biased techniques were shown to outperform OLS in a wide variety of situations. In particular, the technique of principal component regression (PCR) has outperformed or equalled the performance of all the other techniques across the entire range of conditions considered. Furthermore, it has proven to provide reliable predictions even when OLS is subject to the afore-

248

mentioned failure. As such, PCR would appear to be an ideal replacement to OLS as a means of performing stochastic estimation.

While the superior performance of PCR has been proven conclusively in the scenarios considered here, all the tests were based on data taken from a single simulation of a channel flow. Furthermore, all scenarios involved the prediction of a velocity field, using velocity data taken from the same domain. It is possible that these conditions were particularly favourable to PCR, and that other techniques may perform better in other situations. This is particularly relevant in the case of partial least square (PLS), which, theoretically, offers benefits over PCR that may arise when the conditional and unconditional data are taken from distinctly different locations, or are measurements of different properties of the flow. A common stochastic estimation problem involves prediction of a velocity field using pressure measurements, and vice versa, and it would be interesting to see whether this configuration would lead to different results from the ones shown here.

The quantitative performance of OLS has revealed a disparity between the theoretical and actual performance of the technique. While the expression derived for the MSE of OLS correctly identifies the factors which affect prediction accuracy, it nonetheless underestimates the MSE, and does not predict the failure of OLS when $m \approx n$. It has been argued that the discrepancy between theory and practice arises from the finite sample behaviour of the principal components and eigenvalues of $\mathbf{X}$, which was not taken into account in the derivation of the theoretical expression. An intuitive explanation of the cause has been provided, which has been corroborated with empirical evidence. However, a more rigorous mathematical proof would certainly be of use, not just to explain the performance of OLS, but also as a means of further understanding of how and when biased techniques are able to outperform

249

it. It is interesting to observe that the explanations offered here for the poor performance of OLS differ somewhat to those found in literature, especially in the field of chemometrics (e.g. (Næs and Mevik, 2001)). As such, further study into this area may have relevance beyond that of stochastic estimation.

It was realised at a late stage that the simulated channel flow data has been presented using an incorrect grid spacing, which affects some of the plots in this thesis, and means that the vorticity calculations from the data are incorrect. None of the conclusions have been affected by this problem, but it is the author's intention to correct this at the first possible opportunity.

Throughout the thesis, MSE of prediction (in both its theoretical and directly measured form) has been used extensively to investigate competing flow prediction techniques and to assess the accuracy of MP-VSE. This mathematical approach is in contrast to that undertaken in stochastic estimation literature, where prediction performance is generally assessed qualitatively, if at all. However, it must be noted that MSE alone is not sufficient to fully understand the manner in which the error manifests itself in the prediction of a complex flow phenomenon. Intuitively, it is clear that the range of scales within a flow will be predicted with varying degrees of success, depending on the spacing between measurement locations in the experiment, and the fact that stochastic estimation is widely regarded as a technique to predict large scale structure is an implicit recognition of the fact that the smallest scales may be impossible to capture. However, a single error metric such as MSE offers no insight into this behaviour, and further work is required to more fully understand such mechanisms. An initial avenue of investigation could be to compare wavenumber spectra in the original and predicted flow fields, which would give a quantitative measure of how prediction accuracy varies with structure size.

8.2 Volumetric Flow Prediction

The technique of multiple plane volumetric stochastic estimation (MP-VSE) has been proposed, which uses data from multiple stereo PIV measurements to predict a volumetric region of a flow that possesses a homogeneous dimension. The technique provides a means of predicting volumes of a size and resolution that cannot currently be measured by conventional volumetric measurement techniques.

An initial proof of concept was performed by applying the technique in a synthetic experiment using simulated data of a channel flow. As shown by the cross-validation, the resulting predictions accounted for approximately 90% of the variance of the true velocity field, and comparison of the true and predicted data shows excellent qualitative agreement for all three velocity components, as well as vorticity. The practical feasibility of MP-VSE was subsequently demonstrated through the successful application of the technique to the prediction of unstable vortex rings in water. Predictions of the rings during the non-linear instability phase of development have been presented, which closely match the expected behaviour based on theory and previous observations.

To assist in the calibration of the vortex ring experiment, a novel approach to light sheet measurement has been devised. A custom-built target is placed in the path of the light sheet, and an accurate measurement of the light sheet plane position is obtained from a single digital image of the target. Unlike existing approaches to light sheet measurement, the technique is able to provide rapid positional measurement with minimal user interaction. Although the technique is in its early stages of development, the measurement accuracy of the current implementation has been obtained experimentally and

is comparable to that of existing, more time consuming, approaches. However, further improvements to the both the accuracy and speed of the technique should be possible, and the ultimate aim is the removal of any need for user input, thus allowing for true real-time feedback of light sheet position.

The light sheet measurement system was an instrumental part of the calibration of the vortex ring experiment; a process which was subsequently validated from the statistical properties of the PIV measurements. This indicated that the measurement and calibration errors were low, although this was only confirmed qualitatively, and a definitive estimate of the prediction error has not been obtained. Such knowledge will be useful in determining the degree of confidence that can be placed in the predictions, both for the present experiment and for future applications. As such, the development of a reliable approach to obtaining an estimate of prediction error will be an important factor in the adoption of the technique. A preliminary investigation into this task has already been performed (appendix D.6), which shows how the various forms of measurement error propagate through the MP-VSE process. The next step will be to develop a reliable means of estimating the variances of each of these error terms.

Another important avenue of work will be the deployment of MP-VSE in further, more challenging, applications. Presently, there is the possibility of its use as part of an industrial study into the reduction of jet noise, where it would be used to predict the downstream flow of a $\frac{1}{10}$th scale jet engine. Such a project would provide an excellent test of the capabilities of the technique, both in terms of the demanding experimental configuration required to measure the flow, and inherent complexity of predicting flows at high Reynolds numbers.

When considering further applications of the technique, the suitability of MP-VSE to the prediction of vortex rings should not be overlooked. The

252

results from the experiment detailed in this work show that the technique is capable of accurately capturing many aspects of the vortex rings' behaviour. Although the main aim of the experiment was to investigate the performance of MP-VSE, rather than the study of the vortex rings themselves, the success of the experiment suggests that MP-VSE would be a useful tool for this purpose.

Bibliography

Adamcyzk, A. A. and Rimai, L. (1988). 2-dimensional particle tracking velocimetry (PTV): Technique and image processing algorithms, *Experiments in Fluids* **6**: 373–380.

Adrian, R. J. (1977). On the role of conditional averages in turbulence theory, *Proceedings of the 4th biennial symposium on turbulence in liquids*.

Adrian, R. J. (1984). Scattering particle characteristics and their effect on pulsed laser measurements of fluid flow: speckle velocimetry vs. particle image velocimetry, *Applied Optics* **23**(11): 1690–1691.

Adrian, R. J. (1994). Stochastic estimation of conditional structure: a review, *Applied Scientific Research* **53**: 291–303.

Adrian, R. J., Jones, B. G., Chung, M. K., Hassan, Y., Nithianandan, C. K. and Tung, T. C. (1989). Approximation of turbulent conditional averages by stochastic estimation, *Physics of Fluids A* **1**(6): 992–998.

Antonia, R. A. (1981). Conditional sampling in turbulence measurement, *Ann. Rev. Fluid Mech.* **13**: 131–156.

Archer, P. J., Thomas, T. G. and Coleman, G. N. (2008). Direct numeri-

cal simulation of vortex ring evolution from the laminar to early turbulent regime, *J. Fluid Mech.* **598**: 201–266.

Atkinson, C. H. and Soria, J. (2008). Tomographic digital holographic PIV and its application to a water jet, *14th International Symposium on Applications of Laser Techniques to Fluid Dynamics*, Lisbon, Portugal. 7-10 July.

Ausseur, J. M., Pinier, J. T., Glauser, M. N. and Higuchi, H. (2006). Experimental development of a reduced-order model for flow separation control, *44th AIAA Aerospace Sciences Meeting and Exhibit, 9-12 January 2006, Reno, Nevada.*

Bakewell, H. P. and Lumley, J. L. (1967). Viscous sublayer and adjacent wall region in turbulent pipe flow, *Physics of Fluids* **10**(9): 1880–1889.

Belmonte, A., Martin, B. and Goldberg, W. I. (2000). Experimental study of Taylor's hypothesis in a turbulent soap film, *Physics of Fluids* **12**(4): 835–845.

Bergdorf, M., Koumoutsakos, P. and Leonard, A. (2007). Direct numberical simulations at $re_\gamma = 7500$, *J. Fluid Mech.* **581**: 495–505.

Bonnet, J., Cole, D. R., Delville, J., Glauser, M. N. and Ukeiley, L. S. (1994). Stochastic estimation and proper orthogonal decomposition: complimentary techniques for identifying structure, *Experiments in Fluids* pp. 307–214.

Bonnet, J. P., Delville, J., Glauser, M. N., Antonia, R. A., Bisset, D. K., Cole, D. R., Fiedler, H. E., Garem, J. H., Hilberg, D., Jeong, J., Kevlahan, N. K. R., Ukeiley, L. S. and Vincendeau, E. (1998). Collaborative testing of eddy structure identification methods in free turbulent shear flows, *Experiments in Fluids* **25**: 197–225.

Borée, J. (2003). Extended proper orthogonal decomposition: a tool to analyse correlated events in turbulent flows, *Experiments in Fluids* **35**(2): 188–192.

Boulesteix, A. and Strimmer, K. (2005). Predicting transcription factor activities from combined analysis of microarray and ChIP data: a partial least squares approach, *Theoretical Biology and Medical Modelling* **2**: 23.

Brereton, G. J. (1992). Stochastic estimation as a statistical tool for approximating turbulent conditional averages, *Physics of Fluids A* **4**(9): 2046–2054.

Brücker, C. (1996). 3-D scanning-particle image velocimetry: technique and application to a spherical cap wake flow, *Applied Scientific Research* **56**: 157–179.

Cameron, A. C. and Trivedi, P. K. (2005). *Microeconometrics: methods and applications*, Cambridge University Press.

Chatterdee, A. (2000). An introduction to the proper orthogonal decomposition, *Current Science* **78**(7): 808–817.

Citriniti, J. H. and George, W. K. (2000). Reconstruction of the global velocity field in the axisymmetric mixing layer utilizing the proper orthogonal decomposition, *J. Fluid Mech.* **418**: 137–166.

Cole, D. R. and Glauser, M. N. (1998). Applications of stochastic estimation in the axisymmetric sudden expansion, *Physics of Fluids* **10**(11): 2941–2949.

Cole, D. R., Glauser, M. N. and Guezennec, Y. G. (1991). An application of stochastic estimation to the jet mixing layer, *Phys. Fluids* pp. 192–194.

Dazin, A., Dupont, P. and Stanislas, M. (2006a). Experimental characterization of the instability of the vortex rings. Part i: Linear phase, *Experiments in Fluids* **40**: 383–399.

Dazin, A., Dupont, P. and Stanislas, M. (2006b). Experimental characterization of the instability of the vortex rings. Part ii: Non-linear phase, *Experiments in Fluids* **41**: 401–413.

de Jong, S. (1993). SIMPLS: An alternative approach to partial least squares regression, *Chemometrics and Intelligent Laboratory Systems* **18**(3): 251–263.

Delville, J., Ukeiley, L., Cordier, L., Bonnet, J. P. and Glauser, M. (1999). Examination of large-scale structures in a turbulent plane mixing layer. Part 1: Proper orthogonal decomposition, *J.Fluid Mech.* **391**: 91–122.

Druault, P., Delville, J. and Bonnet, J. (2005a). Experimental 3D analysis of the large scale behaviour of a plane turbulent mixing layer, *Flow, Turbulence and Combustion* **74**: 207–233.

Druault, P., Delville, J. and Bonnet, J. (2005b). Proper orthogonal decomposition of the mixing layer flow into coherent structures and turbulent Gaussian fluctuations, *Comptes Rendus Méchanique* **333**: 824–829.

Druault, P., Guibert, P. and Alizon, F. (2005). Use of proper orthogonal decomposition for time interpolation from PIV data, *Experiments in Fluids* **39**: 1009–1023.

Druault, P., Yu, M. and Sagaut, P. (2009). Quadratic stochastic estimation of far-field acoustic pressure with coherent structure events in a 2D compressible plane mixing layer, *Int. J. for Numer. Meth. Fluids* **62**: 906–926.

257

Elsinga, G. E., Scarano, F., Wieneke, B. and van Oudheusden, B. W. (2006). Tomographic particle image velocimetry, *Experiments in Fluids* **41**: 933–947.

Elsinga, G. E., van Oudheusden, B. W. and Scarano, F. (2006). Experimental assessment of tomographic PIV accuracy, *13th International Symposium on Applications of Laser Techniques to Fluid Dynamics*, Lisbon, Portugal. 26-29 June.

Ewing, D. and Citriniti, J. (1999). Examination of of a LSE/POD complimentary technique using single and multi-time information in the axisymmetric shear layer, *Proceedings of the IUTAM Symposium on simulation and identification of organized structures in flows.*

Faber, N. M. (1999). A closer look at the bias-variance trade-off in multivariate calibration, *Journal of Chemometrics* **13**: 185–192.

Fouras, A. and Soria, J. (1998). Accuracy of out-of-plane vorticity measurements derived from in-plane velocity field data, *Experiments in Fluids* **25**: 409–430.

Ganapathisubramani, B. and Lakshminarasimhan, K. (2007). Determination of complete velocity gradient tensor by using cinematographic stereoscopic PIV in a turbulent jet, *Experiments in Fluids* **42**: 923–939.

Geiseke, T. J. and Guezennec, Y. G. (1994). Stochastic estimation of multipoint conditional averages and their spatio-temporal evolution, *Applied Scientific Research* **53**: 305–320.

George, W. K., Arndt, R. E. A. and Corrsin, S. (1989). *Advances in turbulence*, Taylor and Francis.

Giri, N. C. (2004). *Multivariate statistical analysis*, 2 edn, CRC Press.

Glezer, A. (1988). The formation of vortex rings, *Physics of Fluids* **31**(12): 3532–3541.

Graftieaux, L., Michard, M. and Grosjean, N. (2001). Combining PIV, POD and vortex identification algorithms for the study of unsteady swirling flows, *Measurement Science and Technology* **12**: 1422–1429.

Guezennec, Y. G. (1989). Stochastic estimation of coherent structures in turbulent boundary layers, *Physics of Fluids A* **1**(6): 1054–1060.

Harrell Jr., F. E. (2001). *Regression modelling strategies : with applications to linear models, logistic regression, and survival analysis*, Springer.

Hayashi, F. (2000). *Econometrics*, Princeton University Press.

Hinsch, K. D. (2002). Holographic particle image velocimetry, *Measurement Science and Technology* **13**: R61–R72.

Holmes, P., Lumley, J. L. and Berkooz, G. (1996). *Turbulence, coherent structures, dynamical systems and symmetry*, Cambridge University Press.

Hori, T. and Sakakibara, J. (2004). High-speed scanning stereoscopic PIV for 3D vorticity measurement in liquids, *Measurement Science and Technology* **15**: 1067–1078.

Hudy, L. M., Naguib, A. and Humpfreys, W. M. (2007). Stochastic estimation of a separated-flow field using wall-pressure-array measurements, *Physics of Fluids* **19**. DOI:10.1063/1.2472507.

Jackson, J. E. (1991). *A user's guide to principal components*, Wiley.

Jiang, M., Machiraju, R. and Thompson, D. (1990). A novel approach to vortex core region detection, *Joint Eurographics -IEEE TCVG Symposium on Visualisation*, pp. 217–225.

Joilliffe, I. T. (2004). *Principal component analysis*, Springer.

Jordan, P., Schlegel, M., Stalnov, O., Noack, B. R. and Tinney, C. E. (2007). Identifying noisy and quiet modes in a jet, *13th AIAA/CEAS Aeroacoustics Conference*, May 21-23, 2007, Roma, Italy. AIAA Paper 2007–3602.

Kähler, C. J. and Kompenhans, J. (2000). Fundamentals of multiple plane stereo particle image velocimetry, *Experiments in Fluids* **29**(7): S70 – S77.

Keane, R. D. and Adrian, R. J. (1990). Optimization of particle image velocimeters. Part 1: Double pulsed systems, *Measurement Science and Technology* **1**(11): 1202–1215.

Kirby, M. (2001). *Geometric data analysis: An empirical approach to dimensionality reduction and the study of patterns*, Wiley.

Lawley, D. N. (1956). Tests of significance for the latent roots of covariance and correlation matrices, *Biometrika* pp. 128–136.

Lim, T. T. and Nickels, T. B. (1995). Vortex rings, *Fluid Vortices*, Kluwer Academic Publishers, chapter IV, pp. 95–153.

Lumley, J. L. (1967). The structure of inhomogeneous turbulence, *in* A. M. Yaglom and V. I. Tatarsky (eds), *Atmospheric Turbulence and Wave Propagation*, Nauka, Moscow, pp. 166–178.

Lumley, J. L. (1981). Coherent structures in turbulence, *Transition and turbulence; Proceedings of the Symposium on Transition and Turbulence in*

260

Fluids, Madison, WI, October 13-15, 1980. (A82-12438 02-34) New York, Academic Press, 1981, p. 215-242., pp. 215–242.

Maas, H. G., Gruen, A. and Papantoniou, D. (1993). Particle tracking velocimetry in three-dimensional flows: Part 1. Photogrammetric determination of particle coordinates, *Experiments in Fluids* **15**: 133–146.

Matsuda, T. and Sakakibara, J. (2005). On the vortical structure of a round jet, *Physics of Fluids* **17**. DOI:10.1063/1.1840869.

Maxworthy, T. (1977). Some experimental studies of vortex rings, *J. Fluid Mech.* **81**: 465–495.

McGregor, T. J., Spence, D. J. and Coutts, D. W. (2007). Laser-based volumetric color-coded three-dimensional particle velocimetry, *Optics and Lasers in Engineering* **45**: 882–889.

Meng, H., Pan, G., Pu, Y. and Woodward, S. H. (2004). Holographic particle image velocimetry: from film to digital recording, *Measurement Science and Technology* **15**: 673–685.

Mullin, J. A. and Dahm, W. J. A. (2006). Dual-plane stereo particle image velocimetry measurements of velocity gradient tensor fields in turbulent shear flow. ii. Experimental results, *Physics of Fluids* **18**.

Mullin, J. A. and Dahn, W. J. A. (2005). Dual-plane stereo particle image velocimetry (DSPIV) for measuring velocity gradient fields at intermediate and small scales of turbulent flows, *Experiments in Fluids* **38**.

Næs, T. and Mevik, B. (2001). Understanding the collinearity problem in regression and discriminant analysis, *J. Chemometrics* **15**: 413–426.

Naguib, A. M., Wark, C. E. and Juckenöfel, O. (2001). Stochastic estimation and flow sources associated with surface pressure in a turbulent boundary layer, *Physics of Fluids* **13**(9): 2611–2626.

Naitoh, T., Fukuda, N., Gotoh, T., Yamada, H. and Nakajima, K. (2002). Experimental study of axial flow in a vortex ring, *Physics of Fluids* **14**(1): 143–149.

Panik, M. J. (2005). *Advanced statistics from an elementary point of view*, Academic Press.

Papoulis, A. (1985). *Probability, Random Variables, and Stochastic Processes*, 2nd edn, McGraw-Hill.

Pastur, L. R., Lusseyran, F., Fraigneau, Y. and Podvin, V. (2005). Determining the spectral signature of spatial coherent structures in an open cavity flow, *Physical Review E* **72**. DOI:10.1103/PhysRevE.72.065301.

Payne, F. R. and Lumley, J. L. (1967). Large eddy structure of the turbulent wake behind a circular cylinder, *Physics of Fluids* **10**: 194–196.

Pederson, J. M. and Meyer, K. E. (2002). POD analysis of flow structures in a scale model of a ventilated room, *Experiments in Fluids* **33**: 940–949.

Perret, L., Braud, P., Fourmet, C., David, L. and Delville, J. (2006). 3-component acceleration field measurement by dual-time stereoscopic particle image velocimetry, *Experiments in Fluids* **40**(5): 813–824.

Pfadler, S., Dinkelacker, F., Beyrau, F. and Leipertz, A. (2009). High resolution dual-plane stereo-PIV for validation of subgrid scale models in large-eddy simulations of turbulent premixed flames, *Combustion and Flame* **156**: 1552–1564.

Picard, C. and Delville, J. (2000). Pressure velocity coupling in a subsonic round jet, *International Journal of Heat and Fluid Flow* **21**: 359–364.

Pick, S. and Lehmann, F. (2009). Stereoscopic PIV on multiple color-coded light sheets and its application to axial flow in flapping robotic insect wings, *Experiments in Fluids* . DOI:10.1007/s00348-009-0687-5.

Pickering, C. J. D. and Halliwell, N. (1984). Laser speckle velocimetry and particle image velocimetry: photographic film noise, *Applied Optics* **23**(17): 2961–2969.

Post, M. E., Trump, D. D., Goss, L. P. and Hancock, R. D. (1994). Two-color particle-imaging velocimetry using a single argon-ion laser, *Experiments in Fluids* **16**: 23–272.

Prasad, A. K. (2000). Stereoscopic particle image velocimetry, *Experiments in Fluids* pp. 103–116.

Pu, Y. and Meng, H. (2000). An advanced off-axis holographic particle image velocimetry (HPIV) system, *Experiments in Fluids* **29**: 184–197.

Raffel, M., Willert, C., Wereley, S. and Kompenhans, J. (2007). *Particle image velocimetry: a practical guide*, Springer.

Rao, C. R., Toutenburg, H., Shalabh and Heumann, C. (2008). *Linear models and generalizations : least squares and alternatives*, 3rd edn, Springer, Springer.

Ruck, B. (2009). Color-coded tomography in fluid mechanics, *Optics and Laser Technology* . DOI:10.1016/j.optlastec2009.06.004.

263

Saffman, P. G. (1978). The number of rings on unstable vortex rings, *J. Fluid Mech.* **84**(4): 625–639.

Sakai, Y., Tanaka, N. and Kushida, T. (2006). On the development of coherent structure in a plane jet. part 1: Characteristics of two-point correlation and analysis of eigenmodes by the KL expansion, *JSME International Journal, Series B* **49**(1): 115–124.

Scarano, F. (2002). Iterative image deformation methods in PIV, *Measurement Science and Technology* **13**: R1–R19.

Scarano, F., David, L., Bsibsi, M. and Calluaud, D. (2005). S-PIV comparative assessment: image dewarping + misallignment correction and pinhole + geometric back projection, *Experiments in Fluids* pp. 257–266.

Shariff, K., Verzicco, R. and Orlandi, P. (1994). A numerical study of three-dimensional vortex ring instabilities: viscous corrections and early nonlinear stage, *J. Fluid. Mech.* **279**: 351–375.

Shen, G. and Wei, R. (2004). Digital holography particle image velocimetry for the measurement of 3D-3C flows, *Optics and Lasers in Engineering* **43**: 1039–1055.

Sirovich, L. (1987). Turbulence and the dynamics of coherent structures, *Quarterly Journal of Applied Mathematics* **45**: 561–590.

Soria, J. and Atkinson, C. (2008). Towards 3C-3D digital holographic fluid velocity vector field measurement - tomographic digital holographic PIV (tomo-HPIV), *Measurement Science and Technology* **19**. DOI:10.1088/0957-0233/19/7/074022.

Stanislas, M., Okamoto, K., Kähler, C. J., Westerweel, J. and Scarano, F. (2008). Main results of the third international PIV challenge, *Experiments in Fluids* **45**: 27–72.

