Validation of Flow Simulation Model using Particle Image Velocimetry Data and Dimensionality Reduction Techniques

SHEN Li

Jesus College

University of Oxford

Supervisor: Prof. Richard Stone

Co-supervisor: Prof. Ben Williams

A thesis submitted in partial fulfilment of

the requirements for the degree of Doctor of Philosophy

Michaelmas 2021

Acknowledgements

I would like to express my deepest gratitude to Prof. Richard Stone for giving me the opportunity to study at the University of Oxford and for guiding me through my DPhil program as a mentor. His patient supervision and extensive technical insights are invaluable to me and this work. I am very lucky to be his student. I also want to say thank you to Prof. Ben Williams for providing many useful suggestions as my co-supervisor and for allowing me to explore more research directions. I really appreciate Prof. Martin Davy for enrolling me into this excellent research group and providing financial support from his research funding in my first year. My special thank goes to Dr. Chris Willman, who shared his broad scope of knowledge in optical diagnostics and provided daily guidance on the operation of the optical engine. He also helped design the experiment setups. I am very grateful to be part of the Thermal Propulsion Systems Research Group, and my colleagues, Prof. Felix Leach, Dr. Blane Scott, Dr. Nick Papaioannou, Dr. Xiaohang (Leo) Fang, Dr. Laurent Le Page, Mr. Priyav Shah, Ms. Hannah Rana, Mr. Kharthik Chakravarthy, Mr. Maruthi Malladi and other fellow students created pleasant work place and provided enjoyable distractions when necessary. I want to extend my gratitude to both the management and CFD teams in Jaguar Land Rover: Dr. Lyn McWilliam, Dr. Tom Lockyer, Dr. Giuseppe Virelli, Ms. Rachel Magnanon, Dr. Dave Richardson and Mr. Ricardo Santos for their professional management of the project, prompt technical support in building the optical engine and countless efforts in simulating the engine flow fields. The JLR research leaderships also provided huge support, without whom this project would not have been possible. I am very appreciative of my previous advisors, Prof. David Hung and Prof. Kwee-Yan Teh from the University of Michigan–Shanghai Jiao Tong University Joint Institute. Without their recommendations and thorough training, my research progress would otherwise be much less smooth. I had a fruitful discussion with Prof. Adrian Spencer from Loughborough University during the viva, and his insights in modifying the thesis are much appreciated. Last but not least, I would like to thank my family and friends for their endless support in all aspects.

Abstract

In-cylinder air motion has a profound influence on the mixture preparation and subsequent combustion in the internal combustion engines. The particle image velocimetry (PIV) technique provides a quantitative tool to measure the in-cylinder air velocity with high spatial and temporal resolution. Additionally, the in-cylinder flow velocity field can also be simulated by computational fluid dynamics (CFD) tools, of which the Reynolds averaged Navier-Stokes (RANS) method predicts the overall flow structures with acceptable accuracy and affordable computational cost for widespread industrial applications. In order to provide a more accurate and robust prediction, the simulation results need to be validated against the experimental measurements. In this thesis, the crank-angle resolved in-cylinder air flow fields were measured using the PIV technique in multiple planes of a near-production optical engine under different working conditions for a few hundreds of consecutive cycles. The measured flow fields show a high consistency over multiple measurement planes and across different test runs, and hence they can be used to validate the flow simulation model. Vector field comparison metrics were used to compare the measured and the simulated flow fields, and to identify the regions of interest and differences in flow characteristics. Due to the nature of averaging physical parameters in RANS, its validation against experimental results obtained by particle image velocimetry (PIV) requires consideration of how best to average or filter the measured turbulent flows. During the validation process, it was found that the use of the conventional PIV ensemble mean flow field may lead to biased conclusions in some cases. Therefore, two dimensionality reduction techniques, namely the proper orthogonal decomposition (POD) and the kernel principal component analysis (KPCA), were used to preserve the large-scale cycle-to-cycle variations while filtering out the small-scale turbulent fluctuations in experimental measurements, and hence enabled a fair comparison to the RANS simulations. Recommendations are made on how to provide an appropriate validation between measured data and simulation results in highly fluctuating flow fields such as the in-cylinder air motion.

Contents

List of Figures

List of Tables

Chapter 1

Introduction and literature review

In this chapter, the background and motivation of the work in this thesis are introduced. A literature review is provided on different aspects that are related to the work presented in this thesis, including a brief history of optically accessible engines, an overview of optical diagnostic techniques used in engine research (with a specific focus on the particle image velocimetry technique), and an introduction to two dimensionality reduction techniques (proper orthogonal decomposition and kernel principal component analysis). Several parts of this chapter are adapted from the introduction sections of three papers [1–3] that are first-authored (or co-first-authored) by the author of this thesis with additional information and more detailed discussions included.

1.1 Background and motivation

Internal combustion engines (ICE) continue to serve as the main source of power for ground transportation worldwide. A credible study from the International Energy Agency [4] shows that vehicles with an ICE (including both conventional ICE and hybrid vehicles) retained over 90% of the worldwide market share in 2018. Predictions show that the market share of battery electric vehicles (BEVs, without an ICE) may only reach $11 - 28\%$ by 2040, and the increasing numbers of BEVs may lead to instabilities in the power grid [5]. The BEVs do not have tailpipe pollution, however the greenhouse gas (GHG) emitted in the production process is almost doubled compared to that for conventional ICE cars. Their life-cycle GHG emissions highly depends on the source of the electricity and the lifetime of the vehicle [6]. When considering the local air quality (LAQ) benefits, despite the BEVs trim the particulate matter (PM) from the tailpipe, their overall PM pollution may not show a significant reduction either — the BEVs produce more non-exhaust emissions from tire and brake wear than the conventional ICE vehicles due to their higher weight [7]. Therefore, the result of whether the BEVs are more environmental friendly is not yet concluded [8]. On the other hand, hybrid vehicles have been found to be a viable alternative to BEVs in terms of sustainability [9], particularly when the life-cycle GHG emissions are considered [6, 10]. The trend towards hybridisation continues to motivate the study of physical phenomena occurring within the engine cylinder, in order to design more efficient ICEs for future hybrid vehicles that can meet the ever increasingly demanding emissions legislation while at the same time seeking better fuel economy [11].

The gasoline direct fuel injection (GDI) technique, compared to the conventional port fuel injection (PFI), offers a better control to the amount of fuel entering the cylinder, allows for a higher compression ratio, and suffers lower pumping losses [12]. Due to its fuel-economy advantage, GDI is well-adopted by various engine designs. The fuel is injected directly into the cylinder, and it needs to form a homogeneous mixture with the air before the combustion event

to avoid additional emissions caused by unburned or partially burned gases. The homogeneity of the air-fuel mixture, and thus the combustion and emissions characteristics of GDI engines rely greatly on the intake air speed, in-cylinder flow structure and turbulence level [13, 14]. As such, it is important to investigate the in-cylinder air motion using experimental tools and simulations.

Various in-cylinder flow velocity measurement techniques have been developed over the years, and the particle image velocimetry (PIV) technique has become the most widely used for engine in-cylinder flow research since it provides high spatial and high temporal resolution measurements of the rapidly-changing flow [15, 16]. The state-of-the-art PIV developments enable crank-angle resolved in-cylinder flow measurements for a few hundred consecutive cycles with a high spatial resolution (usually millimetre in scale for bulk flow motion).

The development of flow simulation models using computational fluid dynamics (CFD) tools such as the Reynolds averaged Navier-Stokes (RANS) and the large eddy simulation (LES) approaches [17–20] have been accelerating engine design and helping reduce the number of engine tests. Both RANS and LES simulate the large structures of the in-cylinder flow, and LES provides an additional access to the cycle-to-cycle variation (CCV) in the engine [21]. However, the computational cost of LES is considerably higher than RANS, and therefore RANS is still widely used in industrial applications where the time for the engine research & development cycle is critical and limited.

The in-cylinder flow simulation model needs to be validated by experimental measurements to provide more accurate predictions of the highly fluctuating and reacting flow [22]. Although provided as a good predictive tool for global characteristics, it is well known to the modelling community that RANS-based models are unable to capture the stochastic nature of the turbulent flow. Significant progress has been made over the years in RANS models to include the unsteady behaviour of turbulent flows. These techniques vary from adding additional unsteady terms [23] to re-normalised Navier-Stokes equations [24] where small scale perturbations can be included. Despite the improvements, RANS can only provide a single flow

field at a specific crank angle, while the PIV data are collected over multiple cycles. Therefore, it is of interest how to properly compare multi-cycle PIV measurements with a single flow field from RANS-based models, especially when the measured quantities present a strong cycle-to-cycle variation (or more generally, shot-to-shot variation in measurements).

One common approach is averaging the physical quantities that were measured in a controlled test condition to obtain the so-called "ensemble mean",[(1)] and use the experimental ensemble mean as a validation target for simulation models. The measured quantities that are averaged include but are not limited to the flow fields, the temperature fields, and species fields (such as hydroxyl radicals). Yang *et al.* [26] performed a comparative engine cold flow analysis between a RANS flow field and a cycle-averaged LES simulation result using the PIV ensemble mean field as the reference data. RANS simulations were found to predict reasonably good qualitative trends for the mean flow and turbulence, but less accurate descriptions of the mean flow structures and turbulence distribution are found when compared with ensemble-averaged LES and PIV data. Earlier studies from Wu *et al.* [27] used RANS-based combustion models to predict the OH radical fields of a non-premixed methane jet during a combustion event. A poor agreement was found between the predicted RANS OH radical and the converged experimental ensemble average. A higher degree of similarity is, however, seen between the results of the RANS simulation and some of the single shot OH-PLIF images than the similarity between the simulations and the ensemble-averaged OH-PLIF results. From a simulation point of view, one would expect a Reynolds averaged field to have a fairly gradual rise in temperature over a broad area of the flow. However, further statistical analysis suggested that the mean convective velocity term contributes to a rapid rise in the temperature terms, whereas an ensemble-averaged field from the experiment is smoothed both in time and space. The study also suggested that more than 100 realisations of LES would be necessary to enable the

[(1)] In engine research, the ensemble mean typically refers to measured quantities at a fixed crank angle over multiple cycles, assuming the engine operation is a cyclostationary process [25].

comparison with the experimental ensemble-averaged results. Studies have also been carried out by Stein *et al.* [28] comparing LES simulations of in-nozzle flow in turbulent opposed jets with PIV experiments. With a novel statistical method, a comparison of metrics derived from individual cycles of LES and PIV are made possible. However, such LES simulations would still be extremely costly for in-cylinder simulations and hence are not in favour for industrial use due to the demand of shortening the complex engine design process. Recent spray bomb studies on both RANS and LES simulations, using the same combustion model [29, 30], also indicated that the mean flow field from RANS presents higher magnitudes of scalar quantities compared to both LES and experimental ensemble-averaged quantities. The maximum discrepancies between the RANS simulation and LES/experimental ensemble-averaged temperature field are observed for the area of the developed flow where turbulent eddies have a significant effect on flow characteristics. Therefore, when quantitatively comparing in-cylinder experimental ensemble-averaged data with a RANS simulation, methods other than the ensemble-averaging need to be considered.

1.2 Thesis overview

In this thesis, PIV flow fields were measured in multiple planes under various working conditions in a single-cylinder optically accessible engine that is based on the Jaguar Land Rover (JLR) AJ200 gasoline direct injection engine. The PIV data were used to validate the RANS flow model from a commercially available software used by the JLR research and development team. The following sections provide a literature review of various topics that are related to the work in the thesis. The data acquisition apparatus and the data analysis approaches are presented in Chapter 2 and Chapter 3, respectively. Multiple methods were implemented on the measured flow fields to enable comparisons with the simulated data, and their performances on the flow model validation are discussed in the corresponding chapters.

These methods include the ensemble-averaging in Chapters 4–6, the proper orthogonal decomposition in Chapters 5 and 6 and the kernel principal component analysis in Chapter 6. Summary of the findings using each method is provided at the end of each chapter. The key findings and learning points of this thesis are highlighted in Chapter 7.

1.3 Optically accessible engines

The visualisation of the complex flow development in engine cylinder, including the air charge motion, the fuel-air mixing and the combustion processes, have been of interest and have been essential since the invention of the internal combustion engine [31]. Optically accessible engines, or simply "optical engines", have been correcting misunderstandings and initialising new findings ever since the era of Otto, the father of compressed charge internal combustion engines, who was said to hold the false belief that his engine's charge motion was stratified [32, 33], while later on Clerk and Hopkinson [34] built a full-glass model of the Otto cylinder and saw that "the (air-fuel) mixture becomes practically homogeneous even before compression commences". Early notable optical engine studies can be traced back to the 1920–1930s. Ricardo and Glyde [35, 36] measured the flame propagation speed of a side-valve IC engine by installing a disk on the cylinder head with six equally-spaced windows along the radial direction. A slotted disk rotated with the camshaft, serving as a "stroboscope" while combustion occured inside the chamber, and the flame location was determined by counting the numbers of illuminated windows at different phases. It was through the optical access that they confirmed their previous finding in aircraft engines — that detonation is a phenomenon near the end of flame travel. Rassweiler and Withrow [37] clamped a quartz plate onto the cylinder head of a single-cylinder side-valve engine, and with a special motion picture camera, in-cylinder flame imaging within a continuous engine cycle was realised for the very first time. Despite the limitation of the image quality, their design had remarkable time resolution — 5000

frames were recorded every second, corresponding to 2.4 crank angle degree at 2000 revolution per minute (rpm) engine speed.