Stock, J. H. and Watson, M. W. (2002). *Introduction to Econometrics*, first international edn, Addison-Wesley.

Taylor, G. I. (1938). The spectrum of turbulence, *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **164**(919): 476–490.

Taylor, J. and Glauser, M. N. (2002). Towards practical flow sensing and control via POD and LSE based low-dimensional tools, *ASME Fluids Engineering Division Summer Meeting, July 14-18, 2002, Montreal, Quebec, Canada*.

ter Vraak, C. F. and de Jong, S. (1998). The objective function of partial least-squares, *Journal of Chemometrics* **12**: 41–54.

Tiki, M. L. and Akkaya, A. D. (2004). *Robust Estimation and Hypothesis Testing*, New Age International.

Tinney, C. E., Coifet, F., Delville, J., Hall, A. M., Jordan, P. and Glauser, M. N. (2006). On spectral linear stochastic estimation, *Experiments in Fluids* pp. 763–775.

Tinney, C. E., Ukeiley, L. S. and Glauser, M. N. (2008). Low-dimensional characteristics of a transonic jet. Part 2. Estimate and far field prediction, *J. Fluid Mech.* **615**: 53–92.

Townsend, A. A. (1980). *The structure of turbulent shear flow*, Cambridge University Press.

Trefethen, L. N. and D Bau, I. (1997). *Numerical Linear Algebra*, SIAM.

Tropea, C., Yarin, A. L. and Foss, J. F. (eds) (2007). *Springer Handbook of Experimental Fluid Mechanics*, Springer.

Tung, T. C. and Adrian, R. J. (1980). Higher-order estimates of conditional eddies in isotropic turbulence, *Physis of Fluids* **23**: 1469–1470.

Ukeiley, L. and Murray, N. (2005). Velocity and surface pressure measurements in an open cavity, *Experiments in Fluids* **38**: 656–671.

van Doorne, C. W. H. and Westerweel, J. (2007). Measurement of laminar, transitional and turbulent pipe flow using stereoscopic PIV, *Experiments in Fluids* **42**: 259–279.

van Oudheusden, B. W., Scarano, F., van Hinsberg, N. P. and Watt, D. W. (2005). Phase-resolved characterization of vortex shedding in the near wake of a square-section cylinder at incidence, *Experiments in Fluids* **39**: 86–98.

Walker, S. (2002). Two-axis Scheimpflug focusing for particle image velocimetry, *Experiments in Fluids* **13**: 1–12.

Warsi, Z. U. I. (2006). *Fluid dynamics: theoretical and computational approaches*, CRC Press.

Wernet, M. P., John, W. T. and Bridges, J. (2005). Dual PIV systems for space-time correlations in hot jets, *Instrumentation in Aerospace Simulation Facilities, 2003. ICIASF '03. 20th Interational Congress on*, pp. 127–135.

White, H. (1980). Usin least squares to approximate unknown regression functions, *International Economic Review* **21**: 149–170.

Wieneke, B. (2005). Stereo-PIV using self-calibration on particle images, *Experiments in Fluids* pp. 267–280.

Wieneke, B. and Taylor, S. (2006). Fat-sheet PIV with computation of full 3D-strain tensor using tomographic reconstruction, *13th International Symposium on Applications of Laser Techniques to Fluid Dynamics*, Lisbon, Portugal. 26-29 June.

Wold, H. (1975). Soft modelling by latent variables: the non-linear iterative partial least squares (NIPALS) approach, *in* J. Gani (ed.), *Perspectives in Probability and Statistics: Papers, in Honour of M.S. Bartlett on the Occasion of his Sixty-fifth Birthday*, Academic Press, pp. 117 –142.

Wold, S., Sjöström, M. and Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics, *Chemometrics and Intelligent Laboratory Systems* **58**.

Yang, W. (ed.) (2001). *Handbook of flow visualization*, Taylor Francis. Second edition.

Yoshimoto, T. (2009). *Twisted wake of a sphere in a rotating fluid*, Master's thesis, Hokkaido University.

Appendix A

Properties of the Ordinary Least Square Estimator

A.1 Small-sample Unbiasedness of the OLS Estimator

The expectation of the OLS estimator $\hat{\mathbf{b}}_{OLS}$, conditional on $\mathbf{X}$, is given by:

$$\mathrm{E}\left[\hat{\mathbf{b}}_{OLS}|\mathbf{X}\right] = \mathrm{E}\left[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}|\mathbf{X}\right] \tag{A.1}$$

Because $\mathbf{y} = \mathbf{X}\check{\mathbf{b}} + \mathbf{e}$:

$$\begin{aligned}
\mathrm{E}\left[\hat{\mathbf{b}}_{OLS}|\mathbf{X}\right] &= \mathrm{E}\left[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\check{\mathbf{b}} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{e}|\mathbf{X}\right] \\
&= \check{\mathbf{b}} + \mathrm{E}\left[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{e}|\mathbf{X}\right]
\end{aligned} \tag{A.2}$$

Because the expectation is conditioned on $\mathbf{X}$, any instances of the term can be considered as fixed quantities:

$$\mathrm{E}\left[\hat{\mathbf{b}}_{OLS}|\mathbf{X}\right] = \check{\mathbf{b}} + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathrm{E}\left[\mathbf{e}|\mathbf{X}\right] \tag{A.3}$$

From the second tier of OLS assumptions (section 4.1.2), it is known that the $\mathrm{E}[\mathbf{e}|\mathbf{X}] = 0$, leading to:

$$\mathrm{E}\left[\hat{\mathbf{b}}_{OLS}|\mathbf{X}\right] = \check{\mathbf{b}} \tag{A.4}$$

which shows that the OLS estimator is conditionally unbiased. To show that the estimator is unconditionally unbiased, the expectation of both sides of (A.4) is taken:

$$\mathrm{E}\left[\mathrm{E}\left[\hat{\mathbf{b}}_{OLS}|\mathbf{X}\right]\right] = \mathrm{E}\left[\check{\mathbf{b}}\right] \\ = \check{\mathbf{b}} \tag{A.5}$$

The *law of iterated expectations* (LIE) states that for two random variables a and b, the expectation of $\mathrm{E}[a|b]$ is simply the expectation of a, i.e.:

$$\mathbf{E}[a] = \mathrm{E}\left[\mathrm{E}\left[a|b\right]\right] \tag{A.6}$$

Therefore:

$$\mathbf{E}[\hat{\mathbf{b}}_{OLS}] = \check{\mathbf{b}} \tag{A.7}$$

A.2   Small-sample Unbiasedness of the OLS Predictor

The expectation of $y - \hat{y}$, conditional on both $\mathbf{X}$ and $\mathbf{x}$, is given by:

$$
\begin{aligned}
\mathrm{E}\left[y - \hat{y}|\mathbf{x}, \mathbf{X}\right] &= \mathrm{E}\left[\mathbf{x}\check{\mathbf{b}} + e - \mathbf{x}\hat{\mathbf{b}}_{OLS}|\mathbf{x}, \mathbf{X}\right] \\
&= \mathrm{E}\left[\mathbf{x}\check{\mathbf{b}}|\mathbf{x}, \mathbf{X}\right] + \mathrm{E}\left[e|\mathbf{x}, \mathbf{X}\right] - \mathrm{E}\left[\mathbf{x}\hat{\mathbf{b}}_{OLS}|\mathbf{x}, \mathbf{X}\right] \\
&= \mathbf{x}\check{\mathbf{b}} + \mathrm{E}\left[e|\mathbf{x}, \mathbf{X}\right] - \mathbf{x}\mathrm{E}\left[\hat{\mathbf{b}}_{OLS}|\mathbf{x}, \mathbf{X}\right]
\end{aligned}
\tag{A.8}
$$

Under the second tier of OLS assumptions, $\mathrm{E}\left[e|\mathbf{x}, \mathbf{X}\right] = 0$, and from Appendix A.1, $\mathrm{E}\left[\hat{\mathbf{b}}_{OLS}|\mathbf{x}, \mathbf{X}\right] = \check{\mathbf{b}}$. Therefore:

$$
\begin{aligned}
\mathrm{E}\left[y - \hat{y}|\mathbf{x}, \mathbf{X}\right] &= \mathbf{x}\check{\mathbf{b}} - \mathbf{x}\check{\mathbf{b}} \\
&= 0
\end{aligned}
\tag{A.9}
$$

A.3   Efficiency of the OLS estimator

This section demonstrates that the OLS estimator is the best linear unbiased estimator (BLUE) of $\check{\mathbf{b}}$, under the Gauss-Markov assumptions (assumptions 1-3). The following is based on the derivation from Stock and Watson (2002).

The finite matrix of observations $\mathbf{y}$ used to build a linear estimator $\tilde{\mathbf{b}} = \mathbf{A}\mathbf{y}$ can be expressed as $\mathbf{y} = \mathbf{X}\check{\mathbf{b}} + \mathbf{e}$, so the expectation of $\tilde{\mathbf{b}}$, conditional on $\mathbf{X}$, can be written as:

$$
\begin{aligned}
\mathrm{E}\left[\tilde{\mathbf{b}}|\mathbf{X}\right] &= \mathrm{E}\left[\mathbf{A}\mathbf{y}|\mathbf{X}\right] \\
&= \mathrm{E}\left[\mathbf{A}\mathbf{X}\check{\mathbf{b}}|\mathbf{X}\right] + \mathrm{E}\left[\mathbf{A}\mathbf{e}|\mathbf{X}\right] \\
&= \mathbf{A}\mathbf{X}\check{\mathbf{b}} + \mathbf{A}\mathrm{E}\left[\mathbf{e}|\mathbf{X}\right]
\end{aligned}
\tag{A.10}
$$

From assumption 2, $\mathrm{E}[\mathbf{e}|\mathbf{X}] = 0$, so the second term is zero. As $\check{\mathbf{b}}$ is unbiased:

$$\mathrm{E}\left[\tilde{\mathbf{b}}|\mathbf{X}\right] = \mathbf{A}\mathbf{X}\check{\mathbf{b}} = \check{\mathbf{b}}$$

which implies that:

$$\mathbf{A}\mathbf{X} = \mathbf{I} \tag{A.12}$$

Therefore:

$$\tilde{\mathbf{b}} = \mathbf{A}\mathbf{y}$$

$$= \mathbf{A}\mathbf{X}\check{\mathbf{b}} + \mathbf{A}\mathbf{e} \tag{A.13}$$

$$= \check{\mathbf{b}} + \mathbf{A}\mathbf{e}$$

Defining the variance-covariance matrix of $\tilde{\mathbf{b}}$ as:

$$\begin{aligned}
\mathrm{Var}\left[\tilde{\mathbf{b}}|\mathbf{X}\right] &= \mathrm{E}\left[\tilde{\mathbf{b}}\tilde{\mathbf{b}}^{T}|\mathbf{X}\right] \\
&= \mathrm{E}\left[\mathbf{A}\mathbf{e}\mathbf{e}^{T}\mathbf{A}^{T}|\mathbf{X}\right] \\
&= \mathbf{A}\mathrm{E}\left[\mathbf{e}\mathbf{e}^{T}|\mathbf{X}\right]\mathbf{A}^{T} \\
&= \mathbf{A}\sigma^{2}\mathbf{I}\mathbf{A}^{T} \\
&= \sigma^{2}\mathbf{A}\mathbf{A}^{T}
\end{aligned} \tag{A.14}$$

By expressing the OLS estimator $\hat{\mathbf{b}}_{OLS} = (\mathbf{X}^{T}\mathbf{X})^{-1}\mathbf{X}^{T}\mathbf{y}$ as $\hat{\mathbf{b}}_{OLS} = \hat{\mathbf{A}}\mathbf{y}$, the general matrix $\mathbf{A}$ can be written as:

$$\mathbf{A} = \hat{\mathbf{A}} + \mathbf{D} \tag{A.15}$$

where $\mathbf{D}$ is the difference between $\hat{\mathbf{b}}_{OLS}$ and $\tilde{\mathbf{b}}$. This leads to the following

properties:

$$\hat{\mathbf{A}}\hat{\mathbf{A}}^T = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}$$
$$= (\mathbf{X}^T\mathbf{X})^{-1} \tag{A.16}$$

$$\hat{\mathbf{A}}\mathbf{A}^T = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{A}^T$$
$$= (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{A}\mathbf{X}) \tag{A.17}$$
$$= (\mathbf{X}^T\mathbf{X})^{-1}$$

and finally:

$$\hat{\mathbf{A}}\mathbf{D}^T = \hat{\mathbf{A}}(\mathbf{A} - \hat{\mathbf{A}})^T$$
$$= \hat{\mathbf{A}}\mathbf{A}^T - \hat{\mathbf{A}}\hat{\mathbf{A}}^T \tag{A.18}$$
$$= 0$$

The variance matrix of $\tilde{\mathbf{b}}$ can now be written as:

$$\mathrm{Var}\left[\tilde{\mathbf{b}}|\mathbf{X}\right] = \sigma^2\left(\hat{\mathbf{A}} + \mathbf{D}\right)\left(\hat{\mathbf{A}} + \mathbf{D}\right)^T$$
$$= \sigma^2\left(\hat{\mathbf{A}}\hat{\mathbf{A}}^T + \hat{\mathbf{A}}\mathbf{D}^T + \mathbf{D}\hat{\mathbf{A}}^T + \mathbf{D}\mathbf{D}^T\right) \tag{A.19}$$
$$= \sigma^2(\mathbf{X}^T\mathbf{X})^{-1} + \sigma^2\mathbf{D}\mathbf{D}^T$$

For any matrix D, the resulting matrix $\mathbf{D}\mathbf{D}^T$ will be non-negative, which leads to the inequality:

$$\mathrm{Var}\left[\tilde{\mathbf{b}}|\mathbf{X}\right] \geq \mathrm{Var}\left[\hat{\mathbf{b}}|\mathbf{X}\right] \tag{A.20}$$

which will only hold with equality when $\tilde{\mathbf{b}} = \hat{\mathbf{b}}_{OLS}$, in which case:

$$\mathrm{Var}\left[\hat{\mathbf{b}}|\mathbf{X}\right] = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1} \tag{A.21}$$

Hence, under the assumptions 1-3, $\hat{\mathbf{b}}_{OLS}$ is the BLUE of $\check{\mathbf{b}}$.

A.4 MSE of the OLS predictor under the Gauss-Markov Assumptions

The following is based on a related derivation from Stock and Watson (2002). For the unbiased linear estimator $\tilde{\mathbf{b}}$, the MSE of subsequent predictions $\tilde{y} = \mathbf{x}\tilde{\mathbf{b}}$, conditional on $\mathbf{x}$ and $\mathbf{X}$, is expressed as:

$$
\begin{aligned}
\mathrm{MSE}\left[\tilde{y}|\mathbf{x}, \mathbf{X}\right] =& \mathrm{E}\left[(y - \tilde{y})^2 | \mathbf{x}, \mathbf{X}\right] \\
=& \mathrm{E}\left[\left(\mathbf{x}\check{\mathbf{b}} + e - \mathbf{x}\tilde{\mathbf{b}}\right)^2 | \mathbf{x}, \mathbf{X}\right]
\end{aligned}
\tag{A.22}
$$

Substituting the expression for $\tilde{\mathbf{b}}$ in (A.13):

$$
\begin{aligned}
\mathrm{MSE}\left[\tilde{y}|\mathbf{x}, \mathbf{X}\right] =& \mathrm{E}\left(\mathbf{x}\check{\mathbf{b}} + e - \mathbf{x}(\check{\mathbf{b}} + \mathbf{A}e)\right)^2 | \mathbf{x}, \mathbf{X} \\
=& \mathrm{E}\left[(e - \mathbf{x}\mathbf{A}e)^2 | \mathbf{x}, \mathbf{X}\right] \\
=& \mathrm{E}\left[e^2 | \mathbf{x}, \mathbf{X}\right] - 2\mathrm{E}\left[e\mathbf{x}\mathbf{A}e | \mathbf{x}, \mathbf{X}\right] + \mathrm{E}\left[(\mathbf{x}\mathbf{A}e)^2 | \mathbf{x}, \mathbf{X}\right]
\end{aligned}
\tag{A.23}
$$

From OLS assumption 3.b (section 4.1.2), the error terms e and $\mathbf{e}$ are independent, with zero mean, and so the expectation of the second term is zero. The first term corresponds to the variance of e:

$$
\begin{aligned}
\mathrm{MSE}\left[\tilde{y}|\mathbf{x}, \mathbf{X}\right] =& \sigma^2 + \mathrm{E}\left[(\mathbf{x}\mathbf{A}e)^2 | \mathbf{x}, \mathbf{X}\right] \\
=& \sigma^2 + \mathrm{E}\left[\mathbf{x}\mathbf{A}ee^T\mathbf{A}^T\mathbf{x}^T | \mathbf{x}, \mathbf{X}\right] \\
=& \sigma^2 + \mathbf{x}\mathbf{A}\mathrm{E}\left[ee^T | \mathbf{x}, \mathbf{X}\right]\mathbf{A}^T\mathbf{x}^T \\
=& \sigma^2 + \mathbf{x}\mathbf{A}\sigma^2\mathbf{I}\mathbf{A}^T\mathbf{x}^T \\
=& \sigma^2 + \mathbf{x}(\sigma^2\mathbf{A}\mathbf{A}^T)\mathbf{x}^T
\end{aligned}
\tag{A.24}
$$

The expression $\sigma^2 \mathbf{A}\mathbf{A}^T$ is equivalent to $\text{Var}[\tilde{\mathbf{b}}|\mathbf{X}]$, which, from (A.19), can also be written as $\sigma^2(\mathbf{X}^T\mathbf{X})^{-1} + \sigma^2\mathbf{D}\mathbf{D}^T$. Therefore:

$$\text{MSE}\left[\tilde{y}|\mathbf{x}, \mathbf{X}\right] = \sigma^2 + \sigma^2\mathbf{x}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}^T + \sigma^2\mathbf{x}\mathbf{D}\mathbf{D}^T\mathbf{x}^T \tag{A.25}$$

which shows that the MSE of predictions is minimised when $\mathbf{D} = 0$, which occurs only when $\tilde{\mathbf{b}} = \hat{\mathbf{b}}_{OLS}$

A.5 Maximum Likelihood Estimation of the Linear Regression Coefficients

This section obtains the maximum likelihood estimate $\mathbf{b}_{MLE}$ for the linear model $y = \mathbf{x}\mathbf{b} + e$, under OLS assumptions 1-4. It was established in section 4.1.2 that the relevant likelihood function is:

$$\mathcal{L}(\theta|\mathbf{y}, \mathbf{X}) = \left(2\pi\sigma^2\right)^{-\frac{m}{2}} \exp\left(-\frac{(\mathbf{y} - \mathbf{X}\mathbf{b})^T(\mathbf{y} - \mathbf{X}\mathbf{b})}{2\sigma^2}\right) \tag{A.26}$$

where:

$$\theta = \begin{bmatrix} \mathbf{b} \\ \sigma^2 \end{bmatrix} \tag{A.27}$$

The MLE estimate of $\mathbf{b}$ is found by considering the log-likelihood function:

$$\ln\mathcal{L}(\theta|\mathbf{y}, \mathbf{X}) = -\frac{m}{2}\ln\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}\left(\mathbf{y} - \mathbf{X}\check{\mathbf{b}}\right)^T\left(\mathbf{y} - \mathbf{X}\check{\mathbf{b}}\right) \tag{A.28}$$

The first derivative with respect to $\mathbf{b}$ is:

$$\frac{\delta\ln\mathcal{L}(\theta|\mathbf{y}, \mathbf{X})}{\delta\mathbf{b}} = -\frac{1}{\sigma^2}\mathbf{X}^T\mathbf{y} + \frac{1}{\sigma^2}\mathbf{X}^T\mathbf{X}\check{\mathbf{b}} \tag{A.29}$$

Setting to zero and rearranging, yields the Normal equations (equation (4.13)):

$$\mathbf{X}^T\mathbf{X}\mathbf{b} = \mathbf{X}^T\mathbf{y} \tag{A.30}$$

Hence, under assumptions 1-4, the MLE and OLS estimates of $\mathbf{b}$ are identical:

$$\hat{\mathbf{b}}_{MLE} = \hat{\mathbf{b}}_{OLS} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \tag{A.31}$$

A.6 Obtaining the Cramér-Rao Lower Bound of the Linear Regression Coefficients

Having defined the log likelihood, the next stage in the task of determining the Cramér-Rao Lower Bound (CRLB) involves determining the Fisher information matrix $\mathbb{I}(\theta)$. Put simply, the Fisher information matrix is a measure of the information that $\mathbf{y}$ theoretically contains about the estimator θ. The CRLB is then defined as the inverse of $\mathbb{I}(\theta)$.

The Fisher information matrix can be defined in terms of the matrix of partial second derivatives of $\ln \mathcal{L}(\theta)$:

$$\mathbb{I}(\theta) = - \mathbf{E}\left[\frac{\delta^2 \ln \mathcal{L}(\theta)}{\delta\theta\delta\theta^T}\right] \tag{A.32}$$

In the present case, the parameter vector $\theta = [\mathbf{b}^T \sigma^2]^T$, so the matrix of second partial derivatives can be partitioned into four sections:

$$\mathbb{I}(\theta) = \begin{bmatrix} \frac{\delta^2 \ln \mathcal{L}(\theta|\mathbf{X})}{\delta\mathbf{b}\delta\mathbf{b}^T} & \frac{\delta^2 \ln \mathcal{L}(\theta|\mathbf{X})}{\delta\mathbf{b}\delta\sigma^2} \\ \frac{\delta^2 \ln \mathcal{L}(\theta|\mathbf{X})}{\delta\sigma^2\delta\mathbf{b}^T} & \frac{\delta^2 \ln \mathcal{L}(\theta|\mathbf{X})}{\delta(\sigma^2)^2} \end{bmatrix} \tag{A.33}$$

which ultimately leads to the Fisher information matrix[1]:

$$\mathbb{I}(\theta|\mathbf{X}) = \begin{bmatrix} \frac{1}{\sigma^2}\mathbf{X}^T\mathbf{X} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{bmatrix} \tag{A.34}$$

and the CRLB:

$$\text{CRLB} = \mathbb{I}(\theta|\mathbf{X})^{-1} = \begin{bmatrix} \sigma^2(\mathbf{X}^T\mathbf{X})^{-1} & 0 \\ 0 & \frac{2\sigma^4}{n} \end{bmatrix} \tag{A.35}$$

The CRLB for the ith parameter in the vector θ is given by partition i, i in the full CRLB matrix. Hence, under assumptions 1-4, the variance of an unbiased estimator of $\hat{\mathbf{b}}$ is bounded by:

$$\text{Var}(\tilde{\mathbf{b}}|\mathbf{X}) \geq \sigma^2(\mathbf{X}^T\mathbf{X})^{-1} \tag{A.36}$$

From Appendix A.3, the variance of the OLS estimator is:

$$\text{Var}(\hat{\mathbf{b}}_{OLS}|\mathbf{X}) = \sigma^2(\mathbf{X}^T\mathbf{X})^{-1} \tag{A.37}$$

which shows that the OLS estimator achieves the CRLB, and hence is the uniformly minimum variance unbiased estimator (UMVUE).

---

[1]For a full derivation of the Fisher information matrix, see Hayashi (2000)

A.7 Expectation of the Error Term in the Optimum Linear Regression Model

In this section it is shown that for the optimum linear regression model:

$$y = \mathbf{x}\check{\mathbf{b}} + e \tag{A.38}$$

the expectation of the error term is zero, assuming that the model includes a constant term b_c. The optimum model is defined as the linear model that minimises the mean square error of predictions, $\mathrm{E}[e^2] = \mathrm{E}[(y - \mathbf{x}\check{\mathbf{b}})^2]$ over the entire population of $\mathbf{x}$ and y. Such a model is subject to the orthogonality principle, which states that the error term is orthogonal to the independent variables, $\mathbf{x} = [x_1\ x_2\ \ldots\ x_n]$. That is:

$$\mathrm{E}[ex_i] = 0 \tag{A.39}$$

for $i = 1, 2, \ldots, n$. If the constant term is included in the model, then $x_1 = 1$, and so:

$$\mathrm{E}[ex_1] = \mathrm{E}[e] = 0 \tag{A.40}$$

which shows that the expectation of the error is zero.

Appendix B

Bias-Variance Decompositions

B.1 Decomposition of the OLS Prediction Error

In this section, a decomposition of the mean square error of predictions for the OLS predictor performed. The approach employed here is based loosely on the work of Næs and Mevik (2001).

The MSE of prediction is defined as:

$$
\begin{aligned}
\mathrm{MSE}\left[\hat{y}\right] &= \mathrm{E}\left[(y - \hat{y})^2\right] \\
&= \mathrm{E}\left[\left(y - \mathbf{x}\hat{\mathbf{b}}_{OLS}\right)^2\right] \\
&= \mathrm{E}\left[\left(y - \mathbf{x}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}\right)^2\right]
\end{aligned}
\tag{B.1}
$$

Because $y = \mathbf{x}\check{\mathbf{b}} + e$, and $\mathbf{y} = \mathbf{X}\check{\mathbf{b}} + \mathbf{e}$:

$$
\begin{aligned}
\mathrm{MSE}\left[\hat{y}\right] =& \mathrm{E}\left[\left(\mathbf{x}\check{\mathbf{b}} + e - \mathbf{x}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\left(\mathbf{X}\check{\mathbf{b}} + \mathbf{e}\right)\right)^2\right] \\
=& \mathrm{E}\left[\left(\mathbf{x}\check{\mathbf{b}} + e - \mathbf{x}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{X}\check{\mathbf{b}} - \mathbf{x}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{e}\right)^2\right] \\
=& \mathrm{E}\left[\left(\mathbf{x}\check{\mathbf{b}} + e - \mathbf{x}\check{\mathbf{b}} - \mathbf{x}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{e}\right)^2\right] \qquad \text{(B.2)} \\
=& \mathrm{E}\left[\left(e - \mathbf{x}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{e}\right)^2\right] \\
=& \mathrm{E}\left[e^2 - 2e\mathbf{x}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{e} + \left(\mathbf{x}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{e}\right)^2\right]
\end{aligned}
$$

As e and $\mathbf{e}$ are uncorrelated, the term $-2e\mathbf{x}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{e}$ cancels to zero:

$$
\mathrm{MSE}\left[\hat{y}\right] = \mathrm{E}\left[e^2\right] + \mathrm{E}\left[\left(\mathbf{x}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{e}\right)^2\right] \qquad \text{(B.3)}
$$

To simplify the calculation, consider OLS assumptions 3 and 4 to be valid. Therefore, $\mathrm{E}\left[e^2|\mathbf{x}, \mathbf{X}\right] = \sigma^2$, and from the fact that $\mathrm{Cov}(e_i, e_j) = 0$, it follows that $\mathrm{E}\left[\mathbf{e}\mathbf{e}^{\mathbf{T}}|\mathbf{x}, \mathbf{X}\right] = \sigma^2\mathbf{I}$, where $\mathbf{I}$ is the identity matrix:

$$
\begin{aligned}
\mathrm{MSEP}\left[\hat{y}\right] =& \sigma^2 + \mathrm{E}\left[\left(\mathbf{x}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{e}\right)^2\right] \\
=& \sigma^2 + \mathrm{E}\left[\mathbf{x}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{e}\mathbf{e}^T\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{x}^T\right]
\end{aligned} \qquad \text{(B.4)}
$$

Applying the law of iterated expectations to the right hand side:

$$
\begin{aligned}
\text{MSE}\left[\hat{y}\right] =&\sigma^2 + \text{E}\left[\text{E}\left[\mathbf{x}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{e}\mathbf{e}^T\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{x}^T|\mathbf{x},\mathbf{X}\right]\right] \\
=&\sigma^2 + \text{E}\left[\mathbf{x}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\text{E}\left[\mathbf{e}\mathbf{e}^T|\mathbf{x},\mathbf{X}\right]\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{x}^T\right] \\
=&\sigma^2 + \text{E}\left[\mathbf{x}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\sigma^2\mathbf{I}\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{x}^T\right] \\
=&\sigma^2 + \sigma^2\text{E}\left[\mathbf{x}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{x}^T\right] \\
=&\sigma^2 + \sigma^2\text{E}\left[\mathbf{x}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{x}^T\right]
\end{aligned}
\tag{B.5}
$$

The covariance matrix of $\mathbf{X}$ is defined as $\mathbf{S} = \frac{1}{m}\mathbf{X}\mathbf{X}^T$, which allows $(\mathbf{X}^T\mathbf{X})^{-1}$ to be expressed as $(m\mathbf{S})^{-1} = \frac{1}{m}(\mathbf{S})^{-1}$. Therefore:

$$
\text{MSE}\left[\hat{y}\right] =\sigma^2 + \frac{\sigma^2}{m}\text{E}\left[\mathbf{x}(\mathbf{S})^{-1}\mathbf{x}^T\right]
\tag{B.6}
$$

At this point, an alternative approach to that of Næs and Mevik (2001) is taken. Representing $\mathbf{S}$ as a function of it's eigenvectors (i.e. the principal components of $\mathbf{X}$) yields $\mathbf{S} = \mathbf{U}\mathbf{L}\mathbf{U}^T = \sum_{k=1}^{n}l_k\mathbf{u}_k\mathbf{u}_k^T$:

$$
\begin{aligned}
\text{MSE}\left[\hat{y}\right] =&\sigma^2 + \frac{\sigma^2}{m}\text{E}\left[\mathbf{x}(\sum_{k=1}^{n}l_k\mathbf{u}_k\mathbf{u}_k^T)^{-1}\mathbf{x}^T\right] \\
=&\sigma^2 + \frac{\sigma^2}{m}\text{E}\left[\sum_{k=1}^{n}\frac{\mathbf{x}\mathbf{u}_k\mathbf{u}_k^T\mathbf{x}^T}{l_k}\right] \\
=&\sigma^2 + \frac{\sigma^2}{m}\text{E}\left[\sum_{k=1}^{n}\frac{(\mathbf{x}\mathbf{u}_k)^2}{l_k}\right]
\end{aligned}
\tag{B.7}
$$

By making the simplifying assumption that the sample principal components of X are identical to those of the underlying population, then $\text{E}\left[(\mathbf{x}\mathbf{u_k})^2\right] = \text{E}\left[(\mathbf{x}\boldsymbol{\gamma}_k)^2\right] = \lambda_k$, which is the variance accounted for by the kth principal

component:

$$\text{MSE}\,[\hat{y}] = \sigma^2 + \frac{\sigma^2}{m}\text{E}\left[\sum_{k=1}^{n}\frac{\lambda_k}{l_k}\right] \tag{B.8}$$

Also, assuming the sample and population roots to be equal, $l_k = \lambda_k$:

$$\begin{aligned}
\text{MSE}\,[\hat{y}] &= \sigma^2 + \frac{\sigma^2}{m}\sum_{k=1}^{n}1 \\
&= \sigma^2 + \frac{m}{n}\sigma^2 \\
&= \text{unexplained variance} + \text{model variance}
\end{aligned} \tag{B.9}$$

B.2 Simplification of the PCR Estimator

From equation (4.76), the PCR estimator is given by:

$$\hat{\mathbf{b}}_{PCR} = \tilde{\mathbf{U}}\left(\tilde{\mathbf{U}}^T\mathbf{X}^T\mathbf{X}\tilde{\mathbf{U}}\right)^{-1}\tilde{\mathbf{U}}^T\mathbf{X}^T\mathbf{y} \tag{B.10}$$

To simplify this further, the inverse in this expression can be replaced by the pseudo-inverse, which is equivalent in the case of a square, invertible matrix:

$$\hat{\mathbf{b}}_{PCR} = \tilde{\mathbf{U}}\left(\tilde{\mathbf{U}}^T\mathbf{X}^T\mathbf{X}\tilde{\mathbf{U}}\right)^{+}\tilde{\mathbf{U}}^T\mathbf{X}^T\mathbf{y} \tag{B.11}$$

If the matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are all of full rank, then $(\mathbf{ABC})^{+} = \mathbf{C}^{+}\mathbf{B}^{+}\mathbf{A}^{+}$. This leads to:

$$\begin{aligned}
\hat{\mathbf{b}}_{PCR} &= \tilde{\mathbf{U}}\left(\tilde{\mathbf{U}}\right)^{+}\left(\mathbf{X}^T\mathbf{X}\right)^{+}\left(\tilde{\mathbf{U}}^T\right)^{+}\tilde{\mathbf{U}}^T\mathbf{X}^T\mathbf{y} \\
&= \tilde{\mathbf{U}}\tilde{\mathbf{U}}^T\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\tilde{\mathbf{U}}\tilde{\mathbf{U}}^T\mathbf{X}^T\mathbf{y}
\end{aligned} \tag{B.12}$$

For symmetric matrices $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{AB} = \mathbf{BA}$. Therefore:

$$\hat{\mathbf{b}}_{PCR} = \tilde{\mathbf{U}}\tilde{\mathbf{U}}^T \tilde{\mathbf{U}}\tilde{\mathbf{U}}^T \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}$$
$$\hat{\mathbf{b}}_{PCR} = \underline{\mathbf{P}}_{\to\tilde{V}}\,\underline{\mathbf{P}}_{\to\tilde{V}} \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}$$

(B.13)

The matrix $\underline{\mathbf{P}}_{\to\tilde{V}}$ is a projection matrix, which defines the projection onto the subspace $\tilde{V}$, determined by the k principal components in $\tilde{\mathbf{U}}$. Projection matrices possess the property of idempotency, i.e. $\mathbf{AA} = \mathbf{A}$. This leads to the result:

$$\hat{\mathbf{b}}_{PCR} = \underline{\mathbf{P}}_{\to\tilde{V}} \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}$$
$$= \underline{\mathbf{P}}_{\to\tilde{V}}\,\hat{\mathbf{b}}_{OLS}$$

(B.14)

B.3 Bias-Variance Decomposition of the PCR Prediction Error

The MSE of prediction for PCR is given by:

$$\mathrm{MSE}\left[\hat{y}|\right] = \mathrm{E}\left[(y - \hat{y})^2\right]$$
$$= \mathrm{E}\left[\left(y - \mathbf{x}\hat{\mathbf{b}}_{PCR}\right)^2\right]$$
$$= \mathrm{E}\left[\left(y - \mathbf{x}\underline{\mathbf{P}}_{\to\tilde{V}} \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}\right)^2\right]$$

(B.15)

Because $y = \mathbf{x}\check{\mathbf{b}} + e$, and $\mathbf{y} = \mathbf{X}\check{\mathbf{b}} + \mathbf{e}$:

$$
\begin{aligned}
\text{MSE}\left[\hat{y}\right] =& \text{E}\left[\left(\mathbf{x}\check{\mathbf{b}} + e - \mathbf{x}\underrightarrow{\mathbf{P}}_{\tilde{V}}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\left(\mathbf{X}\check{\mathbf{b}} + \mathbf{e}\right)\right)^2\right] \\
=& \text{E}\left[\left(\mathbf{x}\check{\mathbf{b}} + e - \mathbf{x}\underrightarrow{\mathbf{P}}_{\tilde{V}}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{X}\check{\mathbf{b}} - \mathbf{x}\underrightarrow{\mathbf{P}}_{\tilde{V}}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{e}\right)^2\right] \\
=& \text{E}\left[\left(\mathbf{x}\check{\mathbf{b}} + e - \mathbf{x}\underrightarrow{\mathbf{P}}_{\tilde{V}}\check{\mathbf{b}} - \mathbf{x}\underrightarrow{\mathbf{P}}_{\tilde{V}}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{e}\right)^2\right] \\
=& \text{E}\left[\left(e + \mathbf{x}\left(\mathbf{I} - \underrightarrow{\mathbf{P}}_{\tilde{V}}\right)\check{\mathbf{b}} - \mathbf{x}\underrightarrow{\mathbf{P}}_{\tilde{V}}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{e}\right)^2\right]
\end{aligned}
\tag{B.16}
$$

The matrix $\mathbf{I} - \underrightarrow{\mathbf{P}}_{\tilde{V}}$ defines the projection onto the null space of $\tilde{V}$, i.e. the subspace defined by the discarded principal components. Denoting this matrix as $\underrightarrow{\mathbf{P}}_{\tilde{V}_\perp}$:

$$
\text{MSE}\left[\hat{y}\right] = \text{E}\left[\left(e + \mathbf{x}\underrightarrow{\mathbf{P}}_{\tilde{V}_\perp}\check{\mathbf{b}} - \mathbf{x}\underrightarrow{\mathbf{P}}_{\tilde{V}}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{e}\right)^2\right]
\tag{B.17}
$$

When the squared brackets are expanded, the cross-terms cancel due to the presence of the errors e and $\mathbf{e}$, resulting in:

$$
\text{MSE}\left[\hat{y}\right] = \sigma^2 + \text{E}\left[\left(\mathbf{x}\underrightarrow{\mathbf{P}}_{\tilde{V}_\perp}\check{\mathbf{b}}\right)^2\right] + \text{E}\left[\left(\mathbf{x}\underrightarrow{\mathbf{P}}_{\tilde{V}}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{e}\right)^2\right]
\tag{B.18}
$$

Using the LIE, the third term can be written as,

$$\mathrm{E}\left[\left(\mathbf{x}\underline{\mathbf{P}}_{\rightarrow\tilde{V}}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{e}\right)^2\right]$$

$$= \mathrm{E}\left[\mathrm{E}\left[\left(\mathbf{x}\underline{\mathbf{P}}_{\rightarrow\tilde{V}}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{e}\right)^2|\mathbf{x},\mathbf{X}\right]\right]$$

$$= \mathrm{E}\left[\mathrm{E}\left[\mathbf{x}\underline{\mathbf{P}}_{\rightarrow\tilde{V}}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{e}\mathbf{e}^T\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\underline{\mathbf{P}}_{\rightarrow\tilde{V}}\mathbf{x}^T|\mathbf{x},\mathbf{X}\right]\right] \quad\text{(B.19)}$$

$$= \mathrm{E}\left[\mathbf{x}\underline{\mathbf{P}}_{\rightarrow\tilde{V}}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathrm{E}\left[\mathbf{e}\mathbf{e}^T|\mathbf{x},\mathbf{X}\right]\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\underline{\mathbf{P}}_{\rightarrow\tilde{V}}\mathbf{x}^T\right]$$

$$= \mathrm{E}\left[\mathbf{x}\underline{\mathbf{P}}_{\rightarrow\tilde{V}}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\sigma^2\mathbf{I}\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\underline{\mathbf{P}}_{\rightarrow\tilde{V}}\mathbf{x}^T\right]$$

$$= \sigma^2\mathrm{E}\left[\mathbf{x}\underline{\mathbf{P}}_{\rightarrow\tilde{V}}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\underline{\mathbf{P}}_{\rightarrow\tilde{V}}\mathbf{x}^T\right]$$

As before, $(\mathbf{X}^T\mathbf{X})^{-1}$ can be expressed in terms of the covariance matrix $\mathbf{S} = \frac{1}{m}\mathbf{X}^T\mathbf{X} = \mathbf{U}\mathbf{L}\mathbf{U}^T$.

$$\mathrm{E}\left[\left(\mathbf{x}\underline{\mathbf{P}}_{\rightarrow\tilde{V}}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{e}\right)^2\right]$$

$$= \frac{\sigma^2}{m}\mathrm{E}\left[\mathbf{x}\underline{\mathbf{P}}_{\rightarrow\tilde{V}}\left(\mathbf{U}\mathbf{L}\mathbf{U}^T\right)^{-1}\underline{\mathbf{P}}_{\rightarrow\tilde{V}}\mathbf{x}^T\right] \quad\text{(B.20)}$$

$$= \frac{\sigma^2}{m}\mathrm{E}\left[\mathbf{x}\underline{\mathbf{P}}_{\rightarrow\tilde{V}}\mathbf{U}\mathbf{L}^{-1}\mathbf{U}^T\underline{\mathbf{P}}_{\rightarrow\tilde{V}}\mathbf{x}^T\right]$$

Here, $\underline{\mathbf{P}}_{\rightarrow\tilde{V}}\mathbf{U}$ is the projection of $\mathbf{U}$ onto the subspace of $\tilde{V}$, and so is equal to $\tilde{\mathbf{U}}$. Similarly, $\mathbf{U}^T\underline{\mathbf{P}}_{\rightarrow\tilde{V}} = \tilde{\mathbf{U}}^T$. Therefore:

$$\mathrm{E}\left[\left(\mathbf{x}\underline{\mathbf{P}}_{\rightarrow\tilde{V}}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{e}\right)^2\right]$$

$$= \frac{\sigma^2}{m}\mathrm{E}\left[\mathbf{x}\tilde{\mathbf{U}}\mathbf{L}^{-1}\tilde{\mathbf{U}}^T\mathbf{x}^T\right] \quad\text{(B.21)}$$

$$= \frac{\sigma^2}{m}\mathrm{E}\left[\mathbf{x}(\sum_{i=1}^{k}l_i\mathbf{u}_i\mathbf{u}_i^T)^{-1}\mathbf{x}^T\right]$$

Following the same procedure used for OLS (section B.1) leads to:

$$\mathrm{E}\left[\left(\mathbf{x}\underset{\to \tilde{V}}{\mathbf{P}}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{e}\right)^2\right] = \frac{k}{m}\sigma^2 \tag{B.22}$$

Leading to the full bias-variance expression:

$$\mathrm{MSE}\left[\hat{y}\right] = \sigma^2 + \frac{k}{m}\sigma^2 + \mathrm{E}\left[\left(\mathbf{x}\underset{\to \tilde{V}_\perp}{\mathbf{P}}\check{\mathbf{b}}\right)^2\right]$$

$$= \text{unexplained variance} + \text{model variance} + \text{model bias} \tag{B.23}$$

B.4   Overview of the SIMPLS algorithm

(Following from the discussion in section 4.3.1)

To ensure that subsequent score vectors $\mathbf{t}_i$ are orthogonal, the covariance matrix $\mathbf{Y}^T\mathbf{X}$ is "deflated" after every iteration, to ensure that the contribution to $\mathbf{X}$ made by all the previous scores is not present in the covariance matrix for future calculations.

After the ith iteration of the algorithm, the contribution that the scores $\mathbf{T} = [\mathbf{t}_1 \ \mathbf{t}_2 \ \dots \ \mathbf{t}_i]$ make towards $\mathbf{X}$ is given by $\hat{\mathbf{X}} = \mathbf{T}\hat{\mathbf{P}}$, where:

$$\hat{\mathbf{P}} = (\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T\mathbf{X} \tag{B.24}$$

Because the columns of $\mathbf{T}$ are orthogonal, the matrix $\mathbf{T}^T\mathbf{T}$ is diagonal, with the element $(i,i) = \|\mathbf{t}_i\|^2$, i.e. the sum of squared elements of $\mathbf{t}_i$. This orthogonality allows each row of $\hat{\mathbf{P}}$ to be calculated independently at each iteration. Denoting element (i,i) of $\mathbf{T}^T\mathbf{T}$ as c_i, the ith row of $\hat{\mathbf{P}}$ is given by:

$$\hat{\mathbf{p}}_i = \frac{1}{c_i}\mathbf{t}_i^T\mathbf{X} \tag{B.25}$$

and so:

$$c_i \hat{\mathbf{p}}_i = \mathbf{t}_i^T \mathbf{X} \tag{B.26}$$

Orthogonality of $\mathbf{t}$ requires that, for every $i > j$:

$$\mathbf{t}_j^T \mathbf{t}_i = 0 \tag{B.27}$$

As $\mathbf{t}_i = \mathbf{X}\mathbf{r}_i$:

$$\mathbf{t}_j^T \mathbf{X}\mathbf{r}_i = 0 \tag{B.28}$$

and from (B.26):

$$c_j \hat{\mathbf{p}}_j \mathbf{r}_i = 0 \tag{B.29}$$

which shows that $\mathbf{r}_i$ must be orthogonal to all preceding values of $\hat{\mathbf{p}}_j$. This is equivalent to requiring that $\mathbf{r}_i$ is be restricted to the null space of the rows of $\hat{\mathbf{P}}$, denoted as $V_{\hat{\mathbf{P}}_\perp}$. Defining the matrix $\underset{\rightarrow}{\mathbf{P}}_{V_{\hat{\mathbf{P}}_\perp}}$ as the projection onto $V_{\hat{\mathbf{P}}_\perp}$, it is required that:

$$\mathbf{r}_i = \underset{\rightarrow}{\mathbf{P}}_{V_{\hat{\mathbf{P}}_\perp}} \mathbf{r}_i \tag{B.30}$$

which is only true if $\mathbf{r}_i$ is contained entirely within $V_{\hat{\mathbf{P}}_\perp}$. Therefore, for $i > 2$, $\mathbf{r}_i$ is chosen to maximise:

$$
\begin{aligned}
\mathbf{r}_i &= \operatorname{argmax}\left(\left|\mathbf{Y}^T \mathbf{t}_i\right|\right) \\
&= \operatorname{argmax}\left(\left|\mathbf{Y}^T \mathbf{X}\mathbf{r}_i\right|\right) \\
&= \operatorname{argmax}\left(\left|\mathbf{Y}^T \mathbf{X}\underset{\rightarrow}{\mathbf{P}}_{V_{\hat{\mathbf{P}}_\perp}} \mathbf{r}_i\right|\right)
\end{aligned}
\tag{B.31}
$$

which is solved by taking $\mathbf{r}_i$ to be the largest right singular vector of the matrix $\mathbf{Y}^T \mathbf{X}\underset{\rightarrow}{\mathbf{P}}_{V_{\hat{\mathbf{P}}_\perp}}$. In practice, the projection matrix can be calculated as $\underset{\rightarrow}{\mathbf{P}}_{V_{\hat{\mathbf{P}}_\perp}} = \mathbf{I} - \hat{\mathbf{P}}^+\hat{\mathbf{P}}$, where $\hat{\mathbf{P}}^+$ is the pseudo inverse. An overview of the entire SIMPLS process is illustrated below. There are still several efficiency

improvements that can be made to this algorithm (which are provided in full in (de Jong, 1993)), and it is provided here only as an aid to understanding the technique, rather than a basis for a practical implementation.

$$
\begin{aligned}
&\textbf{FOR } i = 1, 2, \ldots, k \\
&\qquad \textbf{IF } i = 1 \\
&\qquad\qquad \text{Calculate SVD of } \mathbf{Y}^T\mathbf{X} \\
&\qquad \textbf{ELSE} \\
&\qquad\qquad \text{Calculate SVD of } \mathbf{Y}^T\mathbf{X}(\mathbf{I} - \hat{\mathbf{P}}^+\hat{\mathbf{P}}) \\
&\qquad \textbf{END} \\
&\qquad \mathbf{r}_i = \text{first right singular vector} \\
&\qquad \mathbf{t}_i = \mathbf{X}\mathbf{r}_i \\
&\qquad \mathbf{p}_i = \frac{1}{\|\mathbf{t}_i\|}{}^2 \mathbf{t}_i\mathbf{X} \\
&\qquad \text{Store } \mathbf{r}_i,\ \mathbf{t}_i \text{ and } \hat{\mathbf{p}}_i \text{ into } \mathbf{R},\ \mathbf{T} \text{ and } \hat{\mathbf{P}} \\
&\textbf{END} \\
&\hat{\mathbf{Q}} = (\mathbf{T}^T\mathbf{T})\mathbf{T}^T\mathbf{Y} \\
&\hat{\mathbf{B}}_{PLS} = \mathbf{R}\hat{\mathbf{Q}}
\end{aligned}
$$