Various optical engines with different access orientations have been designed since then, with the goal of maximising optical access while retaining sufficient similarity to the real engine [22]. Due to the overhead valve design of modern engines, however, most the four-stroke engines no longer achieve their optical accesses from the top. Rather, the Bowditch design (Figure 1.1) [38], with an extended slotted piston, gives a view of the combustion chamber from below. A slotted tubular extension is used to connect a baseline piston (which links to the crankshaft) with the experiment target piston (which moves inside the combustion cylinder). The latter has a transparent window insert, and by placing a $45°$ mirror through the slotted extension and beneath the piston window, the entire cylinder can be observed from the bottom without installing extra devices in the crowded cylinder head. This setup is now adopted by lots of modern optical engines, and with a transparent liner made in quartz, optical access can be achieved both from the side and from the bottom (through the piston window), allowing one of them to be used as the entry/exit for laser beams or light sheets and the other as the viewing accesses for cameras.

1.4 Optical diagnostic techniques in engine research and particle image velocimetry

Early studies on the visualisation of engine in-cylinder behaviour used natural luminosity from the flame generated by the combustion as a light source. Since the 1980s, the accessibility of high power lasers, together with the development of high speed imaging detectors, have allowed a much wider range of optical diagnostic techniques to be implemented in engine research to image and study in-cylinder flow and combustion [15, 39]. These techniques provide an assortment of in-cylinder information with high spatial and time resolutions [15, 22,

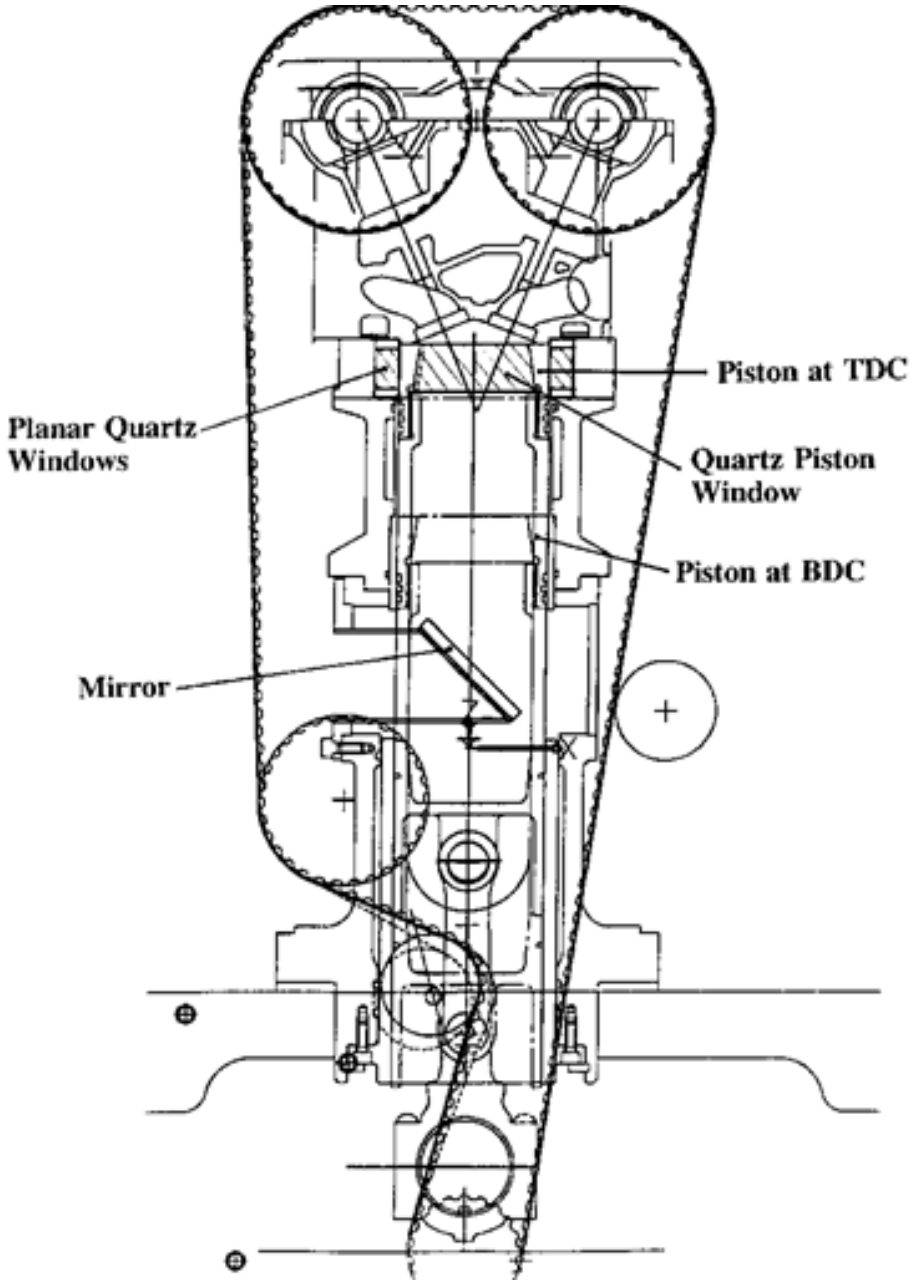


Figure 1.1: A schematic of the Bowditch design [38]. A hollow extension rod connects two pistons, through which a 45° mirror is placed to provide a view of the combustion chamber from below. The upper piston moves inside the combustion cylinder, while the lower piston links to the crankshaft with no cylinder liner installed. This figure was kindly provided by Prof. Richard Stone.

39]. The bulk air motion can be measured by laser doppler anemometry/velocimetry (LDA/LDV) and particle imaging velocimetry (PIV). The fuel spray can be directly imaged by line-of-sight measurements such as shadowgraphy and schlieren photography, or via planar laser induced fluorescence (PLIF) which gives the spatial distribution of the spray on a plane. The laser induced grating spectrography (LIGS) provides accurate point gas temperature measurements [40], and can be used to perform in-situ calibration of the 2-D temperature data obtained by two-colour PLIF (TC-PLIF) [41, 42]. PLIF of hydroxyl radicals (OH-PLIF) is widely used to determine high-temperature reaction zones during combustion. The combustion products, especially those harmful ones such as soot, can be analysed by laser induced incandescence (LII). In the case of the in-cylinder flow field measurement, the PIV technique is discussed in detail in the next several paragraphs.

Particle image velocimetry (PIV) refers to a set of velocity measurement techniques that record the spatial location changes of tracer (seeding) particles inside fluids between different images (i.e., a period of time) [43, 44]. The particles are tiny, usually in the scale of sub-micron in diameter, and assumed to be moving together with the surrounding flow, so their motion reflects the velocity of the neighbouring fluid. Unlike most velocity measurements which give temporal velocity change at a single point, the PIV technique provides velocity measurements over a field (usually a plane) with both high spatial and temporal resolution, and it is, on large scale, non-intrusive so does not affect the measured flow.

The PIV technique has applications in a remarkably wide range of systems – from massive rocket shear layer in supersonic external flows [45], to red blood cells sucked by a female mosquito through its micron-scale proboscis [46]. In engine research, PIV has been used to visualise a variety of in-cylinder flow behaviour – from the bulk air flow during the charging process [47], the tumbling/swirling motion induced by the intake flow [48–50], the large-scale flow structure changes due to asymmetric valve strategy [51] and the flow-spray interactions [52], to smaller-scale structures such as the velocity boundary layer near the

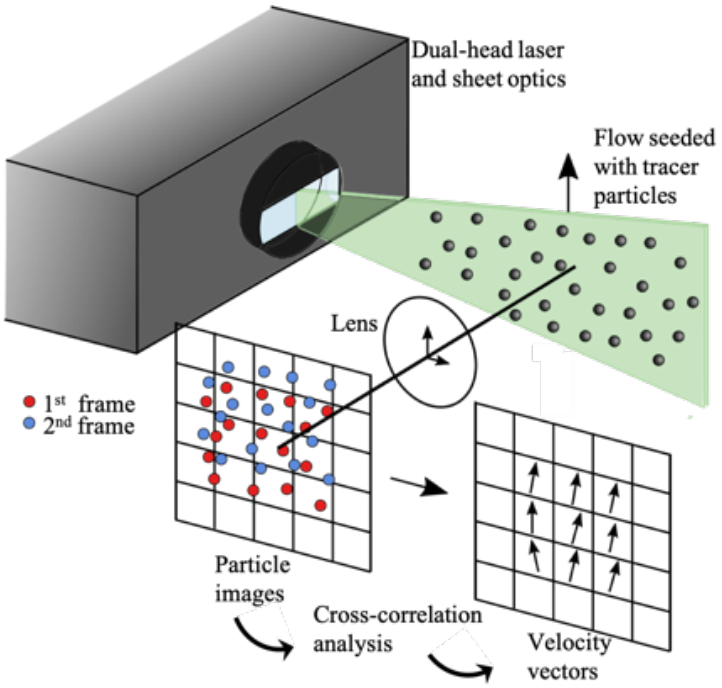


Figure 1.2: A schematic of the planar particle image velocimetry technique. The grids on the frames show interrogation windows, each of which generates one vector. This figure was kindly provided by Dr. Blane Scott [55].

cylinder head [53] and flow fields close to the spark plug [54].

The speed of the engine flow can reach nearly 100 m/s, and the flow direction changes rapidly within a sub-millisecond period. The images thus have to be recorded in a short duration of time, resulting in a short camera exposure time. Therefore, sufficient illumination is required in order to image the particles and find their locations precisely. This is usually achieved by using high power pulsed lasers as the light source due to their high temporal energy density — tens of millijoules of laser energy is concentrated in less than a microsecond. In the conventional planar PIV setup (Figure 1.2), optics (usually a set of cylindrical and spherical lens) turn the laser beam into a light sheet, which illuminates a plane in the engine cylinder. The scattered light from the seeding particles is then captured by a high speed camera, and the velocities of the particles (and therefore the surrounding air flow) are calculated by cross-correlating the particle pattern

change within a small interrogation window (usually 32×32 pixels [56, 57]) between a pair of images that are taken sequentially in several to tens of microseconds. It should be noted that the PIV technique looks for patterns of tens of particles in the interrogation window, as opposed to a similar technique called particle tracking velocimetry (PTV) — the latter tracks individual particles in the flow, providing more robust yet too sparse velocity information.

The pair of images for generating the PIV vectors ought to be taken within a sufficiently small period of time so that the seeding particles stay in the interrogation window. Single-cavity pulsed lasers usually are not fast enough to generate two pulses that illuminate both frames (again separated by only several to tens of microseconds), and those with extremely high repetition rate (higher than 100 kHz) are costly and may not be able to provide adequate illumination (enough pulse energy) in each frame. Therefore, a dual-cavity pulsed laser becomes the preferable and affordable choice as the two cavities can fire independently, providing individual illumination to each frame and allowing flexible frame separation time. The two pulses are aligned such that they follow the same beam path before they leave the laser head, in order to accurately record the in-plane motions of the particles [44].

1.5 Proper orthogonal decomposition and kernel principal component analysis

Principal component analysis (PCA) is a mathematical tool that is based on singular value decomposition (SVD), and it can be used in fluid mechanics to decompose an ensemble of velocity field data into spatial-temporal modes. Based on the concept of PCA, proper orthogonal decomposition (POD) proposed by Lumley [58] has been extensively used to analyze in-cylinder PIV velocity field data [21, 59–62]. POD considers each of the velocity components measured at the same spatial location as a random variable and therefore treats the entire flow field as a multivariate data set. The result of POD is a series of ranked flow

patterns (known as POD modes) which are computed by successively maximising the sample variance within the ensemble and equivalently, the total amount of kinetic energy distributed into each mode. Based on this successive kinetic energy maximisation property, the first several POD modes collectively account for the majority of the high energy flow features within the data set, and therefore the combination of them is regarded as the coherent structure of the ensemble [63, 64]. The higher order POD modes, which are discarded due to their small kinetic energy fractions, are often associated with random Gaussian fluctuations [65], smaller-scale turbulence [66, 67], or measurement noise [59, 68]. As such, it is possible to implement POD as a dimensionality reduction technique to reduce the flow complexity and as a low-order approximation tool to extract the dominant structure of the in-cylinder turbulent flow measured in each cycle (at the same crank angle), and use the POD-reconstructed flow fields to validate the simulation results.

Despite the wide use of POD analysis on in-cylinder flow field studies, one of the fundamental challenges within the method is to yield a clean separation between high variance and low variance modes as no obvious scale separation exists in engine flows. Through a direct numerical simulation (DNS) study, Ma *et al.* [69] suggested that higher order modes obtained from POD of PIV velocity fields are dominated by experimental noise which makes it essential to separate them from lower order modes. However, when using any SVD-based method in a system where many degrees of freedom exist (i.e. turbulent flows), with translation invariant interactions, the variance of each mode decreases monotonically but smoothly with an increase of the mode number; therefore any sharp division between important and unimportant dimensions would be arbitrary [70]. To demonstrate this effect, when comparing LES simulations with PIV experiments, Liu *et al.* [71] found that the number of engine cycles required to extract converged POD modes varied with mode number and phase (i.e., crank angle), making it difficult to set a universal cut-off mode number.