B.5 Bias-Variance Decomposition of the LDR Prediction Accuracy

The LDR prediction $\hat{\mathbf{y}}_{LDR}$ is a p dimensional vector, so the MSE is defined as:

$$
\begin{aligned}
\text{MSE}[\hat{\mathbf{y}}_{LDR}] &= \frac{1}{p}\text{E}\left[(\mathbf{y} - \hat{\mathbf{y}}_{LDR})(\mathbf{y} - \hat{\mathbf{y}}_{LDR})^T\right] \\
&= \frac{1}{p}\text{E}\left[\|(\mathbf{y} - \hat{\mathbf{y}}_{LDR})\|^2\right] \\
&= \frac{1}{p}\text{E}\left[\left\|(\mathbf{x}\check{\mathbf{B}} + \mathbf{e} - \mathbf{x}\hat{\mathbf{B}}_{OLS}\underline{\mathbf{P}}_{\overrightarrow{\tilde{V}}}\right\|^2\right]
\end{aligned}
\tag{B.32}
$$

287

Following the standard procedure leads to:

$$
\begin{aligned}
\text{MSE}[\hat{\mathbf{y}}_{LDR}] =& \frac{1}{p}\text{E}\left[\left\|(\mathbf{e} - \mathbf{x}\check{\mathbf{B}}\left(\mathbf{I} - \underset{\to \tilde{V}}{\mathbf{P}}\right) - \mathbf{x}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{E}\underset{\to \tilde{V}}{\mathbf{P}}\right\|^2\right] \\
=& \frac{1}{p}\left(\sigma^2 + \text{E}\left[\left\|\mathbf{x}\check{\mathbf{B}}\left(\mathbf{I} - \underset{\to \tilde{V}}{\mathbf{P}}\right)\right\|^2\right]\right. \\
& + \left.\text{E}\left[\left\|\mathbf{x}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{E}\underset{\to \tilde{V}}{\mathbf{P}}\right\|^2\right]\right)
\end{aligned}
\tag{B.33}
$$

where $\sigma^2 = \text{E}\left[\|\mathbf{e}\|^2\right] = \sum_{i=1}^{p}\text{E}\left[e_i^2\right]$, which is the sum of the unexplained variance in $\mathbf{y}$. Expanding the third term:

$$
\begin{aligned}
&\text{E}\left[\left\|\mathbf{x}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{E}\underset{\to \tilde{V}}{\mathbf{P}}\right\|^2\right] \\
&= \text{E}\left[\mathbf{x}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{E}\underset{\to \tilde{V}}{\mathbf{P}}\mathbf{E}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}^T\right]
\end{aligned}
\tag{B.34}
$$

Using the LIE, this can be written as:

$$
\begin{aligned}
&\text{E}\left[\left\|\mathbf{x}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{E}\underset{\to \tilde{V}}{\mathbf{P}}\right\|^2\right] \\
&= \text{E}\left[\text{E}\left[\mathbf{x}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{E}\underset{\to \tilde{V}}{\mathbf{P}}\mathbf{E}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}^T|\mathbf{X},\mathbf{x}\right]\right] \\
&= \text{E}\left[\mathbf{x}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\text{E}\left[\mathbf{E}\underset{\to \tilde{V}}{\mathbf{P}}\mathbf{E}^T|\mathbf{X},\mathbf{x}\right]\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}^T\right]
\end{aligned}
\tag{B.35}
$$

The expression $\text{E}\left[\text{E}\left[\mathbf{E}\underset{\to \tilde{V}}{\mathbf{P}}\mathbf{E}^T|\mathbf{X},\mathbf{x}\right]\right]$ can be expressed as $\tilde{\sigma}_{|\mathbf{X}}^2\mathbf{I}$, where $\tilde{\sigma}_{|\mathbf{X}}^2 = \text{E}\left[\left\|\mathbf{e}\underset{\to \tilde{V}}{\mathbf{P}}\right\||\mathbf{X}\right]$. This describes the total unexplained variance of $\mathbf{y}$ contained in the subspace $\tilde{V}$, conditional on $\mathbf{X}$. The need to condition this expression on $\mathbf{X}$ is due to that fact that in both CSE and RRR, the subspace $\tilde{V}$, which

is defined by the columns of $\tilde{\mathbf{U}}$, is dependent on $\mathbf{X}$. This leads to:

$$
\begin{aligned}
\mathrm{E}&\left[\left\|\mathbf{x}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{E}\underaccent{\rightarrow}{\mathbf{P}}_{\tilde{V}}\right\|^2\right]\\
&= \mathrm{E}\left[\tilde{\sigma}^2_{|\mathbf{X}}\mathbf{x}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}^T\right] \\
&= \tilde{\sigma}^2\mathrm{E}\left[\tilde{\sigma}^2_{|\mathbf{X}}\left\|\mathbf{x}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}^T\right\|^2\right]
\end{aligned}
\tag{B.36}
$$

As with the OLS decomposition, this can be written as:

$$
\begin{aligned}
\mathrm{E}&\left[\left\|\mathbf{x}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{E}\underaccent{\rightarrow}{\mathbf{P}}_{\tilde{V}}\right\|^2\right]\\
&= \mathbf{E}[\tilde{\sigma}^2_{|\mathbf{X}}]\frac{n}{m} \\
&= \tilde{\sigma}^2\frac{n}{m}
\end{aligned}
\tag{B.37}
$$

Where $\tilde{\sigma}^2$ is now the total *unconditional* unexplained variance in the subspace $\tilde{V}$. Reconstructing the full MSE gives:

$$
\mathrm{MSE}[\hat{\mathbf{y}}_{LDR}] = \frac{1}{p}\left(\sigma^2 + +\tilde{\sigma}^2\frac{n}{m} + \mathrm{E}\left[\left\|\mathbf{x}\check{\mathbf{B}}\left(\mathbf{I} - \underaccent{\rightarrow}{\mathbf{P}}_{\tilde{V}}\right)\right\|^2\right]\right)
\tag{B.38}
$$

Although this expression has similarities to those derived for the previous techniques, it is not a true bias-variance decomposition. To be considered such, the decomposition must contain:

- a variance term that reduces to zero when the number of modes k in the model is zero, i.e. $\tilde{\mathbf{U}} = 0$.

- a bias term that reduces to zero when the LDR model is identical to the OLS estimator.

The term $\tilde{\sigma}^2\frac{n}{m}$ meets the requirement of a variance term, as it reduces to zero when the subspace $\tilde{V}$, defined by $\tilde{\mathbf{U}}$ is zero. However, the third term,

$\mathrm{E}\left[\left\|\mathbf{x}\check{\mathbf{B}}\left(\mathbf{I}-\underline{\mathbf{P}}_{\rightarrow\tilde{V}}\right)\right\|^{2}\right]$ does not necessarily equal zero when the LDR model is equivalent to the OLS estimator. To see this, the cases when the two techniques are identical. From section 4.3.3, the LDR model is:

$$\hat{\mathbf{y}}_{LDR} = \mathbf{x}\hat{\mathbf{B}}_{OLS}\underline{\mathbf{P}}_{\rightarrow\tilde{V}} \tag{B.39}$$

The most obvious situation where OLS and LDR are equivalent is when $\underline{\mathbf{P}}_{\rightarrow\tilde{V}} = \mathbf{I}$, which arises when $\tilde{\mathbf{U}}$ consists of a complete set of p orthonormal modes. Here, the third term of (B.38) cancels to zero, as hoped. However, in cases where the rank r of $\hat{\mathbf{b}}_{OLS}$ is less than p, the OLS predictor resides in a V_{OLS} which is a smaller than $\mathbb{R}^p$. In these situations, as long as $\underline{\mathbf{P}}_{\rightarrow\tilde{V}}$ projects onto a subspace of size $k < p$, that still contains V_{OLS}, LDR is still identical to OLS even though $\underline{\mathbf{P}}_{\rightarrow\tilde{V}} \neq \mathbf{I}$. Now, the matrix $\mathbf{I} - \underline{\mathbf{P}}_{\rightarrow\tilde{V}}$ will be positive semidefinite, and the third term of (B.38) is unlikely to be zero.

To define a proper bias term, The identity matrix in the third term can be expressed as $\mathbf{I} = \underline{\mathbf{P}}_{\rightarrow V_{OLS}} + \underline{\mathbf{P}}_{\rightarrow V_{OLS_\perp}}$:

$$
\begin{aligned}
\mathrm{E}&\left[\left\|\mathbf{x}\check{\mathbf{B}}\left(\mathbf{I}-\underline{\mathbf{P}}_{\rightarrow\tilde{V}}\right)\right\|^{2}\right] \\
&= \mathrm{E}\left[\left\|\mathbf{x}\check{\mathbf{B}}\left(\underline{\mathbf{P}}_{\rightarrow V_{OLS}} + \underline{\mathbf{P}}_{\rightarrow V_{OLS_\perp}} - \underline{\mathbf{P}}_{\rightarrow\tilde{V}}\right)\right\|^{2}\right] \\
&= \mathrm{E}\left[\left\|\mathbf{x}\check{\mathbf{B}}\underline{\mathbf{P}}_{\rightarrow V_{OLS_\perp}}\right\|^{2}\right] + \mathrm{E}\left[\left\|\mathbf{x}\check{\mathbf{B}}\left(\underline{\mathbf{P}}_{\rightarrow V_{OLS}} - \underline{\mathbf{P}}_{\rightarrow\tilde{V}}\right)\right\|^{2}\right] \\
&\quad - 2\mathrm{E}\left[\mathbf{x}\check{\mathbf{B}}\underline{\mathbf{P}}_{\rightarrow V_{OLS_\perp}}\underline{\mathbf{P}}_{\rightarrow\tilde{V}}\check{\mathbf{B}}^{T}\mathbf{x}^{T}\right]
\end{aligned} \tag{B.40}
$$

Assuming that the subspace $\tilde{V}$ is contained entirely within V_{OLS}, then $\underline{\mathbf{P}}_{\rightarrow V_{OLS_\perp}}\underline{\mathbf{P}}_{\rightarrow\tilde{V}}$ is zero, and the third term disappears. Also, the result from the subtraction of the projection matrices, $\underline{\mathbf{P}}_{\rightarrow V_{OLS}} - \underline{\mathbf{P}}_{\rightarrow\tilde{V}}$ will also be a projection matrix, that describes the projection onto the subspace of V_{OLS} that is not present in $\tilde{V}$:

Therefore:

$$\text{MSE}[\hat{\mathbf{y}}_{LDR}] = \frac{1}{p}\left(\sigma^2 + +\tilde{\sigma}^2\frac{n}{m} + \text{E}\left[\left\|\mathbf{x}\check{\mathbf{B}}\underset{\rightarrow}{\mathbf{P}}_{V_{OLS_\perp}}\right\|^2\right]\right.$$

$$\left. + \text{E}\left[\left\|\mathbf{x}\check{\mathbf{B}}\left(\underset{\rightarrow}{\mathbf{P}}_{V_{OLS}} - \underset{\rightarrow}{\mathbf{P}}_{\tilde{V}}\right)\right\|^2\right]\right) \tag{B.41}$$

This can be thought of as a true bias-variance decomposition, consisting of the unexplained variance term, a model variance terms, and model bias term, which is zero for the OLS case, i.e. $\tilde{V} = V_{OLS}$, and a further term, $\text{E}\left[\left\|\mathbf{x}\check{\mathbf{B}}\underset{\rightarrow}{\mathbf{P}}_{V_{OLS_\perp}}\right\|^2\right]$ which is present when size r of the subspace of V_{OLS} is smaller than p.

It is possible to express this final term in a more readily interpretable form. The MSE of the OLS predictor for multivariate $\mathbf{y}$ is simply:

$$\text{MSE}[\hat{\mathbf{y}}_{OLS}] = \frac{1}{p}\left(\sigma^2 + \sigma^2\frac{n}{m}\right) \tag{B.42}$$

The total unexplained variance can be split into $\sigma^2 = \hat{\sigma}^2 + \hat{\sigma}_\perp^2$, which is unexplained variance in V_{OLS}, and $V_{OLS_\perp}$ respectively. This leads to:

$$=\frac{1}{p}\left(\sigma^2 + (\hat{\sigma}^2 + \hat{\sigma}_\perp^2)\frac{n}{m}\right)$$

$$=\frac{1}{p}\left(\sigma^2 + \hat{\sigma}^2\frac{n}{m} + \hat{\sigma}_\perp^2\frac{n}{m}\right) \tag{B.43}$$

As shown previously, LDR regression is equivalent to OLS regression when $\tilde{V} = V_{OLS}$, i.e:

$$\text{MSE}[\hat{\mathbf{y}}_{LDR}|_{\tilde{V}=V_{OLS}}] = \text{MSE}[\hat{\mathbf{y}}_{OLS}] \tag{B.44}$$

Therefore:

$$\frac{1}{p}\left(\sigma^2 + \hat{\sigma}^2\frac{n}{m} + \mathrm{E}\left[\left\|\mathbf{x}\check{\mathbf{B}}\underset{\rightarrow}{\mathbf{P}}_{V_{OLS_\perp}}\right\|^2\right]\right) = \frac{1}{p}\left(\sigma^2 + \hat{\sigma}^2\frac{n}{m} + \hat{\sigma}_\perp^2\frac{n}{m}\right) \qquad \text{(B.45)}$$

and:

$$\mathrm{E}\left[\left\|\mathbf{x}\check{\mathbf{B}}\underset{\rightarrow}{\mathbf{P}}_{V_{OLS_\perp}}\right\|^2\right] = \hat{\sigma}_\perp^2\frac{n}{m} \qquad \text{(B.46)}$$

This term is a form of model error which is present in the MSE whenever $r < p$. As it is present in the model, regardless of the choice of $\tilde{\mathbf{U}}$, is shall be referred to as the *unremovable model error*.

The full MSE of $\hat{\mathbf{y}}_{LDR}$ can therefore be written as:

$$
\begin{aligned}
\mathrm{MSE}[\hat{\mathbf{y}}_{LDR}] =& \frac{1}{p}\left(\sigma^2 + \hat{\sigma}_\perp^2\frac{n}{m} + \tilde{\sigma}^2\frac{n}{m} + \mathrm{E}\left[\left\|\mathbf{x}\check{\mathbf{B}}\left(\underset{\rightarrow}{\mathbf{P}}_{\hat{U}} - \underset{\rightarrow}{\mathbf{P}}_{\tilde{V}}\right)\right\|^2\right]\right) \\
=& \frac{\sigma^2}{p} + \hat{\sigma}_\perp^2\frac{n}{pm} \\
& + \tilde{\sigma}^2\frac{n}{pm} + \frac{1}{p}\mathrm{E}\left[\left\|\mathbf{x}\check{\mathbf{B}}\left(\underset{\rightarrow}{\mathbf{P}}_{V_{OLS}} - \underset{\rightarrow}{\mathbf{P}}_{\tilde{V}}\right)\right\|^2\right] \\
=& \text{unexplained variance} + \text{unremovable model error} \\
& + \text{removable model variance} + \text{model bias}
\end{aligned}
\qquad \text{(B.47)}
$$

Appendix C

A Novel Approach to Detect Light Sheet Misalignment

C.1 Proposed Technique

The basis of the proposed technique is a purpose-built measurement frame, shown in figure C.1. The frame consists of a square base plate, with three angled sides which overhangs the base plate. It is constructed from black anodised aluminium with dimensions of $290 \times 220 \times 60$mm. The overhanging sides are angled at approximately $30°$, relative to the base plate. It should be noted that the proposed technique only requires a measurement frame of this general form, rather than a frame identical to the one shown here. The dimensions of this particular frame have been chosen for the specific purpose of calibrating the MP-VSE vortex ring experiment described in chapter 7. For other applications, alternative configurations may be necessary to maximise measurement accuracy. A discussion of the aspects of the design that contribute towards accuracy are provided in appendix C.4, along with further

details of construction of the frame described here.

To perform a measurement of the light sheet position, the frame is positioned in the interrogation region of the flow, so that the three angled sides intersect the path of the light sheet. On contact with each of the sides, the light sheet projects a line (which shall be referred to as a *projected light sheet line*, or PLL. Based on the known geometry of the frame, it is possible to use the position of the PLLs to obtain the coordinates of several points on the plane. As a plane can be explicitly defined if at least three non-collinear points are known, it is therefore possible to calculate the exact position of the light sheet relative to the measurement frame.

As can be seen in figure C.1, the edges of the frame are marked by a series of white tracks. The inside edges of these tracks intersect at the eight points $A, B, \ldots, H$ (as shown in figure C.2), which are at known coordinates, and provide a reference for the calibration process. In turn, the consecutive pairs of reference points define reference *lines*, i.e. AB, CD, EF and GH, which will be intersected by at least one of the 3 PLLs (denoted as L_1, L_2 and L_3) at a particular point. In total there are six PLL/reference line intersections, $AB.L_1$, $CD.L_1$, $CD.L_2$, $EF.L_2$, $EF.L_3$ and $GH.L_3$. Theoretically, the intersections $CD.L_1$ and $CD.L'_2$ are identical, as are $EF.L_2$ and $EF.L_3$. However, in practice the calculated intersections are subject to a certain amount of error, which can be reduced to some extent by the inclusion of such redundancy. The aim of the calibration procedure is to calculate the world coordinates of these six points, which are sufficient to define its position of the light sheet plane.

The world coordinates are calculated from an image of the calibration frame. All the reference points, $A, B, \ldots, H$, and PLLs, L_1, L_2, L_3, will have an equivalent in the image plane (denoted by a prime, i.e. $A', B', \ldots, H'$ and

294

Figure C.1: The measurement frame.

L'_1, L'_2, L'_3), although their positions will only be known in pixels and in 2 dimensions. In order to calculate the light sheet position, the relationship between the image coordinates and the corresponding world coordinates must therefore be established. This process is achieved using projective transformations, or homographies, which provide a one-to-one mapping of points between two planes. All the points of interest on the measurement frame are located on one of three planes (i.e. the left, top and right side), which means that three homographies can be used to map all the necessary coordinates from the image plane to each of the world planes.

If the frame is at a known position in flow coordinates, the light sheet position will then be known absolutely. The procedure is also capable of providing the relative position of two or more parallel light sheets, assuming that the spacing of the sheets is sufficiently close that all lights sheets intersect
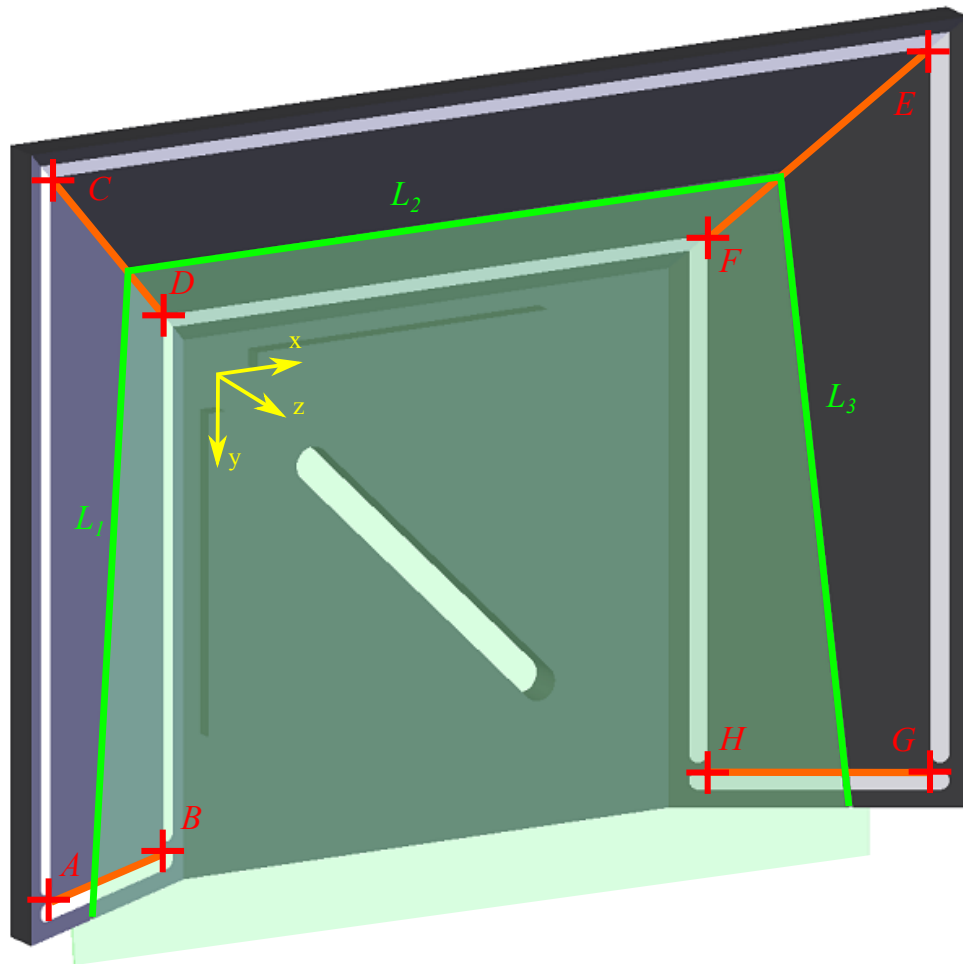
Figure C.2: Position of reference points on the light sheet. The yellow arrows define the origin and orientation of the measurement coordinate system.

the measurement frame. The position of each light sheet (relative to the frame) is obtained separately, which subsequently allows the relative alignment of the planes to be calculated, without the need to know the absolute location of the frame in the flow.

As the measurement process is based on data that is extracted in a digital image, the resolution of the camera sensor stands to limit the accuracy of the feature detection. In order to extend the accuracy of the process to the sub-pixel range, the proposed technique has been designed so that all the necessary features on the target are defined by the intersection of lines. This is motivated by the fact that a discretely sampled image of an edge contains a large amount of information that can be extracted to obtain its position and orientation to sub-pixel accuracy. This concept is demonstrated in the example in figure C.3. The process begins with a synthetically-generated perfect "knife" edge. The line equation that defines this edge has also been used to produce the discrete image in C.3(b), which includes a degree of blurring to simulate lens effects. An edge detection algorithm is used to extract the "edge" pixels, and a least-squares line fit is applied. The resulting line demonstrates excellent agreement with the true position of the edge.

For this approach to obtain the location of the edge to sub-pixel accuracy, it is necessary for the edge to be off-perpendicular to the pixel grid of the sensor. Otherwise, the result of the edge detection provides insufficient information about the edge position. In figure C.4 the same procedure as before is applied to a horizontal edge. Figure C.4(d) clearly shows that the fit suffers in this situation.

In the following sections, the full process of calculating the light sheet position from an image is described in detail. The discussion is split into five sections, each covering a distinct aspect of the procedure.
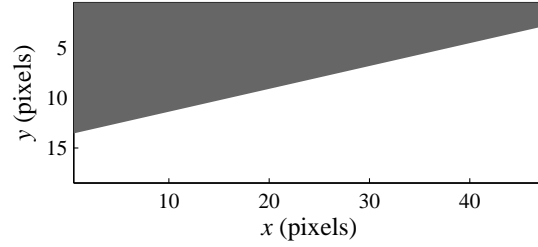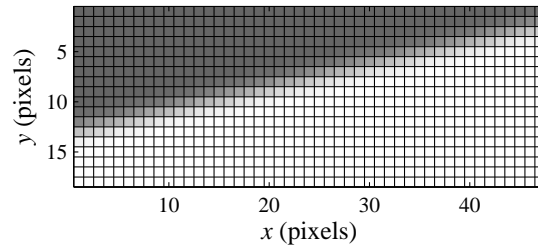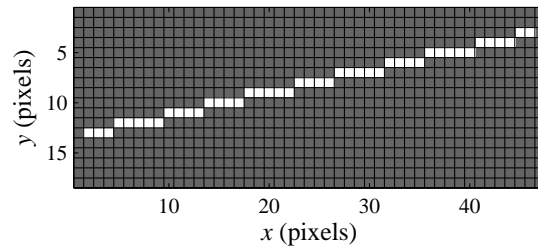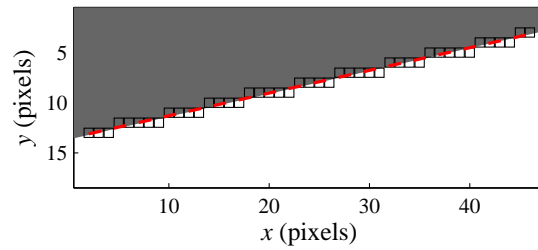
297

(a) Synthetic knife edge
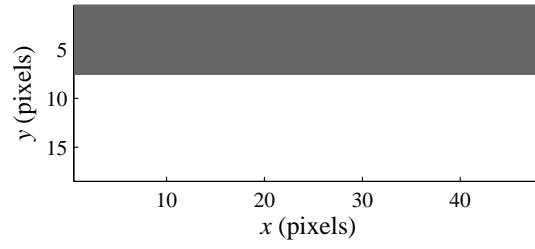

(b) Discretely sampled edge image
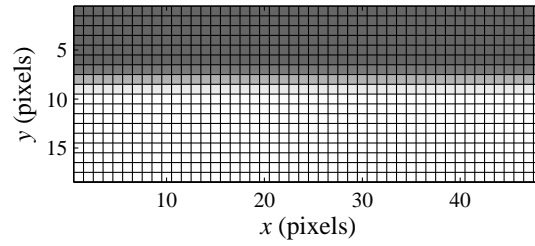

(c) Output of edge detection

(d) Result of the line fitting process. The red line shows the least-squares fit through the edge points (denoted by the black squares)
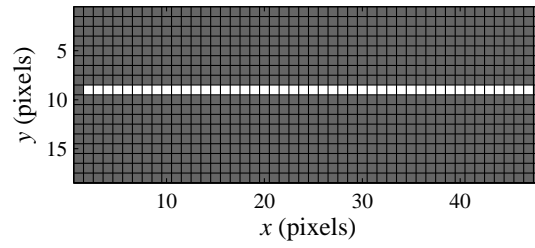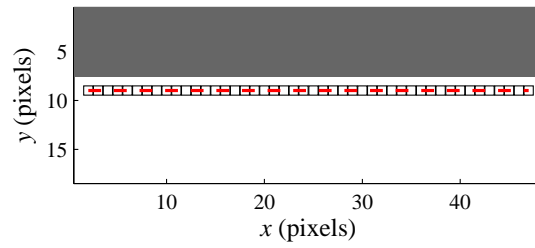
Figure C.3: Stages of the line detection process

(a) Synthetic knife edge

(b) Discretely sampled edge image

(c) Output of edge detection

(d) Result of the line fitting process for a horizontal edge. The red line shows the least-squares fit through the edge points (denoted by the black squares). Note the disparity between the true edge and fitted line.

Figure C.4: Stages of the line detection process

C.1.1   Stage 1: Image Acquisition

The measurement process begins with the acquisition of an image of the measurement frame, with the PLLs visible on the angled sides. If possible, the camera should be placed approximately normal to the target in order to minimise the amount of perspective distortion in the image. Error in the measurement can be minimised further by using a high resolution sensor and a high-quality, distortion-free lens to produce the image. Also, care must be taken to ensure that the image is correctly exposed. This may pose a problem when using high-powered lasers, as the difference in brightness of the PLLs and target tracks may exceed the dynamic range of the camera sensor. In this case, it will be necessary to acquire two images; one of the target by itself and one of the PLLs, possibly with the use of neutral density filters to attenuate the light intensity. As long as the camera and target are not moved between images this does not pose a problem for the technique, as the detection of the target tracks and PLLs is carried out independently. For simplicity, the work in this chapter will consider the case where only a single image is required. The process will be illustrated using the example image shown in figure C.5. Note that the camera has been rotated approximately 30 degrees. This is a precautionary measure to avoid the possibility that edges in the image lie perpendicular to the pixel grid.

C.1.2   Stage 2: Identification of Reference Points

Once a suitable image of the calibration frame has been acquired, the first goal is to recover the coordinates of the reference points in image space. As described earlier, the reference points are located at the intersections of the edges of the white track lines that run around the sides of the frame.
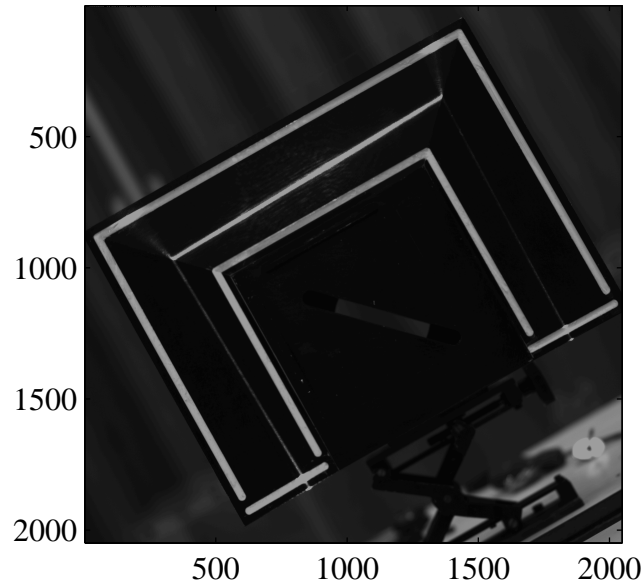
Figure C.5: Original image of the calibration plate

The process of recovering the reference points involves the following stages:

1. Extracting the regions of the image that contain the track line edges.

2. Detecting the edge of the tracks.

3. Fitting lines to the edges.

4. Finding the intersections of the lines that define the reference points.

For the line fitting to work accurately, it is important that the line is fitted only to points from the edge in question, as any other detected edge features in the image may reduce the accuracy of the line fitting stage. Therefore the first step towards the recovery of the reference points is to accurately define the regions of the image that contain the track edges, so that the edge
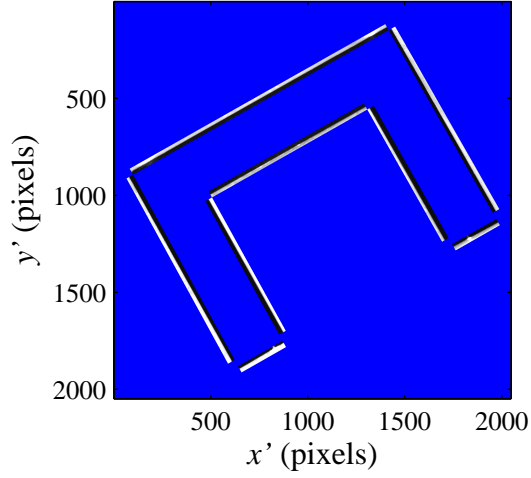
301

Figure C.6: Calibration plate image with the edge region mask applied. The mask identifies the eight separate track edges of the input image (C.5).

detection process is applied only to the edge of interest. In the current implementation, the user is required to manually select the four extreme reference points in the image (A',C',E' and G'), which are used to map a mask on to the image. The result of this process is shown in figure C.6. Each of the eight edges can now be processed individually. Edge detection is carried out using a simple differential based approach. An edge is characterised by a sharp change in image intensity, which will manifest itself as a region of high magnitude in the derivative of the edge image. The position of the edge can therefore be found by a simple thresholding of the derivative magnitude. In the current implementation, the *Sobel* convolution kernels, $\mathbf{G}_x$ and $\mathbf{G}_y$, are used to provide a simple approximation of the image derivative:

$$\mathbf{G}_x = \begin{bmatrix} +1 & 0 & -1 \\ +2 & 0 & -2 \\ +1 & 0 & -1 \end{bmatrix}, \quad \mathbf{G}_y = \mathbf{G}_x^{\mathbf{T}} = \begin{bmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \tag{C.1}$$

302

The subsequent convolutions of $\mathbf{G}_x$ and $\mathbf{G}_y$ with the image approximate the derivatives along the horizontal and vertical directions respectively. Conventionally, the full approximation of the image derivative magnitude is given by the norm of the two resulting derivatives. However, for this particular application, only a single kernel is used. This is necessary because the detection of the PLLs, described later in section C.1.3, is based on the sign of the gradient, which is not available if the norm of the derivatives is taken. The same approach is used here for convenience.

The orientation of each edge is known in advance to be either approximately horizontal or vertical, which allows the derivative to be calculated using the appropriate kernel. After the convolution, a threshold is applied to the gradient image to identify the pixels that lie along the edge. In certain cases, there will be spurious pixels in the thresholded image, as shown in the example given in figure C.7 (a); here, part of the projected light sheet is present in the image of the track edge, which will show up in the thresholded edge image. If these points are included in the subsequent line fitting, the resulting line will deviate from the true edge position. To avoid this problem, a robust line fitting algorithm is used to identify and remove any points that are not part of the true edge.

The process begins by extracting the coordinates of all the edge points from the thresholded image, then carrying out a least-squares line fit to these points. The distance between the line and each of the points is calculated, and any points that fall outside of a specified distance threshold are discarded. This process is repeated for a predetermined number of iterations, each time reducing the distance threshold. Continuing the example from figure C.7 (a), the resulting line fitting procedure is illustrated in C.7 (b). The red pixels indicate points from the output of the edge detection process that are removed
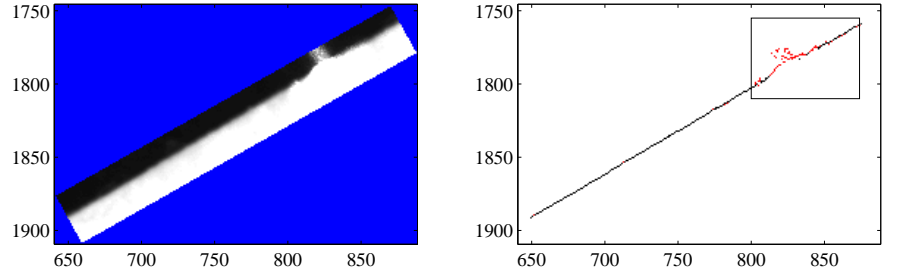
(a) Original Image of the edge

(b) Output of edge detection routine. The red pixels are subsequently removed by the robust line fitting routine.
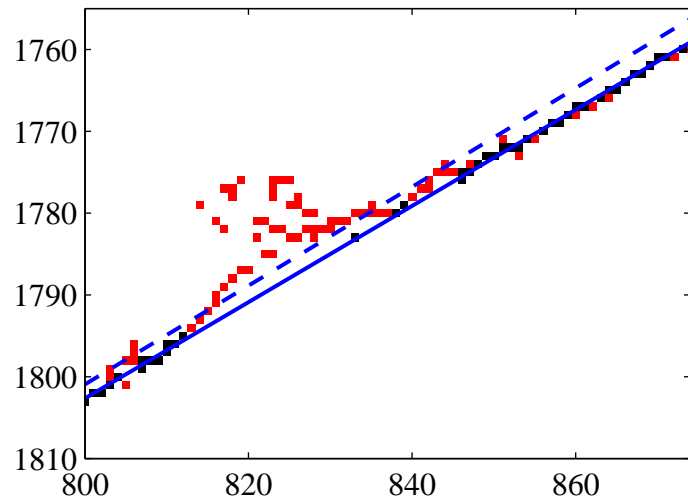
(c) Closeup of detected edge points. The dashed line corresponds to the position of the line when the spurious pixels are included in the line fitting; the solid line shows the fit when the spurious pixels are removed.

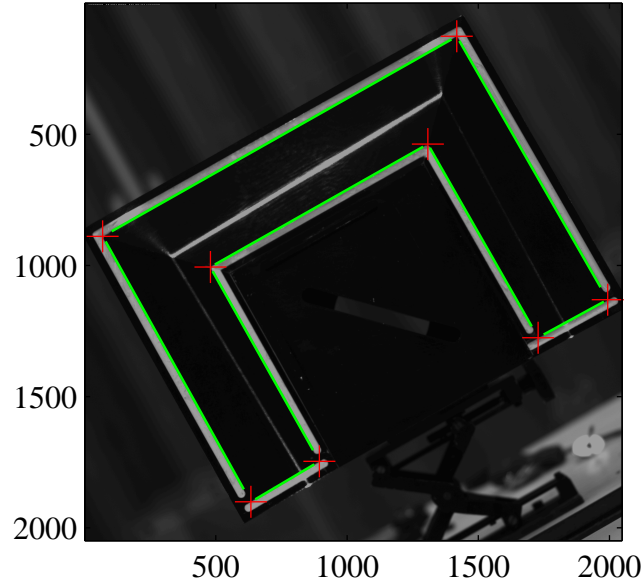Figure C.7: Stages involved in the line fitting process

304

Figure C.8: Detected track edge locations (green lines), and resulting reference points (red crosses)

in subsequent passes of the line fitting algorithm. The closeup in figure C.7 (c) clearly shows the improvement in accuracy due to the robust approach; the dashed line indicates the position of the original line fit, while the solid line shows its final position.

Once lines have been fitted to all the track edges, the intersections of the relevant lines are calculated, yielding the coordinates of the image reference points, $A', B', \ldots, G'$. The calculated positions of the edges and reference points on the example image are given in figure C.8.

C.1.3 Stage 3: Light Sheet Line Detection

Detection of the light sheet lines uses a similar approach to the one described in the previous section. The process begins by selecting the region of the

image that contains the PLL in question. This is straightforward, as each of the PLLs falls within a region defined by four reference points whose positions in the image are already known. The process of edge detection is slightly more sophisticated than before as the PLLs are of a finite thickness, and so have two edges which must be detected separately[1]. The relevant Sobel differential kernel is again applied to the image. This time, the magnitude of the differentiated image is not calculated, as the sign of the derivative provides a means of distinguishing between the two edges. One edge of the PLL will be characterised by a sharp change from low to high pixel intensity, producing a large positive gradient. For the second edge, the change from high to low intensity will produce a large negative gradient.

From the image derivative, two separate edge images are produced by applying both a positive and negative threshold in turn. The robust line fitting method (described in section C.1.2) is then applied to each of the edge images. The intersections of the resulting lines and the reference lines (i.e $A'B', C'D', E'F', G'H'$) are then found. Figure C.9 shows the result of this procedure for the first PLL, L'_1. Both the left edge, L'_{1a} and right edge, L'_{1b} have been identified, and their intersections with the reference lines $A'B'$ and $C'D'$ are marked by red crosses.

At the end of this stage, the pixel coordinates are known for the six

---

[1]The term "edge" is used loosely here, as a light sheet will commonly have a Gaussian-like profile, which makes it difficult to define where the edge actually starts. However, this should not pose a problem in the current work as long as the same definition is used for both edges
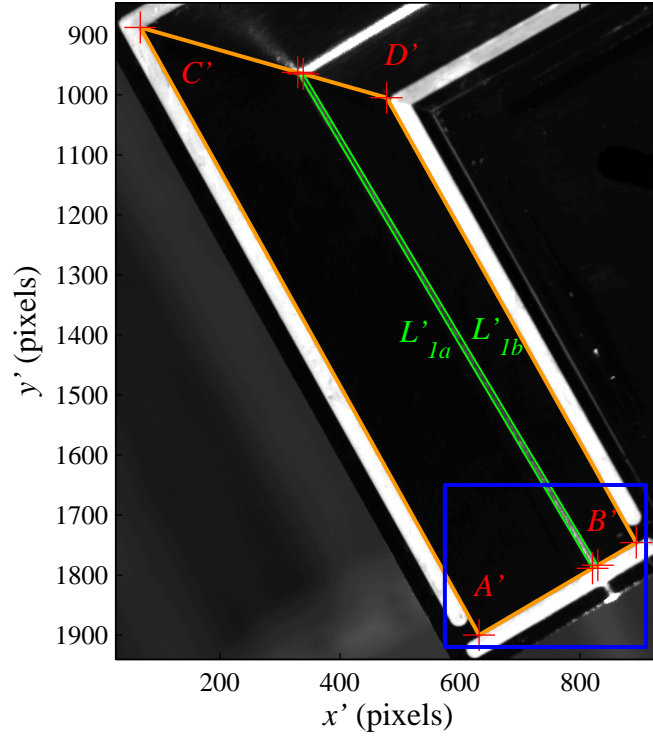
(a) Position of detected PLL/reference line intersections
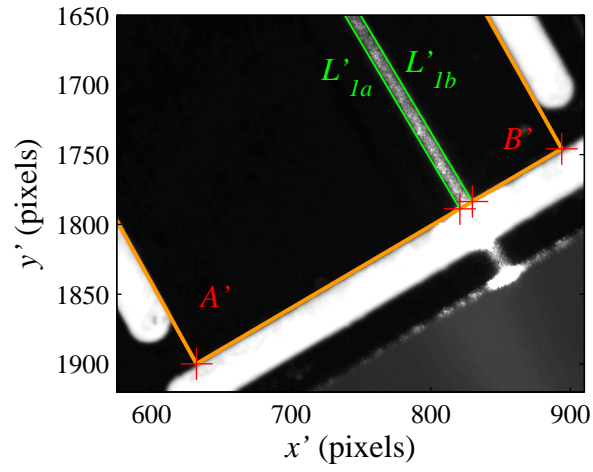
(b) Detail of the intersection between the two edges of the PLL, L'_{1a} and L'_{1b}, and the reference line $A'B'$

Figure C.9: Result of the PLL detection process

307

pairs of PLL/reference line intersections, i,e:

$$A'B'.L'_{1a} \, , \; A'B'.L'_{1b}$$
$$C'D'.L'_{1a} \, , \; C'D'.L'_{1b}$$
$$\vdots$$
$$F'G'.L'_{3a} \, , \; F'G'.L'_{3b}$$

(C.2)

## C.1.4 Stage 4: Reprojection of image-space coordinates

With knowledge of the positions of the reference points, both in the image and on the measurement frame, it is possible to calculate the positions of the PLL/reference line intersections in world coordinates. Fundamental to this is the concept of homography. If points on a plane π are projected through a common point in space onto a second plane π', the corresponding points on the planes (x, y) and (x', y') are related by a homography, which provides a transformation from the points on one plane to the corresponding points on the other, and vice versa (see figure C.10). This is precisely the situation when an image of the measurement target is acquired; points on a plane of the target (i.e. either the left, top and right side of the frame) are projected through the camera lens and onto the image sensor (figure C.10). Hence, if the homography between the target plane and the image sensor is known, then the image coordinates of the PLL/reference line intersections can be transformed to their real-world equivalents.

A homography is described by a 3×3 matrix, $\mathbf{H}$:

$$\begin{bmatrix} kx \\ ky \\ k \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x' \\ y' \\ 1' \end{bmatrix}$$

(C.3)

308
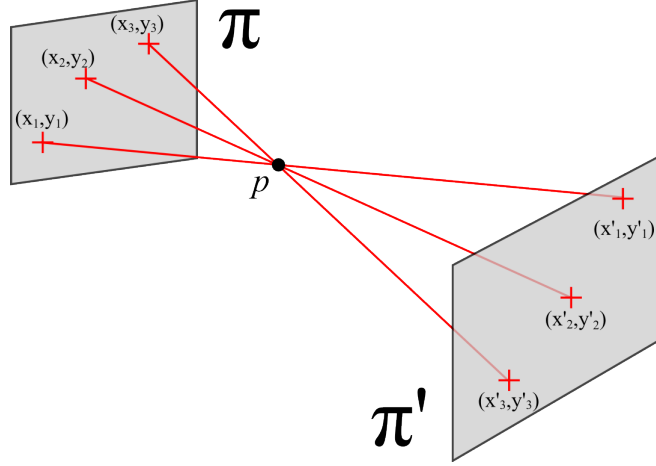
(a) The points on the plane π' can be obtained by projecting the points on plane π though the point p, and vice-versa. This relationship can be expressed as a homography.
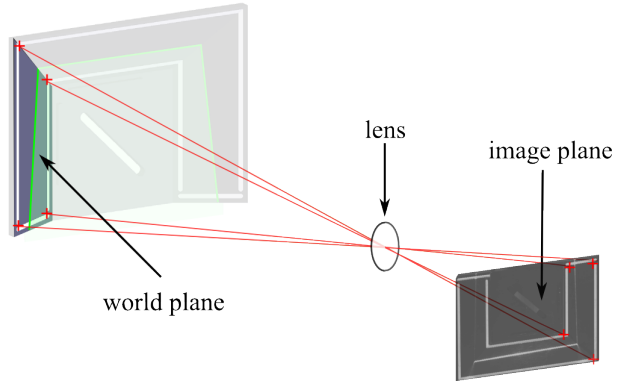
(b) The camera image is formed by the projection of the points on the calibration target though the lens and on to the sensor. Given a point in the image, knowledge of the homography allows the corresponding location on the calibration target to be calculated.

Figure C.10: The concept of homography and its use in the current application

or simply:

$$\mathbf{x} = \mathbf{Hx} \tag{C.4}$$

where k is a non-zero scalar. The matrix $\mathbf{H}$ can be obtained if at least four point correspondences are known (i.e. a point whose coordinates on both planes are known). In the present implementation, the solution is obtained using the `maketform` function in the MATLAB image processing toolbox. For the current application, four point correspondences are known for each of the three calibration frame planes:

- on the left plane, the world reference points A, B, C and D correspond to the image reference points A', B', C' and D'

- on the top plane, the world reference points C, D, E and F correspond to the image reference points C', D', E' and F'

- on the right plane, the world reference points E, F, G and H correspond to the image reference points E', F', G' and H'

In their original form, the coordinates of the world reference points describe their position in 3-dimensional space. Note that in the transformation defined in (C.4), all coordinates are expressed in only 2 dimensions, which describes their position on the plane. Before a transformation can be computed, the coordinate system of each plane on the measurement frame is rotated, in order to make one of the 3 components redundant.

Once the three homographies are computed, the coordinates of the light sheet/reference line intersection boundaries ($A'B'.L'_{1a}$, $A'B'.L'_{1b}$, etc.) are projected onto the world plane, and their 3-D coordinates are recovered.

C.1.5   Stage 5: Calculation of Light Sheet Position

After the reprojection stage, the position of the light sheet, which is of a finite thickness, is represented by 12 points, with six points describing each side of the sheet. In the current implementation of the technique, the aim is to represent the position of the light sheet in terms of a single plane, which is assumed to pass directly though the centre of the light sheet's width. Therefore, the coordinates of each point pair are averaged to produce an estimate of six points that lie in the centre of the sheet, i.e: a

$$
\begin{aligned}
A'B'.L_1' &= \frac{A'B'.L_{1a}' + A'B'.L_{1b}'}{2} \\
C'D'.L_1' &= \frac{C'D'.L_{1a}' + C'S'.L_{1b}'}{2} \\
&\vdots \\
G'H'.L_3' &= \frac{G'H'.L_{3a}' + G'H'.L_{3b}'}{2}
\end{aligned}
\tag{C.5}
$$

A plane in x-y-z space is defined by the homogeneous equation:

$$
ax + by + cz + d = 0 \tag{C.6}
$$

The equation has four unknowns, a, b, c and d, but only three degrees of freedom. Hence, three points on the plane are required to estimate the unknowns. Six points on the light sheet are known, leading to an over-determined system of equations:

$$
\mathbf{Xa} = 0 \tag{C.7}
$$

where $\mathbf{X}$ is a 6×4 matrix of points:

$$\mathbf{X} = \begin{bmatrix} x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ x_6 & y_6 & z_6 & 1 \end{bmatrix} \tag{C.8}$$

and $\mathbf{a} = [a\,b\,c\,d]^T$. The least-squares solution, $\hat{\mathbf{a}}$, is given by the smallest right singular vector of $\mathbf{X}$, which minimises $\|\mathbf{Xa}\|$ subject to the constraint that $\|\mathbf{a}\| = 1$.