Various studies have since used different criteria to separate low-order and high-order

modes generated by SVD-based methods. Roudnitzky *et al.* [65] proposed a triple POD decomposition algorithm and the flow fields were separted into the mean component, the coherent structures and the random Gaussian fluctuation. Qin *et al.* [66] and Zhuang and Hung [67] suggested dividing POD modes into four quadrants (the dominant structure, the coherent structure, the turbulence structure and the noise structure) based on the flow structure collected by the summation of sequential POD components. Buhl *et al.* [72] proposed using POD-based conditional averaging to identify the large-scale structure fluctuations in IC engines. Butcher and Spencer [73] applied a cross-correlation to the POD modes that are respectively derived from two halves of a data set, and used a cut-off threshold to separate the coherent structures and turbulent fields. Epps and Techet [68] proposed a threshold criterion based on the root mean square error of PIV data that can be used as a rough limit to separate high order modes. Modes with singular values greater than the criteria are found to be less affected by noise. The error was also found to be associated with a random Gaussian distribution. Epps and Krivitzky [59, 74] further studied the effectiveness of SVD in filtering out the noise and reconstructing the data. These studies explored the deeper applicability of SVD-based methods in fluid mechanics. Despite these methods are useful tools for researchers to perform qualitative and quantitative descriptions of in-cylinder flow fields, they are still being criticised due to a certain lack of objectiveness in their application [60]. It should be noted that, when there exists a clear separation of high and low singular values in the data set, the SVD-based method can ensure a good noise filtering with an error threshold. However, when the singular values are continuous (which is likely to be the case for the in-cylinder flow fields due to small sample sizes), any sharp division between the important and unimportant dimensions would be arbitrary [70], and hence the number of modes used for flow field reconstruction (i.e., the "cut-off mode number") may be critical to improve the original noisy data. Therefore it is worth exploring additional numerical techniques that are applicable to the PIV data analysis, while allow a less ambiguous separation in modes to be achieved.

SVD-based models perform an orthogonal basis transformation on data with an inherent assumption that linearity frames as a change of basis [75]. In the case of in-cylinder flow fields, however, there may exist some non-linear correlations between the flow fields since they are highly turbulent. This may be the reason why a clear cut-off POD mode number is difficult to achieve when the POD technique[(2)] is applied to a non-linear flow data set. Therefore, a non-linear manifold reduction technique may need to be considered when extracting coherent structures in the data. Kernel principal component analysis (KPCA) [76] is a generalisation of the POD (or the linear PCA) technique, and has been widely used in classification, feature extraction and de-noising applications [77, 78]. The main idea for KPCA is based on Cover's theorem, where the non-linear data structure in the input space is more likely to be linear after a high-dimensional non-linear mapping [79]. Studies have found that when non-linear structures exist in the original data, KPCA would be significantly better in de-noising or separating dominating features from the original data [77]. Although widely used in other fields for data analysis, only a limited application has been found in fluid mechanics. Mirgolbabaei and Echekki [80] reconstructed the thermo-chemical scalars based on KPCA methods, and the results showed a good match with original data of species and temperature with the use of only two principal components. A further study of a turbulent combustion closure model using experimental non-linear principal component reconstruction, showed good agreement for the mean and root mean square (RMS) of temperature and measured species mass fractions [81]. The successful implementation of KPCA in the complex thermo-chemical data inspired the idea that it is worth investigating whether this technique can also perform de-noising and reconstruction to the highly-turbulent in-cylinder PIV flow data, of which the conclusion is unknown to the thermo-fluid community before the thesis was written.

[(2)] POD is a variant of linear PCA when applied to thermo-fluid analyses.

Chapter 2

Data acquisition apparatus and methods

In this chapter, the apparatus and methods used to obtain in-cylinder flow data presented in this thesis are introduced, including the setups for the PIV experiments and the flow simulation model. The PIV supplementary systems and the setups for the multi-plane PIV measurements are firstly presented. Velocity measurements were accomplished at a variety of specified test points to allow comparison between PIV measurements and simulation results from commercialised computational fluid dynamics (CFD) software. A brief description on the flow simulation model is also included. The setup of the single cylinder optically accessible engine on which the tests were performed can be referred to in Appendix A. It should be noted that the PIV optical setups and tests were conducted by the author of this thesis, while the optical engine was assembled by previous group members (see Appendix A for more details), and the flow simulations were performed by engineers from Jaguar Land Rover.

2.1 PIV supplementary systems

The transparent liner and piston window (details in Appendix A) provide optical access from the side and bottom of the cylinder, yet several other supplementary systems are also required to accomplish the PIV measurements. In this section, the PIV particle seeding system is first introduced, followed by descriptions of the laser system that illuminates the cylinder and the high-speed camera that records the pairs of images for cross-correlation. The synchronisation system is also presented.

2.1.1 Particle seeding system

Seeding particles are introduced to the intake air as tracers that scatter the light from the illumination source, and are recorded by imaging system, to provide local velocity of the surrounding gases. The seeding system in our experiment (Figure 2.1) involves a pressure regulator, a mass flow controller (MKS mass flow controller), two solenoid valves and an atomiser (LaVision aerosol generator). Compressed air at a fixed pressure of 4 bar is supplied to the controller, which regulates the exiting mass flow rate of the seeding air and thus the seeding density inside the engine cylinder. The two solenoid valves separate the circuit into two branches: when valve A opens, the compressed air bypasses the atomiser and directly enters the intake plenum downstream of the throttle without any seeding particles; while valve B enables the atomiser, providing oil droplets first to the intake plenum, and then join the mainstream air into the cylinder. It should be noted that the bypass design is necessary, otherwise the sudden introduction of the seeding air into the intake plenum would affect the main flow rate, and may further cause in-cylinder pressure oscillation [55]. During the experiment, the seeding air (without particles) is supplied together with main air flow, and once the engine reaches the desired test condition, a switch closes valve A and opens valve B, allowing the seeding particles to enter the cylinder.

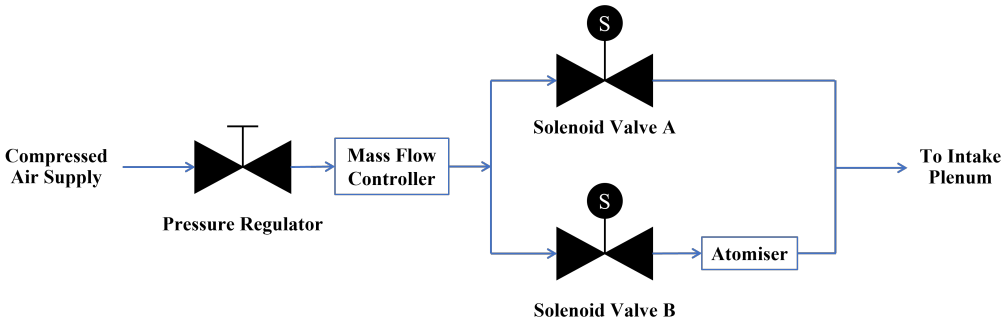


Figure 2.1: A schematic of the particle seeding system for the PIV experiment.

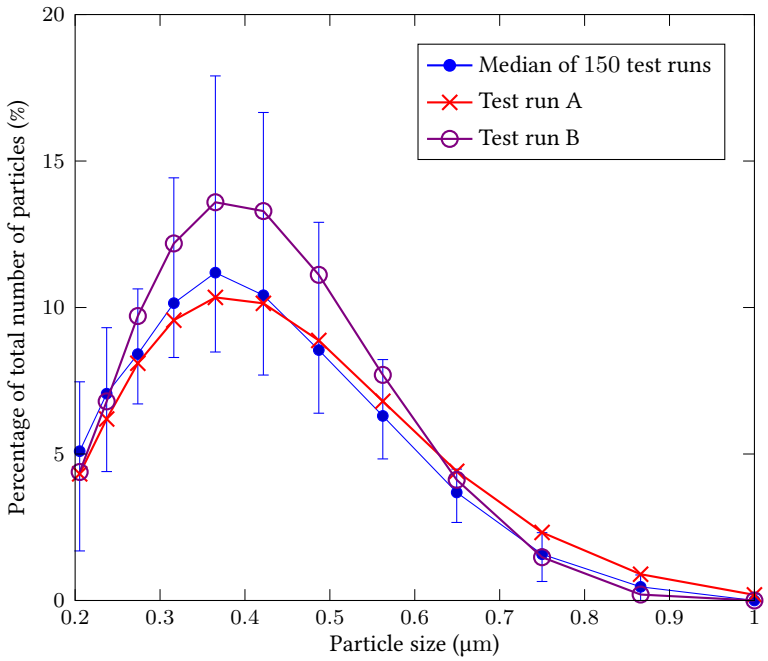


Figure 2.2: Seeding particle size distribution from the LaVision aerosol generator measured by a Cambustion DMS500 fast particulate spectrometer. The blue dots show the median values of 150 test runs, two of which are respectively shown in red crosses and purple circles. The blue error bars show the 25% (lower bound) and 75% (upper bound) quantiles of the 150 test runs, respectively.

The particles should be small enough to follow the flow without influencing it (i.e., having a small enough Stokes number), yet sufficiently large to scatter the light and provide a strong signal. In engine operation, they should also be easy to clean so that the components would not be damaged, and also have a high enough evaporation point since the in-cylinder temperature can easily exceed $110°C$ at late-compression stroke even without combustion [42]. Adrian and Westerweel [44] suggest that particles with diameters of about a micron are a suitable choice for in-cylinder flow measurements. Olive oil was used as the seeding source in all the PIV experiments presented in this thesis. Figure 2.2 shows the size distribution of the oil droplets generated from the atomiser. The particle size distribution was measured by a particulate analyser (Cambustion DMS500) under the same seeding air pressure and mass flow rate as used in the experiments. The measured size distribution also matches with the specifications from the atomiser manufacturer.

2.1.2 Illumination and image recording systems

The crank angle-resolution planar PIV measurement requires sufficient (in the range of $5 - 500$ mJ per pulse) and stable illumination of the sub-micron-sized seeding particles. The light needs to be concentrated within a short period of time (in the scale of tenths of a microsecond) and with a high repetition rate (in the order of kHz). Solid-state pulsed lasers with Neodymium-doped crystals are suitable as they provide easy-collimated and high-energy-density beam at high repetition rates (commonly up to 10 kHz) [44]. Frequency-doubling crystals are installed in the laser cavity to perform a second-harmonic generation process [82], and to convert the fundamental infrared emission from the Neodymium-doped crystals (1053 nm for Nd:YLF[(1)] and 1064 nm for Nd:YAG[(2)]) into visible light wavelengths (respectively at 526.5 nm and 532 nm, both green), which then allow the scattered light from the seeding particles to be captured by

[(1)] Nd:YLF stands for neodymium-doped yttrium lithium fluoride.
[(2)] Nd:YAG stands for neodymium-doped yttrium aluminum garnet.

Table 2.1: Nd:YLF laser specifications

Parameters	Descriptions
Wavelength	527 nm
Pulse energy	40 mJ @ 1 kHz
Pulse width	170 ns @ 1 kHz
Repetition rate	up to 8 kHz
Beam diameter	5.0 mm nominal

normal camera. The light source of the PIV experiment was a diode-pumped, dual-cavity Nd:YLF laser (Photonics Industries DM20-527-DH), whose specification is listed in Table 2.1.

A high-speed camera is used to record the scattered light from the seeding particles. CMOS[3] sensor cameras have become the norm in PIV imaging [43]. Unlike CCD[4] sensors which read out pixels on a row-by-row basis, each pixel is individually read out in a CMOS sensor and hence one can trade image resolution in both row and column directions for a higher frame rate. This feature is very useful when high time resolution is required to cooperate with high-repetition-rate lasers. The reduction of numbers of pixels in each image can have a negligible effect when the imaged object does not follow the aspect ratio of the sensor, since in this case some pixels would otherwise inevitably be wasted. Table 2.2 presents the specification of the CMOS camera (Vision Research Phantom VEO 710L) that was used in all the PIV measurements in this thesis.

2.1.3 Synchronisation system

The laser and camera need to be synchronised with the crankshaft in order to image the in-cylinder flow at the desired engine phase. This is achieved using the DaVis® (LaVision, Ver. 8.4.0) software with a programmable timing unit (PTU). The PTU first receives a "kick-off" TTL pulse

[3] CMOS stands for complementary metal oxide semiconductor.
[4] CCD stands for charge coupled devices.

Table 2.2: Camera specifications

Parameters	Descriptions
Sensor resolution	1280×800 pixels
Frame rate	7400 fps at full resolution
Pixel size	$20 \times 20\,\mu m$
RAM size	18 GB
Maximum number of images	12025 frames at full resolution

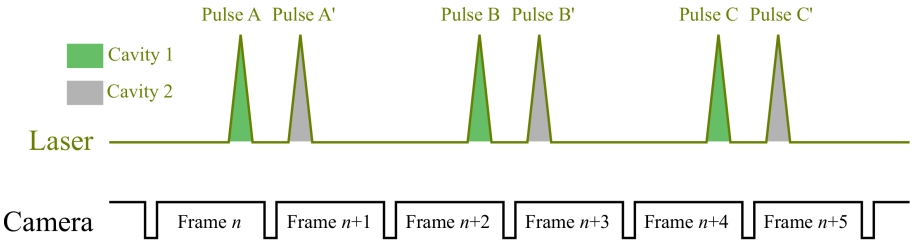


Figure 2.3: Timing diagram of the laser and the camera (not to scale).

from an in-house LabView program (called "ETCS", see Appendix A for detail) and starts reading the crank encoder to check the engine phase. The laser and camera are then controlled by the PTU to fire and to record images at specified crank angles. The images are transferred from camera memory to DaVis for PIV processing after the test. The system is operated in "double-frame" mode to provide pairs of images for cross-correlation, and therefore the laser and the camera need to be carefully timed. One widely-used method is the frame straddling approach [83, 84]. The timing diagram, as presented in Figure 2.3, shows that the first illumination pulse A (from laser cavity 1) is captured at the end of the camera exposure n, where the second pulse A′ (from laser cavity 2) is at the beginning of the next camera exposure $n + 1$ — same procedure continues for frames $n + 2$ and $n + 3$. The resulting camera frame rate is thus twice the laser repetition rate (for each cavity). The separation time between pulses from each of the two laser cavities (for instance, A and A′) has to change according to the measured flow speed to maintain

a suitable particle moving distance between two camera frames (usually less than a quarter of the size of the interrogation window [43, 44]).