The resulting values of a,b, c and d give the position of the light sheet in space, defined in the coordinate system used by the reference point coordinates.

C.2   Assessment of Measurement Performance

A simple experiment was performed to quantify the accuracy of the light sheet measurement technique, using the setup shown in figure C.11. The measurement frame was mounted onto a micrometer traverse which allowed the target to be accurately traversed along the z-axis of the frame's coordinate system (as defined in figure C.2). A continuous wave diode laser was passed through a cylindrical lens to generate a light sheet, which was projected onto the frame from below. Images of the frame were acquired using a PCO.2000 camera (described in appendix D.4) fitted with a Nikon 50mm f1.4 lens.

Using the traverse, the measurement frame was moved forward in 2mm increments to six different locations along the z axis, and an image of the target was obtained at each position. This process was repeated three times; in each case, the mirror and cylindrical lens were adjusted slightly to change
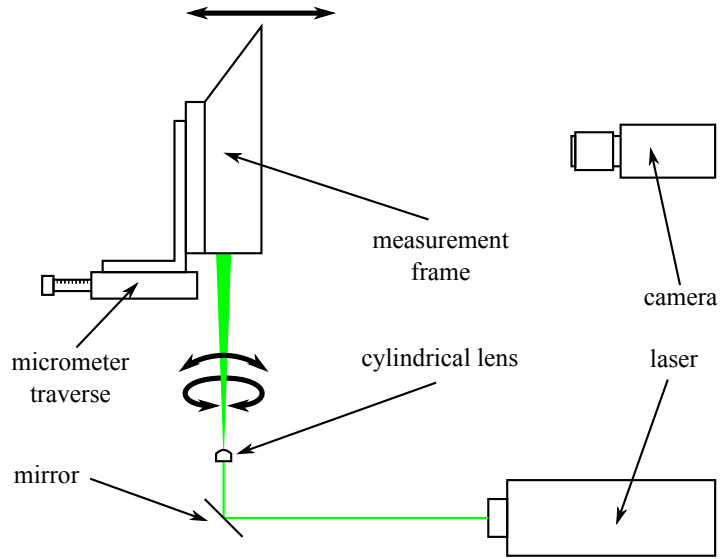
Figure C.11: Experimental setup used to assess the performance of the proposed technique

the orientation of the light sheet.

Relative to the measurement frame coordinates, each run effectively created six parallel light sheets, separated by 2mm along the z-axis. The measurement procedure was applied for each effective light sheet, and the resulting plane equations were compared with the true situation to assess the performance of the technique. Two separate error measurements were made. Firstly, the z position of each light sheet was calculated at an x, y position in the centre of the measurement target. For each possible combination of light sheet pairs, the distance between the two planes was calculated, and the deviation from the true distance (which was a multiple of 2mm) was recorded in each case. Secondly, the vector normal to each plane were obtained. Again, for each combination of light sheet pairs, the angle between these vectors was calculated. As the sheets were parallel, the true plane-normal vectors were identical, so the true angle between the vectors was zero. This procedure was

performed separately for the three runs, which provided a total of 45 error measurements for both the plane z position and the plane angle.

Over the three runs, the root mean square error in the measured z positions was 40μm, and the RMS error in the plane angle was $0.03°$. These figures represent the error in the measurement of the *relative* position of two light sheets, and so include the error from two separate absolute position measurements. As such, the error of an individual measurement may be lower than given here. On the other hand, the possible presence of systematic bias, which would be present in an absolute measurement of position, would be cancelled out in the relative measurements, and would not be reflected in these results. Further testing would be necessary to establish the performance in the absolute case. However, for the purposes of calibrating the MP-VSE experiment described in chapter 7, only the relative position of the light sheets was required, so these results are sufficient to validate the use of the technique for this purpose.

The majority of competing approaches described in section **??** do not report measurement accuracy, and so a comparison of performance is only possible in a couple of cases. A further problem is that the measure of accuracy obtained here does not apply to the proposed technique in general, only to the specific implementation used in the experiment. For example, different sized measurement frames (i.e. for use with larger or smaller flow measurement regions) would likely produce different results. Also, the effect of camera resolution has not been considered. Fortunately, in the following cases in literature, the reported accuracies apply to experimental configurations that correspond closely to the MP-VSE experiment, which the measurement frame has been specifically designed for. Therefore, at the very least, the performance of the techniques can be compared for this particular scenario.

For the technique described by Hori and Sakakibara (2004), an error of 100μm is reported for the measurement of the z position of the light sheet. This approach was utilised in a scanning-light sheet PIV experiment, where each PIV measurement consisted of a 10×10mm region. This is equivalent to the region employed in the MP-VSE experiment, and so a direct comparison of the error is valid. It is noted, however, that the figure reported by Hori and Sakakibara (2004) is an estimate, rather than a direct measurement, of the error.

An experiment similar to the one employed here was made by Wieneke (2005) to assess the accuracy of the PIV self-calibration approach outlined in the same paper. This involved applying self-calibration to images of a random dot target, which was mounted on a traverse. The resulting accuracy in terms of z location was found to be within 4μm. Although this is an order of magnitude lower than the accuracy obtained here, a comparison is perhaps slightly unfair, as the experiment used a synthetic target rather than actual particle images, and may not completely reflect the true performance. It is noted that in practical applications, self-calibration approaches are reported as being able to identify the position of the light sheet to within $0.2 - 0.5$ pixels (Wieneke and Taylor, 2006). Assuming a 100×100mm measurement region, this would correspond to an absolute error of $20 - 50\mu$m for standard 1 megapixel cameras, or $10 - 25\mu$m for 2 megapixel cameras. However, to relate the performance of self-calibration to the *exact* experimental setup described in chapter 7, the error would be $42 - 104\mu$m due to the use of low resolution cameras (Point Gray Research Inc. Firefly MV board cameras with a resolution of 640×480, described in appendix D.4). For this application at least, the proposed technique may well offer the best performance.

C.3   Conclusions and Further Work

A novel technique has been proposed to allow the rapid measurement of a light sheet's position, both absolutely and relative to another sheet. The required hardware and software has been created and has been successfully applied to measure the relative positions of multiple parallel light sheets. An experiment has been performed to estimate the technique's measurement error.

At present, the task of obtaining a position measurement from an image still requires a small amount of user input. This arises from the need to correct for differences in brightness of the three projected light sheet lines. The current geometry of the target means that the top plane of the frame is directly in the path of the light sheet, while the two side planes are intersected by the light sheet at a shallow angle. As a result, the top PLL is far brighter than the others. Presently, this requires the manual adjustment of the image contrast in order to aid the process of line detection, although in future iterations of the software, it would be simple to automate this process.

Nonetheless, the problem of differences in light intensity may still be a problem that must be dealt with directly. This was apparent from the use of the target in the MP-VSE experiment described in chapter 7. In this application, the light sheets had a narrow angle of divergence, which meant that the difference in the intensities of the PLLs was larger than the dynamic range of the camera. Therefore, two separate images had to be acquired; one for the top PLL (which was attenuated with a neutral density filter), and one for the side PLLs.

To address this problem, it may be necessary to reconsider the current design of the measurement target. For example, a target of the form shown in figure C.12 would produce two PLLs of roughly identical intensity. The
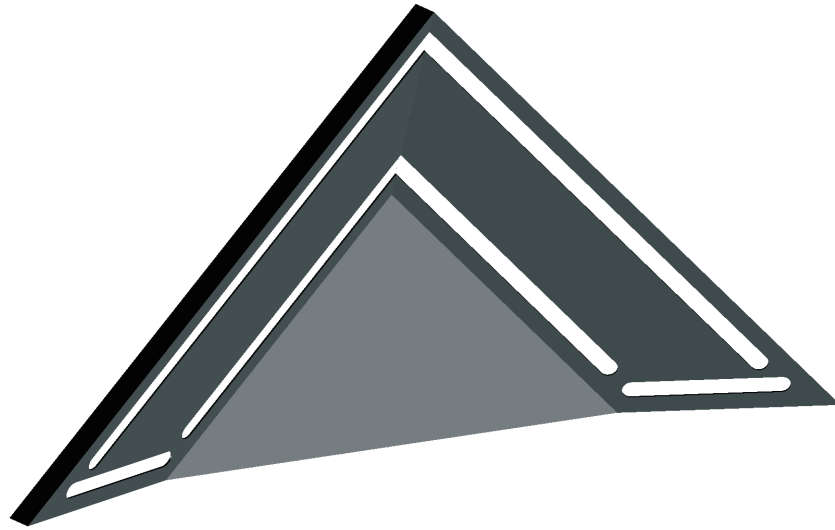
Figure C.12: Potential design to produce uniform intensity of projected light sheet lines

measurement procedure would be carried out in the same manner as before, although the number of points available to fit the plane will be reduced, at the possible expense of accuracy of the measurement.

An alternative approach may be to consider the use of materials or coatings with more appropriate reflectance properties. For example, a partially-transparent material such as ground glass may provide flexible control over the reflected light intensity, and different reflectivities could be employed for the top and side panels.

Once all user intervention is removed from the procedure, the time to produce a measurement would by dictated solely by the speed of the software. With efficient code, it would therefore be possible to provide near real-time feedback of the light sheet position, which would offer an unrivalled degree of assistance to the task of light sheet alignment.

The changes considered so far relate mainly to the usability of the approach. However, there are potentially a number of changes that may bring

about improvements in accuracy of the measurements. As such, it may be useful to investigate the following aspects of the procedure:

- The effects of edge angle, image resolution, under/over exposure, etc. on the accuracy of the edge detection routine.

- The use of a more sophisticated edge detection routine.

- The validity of defining the light sheet position with a plane mid-way between the detected edges of the finite thickness sheet.

- The effect of curvature of the light sheet.

- The relationship between measurement accuracy and frame size.

A final, but no less important, avenue for work is the further assessment of the technique's practical accuracy. The tests described in the previous section only addressed the accuracy of the measurement of one light sheet relative to another. Before using the technique for the purpose of absolute measurements, it would be useful to first assess its accuracy for this purpose.

Although a wide variety of potential improvements have been considered in this section, this should not detract from the current performance of the technique. The experimental assessment of the technique in section C.2 has shown that in terms of accuracy alone, the technique compares favourably with similar approaches. Of course, this does not account for the significant advantage of the proposed technique, with regards to the speed and simplicity in which measurements can be made. Furthermore, it is likely that the current performance is already sufficient for many applications. For example, in the dual-plane PIV experiment described in (Mullin and Dahn, 2005), a misalignment of just under 1° was detected, and was deemed sufficiently small for

318

the task of computing the velocity gradient tensor field. Clearly, this level of misalignment is easily detectable with the current setup. A similar argument is valid in the case of the MP-VSE experiment, where the sensitivity of the sheet-forming adjustment mechanism meant that the obtainable resolution of the light sheet position was far greater that the accuracy of the measurement technique (see section 7.3.2). It was therefore necessary to accept a misalignment of approximately 100μm in z, and a rotation of $0.2°$; again, this was well within the capability of the technique.

C.4   Construction of the Light Sheet Measurement Frame

Certain elements of the measurement frame design can be chosen by the user to suit the required application. In this section, some of these choices will be considered, with reference to the design of the particular frame used in this thesis. An aspect of particular importance is the angle of the sides of the frame, which makes a major contribution to the overall measurement accuracy. This can be shown by considering how error manifests itself in the measurement of a light sheet's position. The aim of the proposed technique is to determine the world coordinates of enough points on the light sheet to define its location in space. This is achieved by first calculating the position of these points within an image of the frame, and then projecting the points into world coordinates. Any error in the calculation of the points in the image will propagate into the world coordinates, which will no longer lie exactly on the light sheet plane. This leads to error in the calculated light sheet position (of course, there are other sources of error, such as in the physical construction of the target, but

319

these will not be considered here). In this context, the angle of the frame sides can be shown to have a considerable effect on the measurement accuracy. For simplicity, a 2-dimensional approximation to the measurement process will be considered, as shown in figure C.13. This example concerns a single point on the light sheet, which is determined by the intersection of the sheet with the frame side (represented by a red dot). This point is projected onto the image plane (blue dot), but due to various error sources, the position of the true point is not determined exactly; instead, the perceived position (blue cross) is subject to a degree of error, Δx. Projection of the erroneous point back onto the target results in a point that no longer lies on the light sheet plane (red cross). The amount of error introduced into the overall measurement is a function of e, which is the closest distance from the calculated point to the light sheet. Comparing figures C.13 (a) and (b), it is clear that for a given Δx, the error e becomes smaller as the angle of the frame sides is reduced.

This suggests that the measurement frame should be designed with as shallow an angle as possible. In practice, however, there are several factors that limit the minimum achievable angle. Firstly, when the aim is to measure the relative positions of multiple light sheets, the depth of the frame must be large enough to allow all the light sheets to project onto the sides. As such, a shallow angle on the sides will require a much larger frame than otherwise. Aside from the fact that more material is required for the construction, the light sheet(s) that project onto the outermost part of the frame may potentially have to be much wider than the flow measurement region requires. This is undesirable, as the full efficiency of the light source can no longer be used for the flow measurement.

A secondary feature of the current measurement frame is the ability to mount a PIV calibration target to the base plate, as shown in figure C.14.
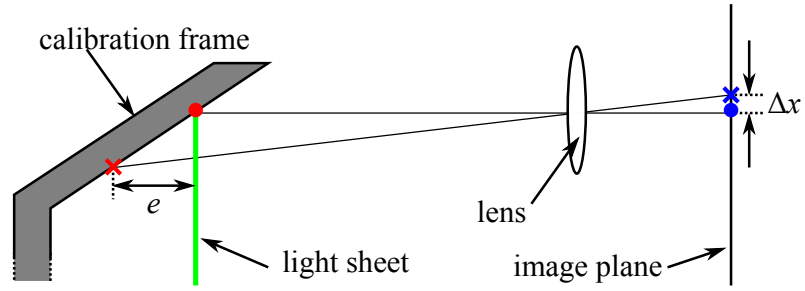
(a) Large angle frame, leading to a large error e in the calculated light sheet point.
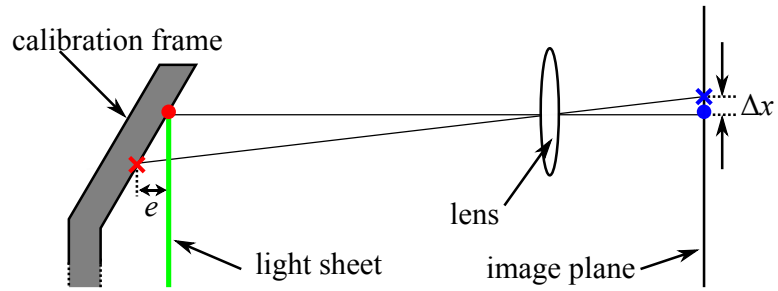
(b) Shallow angle frame, with a smaller error e.

Figure C.13: Comparison of error for different angles of calibration frame sides. The red dot corresponds to the true position of the light sheet point, and the blue point is the position of the true point when projected onto the image sensor. Due to errors in the calibration process, the perceived position light sheet point (blue cross) on the image plane is subject to a degree of error Δx. The perceived position, reprojected in world coordinates (red cross) now lies a distance e from the light sheet.
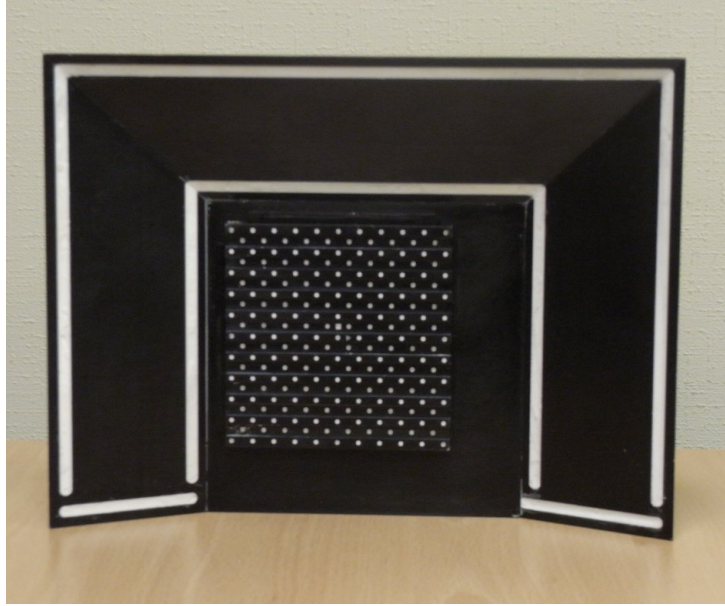
Figure C.14: Measurement frame with PIV calibration target fitted to base plate

This is beneficial from a time-saving point of view, as it allows the tasks of light sheet alignment and PIV system calibration to be carried out with the same equipment. However, this feature adds further size to the frame, which will increase the effect of the aforementioned light source efficiency problem.

The current frame was designed for the specific purpose of calibrating the vortex ring experiment, described in chapter 7. This necessitated a frame that was capable of measuring two parallel light sheets, approximately 4mm thick, with a spacing of 16mm. For each light sheet, the PIV measurement region was approximately 100×100mm. Based on the issues described here, an angle of approximately 30° was chosen for the frame sides, which was deemed to be the best compromise between accuracy and overall frame size.

The frame was fabricated from four separate aluminium pieces, (the base plate and the three individual sides), which were bolted together and

subsequently anodised to provide a low reflectivity surface suitable for use with high-power lasers. 5 mm reference tracks were milled into the anodised frame, and filled with white silicone sealant. The optional PIV calibration plate is mounted by bolting the plate to the frame through the diagonal slot in the centre of the base plate. Two reference rails on the base plate are used to accurately align the PIV target to a known position on the frame. A technical drawing of the frame is included in figure C.15.

DETAIL A
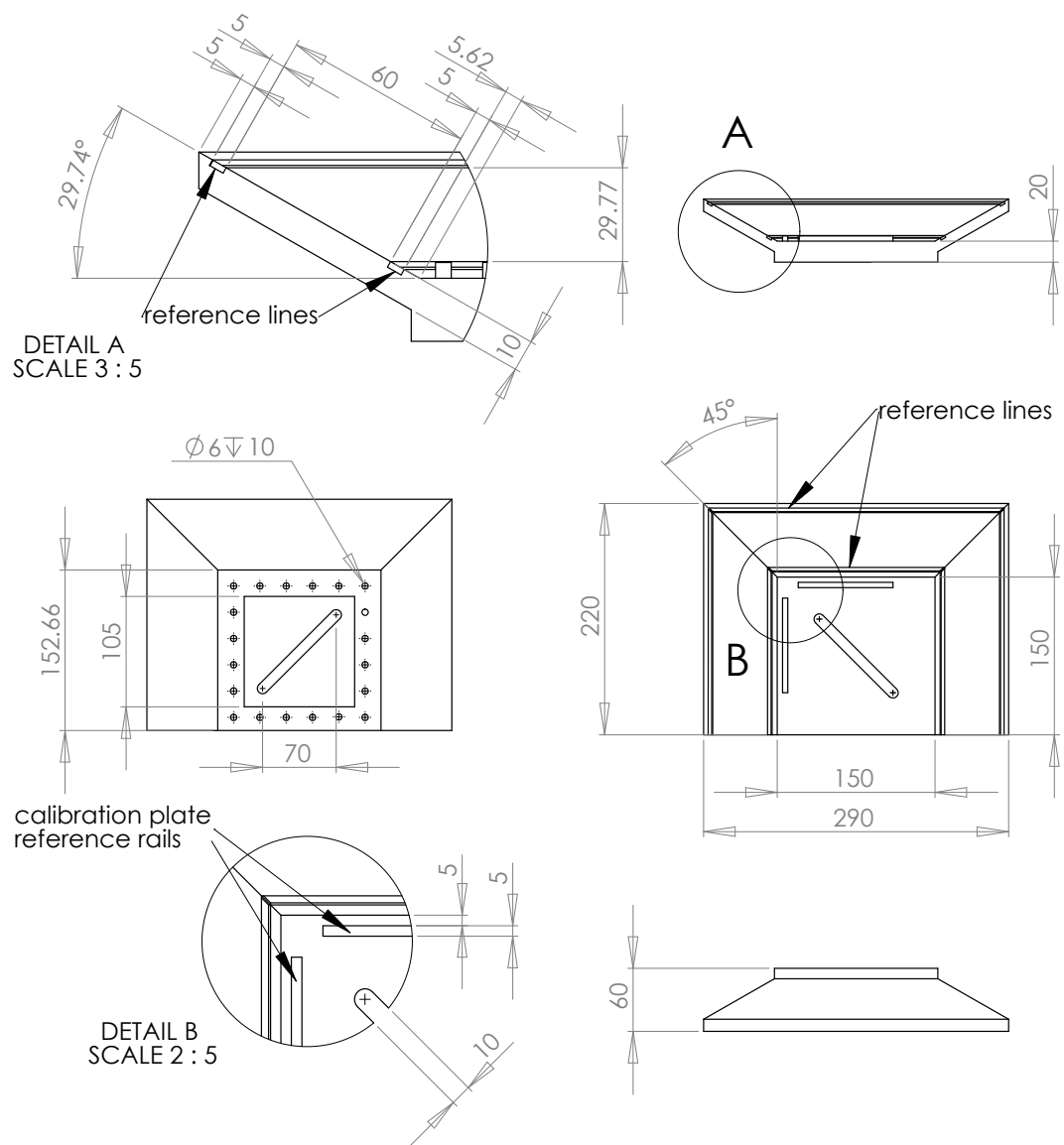SCALE 3 : 5

DETAIL B
SCALE 2 : 5

Figure C.15: Technical drawing of calibration frame. *Drawing has been resized to 60% of original, so references to scale are no longer valid.*

Appendix D

Miscellanea

D.1 An Efficient Singular Value Decomposition of a Rank Deficient Covariance Matrix

For an $m \times n$ matrix $\mathbf{X}$ and an $m \times p$ matrix $\mathbf{Y}$, the size of the covariance matrix $\mathbf{C} = \mathbf{Y}^T\mathbf{X}$ is $p \times n$, which can often be very large in the test scenarios considered in chapter 5. An extreme example of this situation occurs in appendix D.2, where the dimensions the covariance matrix reach 3072×18721, and the computation of SVD would not possible with the available computing resources. However, in situations where $m < n, p$, the covariance matrix will be rank deficient, which can be exploited to reduce the computational complexity when the SVD of $\mathbf{C}$ is sought. As described in section 3.4, the computation of the SVD of $\mathbf{C}$ can potentially be calculated from an EVD of either $\mathbf{C}^T\mathbf{C}$, or $\mathbf{C}\mathbf{C}^T$, but in the current case, this would still only reduce the problem to the EVD of a 3072×3072 matrix, which was still infeasible with

the resources at hand.