2.2 Optical setup for multi-plane PIV measurements

The laser beam needs to form a light sheet in order to illuminate a certain plane in the engine for planar velocity measurements. In our experiment, data from multiple planes are necessary to validate the simulation results. In this section, the optical alignment setups for two cross-sections of the engine cylinder are discussed.

2.2.1 Swirl plane

The swirl plane is normal to the cylinder axis, of which the height (along the engine stroke) is yet to be determined. The setup needs to have flexibility in height adjustment to allow measurements in multiple swirl planes for better examination of flow structures and thorough validation of the simulation model. Figure 2.4 illustrates the swirl plane PIV setup in a top view (from the cylinder head) and in a front view (from the camshaft drive belt).

The beams from the two laser cavities have the same wavelength, but their polarisation states are orthogonal to each other — one is vertically polarised, and the other is horizontally polarised. Despite the careful alignment, the resulting light sheet would inevitably hit the metal piston surface and cause different background scatter between the two frames in the same PIV image pair, as the reflectance of metal surfaces varies with the incoming polarisation state. To avoid different intensities between two frames and errors when calculating vectors, a quarter wave plate is placed at the exit of the laser head with its fast axis aligned at 45° to the vertical and horizontal polarisations of the light from the two cavities. For the laser beam from each laser cavity, the plate introduces a 90° phase shift between the components aligned with the fast and slow axes, resulting in left circular and right circular polarisations. Both circularly polarised

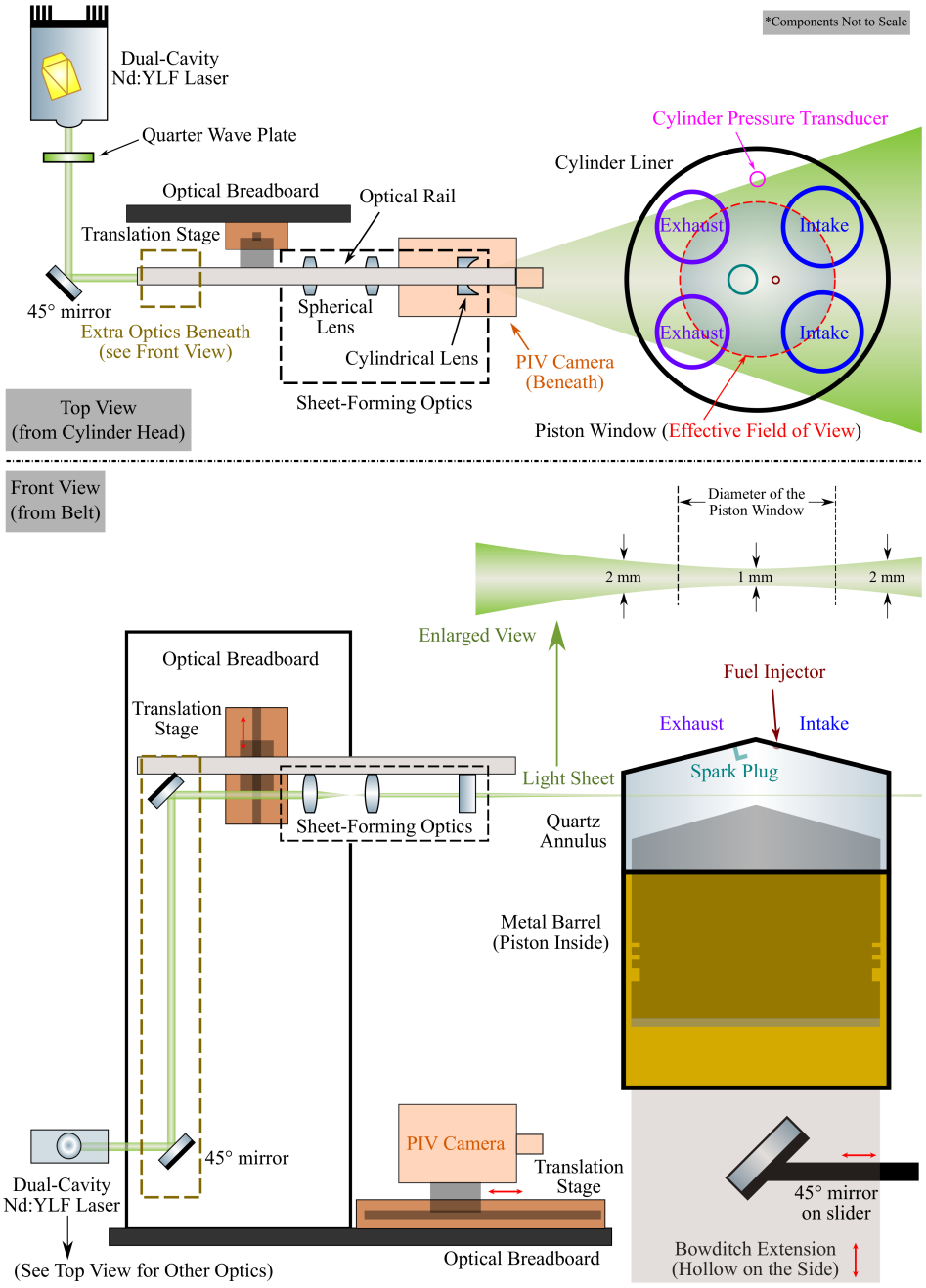


Figure 2.4: Swirl plane PIV setup. Upper: top view (from the cylinder head); Lower: front view (from the belt). An enlarged front view of the light sheet across the quartz annulus is also shown.

beams have the same magnitude of vertical and horizontal components, and therefore they would have the same reflectance from metal surfaces.

Three $45°$ mirrors are used to align the beam with the cylinder centreline. The upper $45°$ mirror, together with the set of sheet-forming optics (detailed discussion in the next paragraph), is mounted on an optical rail to ensure the correct alignment. Two translation stages (Zaber T-LSR150B) respectively move the entire rail to any desired height and adjust the position of the camera to make the image in focus at the location of the light sheet, so that the velocities in different swirl planes can be measured.

The sheet-forming optics set consists of two spherical lenses that shrink the beam diameter from around 20 mm down to 10 mm,[(5)] and a concave cylindrical lens which expands the beam in one direction and forms the light sheet (whose cross-section is a very flat ellipse). It should be noted that convex cylindrical lenses can also be used to form a light sheet, but it will introduce a focal line with extreme high energy that heats up the surrounding air and may cause damage to the neighbouring devices [43]. Therefore, a concave lens ought to be used for safety reasons, as its focal point is before the lens (and thus the sheet will not reach a focus).

The light sheet would not stay collimated from the optics to the far-side of the quartz annulus due to its natural divergence, and the sheet thickness would thus increase with the distance of travel. To minimise the effect of the divergence, the two spherical lenses are slightly distanced so that the sheet thickness first decreases and then increases. The sheet-forming optics system is placed such that the resulting light sheet reaches its thinnest at the cylinder centre axis with a 1 mm nominal thickness,[(6)] and the sheet is about 2 mm thick at both sides of the quartz annulus. An enlarged front view of the light sheet across the annulus is provided in Figure 2.4, where the dashed lines show the light sheet thickness across the piston window (PIV field of view) has small but acceptable variation.

[(5)] The beam is 5 mm in diameter at the exit of the laser head, but naturally diverges as it travels.
[(6)] The laser beam has a Gaussian profile [44].

Another 45° mirror is placed beneath the piston window by mounting it onto a slider, which slides through the hollow Bowditch extension (side view of the extension was shown in Figure A.5). The height of the slider is carefully calculated so that the fast-moving piston assembly would not hit it during engine operation. The mirror gives the bottom view of the cylinder, which allows the camera to capture particle movement in the illuminated swirl planes. The camera is also mounted on a translation stage to adjust the focal plane location according to the height of the light sheet. The effective PIV field of view is thus a circle with the diameter of the piston window.

2.2.2 Cross-Tumble plane

In addition to the planar velocity in the swirl planes, the flow behaviour in the rectangular cross-sections of the cylinder, namely the tumble and cross-tumble planes, need to be examined as well. Measurements in the tumble plane[7] were accomplished by Scott [55]. The cross-tumble plane PIV setup is presented in Figure 2.5.

In the cross-tumble planes, the intake and exhaust valves are located on each side of the measurement plane. The setup is similar to the swirl plane one, but with small changes due to different planes of measurement. Chief among them is that the camera and sheet-forming optics are swapped in location — the light sheet now enters the cylinder from beneath (through the piston window) and the camera is placed at a higher location viewing from the side (through the quartz annulus) to capture the scattered light from the seeding droplets.

As illustrated in the front view (bottom plot in Figure 2.5), the 45° mirror mounted on the slider is now used to turn the light sheet propagation direction from horizontal (exhaust to intake) to vertical (piston to cylinder head). Measurements in different cross-tumble planes can

[7] There are one intake valve and one exhaust valve located on each side of the tumble plane. The plane has a 1 mm offset to the flywheel side to reduce laser light scatter from the cylinder head. The tumble plane tests were conducted before the author joined the group, and therefore they should not be considered as part of the author's work.

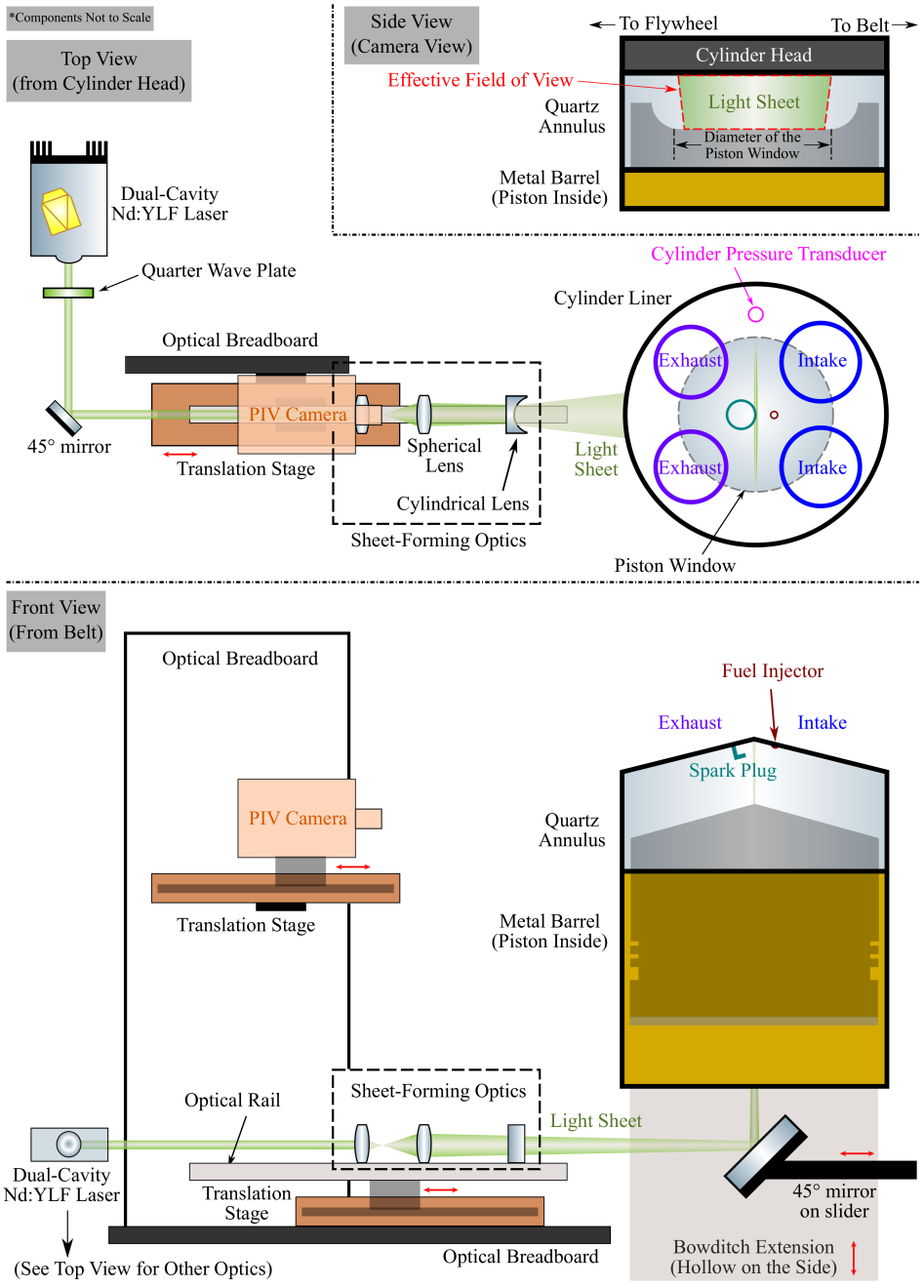


Figure 2.5: Cross-Tumble plane PIV setup. From top to bottom: side view (from the PIV camera); top view (from the cylinder head); front view (from the belt).

be achieved by moving the slider to illuminate various planes, and adjusting the camera as well as sheet-forming optics positions accordingly using translation stages. The resulting effective field of view (see the camera view) is a trapezoid whose bottom edge is slightly smaller than the piston window diameter, in order to avoid strong scatter caused by the metal retaining ring beneath the piston window.

2.3 Engine test points and experiments conducted

The setups discussed in the previous section enable PIV measurements in multiple planes within the cylinder, and it is also important that the measurements are conducted when the engine is running under specified operation conditions. The test points for our experiment are selected from JLR's production engine load map, in order to mimic the actual engine behaviour to the greatest extent. Due to the limitation of material strength, the optical engine can only operate under light to medium load (maximum speed 2000 rpm, *cf* Table A.1). Four baseline loads (Table 2.3) were used in our experiment: T7 has the minimum load in terms of both the engine speed and intake manifold absolute pressure (MAP); T1 operates at medium engine speed but the air is throttled slightly — its MAP is about 80% of atmosphere pressure; T2 and T8 both run under nearly-fully-opened throttle, and the T8 point has a higher speed. The intake air and coolant temperatures are kept constant across different engine loads.

Different engine hardware was also tested in order to examine flow structure under various conditions. Two pistons were used to provide different compression ratios — Piston B is lower (designed value $11:1$) compared to Piston C ($12.5:1$). The intake valves were phased to an optimal operation point using the CVVL system (described in Appendix A).

Besides the pistons, two different camshafts, named "C33" and "Bootleg", were also tested. The C33 camshaft has a standard profile used in production engines, while the Bootleg camshaft introduces a small and early opening of the intake valves in the exhaust stroke, which provides

Table 2.3: Baseline engine loads

Parameters	Engine Load			
	T1	T2	T7	T8
Engine speed (rpm)	1500	1500	1000	1750
Intake manifold absolute pressure (kPa)	80	95	64	96
Intake air volume flow rate (L/s)	1.57	3.65	1.00	3.40
Intake air temperature (°C)	45	45	45	45
Coolant temperature (°C)	23	23	23	23

a longer valve overlap (i.e., time when both the intake and exhaust valves are open) and thus influences the gas exchange process. The throttle was slightly adjusted to alter the MAP to match the baseline engine (equipped with the C33 camshaft) intake air volume flow rate when using the Bootleg camshaft.

The engine normally operates with both intake valves activated (denoted as dual valve actuation, or DVA), which should create a symmetric flow in the swirl plane. It was also of special interest to induce a strong swirling in-cylinder flow to promote fuel-air mixing. Therefore, the flywheel-side intake valve was disabled at several test points, and the engine was running under so-called single valve actuation (SVA) conditions, where two methods were used to match the intake air volume flow rate achieved in DVA cases. The first method introduced less throttle (denoted as SVA Throttle) and thus raised the MAP, and the second increased the maximum inlet valve lift while keeping the MAP unaltered.

Table 2.4 lists the test matrix with the accomplished test points highlighted in colour. For the 18 yellow-highlighted points, the engine was motored without fuel injection or ignition. For the three blue-highlighted points, operations with and without fuel were measured for each cycle, but the fuel-air mixture was not ignited. The measurements were conducted in three different planes: the lower and higher swirl planes (denoted L-SP and H-SP) which are 3.5 mm and 0.5 mm

Table 2.4: Test matrix and conducted test points

Load	Camshaft equipped	Intake valve activated	Piston installed					
			Piston B			Piston C		
T1	Bootleg	DVA [a]	L-SP	H-SP	CTP	L-SP	H-SP	CTP
		SVA throttle [a]	L-SP	H-SP	CTP	L-SP	H-SP	CTP
		SVA peak lift [a]	L-SP	H-SP	CTP	L-SP	H-SP	CTP
	C33	DVA [a, b]	L-SP	H-SP	CTP	L-SP	H-SP	CTP
T2	C33	DVA [a, b]	L-SP	H-SP	CTP	L-SP	H-SP	CTP
		SVA throttle	L-SP	H-SP	CTP	L-SP	H-SP	CTP
T7	C33	DVA [b]	L-SP	H-SP	CTP	L-SP	H-SP	CTP
		SVA throttle	L-SP	H-SP	CTP	L-SP	H-SP	CTP
T8	C33	DVA [b]	L-SP	H-SP	CTP	L-SP	H-SP	CTP

[a] Data for Piston B in the tumble plane (TP) is also available.
[b] Data for Piston C in the tumble plane (TP) is also available.
[*] Operations with and without fuel injection (no ignition) were measured in the blue-highlighted test points. Fuel was not injected in the yellow-highlighted test points.

below the fire deck, whereas the cross-tumble plane (denoted CTP) is the middle plane of the cylinder. At least 300 engine cycles were recorded at each test point, and depending on the number of measurements per cycle, the tests were separated into two or three runs per point due to the camera memory limit (i.e., 150 or 100 cycles per run). Pairs of PIV images were recorded for consecutive engine cycles at five CAD intervals in each test run, and the corresponding valve lifts and pressure traces were logged at every crank angle degree. The lower swirl plane and cross-tumble plane data were collected from 330° to 30° bTDCf[(8)] during the intake and compression

[(8)] bTDCf refers to "before the top dead centre at which the firing occurs".

strokes[9] in order to examine how the in-cylinder flow evolves during the mixture formation process and before the upcoming combustion event.[10] The higher swirl plane data focuses on near-TDC flow structure from mid-compression (90° bTDCf) to mid-expansion (90° aTDCf[11]) engine phases. It should be noted again (see also Section 2.2.2) that Scott [55] conducted the tumble plane (denoted TP) PIV measurements using this optical engine at the same set of test points from 330° to 30° bTDCf, and therefore four of the listed test conditions have data in four different planes (the lower and higher swirl, cross-tumble and tumble planes).

2.4 Image processing and vector generation

The pairs of images were processed in DaVis® (LaVision, Ver. 8.4.0) to generate vectors. As mentioned in the setup section (Section 2.2.1), the light sheet hit the metal surfaces (pistons, cylinder head and valves) in the cylinder, and caused strong background scatter as illustrated in an example pair of raw images (Figure 2.6(a)) in the lower swirl plane. The intake valves are clearly visible, and therefore background subtraction needs to be applied so that the PIV calculation only accounts for the droplet (and thus the surrounding air) motion.

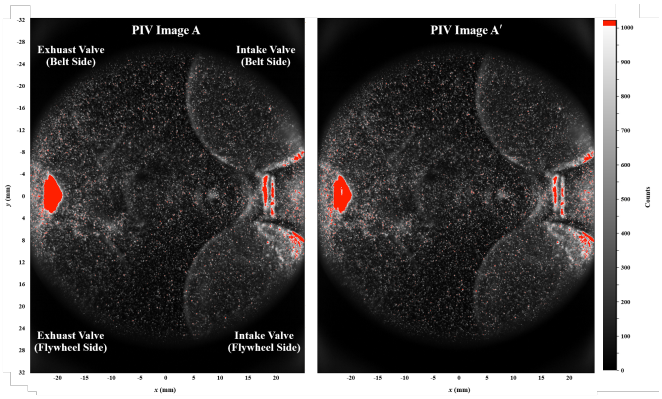
Two different background subtraction methods were attempted. The first subtracted the Gaussian average based on image intensities at the nearest four measurements (± 5 and 10 crank angle degrees), assuming the scatter condition has minimal change within one millisecond. This "crank angle filter" method worked effectively at many crank angles, but when the intake valves are open, due to the location differences of the valves viewed from the piston window, the background subtracted images (Figure 2.6(b)) still show the edges of the

[9] When the piston is at a high position (less than 30 crank angle degrees away from the top dead centre), it blocks the camera view in the cross-tumble and the tumble plane, and the light sheet in the lower swirl plane. Hence the flow fields were measured from 330° to 30° bTDCf.

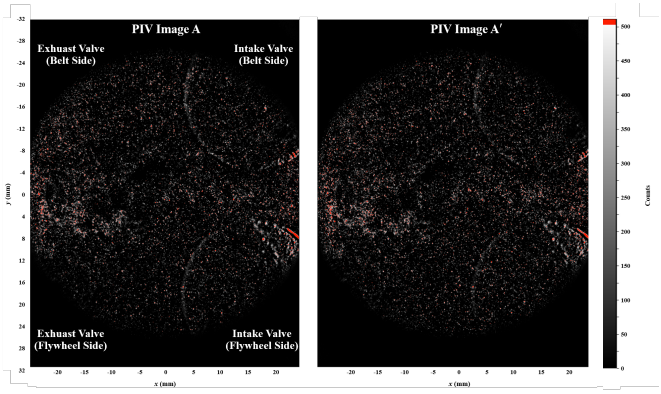
[10] Although the tests were mostly conducted under motored conditions, as mentioned in Section 1.1, the in-cylinder air motion has a profound effect on the mixture formation and the combustion characteristics.

[11] aTDCf refers to "after the top dead centre at which the firing occurs".

(a) PIV raw images



(b) Background subtracted images (crank angle filter)



(c) Background subtracted images (cycle filter)

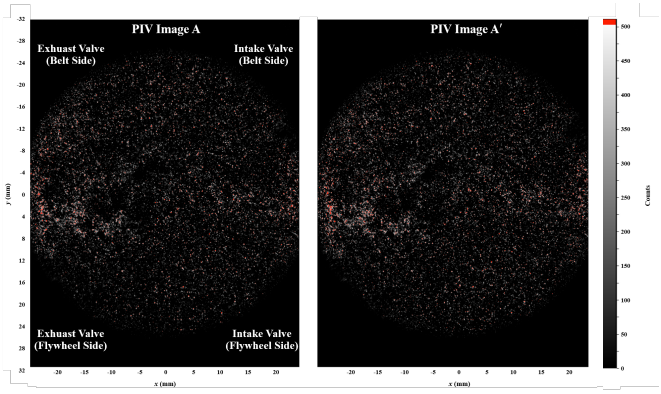


Figure 2.6: Example PIV image pair in the lower swirl plane. Image A (left column) refers to the first image in the pair, and Image A′ (right column) is the second image taken several microseconds later (adjusted for various flow speeds). Note that different colour (intensity count) scales are used in the raw images in (a) versus the background subtracted images in (b) and (c).

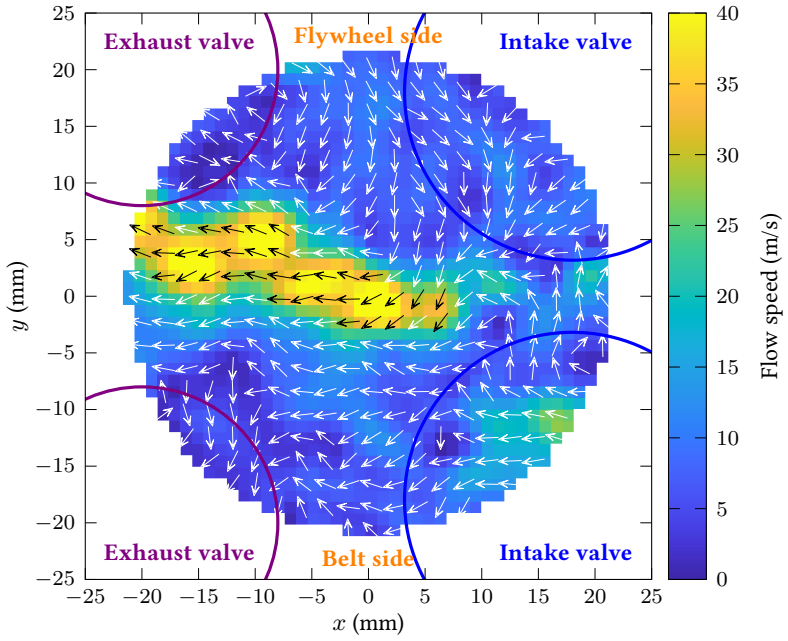


Figure 2.7: Example PIV flow field in the lower swirl plane generated by the pair of PIV images shown in Figure 2.6(c).[12]

intake valves, which may lead to a biased flow speed calculation since the PIV cross-correlation process in this case accounts for both the particle motions and the movement of the valve edges. An alternative filtering method was proposed — the filter still applied a Gaussian average, but instead of referencing the nearest measurements in the same cycle but at different crank angles, it used the data at the same crank angle of the two previous and two following cycles, and therefore the valve location is fixed among these cycles. As shown in Figure 2.6(c), this "cycle filter" effectively filtered out the background scatter from the edge of the valves.

After the background subtraction, an universal particle intensity normalisation algorithm was also enabled to compensate for the intensity changes between frames due to the energy

[12] The vectors are in unit length and only illustrate the flow direction, while the magnitude of the vector at each location are shown by the colour map. For clarity only every second vector is plotted in each direction, while the false colour background uses the actual spatial resolution from the PIV experiment. The colour (black/white) of each vector is chosen to contrast with the false colour background. This plot style is adopted in this thesis for all the plots containing vectors, unless otherwise specified.

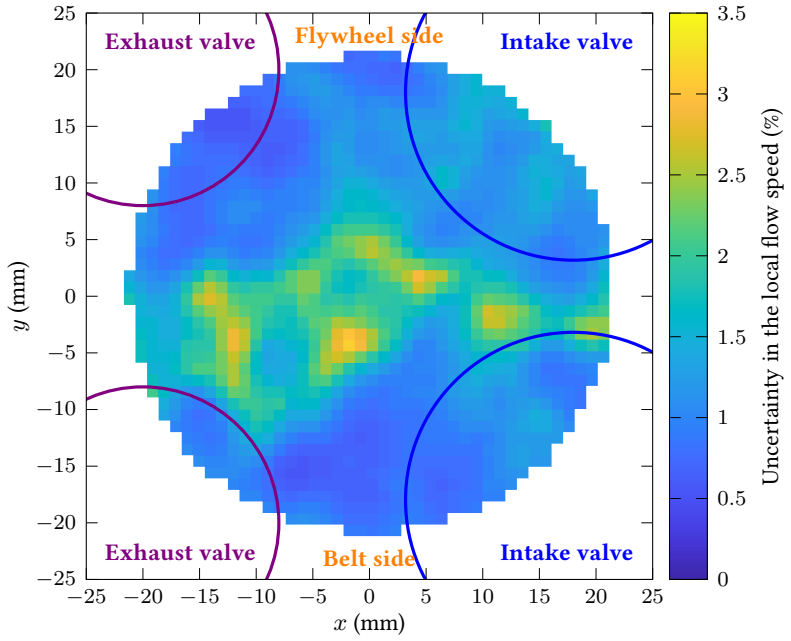


Figure 2.8: Uncertainty in the local flow speed for the example PIV flow field shown in Figure 2.7. The uncertainty at each location is normalised by the local flow speed and is shown in %.

differences in laser pulses. The pixels in which the light sheet did not reach were masked out. The vector generation process used the multi-pass algorithm, which decreases the interrogation window from 128 × 128-pixel to 32 × 32-pixel with 50% overlap [56, 57]. Local spurious vectors whose peak ratios (correspond to the ratio of correlation values between the highest and the second highest correlation peaks [85]) smaller than two were removed, and were not interpolated nor replaced (therefore they were left as a hole in the flow fields). The total number of removed vectors was less than 0.1% in all the data sets. Figure 2.7 shows the flow field generated by Figure 2.6, and the its uncertainty in the local flow speed is reported in Figure 2.8. The uncertainty at each location is lower than 3.2% in this example flow field, and is lower than 5% in any measured flow field in all the data sets. The images obtained at all the test points (Table 2.4) were processed using the same algorithm to ensure consistency.