As a solution, the task of computing the SVD of $\mathbf{C}$ can be reduced to a process involving 3 separate EVDs of $m \times m$ matrices (which corresponds to a size 500×500 in the example in appendix D.2). To achieve this, a specific form of the SVD, known as the *thin* SVD, is used, which only computes the singular vectors that have a non-zero singular value. To illustrate the difference between the thin and *full* SVD, the resulting decompositions can be represented visually. Beginning with the *full* SVD, the resulting decomposition of the $m \times n$ matrix $\mathbf{A}$ is:

$$\underset{m}{\overset{n}{\mathbf{A}}} = \underset{m}{\overset{m}{\mathbf{U}_L}} \ \underset{m}{\overset{n}{\mathbf{S}}} \ \underset{n}{\overset{n}{\mathbf{V}_R^T}}$$

while the thin SVD is:

$$\underset{m}{\overset{n}{\mathbf{A}}} = \underset{m}{\overset{n}{\mathbf{U}_L}} \ \underset{n}{\overset{n}{\mathbf{S}}} \ \underset{n}{\overset{n}{\mathbf{V}_R^T}}$$

The proposed approach begins with the separate SVD decompositions of the $m \times n$ matrix $\mathbf{X}$ and $m \times p$ matrix $\mathbf{Y}$, where $m << n, p$. For efficiency, this is obtained from the EVD of $\mathbf{X}\mathbf{X}^T$ and $\mathbf{Y}\mathbf{Y}^T$, which are both $n \times n$ matrices. Continuing with the visual approach, matrix $\mathbf{C}$ can be represented as:

326

$$\underset{m}{\boxed{\mathbf{C}}^{\,n}} = \underset{p}{\boxed{\mathbf{Y}^T}} \; \underset{m}{\boxed{\mathbf{X}}^{\,n}}$$

Replacing $\mathbf{X}$ and $\mathbf{Y}$ with their corresponding SVDs:

$$\underset{m}{\boxed{\mathbf{C}}^{\,n}} = \underset{p}{\boxed{\mathbf{V_Y}}} \; \underset{m}{\boxed{\mathbf{S_Y}}^{\,m}} \; \underset{m}{\boxed{\mathbf{U_Y}^T}^{\,m}} \; \underset{m}{\boxed{\mathbf{U_X}}^{\,m}} \; \underset{m}{\boxed{\mathbf{S_X}}^{\,m}} \; \underset{m}{\boxed{\mathbf{V_X}^T}^{\,n}}$$

Multiplying the final 5 terms together produces a $m \times n$ matrix, denoted $\mathbf{A}$:

$$\underset{m}{\boxed{\mathbf{C}}^{\,n}} = \underset{p}{\boxed{\mathbf{V_Y}}} \; \underset{m}{\boxed{\mathbf{A}}^{\,n}}$$

Representing the matrix $\mathbf{A}$ as it's SVD (again, computed via the EVD of the $m \times m$ matrix $\mathbf{A}\mathbf{A}^T$):

$$\underset{m}{\boxed{\mathbf{C}}^{\,n}} = \underset{p}{\boxed{\mathbf{V_Y}}} \; \underset{m}{\boxed{\mathbf{U_A}}^{\,m}} \; \underset{m}{\boxed{\mathbf{S_A}}^{\,m}} \; \underset{m}{\boxed{\mathbf{V_A}^T}^{\,n}}$$

A property of an orthogonal matrix is that it preserves the inner product, i.e. if $\mathbf{A}$ is orthogonal, then for two real vectors, $\mathbf{x}$ and $\mathbf{y}$ the inner product:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}\mathbf{A}, \mathbf{y}\mathbf{A} \rangle \tag{D.1}$$

Also, an orthonormal matrix preserves length:

$$\|\mathbf{x}\| = \|\mathbf{x}\mathbf{A}\| \tag{D.2}$$

The matrix $\mathbf{U_A}$ is orthonormal, and the matrix $\mathbf{V_Y}$ has orthonormal columns. Therefore, for any two different columns i, j of $\mathbf{V_Y}$:

$$\langle \mathbf{v}_{\mathbf{Y}i}, \mathbf{v}_{\mathbf{Y}j} \rangle = \langle \mathbf{v}_{\mathbf{Y}i}\mathbf{U_A}, \mathbf{v}_{\mathbf{Y}j}\mathbf{U_A} \rangle = 0 \tag{D.3}$$

and also:

$$\|\mathbf{v}_{\mathbf{Y}i}\| = \|\mathbf{v}_{\mathbf{Y}i}\mathbf{U_A}\| = 1 \tag{D.4}$$

Hence, the orthonormality of the columns of $\mathbf{V_Y}$ is preserved in the matrix $\mathbf{V_Y}\mathbf{U_A}$, which shall be denoted $\mathbf{U}$. The resulting decomposition can then be written as:

$$\tag{D.5}$$

Here, the matrices $\mathbf{U}$ and $\mathbf{V}$ both have orthonormal columns, and the matrix $\mathbf{S}$ is a diagonal matrix with positive elements of decreasing size. A factorisation of this form is unique, and corresponds to the SVD.

This approach is equally applicable when finding the SVD of the deflated covariance matrix, $\mathbf{Y}^T\mathbf{X}\underset{\rightarrow\hat{\mathbf{P}}_\perp}{\mathbf{P}}$. Here, rather than beginning the procedure by finding the SVDs of $\mathbf{X}$ and $\mathbf{Y}$, the SVDs of $\mathbf{X}\underset{\rightarrow\hat{\mathbf{P}}_\perp}{\mathbf{P}}$ and $\mathbf{Y}$ are found instead.

As the dimensions of $\mathbf{X}$ and $\mathbf{X}\underset{\rightarrow}{\mathbf{P}}_{\hat{\mathbf{P}}_\perp}$ are identical, the procedure can then continue as before.

D.2 The Inclusion of Higher Order Terms in Regression Models Using the Simulated Channel Flow Data

This section considers whether higher order terms are required to correctly specify regression models built from the simulated channel flow data described in section 5.5. Two approaches are used for this purpose, both of which are outlined in section 4.2.1. In the first, the residuals of an OLS regression model are inspected, and in the second, the MSE of prediction is compared for two models; one containing only 1st order terms, and the other containing zeroth, first and second order terms. Both approaches indicate that only first order terms are necessary.

D.2.1 Inspection of Residuals

As explained in section 4.2.1, a correctly specified linear model can be identified by plotting the predicted values $\hat{y}$ against the residuals $y - \hat{y}$. The model can be assumed to be correctly specified if the residuals are trendless, and centred around zero. This approach was applied using the predicted values and residuals from a first-order OLS regression. In line with the other scenarios in chapter 5, 64 evenly spaced conditional vectors (figure D.1) were used to predict the full 32×32 field, and the model was built from 500 observations.

Due to the large number of elements in $\mathbf{y}$ in the resulting model ($p =$
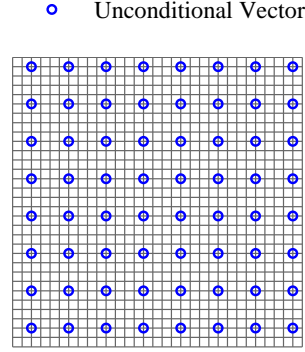
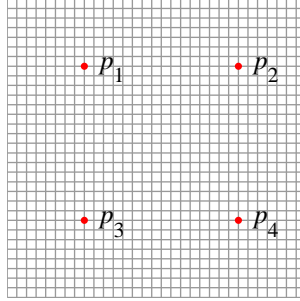Figure D.1: Position of unconditional vectors used in regression model

Figure D.2: Position of predicted vectors used to assess linearity

3072), an inspection of the residuals for each element is not possible. Instead, a small selection were chosen for analysis, which correspond to the u, v and w velocity components of the vectors shown in figure D.2. The resulting plots are shown in figure D.3.

The plots clearly meet the requirements of a correctly specified model, suggesting that the relationship between the unconditional and conditional vectors is entirely linear. Of course, only 4 vectors out of a total of 1024 have been inspected here, so it possible that some non-linearity in the data exists elsewhere. Therefore, further reassurances can be had from a direct comparison of the performance of first and second order models, which is undertaken in the next section.

(a) predicted vs. residual velocity components at p_1

(b) predicted vs. residual velocity components at p_2

(c) predicted vs. residual velocity components at p_3
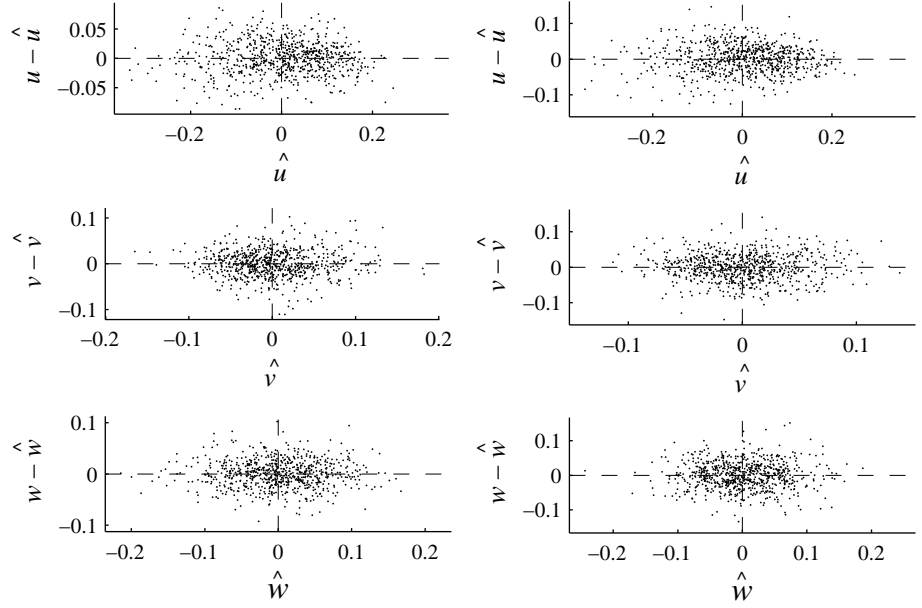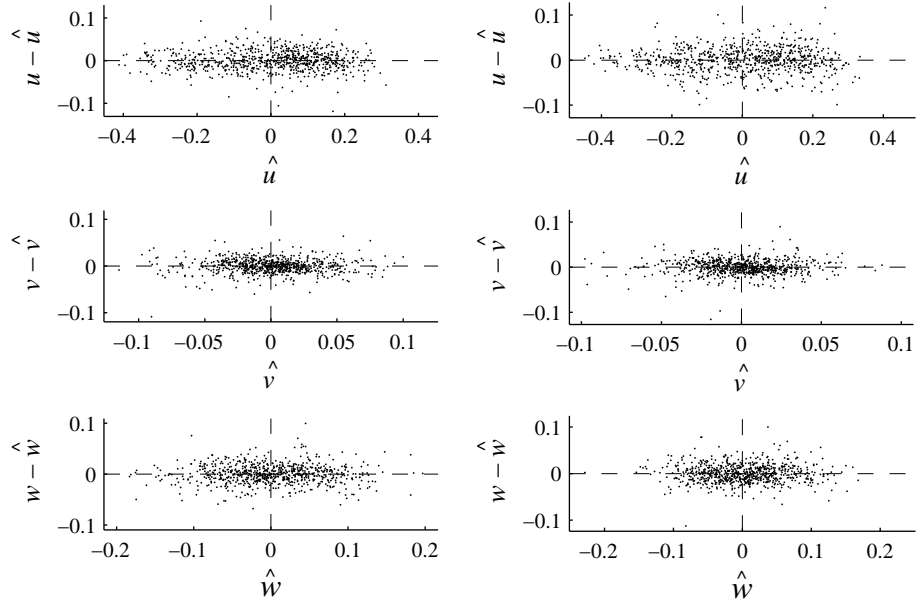
(d) predicted vs. residual velocity components at p_4

Figure D.3: Scatter plots of predicted data vs. residuals

D.2.2 Comparison of Prediction Performance for First and Second Order Regression Models

Once again, 64 evenly spaced conditional vectors were used to predict the full 32×32 field, using 500 observations. The number of independent variables depends on the order of the linear model in question. Starting with $\tilde{n}$ original (i.e. directly observable) independent variables, $\tilde{\mathbf{x}} = \tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_{\tilde{n}}$, the full set of independent variables $\mathbf{x}$ for a jth order linear model will consist of polynomial and cross terms of $\tilde{\mathbf{x}}$, up to the order of the model. The the total number of independent variables, n, is given by the binomial coefficient:

$$n = \frac{(k+j)!}{k!j!} \tag{D.6}$$

In the present case, $k = 192$, resulting in a value of $n = 18721$ for the 2nd order model. Due to the size of n, the cross-validation procedure could not be carried out on the original computer (described in section 5.3), as the memory requirements of cross-validating the full second order models exceeded the 1.2GB of addressable memory available to MATLAB on the 32-bit machine. The validation was instead performed on a Linux machine equipped with a 1.86GHz Core 2 duo processor and 8GB RAM. The extra memory and processing power afforded by this machine allowed the full second order models to be implemented, although the execution time of the cross-validation was far greater than usual, taking approximately 1 hour for the full validation of CSE, RRR and PCR, and 24 hours for PLS. The summary of parameters used in this investigation are shown in table D.1, and the results are shown in figures D.4 and D.5. The number of modes used in the biased models are given in table D.2.

Parameter	Value
Sample size, m	500
Unconditional Vectors	64
Independent Variables, n	192,18721
Conditional Vectors	1024
Dependent Variables, p	3072

Table D.1: Summary of parameter values
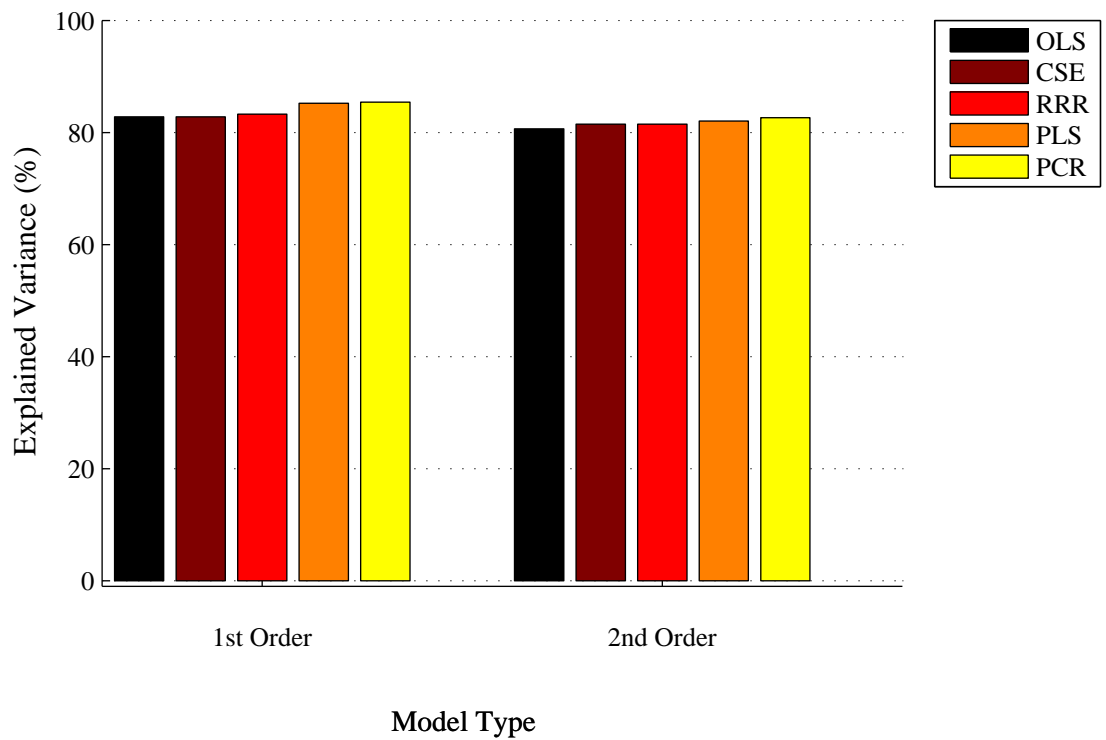
Figure D.4: Average explained variance

Model	CSE	RRR	PLS	PCR
1st Order	500 (*500*)	117 (*192*)	103 (*192*)	113 (*192*)
2nd order	120 (*500*)	120 (*500*)	142 (*500*)	138 (*500*)

Table D.2: Optimum number of modes for the biased regression models

Figure D.5: Percentage improvement over OLS prediction

Comparison of the first and second order models shows that no improvement results from the inclusion of higher order terms. In fact, the second order model actually performs worse, which will arise the due to an decrease in s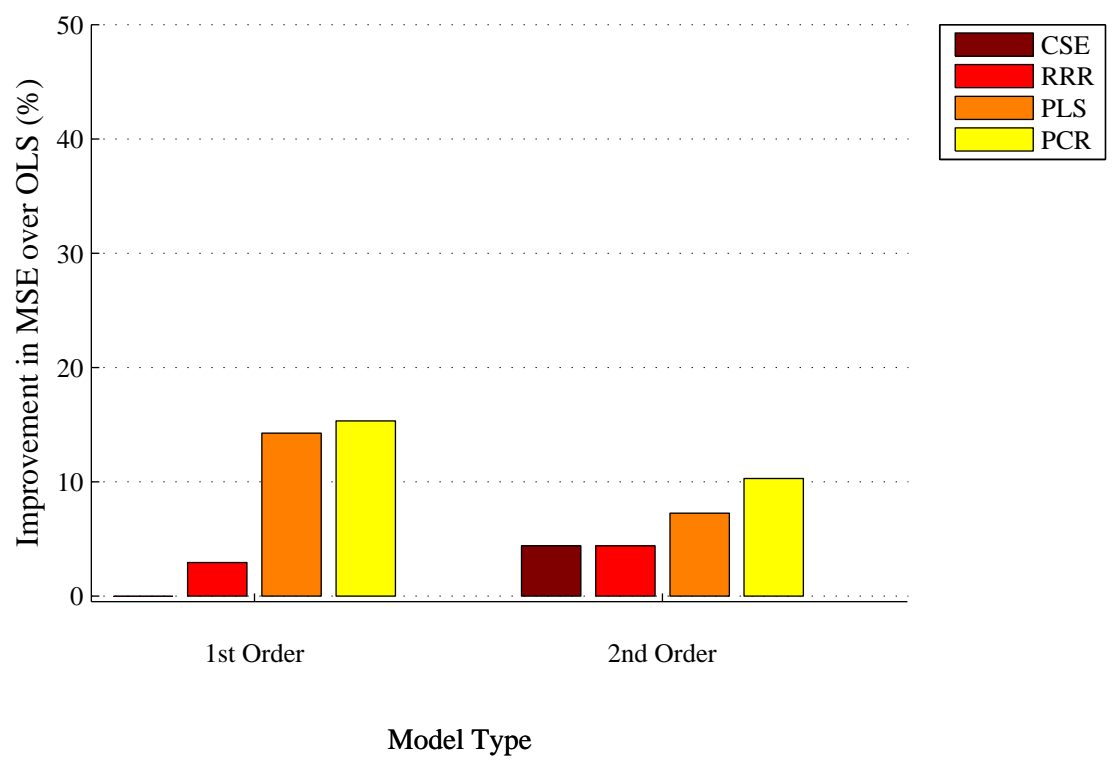tability resulting from the extra terms in the model. The relative performance of the competing techniques is consistent with the results for the four test scenarios in chapter 5. As such, discussion of these results will not be undertaken here.

D.3 Estimation Error in the Principal Components and Eigenvalues of X, and the Effect on Prediction Error

The results from test scenario 1 demonstrate that the theoretical MSE decomposition of the OLS prediction is unreliable in situations when $m < n$ and/or $m \approx n$. The work in this section suggests that this discrepancy is likely to be caused by the presence of estimation error in the principal components and eigenvalues of $\mathbf{X}$, that has not been taken into account in the theoretical expression. In the original form of the MSE decomposition, the empirical principal components, $\mathbf{u}_k$, and eigenvalues, l_k, are assumed to equal the underlying population principal components, $\boldsymbol{\gamma}_k$, and eigenvalues, λ_k. The implications of these assumptions are recapped by returning to the derivation of the MSE decomposition, beginning at equation (4.66):

$$\text{MSE}\left[\hat{y}\right] = \sigma^2 + \sigma^2 \text{E}\left[\mathbf{x}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{x}^T\right] \tag{D.7}$$

Which can be expressed in terms of the principal components and corresponding eigenvalues of $\mathbf{X}$ as:

$$\text{MSE}\,[\hat{y}] = \sigma^2 + \frac{\sigma^2}{m} \sum_{k=1}^{n} \text{E}\left[\frac{(\mathbf{x}\mathbf{u}_k)^2}{l_k}\right] \tag{D.8}$$

However, as noted in section 5.4, for situations where $m < n$, the number of principal components in $\mathbf{X}$ is limited to m, rather than n. Hence, equation (D.8) can be generalised to account for any m:

$$\text{MSE}\,[\hat{y}] = \sigma^2 + \frac{\sigma^2}{m} \sum_{k=1}^{q} \text{E}\left[\frac{(\mathbf{x}\mathbf{u}_k)^2}{l_k}\right] + \text{E}\left[\left(\mathbf{x}\underline{\mathbf{P}}_{\xrightarrow{\tilde{V}_\perp}} \check{\mathbf{b}}\right)^2\right] \tag{D.9}$$

where $q = \min(n, m)$, and the final term is the bias introduced into the prediction when $m < n$ (see section 5.4). Given the assumption that $\mathbf{u}_k = \boldsymbol{\gamma}_k$, and $l_k = \lambda_k$, this becomes:

$$\text{MSE}\,[\hat{y}] = \sigma^2 + \frac{\sigma^2}{m} \sum_{k=1}^{q} \text{E}\left[\frac{(\mathbf{x}\boldsymbol{\gamma}_k)^2}{\lambda_k}\right] + \text{E}\left[\left(\mathbf{x}\underline{\mathbf{P}}_{\xrightarrow{\tilde{V}_\perp}} \check{\mathbf{b}}\right)^2\right] \tag{D.10}$$

The expected value of $(\mathbf{x}\boldsymbol{\gamma}_k)^2$ gives the variance accounted for by the kth population principal component, which is the kth eigenvalue λ_k. Therefore:

$$\text{MSE}\,[\hat{y}] = \sigma^2 + \frac{\sigma^2}{m} \text{E}\left[\sum_{k=1}^{q} \frac{\lambda_k}{\lambda_k}\right] + \text{E}\left[\left(\mathbf{x}\underline{\mathbf{P}}_{\xrightarrow{\tilde{V}_\perp}} \check{\mathbf{b}}\right)^2\right] \tag{D.11}$$

and so:

$$\text{MSE}\,[\hat{y}] = \sigma^2 + \frac{q}{m}\sigma^2 + \text{E}\left[\left(\mathbf{x}\underline{\mathbf{P}}_{\xrightarrow{\tilde{V}_\perp}} \check{\mathbf{b}}\right)^2\right] \tag{D.12}$$

The steps outlined here imply that each principal component in $\mathbf{X}$ contributes an equal amount to the model variance $\frac{q}{m}\sigma^2$. In reality, the assumptions that

$\mathbf{u}_k = \boldsymbol{\gamma}_k$, and $l_k = \lambda_k$ do not hold exactly, as both $\mathbf{u}_k$ and l_j are subject to a degree of estimation error. In the following discussion, it is assumed that $m \geq n$ (i.e. $q = n$), which ensures that the full set of n population principal components and eigenvalues are estimated from $\mathbf{X}$.