2.5 Reynolds averaged Navier-Stokes simulation

The in-cylinder flow fields were simulated by engineers from Jaguar Land Rover using the Siemens STAR-CD® software with the es-ice™ package (ver. 4.26.022) and the k-ϵ RNG turbulence model [24, 86]. The time step was set to 0.1 CAD, with refinements of 0.05 CAD close to the valve opening and closing, as well as during valve overlap. The automatically trimmed mesh with 0.7 mm mesh size generated over 3.3 million cells when the piston is at the bottom dead centre (Figure 2.9), and the mesh was adjusted among different crank angles to account for the piston and valve motions. The mesh was refined to 0.3 mm both upstream and downstream of the intake valves when they were open to better model the intake process.

The RANS simulation used the same engine geometry as the optical engine, and was initialised before the exhaust valve opening, i.e., in a closed control volume, during the late expansion stroke. The following parameters measured during the experiment were used as the boundary conditions of the model (all the parameters were averaged for 300 measurements):

1. crank angle-resolved intake (purple solid line in Figure 2.10) and exhaust valve lifts (purple dashed line in Figure 2.10),
2. crank angle-resolved pressures inside the inlet and the exhaust runners,
3. intake, exhaust and main engine coolant temperatures, and
4. intake air volume flow rate.

Due to the unknown amount of blow-by loss during the optical engine operation, the piston crevice length (distance between the piston surface and the upper piston ring) was tuned in the model to match the in-cylinder pressure traces with the experiments. There were no other physical parameters adjusted during the simulation. The difference of in-cylinder pressure between the experiment (300-cycle average, blue solid line in Figure 2.10) and the simulation (red dashed line in Figure 2.10) was less than 1.8% at all crank angles at the T2, Piston B, C33, DVA test point, and was less than 3% for all the other test points (*cf* Table 2.4).

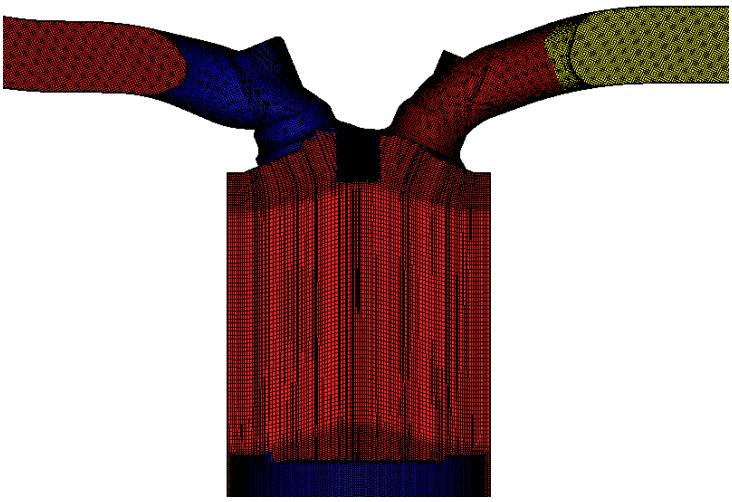


Figure 2.9: Example CFD mesh in the tumble plane when the piston is at the bottom dead centre.

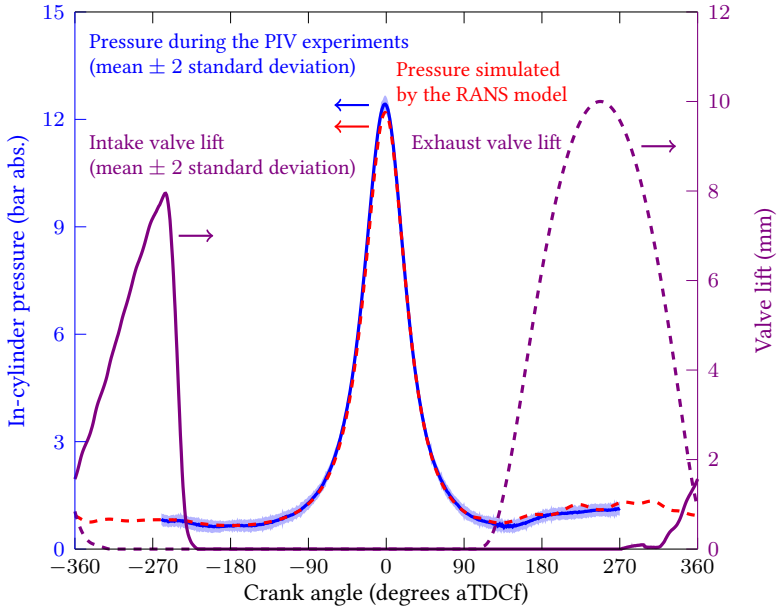


Figure 2.10: Engine operation parameters at the T2, Piston B, C33, DVA test point. Left y axis: in-cylinder pressure trace recorded during the PIV experiments (blue solid line with shaded error bar showing $\pm$ 2 standard deviation) and pressure trace simulated by the RANS flow model (red dashed line). Right y axis: intake (purple solid line with shaded error bar) and exhaust (purple dashed line) valve lifts measured during the PIV experiments and used as boundary conditions in the RANS flow model.

Chapter 3

Data analysis approaches

In this chapter, the data analysis approaches used for various purposes throughout the thesis are presented to facilitate discussion in later chapters. Two different PIV flow field averaging metrics are firstly introduced. The algorithms for two dimensionality reduction techniques, namely proper orthogonal decomposition (POD) and kernel principal component analysis (KPCA), are discussed in detail. Four different metrics are then presented to provide a quantitative comparison of vector fields.

3.1 PIV flow field averaging metrics

In this section, two metrics that are used to average the PIV flow fields are defined. Figures 3.1(a)–(c) provide flow fields that were measured at an example test point (T2 load, Piston B, C33, DVA) in three consecutive cycles at 290° bTDCf, when both intake valves were open. The blue and purple circles show the edges of the intake and exhaust valves as viewed by the PIV camera. This 300-cycle data set will be used throughout this chapter to illustrate the results of the corresponding mathematical operations.

The PIV measurements in multiple engine cycles need to be averaged into a single field in order to enable a comparison with the corresponding RANS-simulated flow field (Figure 3.1(d)), as the RANS model only provides an averaged flow field at a certain engine phase (i.e., crank angle) and excludes small-scale turbulence. One common approach is to apply the Reynolds decomposition [87] to the data set, and obtain the so-called ensemble mean flow field. Assuming the engine operation is a cyclostationary process [25], the term "ensemble" commonly refers to flow fields that were measured in different cycles but at a specific crank angle on the same measurement plane under a controlled test condition.

Given a set of PIV flow fields $\{\mathbf{V}_n\}$ that were measured at the same crank angle in different cycles ($n = 1, 2, \cdots, N$) at various locations within the engine cylinder ($\ell = 1, 2, \cdots, L$), each flow field $\mathbf{V}_n$ has two measured velocity components: the horizontal components $\mathbf{u}_n$ and the vertical components $\mathbf{v}_n$, and the two components can each be represented by a column array with a size of $L \times 1$:

$$\mathbf{u}_n = \begin{bmatrix} u_{1,n} & u_{2,n} & \cdots & u_{\ell,n} & \cdots & u_{L,n} \end{bmatrix}^{\mathsf{T}}, \tag{3.1}$$

$$\mathbf{v}_n = \begin{bmatrix} v_{1,n} & v_{2,n} & \cdots & v_{\ell,n} & \cdots & v_{L,n} \end{bmatrix}^{\mathsf{T}}, \tag{3.2}$$

where the elements $u_{\ell,n}$ and $v_{\ell,n}$ represent, respectively, the horizontal and vertical velocity components at location ℓ for cycle n at the measured crank angle, and the symbol $^{\mathsf{T}}$ denotes the

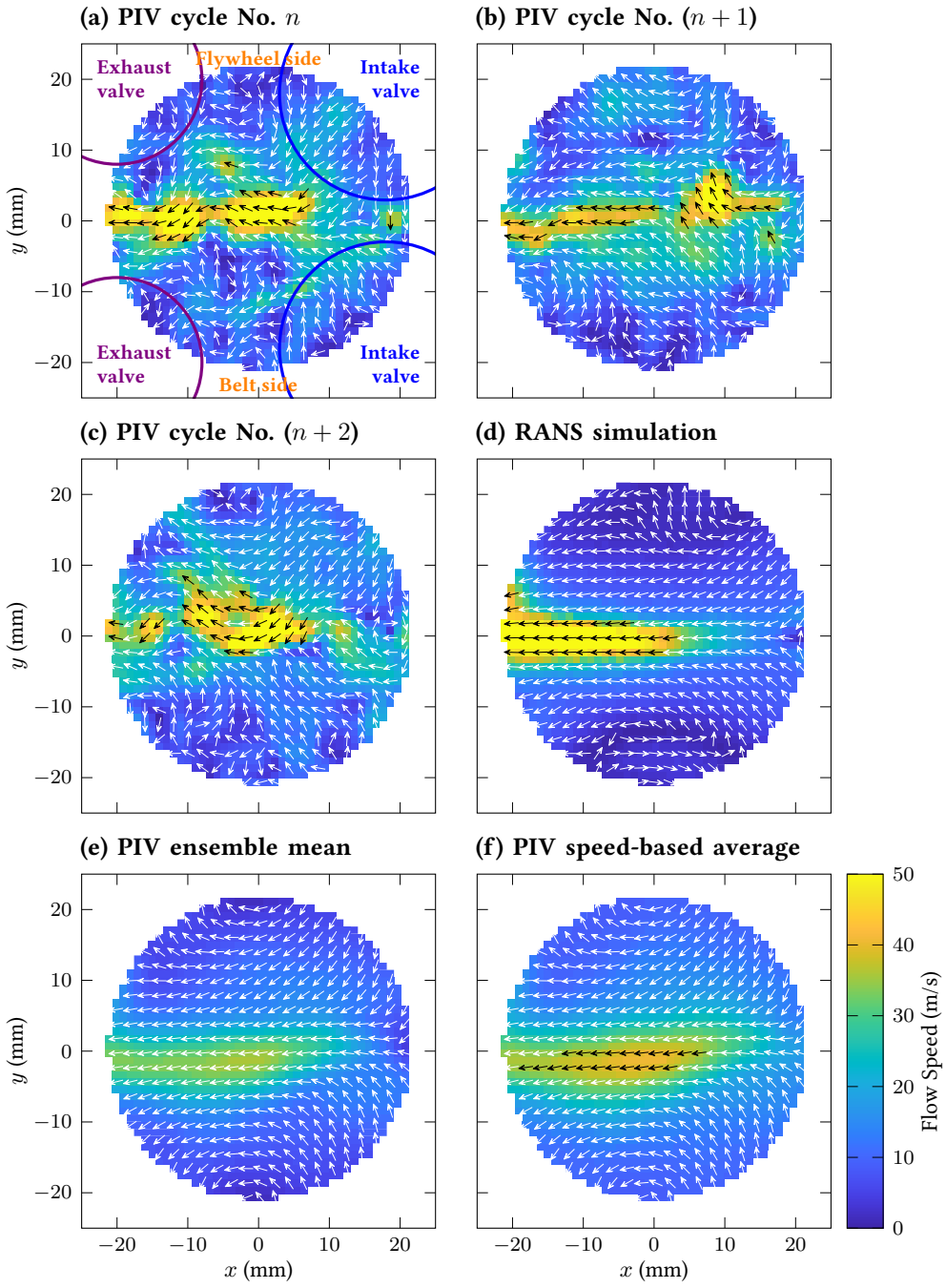


Figure 3.1: Example flow fields illustrating the PIV flow field averaging techniques. Data were collected in the lower swirl plane at 275° bTDCf at the T2, Piston B, C33, DVA test point. PIV flow fields in (a)–(c) were recorded in three consecutive cycles. (e) & (f) were computed using Equations (3.5) and (3.8), respectively.

matrix transpose operation.[1] The order of the location index ℓ can be arbitrarily chosen as long as the same order is used in the representation of different fields. Each flow field $\mathbf{V}_n$ is therefore represented by a column array with a size of $2L \times 1$:

$$\mathbf{V}_n = \begin{bmatrix} \mathbf{u}_n \\ \mathbf{v}_n \end{bmatrix}, \tag{3.3}$$

and the entire data set $\{\mathbf{V}_n\}$ can be organised into a matrix $\mathbf{V}$ with a size of $2L \times N$:

$$\begin{aligned} \mathbf{V} &= \begin{bmatrix} \mathbf{V}_1 & \mathbf{V}_2 & \cdots & \mathbf{V}_n & \cdots & \mathbf{V}_N \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n & \cdots & \mathbf{u}_N \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n & \cdots & \mathbf{v}_N \end{bmatrix} \\ &= \begin{bmatrix} u_{1,1} & u_{1,2} & \cdots & u_{1,n} & \cdots & u_{1,N} \\ u_{2,1} & u_{2,2} & \cdots & u_{2,n} & \cdots & u_{2,N} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ u_{\ell,1} & u_{\ell,2} & \cdots & u_{\ell,n} & \cdots & u_{\ell,N} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ u_{L,1} & u_{L,2} & \cdots & u_{L,n} & \cdots & u_{L,N} \\ v_{1,1} & v_{1,2} & \cdots & v_{1,n} & \cdots & v_{1,N} \\ v_{2,1} & v_{2,2} & \cdots & v_{2,n} & \cdots & v_{2,N} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ v_{\ell,1} & v_{\ell,2} & \cdots & v_{\ell,n} & \cdots & v_{\ell,N} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ v_{L,1} & v_{L,2} & \cdots & u_{L,n} & \cdots & v_{L,N} \end{bmatrix}. \end{aligned} \tag{3.4}$$

[1] The elements of all the matrices presented in this thesis are real numbers, hence the notations of non-conjugate and conjugate matrix transpose operations are not separated for simplicity. The complex number representation for flow fields is beyond the scope of this thesis, but a brief note is provided here. If the vectors in the flow field are represented in a complex number form (i.e., $u_{\ell,n}+i\cdot v_{\ell,n}$, where i is an indeterminate such that $i^2 = -1$), it is recommended to use the non-conjugate matrix transpose for data matrix re-organising (for instance in Equation (3.1)) so that the original flow direction remains the same after the re-organising. On the other hand, the conjugate matrix transpose should be used for specific mathematical operations (such as the singular value decomposition in Equation (3.15)).

The two velocity components ($\overline{\mathbf{u}}_{\text{em}}$ and $\overline{\mathbf{v}}_{\text{em}}$) of the phase-averaged ensemble mean $\overline{\mathbf{V}}_{\text{em}}$ can be found by taking the element-wise average of the fields of $\mathbf{u}_n$ and $\mathbf{v}_n$:

$$\overline{\mathbf{V}}_{\text{em}} = \begin{bmatrix} \overline{\mathbf{u}}_{\text{em}} \\ \overline{\mathbf{v}}_{\text{em}} \end{bmatrix}, \qquad \overline{\mathbf{u}}_{\text{em}} = \frac{1}{N} \cdot \sum_{n=1}^{N} \mathbf{u}_n, \qquad \overline{\mathbf{v}}_{\text{em}} = \frac{1}{N} \cdot \sum_{n=1}^{N} \mathbf{v}_n, \tag{3.5}$$

where the subscript "em" denotes the ensemble mean. The averaging process in Equation (3.5) can be achieved by taking the mean of each row of the data set matrix $\mathbf{V}$ (Equation (3.4)). The resulting $\overline{\mathbf{V}}_{\text{em}}$ is a column array with a size of $2L \times 1$; $\overline{\mathbf{u}}_{\text{em}}$ and $\overline{\mathbf{v}}_{\text{em}}$ both have a size of $L \times 1$. Figure 3.1(e) shows the ensemble mean of 300 cycles at the example test point. The magnitude (speed) field of the ensemble mean $\overline{\mathbf{M}}_{\text{em}}$, which was plotted as the colour map in Figure 3.1(e), is calculated by:

$$\begin{aligned} \overline{\mathbf{M}}_{\text{em}} &= \left(\overline{\mathbf{u}}_{\text{em}} \odot \overline{\mathbf{u}}_{\text{em}} + \overline{\mathbf{v}}_{\text{em}} \odot \overline{\mathbf{v}}_{\text{em}}\right)^{\odot \frac{1}{2}} \\ &= \begin{bmatrix} (M_{\text{em}})_1 & (M_{\text{em}})_2 & \cdots & (M_{\text{em}})_\ell & \cdots & (M_{\text{em}})_L \end{bmatrix}^{\mathsf{T}}, \end{aligned} \tag{3.6}$$

where $\mathbf{A} \odot \mathbf{B}$ denotes the Hadamard product (element-wise product) of two matrices $\mathbf{A}$ and $\mathbf{B}$, $(\mathbf{A})^{\odot \frac{1}{2}}$ denotes the Hadamard square root (element-wise square root) of matrix $\mathbf{A}$. $\overline{\mathbf{M}}_{\text{em}}$ is a column array with a size of $L \times 1$. The element of $\overline{\mathbf{M}}_{\text{em}}$ at location ℓ in the flow field, $(M_{\text{em}})_\ell$, can be computed using:

$$(M_{\text{em}})_\ell = \sqrt{\left(\frac{1}{N} \cdot \sum_{n=1}^{N} u_{\ell,n}\right)^2 + \left(\frac{1}{N} \cdot \sum_{n=1}^{N} v_{\ell,n}\right)^2}. \tag{3.7}$$

As shown by Equation (3.7), the computation of the ensemble mean inevitably cancels out components with opposing signs at a specified location in the flow field (for instance, $u_{\ell,n}$ can have opposing signs in different cycles, i.e., different n). The resulting ensemble mean flow field, as shown in Figure 3.1(e), has a lower flow speed in the central region (near the $y = 0$ line) compared to the individual cycles (Figures 3.1(a)–(c)). Such a feature may lead to misinterpretations when the ensemble mean is used to make comparisons with the RANS simulation (Figures 3.1(d)) — one would think the RANS simulation over-predicts the flow

speed, but it is the ensemble mean (Figure 3.1(e)) which indeed under-represents the actual measured flow speed in individual cycles (Figures 3.1(a)–(c)).

As an alternative approach, the magnitude (speed) of the ensemble mean can be replaced by the arithmetic mean of the magnitude (average of speed) for all cycles at the same crank angle, and this averaging method is termed the "speed-based average". The magnitude (speed) field of the speed-based average is defined by:

$$\begin{aligned}\overline{\mathbf{M}}_{\text{sa}} &= \frac{1}{N}\cdot\sum_{n=1}^{N}\left(\mathbf{u}_n\odot\mathbf{u}_n+\mathbf{v}_n\odot\mathbf{v}_n\right)^{\odot\frac{1}{2}}\\ &= \begin{bmatrix}(M_{\text{sa}})_1 & (M_{\text{sa}})_2 & \cdots & (M_{\text{sa}})_\ell & \cdots & (M_{\text{sa}})_L\end{bmatrix}^{\mathsf{T}},\end{aligned} \tag{3.8}$$

where the subscript "sa" denotes the speed-based average. $\overline{\mathbf{M}}_{\text{sa}}$ is a column array with a size of $L\times 1$. The element of $\overline{\mathbf{M}}_{\text{sa}}$ at location ℓ in the flow field, $(M_{\text{sa}})_\ell$, can be computed using:

$$(M_{\text{sa}})_\ell = \frac{1}{N}\cdot\sum_{n=1}^{N}M_{\ell,n} = \frac{1}{N}\cdot\sum_{n=1}^{N}\sqrt{u_{\ell,n}^2+v_{\ell,n}^2}, \tag{3.9}$$

where $M_{\ell,n}=\sqrt{u_{\ell,n}^2+v_{\ell,n}^2}$ is the flow speed (magnitude of the velocity vector) at location ℓ for cycle n. The values of $M_{\ell,n}$ at different locations (different ℓ) were plotted as the colour map for Figures 3.1(a)–(c) for the corresponding cycle (the same n).

It should be noted that "the cycle-averaged field of the flow speed field" (i.e., the speed-based average, *cf* Equations (3.8) and (3.9)) is different from "the speed field of the cycle-averaged flow field" (i.e., the ensemble mean, *cf* Equations (3.6) and (3.7)). Since the flow speed at any location in the individual cycles ($M_{n,\ell}$) is always non-negative, the averaging process in the speed-based average flow field (*cf* Equation (3.9)) does not have the same drawback (opposed sign components cancellation) as it had in computing the ensemble mean (*cf* Equation (3.7)). Therefore, the flow speeds in the speed-based average flow field (Figure 3.1(f)) are closer to the ones of the individual cycles (Figure 3.1(a)–(c)), while the ensemble mean (Figure 3.1(e)) has a much lower flow speed.

On the other hand, despite the ensemble mean (Figure 3.1(e)) cannot illustrate the extent of the cycle-to-cycle variation, it still provides a good estimate of the averaged flow directions in the

individual cycles (Figure 3.1(a)–(c)). Therefore, the flow directions in the speed-based average flow field (Figure 3.1(f)) are directly adopted from the ensemble mean.

3.2 Dimensionality reduction techniques

One of the key features of the in-cylinder flow is that its structure may vary from cycle-to-cycle — for instance, the high-speed intake jet[(2)] in Figure 3.1(a)–(c) changes both in location and direction in three consecutive cycles, albeit that the measurement was taken at the same crank angle and in the same plane. The averaging metrics introduced in the previous section (Section 3.1) fail to maintain the cycle-to-cycle variation (CCV) information. Therefore, it is necessary to find data filtering techniques that can incorporate CCV, while still being able to exclude the small-scale turbulence and noise in the measured flow fields and provide validation targets for the RANS flow model. The filtering process can be fulfilled by the dimensionality reduction techniques, two of which, namely the proper orthogonal decomposition (POD) and the kernel principal component analysis (Kernel PCA, or simply KPCA), are introduced in this section, together with their corresponding mathematical notations and algorithms.

3.2.1 Proper orthogonal decomposition

Proper orthogonal decomposition (POD) [58] is a variant of the linear principal component analysis (PCA), and is used to impose orthogonal projection of data onto lower dimensional spaces (known as the principal subspace), where the variance of the projected data is maximised [79]. The projected data, also known as the reconstructed data or the low-order approximation of the original data, can therefore have a smaller data structure and contain less noise or error. Various algorithms can be used to perform POD, one of which that is commonly

[(2)] The air streams from both intake valves collide in the cylinder, and the resulting high-speed three-dimensional flow structure is usually in a jet-like shape when viewed in two-dimensional planes (such as the lower swirl plane in Figure 3.1 and in the cross-tumble plane in Figure 5.1). This flow feature is called the "intake jet" in this thesis.

used in turbulent flow field reconstruction is based on the singular value decomposition (SVD) approach [88–90]. The SVD-based POD algorithm is discussed in the following paragraphs.

Following the Reynolds decomposition approach [87], the original flow fields can be separated into the ensemble mean and the fluctuations around the ensemble mean. This data centring process is achieved by subtracting the phase-averaged ensemble mean $\overline{\mathbf{V}}_{\mathrm{em}}$ (*cf* Equation (3.5)) from each flow field $\mathbf{V}_n$, leaving the fluctuation flow field $\mathbf{V}'_n$:

$$\mathbf{V}'_n = \mathbf{V}_n - \overline{\mathbf{V}}_{\mathrm{em}}. \tag{3.10}$$

Analogous to the approach in Equation (3.4), the entire set of the fluctuation flow fields, $\{\mathbf{V}'_n\}$, can be expressed by a matrix $\mathbf{V}'$ with a size of $2L \times N$:

$$\begin{aligned}
\mathbf{V}' &= \begin{bmatrix} \mathbf{V}'_1 & \mathbf{V}'_2 & \cdots & \mathbf{V}'_n & \cdots & \mathbf{V}'_N \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{u}'_1 & \mathbf{u}'_2 & \cdots & \mathbf{u}'_n & \cdots & \mathbf{u}'_N \\ \mathbf{v}'_1 & \mathbf{v}'_2 & \cdots & \mathbf{v}'_n & \cdots & \mathbf{v}'_N \end{bmatrix} \\
&= \begin{bmatrix}
u'_{1,1} & u'_{1,2} & \cdots & u'_{1,n} & \cdots & u'_{1,N} \\
u'_{2,1} & u'_{2,2} & \cdots & u'_{2,n} & \cdots & u'_{2,N} \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
u'_{\ell,1} & u'_{\ell,2} & \cdots & u'_{\ell,n} & \cdots & u'_{\ell,N} \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
u'_{L,1} & u'_{L,2} & \cdots & u'_{L,n} & \cdots & u'_{L,N} \\
v'_{1,1} & v'_{1,2} & \cdots & v'_{1,n} & \cdots & v'_{1,N} \\
v'_{2,1} & v'_{2,2} & \cdots & v'_{2,n} & \cdots & v'_{2,N} \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
v'_{\ell,1} & v'_{\ell,2} & \cdots & v'_{\ell,n} & \cdots & v'_{\ell,N} \\
\vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
v'_{L,1} & v'_{L,2} & \cdots & u'_{L,n} & \cdots & v'_{L,N}
\end{bmatrix},
\end{aligned} \tag{3.11}$$

where $u'_{\ell,n}$ and $v'_{\ell,n}$ are, respectively, the horizontal and vertical velocity components of the fluctuation flow field at location ℓ for cycle n. The values of $u'_{\ell,n}$ and $v'_{\ell,n}$ are computed by:

$$u'_{\ell,n} = u_{\ell,n} - \frac{1}{N} \cdot \sum_{n=1}^{N} u_{\ell,n}, \qquad v'_{\ell,n} = v_{\ell,n} - \frac{1}{N} \cdot \sum_{n=1}^{N} v_{\ell,n}. \tag{3.12}$$

When the number of parameters ($2L$) is no less than the number of observations (N) in the data matrix $\mathbf{V}'$,[3] the proper orthogonal decomposition to the set of fluctuation flow fields separates each fluctuation flow field into a linear combination of M principal components (also known as the POD modes $\boldsymbol{\phi}^{(m)}$, where $m = 1, 2, \cdots, M$):

$$\mathbf{V}' = \boldsymbol{\Phi} \cdot \mathbf{C}, \qquad \mathbf{V}'_n = \sum_{m=1}^{M} \left(\boldsymbol{\phi}^{(m)} \cdot C_n^{(m)} \right), \tag{3.13}$$

where M is the rank of the data matrix $\mathbf{V}'$, and $M \leq N$;

$\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\phi}^{(1)} & \boldsymbol{\phi}^{(2)} & \cdots & \boldsymbol{\phi}^{(m)} & \cdots & \boldsymbol{\phi}^{(M)} \end{bmatrix}$ is a $2L \times M$ matrix containing M modes;

$\boldsymbol{\phi}^{(m)}$ is a $2L \times 1$ column array whose ℓth and $(\ell + L)$th entries are, respectively, the mth principal component for the horizontal and vertical velocity components at location ℓ;

$\mathbf{C}$ is a $M \times N$ matrix whose mth row and nth column element $C_n^{(m)}$ is the "POD mode m coefficient of cycle n", i.e.,

$$\mathbf{C} = \begin{bmatrix} C_1^{(1)} & C_2^{(1)} & \cdots & C_n^{(1)} & \cdots & C_N^{(1)} \\ C_1^{(2)} & C_2^{(2)} & \cdots & C_n^{(2)} & \cdots & C_N^{(2)} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ C_1^{(m)} & C_2^{(m)} & \cdots & C_n^{(m)} & \cdots & C_N^{(m)} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ C_1^{(M)} & C_2^{(M)} & \cdots & C_n^{(M)} & \cdots & C_N^{(M)} \end{bmatrix}. \tag{3.14}$$

[3] This $2L \geq N$ assumption is usually valid for the up-to-date in-cylinder flow applications, where L is in the order of thousands and N is in the order of tens (for LES simulation) or hundreds (for PIV measurements). For other turbulent flow applications, the number of measured locations L is also usually larger than the number of observations N. If POD is applied to a data set where $2L < N$, the algorithm to be discussed and recommendations to be made in later paragraphs may need to change, and this condition is beyond the scope of this thesis.