Firstly, it is known that the small-sample estimate $\mathbf{u}_k$ is an unbiased estimate of $\boldsymbol{\gamma}_k$ but is subject to a degree of variance, which increases as the sample size decreases. This has implications on the assumption that $\mathrm{E}[(\mathbf{x}\mathbf{u}_k)^2] = \mathrm{E}[(\mathbf{x}\boldsymbol{\gamma}_k)^2] = \lambda_k$. Considering the largest principal component, $\mathbf{u}_1$, and assuming that the λ_1 is larger than λ_2, the quantity $\mathrm{E}[(\mathbf{x}\mathbf{u}_1)^2]$ will only equal λ_1 when $\mathbf{u}_1 = \boldsymbol{\gamma}_1$. As $\boldsymbol{\gamma}_1$ is the vector that passes through the direction of highest variance in the distribution of $\mathbf{x}$,, any vector $\mathbf{v}$ will will necessarily produce $\mathrm{E}[(\mathbf{x}\mathbf{v})^2] < \lambda_1$, when $\mathbf{v} \neq \boldsymbol{\gamma}_1$. Due to the small-sample variance of $\mathbf{u}_1$, on average $\mathbf{u}_1 \neq \boldsymbol{\gamma}_1$, and so $\mathrm{E}[(\mathbf{x}\mathbf{u}_1)^2] < \lambda_1$. Similarly, the smallest population principal component $\boldsymbol{\gamma}_n$ passes through the direction of lowest variance, which means that $\mathrm{E}[(\mathbf{x}\mathbf{v})^2] > \lambda_n$ when $\mathbf{v} \neq \boldsymbol{\gamma}_n$, assuming that the λ_n is smaller than λ_{n-1}. As such, $\mathrm{E}[(\mathbf{x}\mathbf{u}_n)^2] > \lambda_n$. In general then, it appears that the expected contribution of $\mathrm{E}[(\mathbf{x}\mathbf{u}_k)^2]$ will be smaller than λ_k for the largest principal components, and larger for the smallest.

For small-sample estimates of λ_k, Lawley (1956) showed that there is bias of the form,

$$\mathrm{E}[l_k] = \lambda_k \left(1 - \frac{1}{m} \sum_{h \neq k}^{n} \left(\frac{\lambda_h}{\lambda_k - \lambda_h} \right) \right) + \mathrm{O}(1/n^2) \tag{D.13}$$

assuming that the elements of $\mathbf{X}$ are normally distributed, and the population eigenvalues are distinct. The exact nature of this bias is rather difficult to predict exactly, as each $\mathrm{E}[l_k]$ is determined by a complex interaction between all the underlying population roots. However, contrary to the behaviour of

337

$E[(\mathbf{x}\mathbf{u}_k)^2]$, the general effect is that the largest roots will be larger than the equivalent population roots, and the smallest roots will be smaller. As before, the overall severity of bias increases as the sample size, m, decreases.

This combination of small-sample behaviour of $\mathbf{u}_k$ and l_k is likely to have an effect on the model variance component of equation (D.9). Writing this term as:

$$\frac{\sigma^2}{m} \sum_{k=1}^{q} E\left[\frac{(\mathbf{x}\mathbf{u}_k)^2}{l_k}\right] = \frac{\sigma^2}{m} \sum_{k=1}^{q} C_k \tag{D.14}$$

Under the original assumptions, $C_k = 1$ for all k. However, given the behaviour described above, it is expected that $C_k < 1$ for the largest principal components, and $C_k > 1$ for the smallest. This suggests that the smallest principal components of $\mathbf{X}$ will contribute a disproportionate amount towards the model variance. Moreover, it is noted that as m decreases, the lowest possible value of that C_k can take for the largest principal components is 0, whereas for the smallest components, the term can increase indefinitely. Therefore, it is possible that overall, the model variance will be larger than the $\sigma^2 \frac{q}{m}$ quoted in the original derivation.

This line of reasoning would explain the fact that the theoretical expression for the MSE is an underestimate of the true error for m, and why it becomes increasingly erroneous with decreasing m, However, the argument is, at best, only a hypothesis, and it does not directly explain the behaviour of the MSE when $m < n$. Furthermore, the above discussion has considered the behaviour of $E[(\mathbf{x}\mathbf{u}_k)^2]$ and $E[l_k]$ independently, even though these two terms are not actually independent in the model variance term, i.e:

$$\frac{\sigma^2}{m} \sum_{k=1}^{n} E\left[\frac{(\mathbf{x}\mathbf{u}_k)^2}{l_k}\right] \neq \frac{\sigma^2}{m} \sum_{k=1}^{n} \frac{E[(\mathbf{x}\mathbf{u}_k)^2]}{E[l_k]} \tag{D.15}$$

In the absence of a more rigorous mathematical explanation, it is necessary to rely on empirical evidence to support this argument. For this purpose, a simple regression problem was created using simulated observations of the univariate variable y and the 50 element vector $\mathbf{x}$. The behaviour of y was determined by the model $y = \mathbf{x}\check{\mathbf{b}} + e$, where $\check{\mathbf{b}} = [1\ 1\ \dots\ 1]^T$, and both the error term e, and the elements of $\mathbf{x}$ were normally distributed. The underlying distribution of $\mathbf{x}$ was designed so that the population eigenvalues linearly decreased from $\lambda_1 = 5$ to $\lambda_{50} = 1$.

Using data generated from this underlying model, the MSE of prediction $\text{MSE}[\hat{y}]$ for OLS regression was measured, for a model built using a sample size ranging from $m = 1, 2, \dots, 500$. Alongside the regression procedure, the average values of $\text{E}[(\mathbf{x}\mathbf{u}_k)^2]$, $\text{E}[l_k]$ and $\text{E}[C_k]$ were obtained. Also, when $m < 50$, the bias term from equation (D.8):

$$\text{E}\left[\left(\mathbf{x}\underset{\rightarrow}{\mathbf{P}}_{\tilde{V}_\perp}\check{\mathbf{b}}\right)^2\right] \tag{D.16}$$

was also calculated. To ensure representative statistics, this process was repeated 200 times, and the results averaged.

Figure D.9 shows the resulting average behaviour of $\text{E}[(\mathbf{x}\mathbf{u}_k)^2]$ versus k, for $m = 25, 50, 100, 500$. The dashed line on the plot corresponds to λ_k, which is the value of $\text{E}[(\mathbf{x}\mathbf{u}_k)^2]$ that would arise if $\mathbf{u}_k = \boldsymbol{\gamma}_k$. As predicted, the small-sample variance of $\mathbf{u}_k$ ensures that $\text{E}[(\mathbf{x}\mathbf{u}_k)^2]$ is less than λ_k for the largest principal components, and larger than λ_k for the smallest. Also, the deviation from λ_k increases with decreasing m. Note that for $m = 25$, only 25 principal components exist, and so $\text{E}[(\mathbf{x}\mathbf{u}_k)^2]$ is truncated at this value.

Figure D.7 shows how $\text{E}[l_k]$ varies with k for the same choices of m. Again, this corresponds to the expected behaviour, where $E[l_k]$ is too large
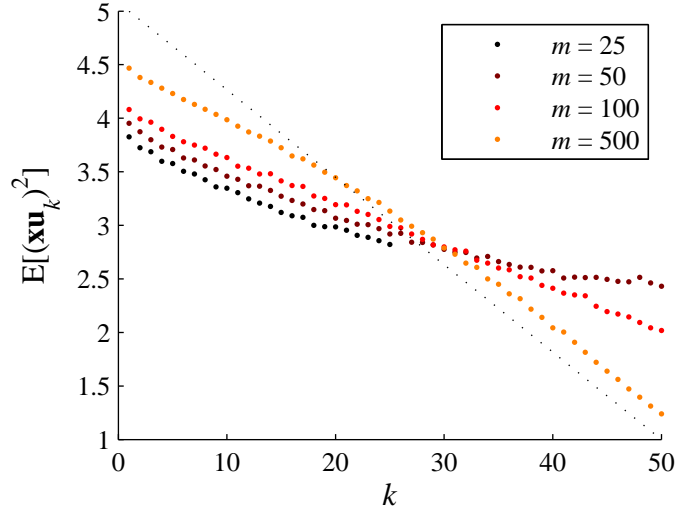
339

Figure D.6: Size of $E[l_k]$ for differing sample sizes. The true values of the eigenvalues are indicated by the dashed line.

for the largest principal components, and too low for the smallest. In the case of $m = 50$, that bias is so large that many of the smallest eigenvalues are almost zero. However, note that for $m = 25$, the omission of the lowest 25 principal components ensures that these extremely small eigenvalues are no longer present. The effect of this is illustrated in the plot of C_k versus k in figure D.8. Here, the horizontal dashed line corresponds to $C_k = 1$, which was the assumption in the original derivation of the OLS MSE expression.

The behaviour illustrated here corresponds exactly with the proposed hypothesis. For small k, the value of C_k is less than 1, and for large k, it is larger. Also increase in C_k for the smallest principal components can exceed the corresponding decrease in C_k for the largest components by many orders of magnitude, which is likely to inflate the model variance far beyond the original value of $\sigma^2 \frac{q}{m}$. However, for $m = 25$, the omission of the smallest components appears to avoid this problem to a large extent. These concepts are made
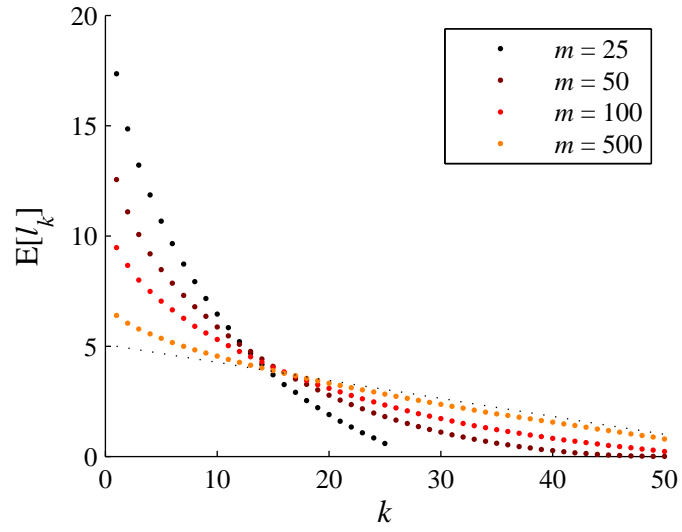
340

Figure D.7: Size of $E[(\mathbf{x}\mathbf{u}_k)^2]$ for differing sample sizes. The true values of the eigenvalues are indicated by the dashed line.
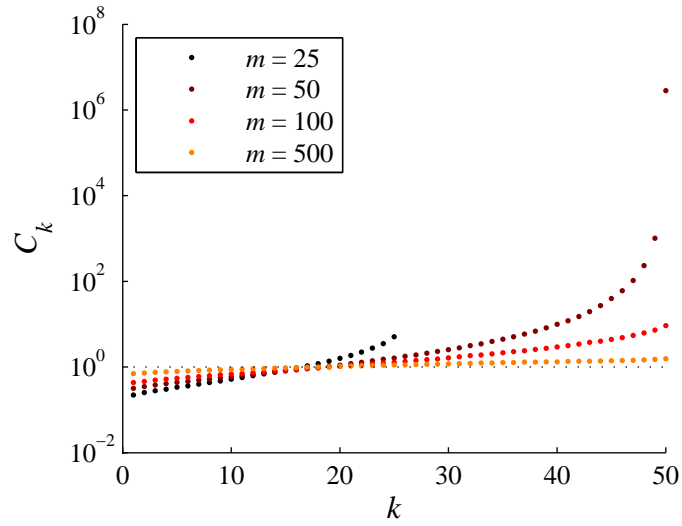

Figure D.8: Size of C_k for differing sample sizes. The dashed line indicates $C_k = 1$.
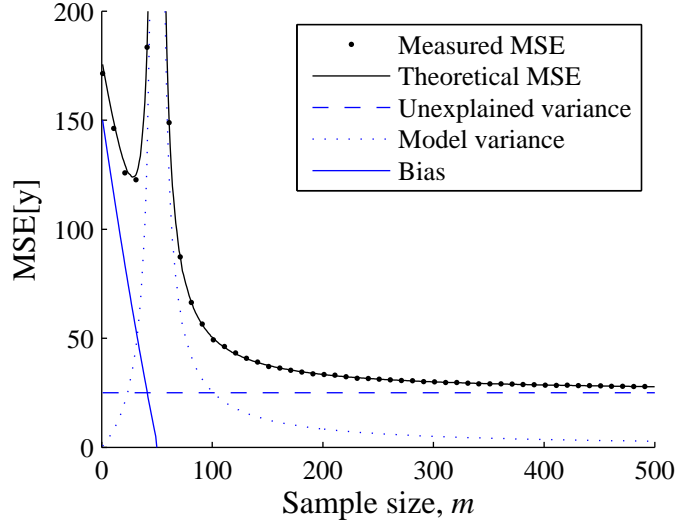
Figure D.9: Small-sample biasedness of principal component eigenvalues for differing sample sizes. The true values of the eigenvalues is indicated by the dashed line.

explicit in the final plot, which shows how the theoretical MSE, based on the expression:

$$\text{MSE}\left[\hat{y}\right] = \sigma^2 + \frac{\sigma^2}{m} \sum_{k=1}^{q} \text{E}[C_k] + \text{E}\left[\left(\mathbf{x}\underrightarrow{\mathbf{P}}_{\tilde{V}_\perp} \check{\mathbf{b}}\right)^2\right] \tag{D.17}$$

varies with m. Also shown on the plot are the individual components of the theoretical MSE, as well as the actual MSE based on predictions of y. First of all, it can be seen that the theoretical MSE, which is derived purely from the behaviour of $\mathbf{X}$, and knowledge of the true underlying model, matches the actual MSE precisely. This confirms that equation (D.17) is exact. As hypothesised, when $m > n$ and $m \approx n$ the model variance, $\sum_{k=1}^{q} C_k$ becomes vastly inflated due to the contribution of the smallest principal components, to the point where at $m = n$ (i.e. $m = 50$), the model variance is in the order of 10^6. When $m < n$, and with decreasing m, the problematic principal com-

Parameter	Value
Sensor type	15.6×15.3mm Peltier cooled CCD
Sensor resolution	2048×2048 pixels
Pixel size	7.4μm
Analog-to-digital converter	14-bit
Maximum dynamic range	72dB
Minimum exposure time	5μs (first frame only)
Minimum inter-frame time	500ns
Maximum frame rate	14.7 Hz
Interface	CameraLink

Table D.3: Specification of PCO.2000 cameras.

ponents are removed consecutively, which begins to reduce the model variance again. However, as shown in the plot, this increases the model bias, leading to a distinct local minimum in the MSE at around $m = 30$.

The results from this simulation provide convincing evidence that the proposed hypothesis is correct. It remains to be seen how these results would differ if the behaviour of the underlying data was changed; certainly, the precise values of $\mathrm{E}[(\mathbf{x}\mathbf{u}_k)^2]$ and $\mathrm{E}[l_k]$ appear to be determined by the complex interaction of the underlying eigenvalues of $\mathbf{x}$. However, it seems likely that for most practical applications, the smallest principal components of $\mathbf{X}$ will always be a serious problem for OLS regression in situations where $m \approx n$.

D.4   Specification of PIV Cameras

The specification of the PCO Gmbh. PCO.2000 and Point Gray Research FireFly MV cameras are provided in tables D.3 and D.4, respectively.

Parameter	Value
Sensor type	1/3" progressive scan CMOS
Sensor resolution	752×480 pixels
Pixel size	6μm
Analog-to-digital converter	10-bit
Maximum dynamic range	> 50dB
Minimum exposure time	120μs
Minimum inter-frame time	≈ 17ms
Maximum frame rate	60Hz
Interface	Firewire 400

Table D.4: Specification of Firefly MV cameras.

D.5 Prediction of the Number of Waves on the Experimental Vortex Rings

This section shows how the number of waves on the experimental vortex rings can be calculated, using the model proposed by Saffman (1978). The working here is adapted from the approach outlined by Dazin *et al.* (2006a), and uses the average vortex ring parameters from the initial PIV measurements (summarised in table 7.4) to calculate the prediction.

Saffman (1978) describes how a *confluent hypergeometric* function is well suited for modelling the vorticity distribution within an experimental vortex ring. This behaviour of this function is determined by the parameter ϵ, which can be calculated from the inner core radius, a_i, and the effective core radius, a_e:

$$\frac{a_i}{a_e} = \frac{1.45\epsilon}{0.47 + 0.63\epsilon} \tag{D.18}$$

The inner core radius a_i is defined as the distance from the core where the tangential velocity is maximum, the average of which was calculated from the

initial PIV study. The effective core radius is calculated from:

$$V = \frac{\Gamma}{4\pi R} \left(\ln \left(\frac{8R}{a_e} \right) - 0.25 \right)$$
(D.19)

where the ring translational velocity V, radius R, and circulation Γ are obtained from the PIV measurements. The instability on a vortex ring is a result of an infinite number of unstable modes, each with a different number of waves, n_j (although the importance of the modes quickly falls with mode number). The number of waves, n_j for mode j is given by:

$$n_j = \frac{\kappa_j s_i}{8} \frac{a_e}{a_i} e^{(\tilde{V}+0.25)}$$
(D.20)

where $\tilde{V}$ is the nondimensional velocity:

$$\tilde{V} = \frac{V 4\pi R}{\Gamma}$$
(D.21)

For a given ϵ, the values of $\kappa_j s_i$ are found numerically. Solutions for a variety of different ϵ are given in Saffman (1978), which are used here. From equation (D.18), a value of $\epsilon = 0.4$ is obtained, which gives values of $\kappa_1 s_i = 2.37$, $\kappa_2 s_i = 2.37$ and $\kappa_3 s_i = 5.74$. The resulting number of waves on the first three modes are $n_1 = 7.8$, $n_2 = 13.5$ and $n_3 = 18.2$.

D.6 Propagation of Error in the MP-VSE Procedure

This section shows how the various sources of error propagate through the MP-VSE procedure, and dictate the ultimate prediction accuracy of the technique

for a given experiment.

In practice, the prediction error will differ for each predicted velocity component in the volume Ω. For simplicity, the prediction of a single velocity component $\check{y}$ will be considered here, which lies at a point (x_1, y_1, z_1) in the prediction volume. The flow is assumed to be homogeneous in z.

The MP-VSE process begins by making a series of measurements at a plane $\tilde{P}$, positioned as closely as possible to the hypothetical plane P, which lies exactly normal to the homogeneous dimension (i.e. the $x - y$ plane) at a location z_P. It is assumed that z_1 and z_P are at different locations in z. In the actual prediction model, an $i \times j$ grid of vectors is predicted, but for this example, it is the variable $\check{y}_P$, which is of importance. This is the variable on the P plane that corresponds to the true value $\check{y}$; i.e the velocity component that has the same (x, y) location as $\check{y}$, but lies at z_P. However, due to PIV measurement error and any misalignment of the $\tilde{P}$ plane, there will be a discrepancy between the measured data $\tilde{y}_P$, and the true value $\check{y}_P$. To account for this, $\check{y}_P$ can be expressed as:

$$\check{y}_P = \tilde{y}_P + e_{\tilde{y}_P} \tag{D.22}$$

Using the measured data $\tilde{y}_P$ and $\tilde{\mathbf{x}}_P$, a prediction model of the form $\hat{y}_P = \tilde{\mathbf{x}}_\mathbf{P}\hat{\mathbf{b}}$ is built. To account for the prediction error present in $\hat{y}_P$ (i.e. the combination of unexplained error and model error), $\tilde{y}_P$ is expressed as:

$$\begin{aligned} \tilde{y}_P &= \hat{y}_P + e_{\hat{y}_P} \\ &= \tilde{\mathbf{x}}_P\hat{\mathbf{b}} + e_{\hat{y}_P} \end{aligned} \tag{D.23}$$

It is noted that after the process of cross-validation, an estimate of the MSE

of prediction, $\mathrm{E}[(e_{\hat{y}_P})^2]$ is known. For the final reconstruction, $\tilde{\mathbf{x}}_Q$ is extracted from the measurements of the $\tilde{Q}$ planes, (which are aligned, as best as possible, to lie parallel to the homogeneous dimension z), and used to predict $\tilde{y}_P$. This introduces further error, due to the fact that measurement error and misalignment in the $\tilde{Q}$ planes results in a disparity between $\tilde{\mathbf{x}}_P$ and $\tilde{\mathbf{x}}_Q$:

$$\tilde{\mathbf{x}}_P = \tilde{\mathbf{x}}_Q + \mathbf{e}_{\tilde{\mathbf{x}}_Q} \tag{D.24}$$

Finally, because z_P and z_1 lie at different locations, any inhomogeneity in the flow means that the prediction model, built at z_P, will introduce an error term e_H when applied at z_1. The various sources of error considered here can be combined into a single expression for the MSE of prediction:

$$
\begin{aligned}
\mathrm{MSE}[\hat{y}] &= \mathrm{E}\left[(\breve{y} - \hat{y})^2\right] \\
&= \mathrm{E}\left[(\breve{y}_P - \hat{y} + e_H)^2\right] \\
&= \mathrm{E}\left[\left(\breve{y}_P - \tilde{\mathbf{x}}_Q\hat{\mathbf{b}} + e_H\right)^2\right] \\
&= \mathrm{E}\left[\left(\breve{y}_P - \tilde{\mathbf{x}}_P\hat{\mathbf{b}} - \mathbf{e}_{\tilde{\mathbf{x}}_Q}\hat{\mathbf{b}} + e_H\right)^2\right] \\
&= \mathrm{E}\left[\left(\tilde{y}_P - e_{\tilde{y}_P} - \tilde{\mathbf{x}}_P\hat{\mathbf{b}} - \mathbf{e}_{\tilde{\mathbf{x}}_Q}\hat{\mathbf{b}} + e_H\right)^2\right] \\
&= \mathrm{E}\left[\left(e_{\hat{y}_P} - e_{\tilde{y}_P} - \mathbf{e}_{\tilde{\mathbf{x}}_Q}\hat{\mathbf{b}} + e_H\right)^2\right]
\end{aligned}
\tag{D.25}
$$

Assuming that each of these error terms is independent, the MSE can be expressed as:

$$\mathrm{MSE}[\hat{y}] = \mathrm{E}[(e_{\hat{y}_P})^2] + \mathrm{E}[(e_{\tilde{y}_P})^2] + \mathrm{E}[(\mathbf{e}_{\tilde{\mathbf{x}}_Q}\hat{\mathbf{b}})^2] + \mathrm{E}[(e_H)^2] \tag{D.26}$$

which shows that the final MSE is a relatively simple sum of error variances,

one of which, $\mathrm{E}[(e_{\tilde{y}_P})^2]$, will already be known from the cross-validation process. Therefore, an estimate of the variance of the remaining error terms will provide an estimate of the final prediction error, although it should be remembered that each of these error terms is in fact a combination of various errors that arise at each stage of the process. For some of these errors, such as the PIV measurement error, it may be possible to obtain a reasonable estimate of the variance from PIV literature. Variances relating to light sheet misalignment may be more difficult to obtain. One approach would be to estimate the distances that the light sheets are misaligned by, and then estimate the resulting error by calculating the drop in the correlation of velocity over this distance. It may be possible to achieve this using the P and Q plane data, or it may require a further set of measurements to be made.

The total number of error terms is likely to be far too large to be estimated individually, and so it will be necessary to make further assumptions and generalisations in order to make the process remotely feasible, especially considering that the above expression only applies to a single predicted component. Of course, the introduction of further assumptions may reduce that accuracy of the estimate. Already, the assumption that the error terms are independent is unlikely to hold, but accounting for the covariances between all the error terms would be impossible. Nonetheless, the results in this section indicate that with further work, an estimate of the overall MSE can be obtained. This will be of considerable use in determining the degree of confidence that can be placed in the predictions for a given application.