Note that all the cycles share the same set of POD modes, while their coefficients $C_n^{(m)}$ differ from cycle to cycle for a single mode, and vary among different POD modes for a single cycle. Each term $\left(\boldsymbol{\phi}^{(m)} \cdot C_n^{(m)}\right)$ can be regarded as one "POD component" of the fluctuation field.

The POD modes and mode coefficients can be obtained by SVD-based approaches. The most straight-forward method (known as the "direct method") is by performing a singular value decomposition operation to the fluctuation flow field matrix $\mathbf{V}'$, which yields:

$$\mathbf{V}' = \mathbf{P} \cdot \mathbf{\Gamma} \cdot \mathbf{Q}^{\mathsf{T}}, \tag{3.15}$$

where $\mathbf{P}$ is a $2L \times 2L$ unitary matrix (i.e., $\mathbf{P}^{\mathsf{T}}\mathbf{P} = \mathbf{P}\mathbf{P}^{\mathsf{T}} = \mathbf{I}_{2L}$), $\mathbf{\Gamma}$ is a $2L \times N$ rectangular diagonal matrix (known as the singular value matrix of $\mathbf{V}'$), and $\mathbf{Q}$ is a $N \times N$ unitary matrix. The rank of $\mathbf{\Gamma}$ is the same as the rank of the data matrix $\mathbf{V}'$, and thus only the first M entries in the main diagonal of $\mathbf{\Gamma}$ have non-zero values. The non-zero elements of $\mathbf{\Gamma}$ are sorted in a decreasing order, i.e., $\gamma_1 \geq \gamma_2 \geq \cdots \geq \gamma_m \geq \cdots \geq \gamma_M$. The POD mode matrix $\mathbf{\Phi}$ is the first M columns of the unitary matrix $\mathbf{P}$,[4] and hence has a size of $2L \times M$. The POD modes serve as an orthonormal basis of the set of flow fields, i.e.,

$$\mathbf{\Phi}^{\mathsf{T}} \cdot \mathbf{\Phi} = \mathbf{I}_M, \qquad \left(\boldsymbol{\phi}^{(i)}\right)^{\mathsf{T}} \cdot \boldsymbol{\phi}^{(j)} = \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}, \tag{3.16}$$

where $\mathbf{I}_M$ denotes the $M \times M$ identity matrix, and δ_{ij} is the Kronecker delta. Therefore, from Equations (3.13) and (3.16), the mode coefficient matrix $\mathbf{C}$ is given by $\mathbf{C} = \mathbf{\Phi}^{\mathsf{T}} \cdot \mathbf{V}'$, and has a size of $M \times N$.

Sirovich [91] suggested that performing SVD to a $2L \times N$ matrix (*cf* Equation (3.15)) may lead to a high computational cost especially when the number of measured locations (L) are large, and proposed the use of a so-called "snapshot method". To apply the snapshot method, a

[4] Note that the $(M+1)$th to the $(2L)$th rows of $\mathbf{\Gamma}$ are all zeros, and therefore the $(M+1)$th to the $(2L)$th columns of $\mathbf{P}$ do not contribute to the SVD of $\mathbf{V}'$ (*cf* Equation (3.15)).

correlation matrix $\mathbf{G}$ is firstly defined as:

$$\mathbf{G} = \left(\mathbf{V}'\right)^{\mathsf{T}} \cdot \mathbf{V}', \tag{3.17}$$

where the size of the correlation matrix is $N \times N$. The eigenvalue problem of the correlation matrix $\mathbf{G}$ can be solved by a singular value decomposition to the correlation matrix, which yields:

$$\mathbf{G} = \mathbf{Q} \cdot \mathbf{\Lambda} \cdot \mathbf{Q}^{\mathsf{T}}, \tag{3.18}$$

where $\mathbf{Q}$ is the same $N \times N$ unitary matrix as in Equation (3.15), and $\mathbf{\Lambda} = \mathbf{\Gamma}^{\mathsf{T}}\mathbf{\Gamma} = \mathbf{\Gamma}^{\odot 2}$ is a $N \times N$ square diagonal matrix. The rank of $\mathbf{\Lambda}$ is M, and the non-zero elements of $\mathbf{\Lambda}$ are sorted in a decreasing order, i.e., $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \geq \cdots \geq \lambda_M$. Therefore, only the first M entries on the main diagonal of $\mathbf{\Lambda}$ have non-zero values. Combining Equations (3.13), (3.16), (3.17) and (3.18) gives:

$$\begin{aligned}
\mathbf{G} = \left(\mathbf{V}'\right)^{\mathsf{T}} \cdot \mathbf{V}' &= \mathbf{Q} \cdot \mathbf{\Lambda} \cdot \mathbf{Q}^{\mathsf{T}} \\
(\mathbf{\Phi} \cdot \mathbf{C})^{\mathsf{T}} \cdot \mathbf{\Phi} \cdot \mathbf{C} &= \mathbf{Q} \cdot \mathbf{\Lambda}^{\odot \frac{1}{2}} \cdot \mathbf{\Lambda}^{\odot \frac{1}{2}} \cdot \mathbf{Q}^{\mathsf{T}} \\
\mathbf{C}^{\mathsf{T}} \cdot \mathbf{\Phi}^{\mathsf{T}} \cdot \mathbf{\Phi} \cdot \mathbf{C} &= \mathbf{Q} \cdot \left(\mathbf{\Lambda}^{\odot \frac{1}{2}}\right)^{\mathsf{T}} \cdot \mathbf{\Lambda}^{\odot \frac{1}{2}} \cdot \mathbf{Q}^{\mathsf{T}} \\
\mathbf{C}^{\mathsf{T}} \cdot \mathbf{I}_M \cdot \mathbf{C} &= \left(\mathbf{\Lambda}^{\odot \frac{1}{2}} \cdot \mathbf{Q}^{\mathsf{T}}\right)^{\mathsf{T}} \cdot \mathbf{I}_N \cdot \left(\mathbf{\Lambda}^{\odot \frac{1}{2}} \cdot \mathbf{Q}^{\mathsf{T}}\right).
\end{aligned} \tag{3.19}$$

Recall that only the first M ($M \leq N$) entries in the $N \times N$ diagonal matrix $\mathbf{\Lambda}$ have non-zero values. In other words, the $(M+1)$th to the Nth rows and the $(M+1)$th to the Nth columns of $\mathbf{\Lambda}$ are all zeros. The $M \times N$ mode coefficient matrix $\mathbf{C}$ can therefore be obtained by:

$$\mathbf{C} = \left(\mathbf{\Lambda}_{1:M,1:N}\right)^{\odot \frac{1}{2}} \cdot \mathbf{Q}^{\mathsf{T}}, \tag{3.20}$$

where $\mathbf{\Lambda}_{1:M,1:N}$ denotes a sub-matrix of matrix $\mathbf{\Lambda}$, composed of the first M rows and the first N columns (i.e., all the columns) of matrix $\mathbf{\Lambda}$.

The $2L \times M$ POD mode matrix $\mathbf{\Phi}$ can be computed using:

$$\begin{aligned}
\mathbf{\Phi} \cdot \mathbf{C} &= \mathbf{V}' \\
\mathbf{\Phi} \cdot \left(\mathbf{\Lambda}_{1:M,1:N}\right)^{\odot\frac{1}{2}} \cdot \mathbf{Q}^{\mathsf{T}} &= \mathbf{V}' \\
\mathbf{\Phi} \cdot \left(\mathbf{\Lambda}_{1:M,1:N}\right)^{\odot\frac{1}{2}} \cdot \mathbf{Q}^{\mathsf{T}} \cdot \mathbf{Q} &= \mathbf{V}' \cdot \mathbf{Q} \\
\mathbf{\Phi} \cdot \left(\mathbf{\Lambda}_{1:M,1:N}\right)^{\odot\frac{1}{2}} &= \mathbf{V}' \cdot \mathbf{Q} \\
\mathbf{\Phi} \cdot \left(\mathbf{\Lambda}_{1:M,1:N}\right)^{\odot\frac{1}{2}} \cdot \left(\mathbf{\Lambda}_{1:N,1:M}\right)^{\odot(-\frac{1}{2})} &= \mathbf{V}' \cdot \mathbf{Q} \cdot \left(\mathbf{\Lambda}_{1:N,1:M}\right)^{\odot(-\frac{1}{2})} \\
\mathbf{\Phi} &= \mathbf{V}' \cdot \mathbf{Q} \cdot \left(\mathbf{\Lambda}_{1:N,1:M}\right)^{\odot(-\frac{1}{2})} .
\end{aligned} \tag{3.21}$$

It should be noted that although the snapshot method performs SVD on a smaller matrix (*cf* Equation (3.18)) compared to the direct method (*cf* Equation (3.15)),[(5)] the amount of time saved by using the snapshot method is negligible thanks to the high computational power of modern computers — for a data set measured at 1870 locations in 300 cycles, the direct method takes 189 ms using a normal personal computer, while the snapshot method takes 83 ms using the same computer. Therefore, the two methods are equally suitable for performing POD.

Several important properties of the POD results are highlighted in the following paragraphs. In addition to the velocity field decomposition (Equation (3.13)), the (specific) kinetic energy content of each fluctuation flow field (denote KE_n for cycle n) is partitioned into each mode (recall the mode orthonormality (Equation (3.16))):

$$\mathrm{KE}_n = \frac{\left\|\mathbf{V}'_n\right\|^2}{2} = \sum_{m=1}^{M} \frac{1}{2}\left(C_n^{(m)}\right)^2 = \sum_{m=1}^{M} \mathrm{KE}_n^{(m)}, \tag{3.22}$$

where $\|\cdot\|$ denotes the $\mathbb{L}^2$ norm, and $\mathrm{KE}_n^{(m)}$ is the amount of kinetic energy of cycle n collected by the mth mode, which is equal to half of the squared mode m coefficient of Cycle n ($C_n^{(m)}$), i.e.:

$$\mathrm{KE}_n^{(m)} = \frac{1}{2}\left(C_n^{(m)}\right)^2. \tag{3.23}$$

[(5)] The covariance matrix is smaller only when $2L > N$.

From Equation (3.20), one can deduce the following equations:

$$\mathbf{C}\cdot\mathbf{C}^{\mathsf{T}} = \left(\mathbf{\Lambda}_{1:M,1:N}\right)^{\odot\frac{1}{2}}\cdot\mathbf{Q}^{\mathsf{T}}\cdot\mathbf{Q}\cdot\left(\mathbf{\Lambda}_{1:N,1:M}\right)^{\odot\frac{1}{2}} = \mathbf{\Lambda}_{1:M,1:M}, \tag{3.24}$$

or in element form (*cf* Equation (3.14)):

$$\sum_{n=1}^{N}\left(C_n^{(m)}\right)^2 = \lambda_m. \tag{3.25}$$

Recall that the non-zero elements of the rank-M diagonal matrix $\mathbf{\Lambda}$ are sorted in a decreasing order, i.e., $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m \geq \cdots \geq \lambda_M$ (Equation (3.18)), the sum of the squared POD coefficients are successively maximised from the first mode to the last (mode M):

$$\sum_{n=1}^{N}\left(C_n^{(1)}\right)^2 \geq \sum_{n=1}^{N}\left(C_n^{(2)}\right)^2 \geq \cdots \geq \sum_{n=1}^{N}\left(C_n^{(m)}\right)^2 \geq \cdots \geq \sum_{n=1}^{N}\left(C_n^{(M)}\right)^2. \tag{3.26}$$

Combining with Equation (3.23), the kinetic energy content collected by each mode is also successively maximised:

$$\sum_{n=1}^{N}\mathrm{KE}_n^{(1)} \geq \sum_{n=1}^{N}\mathrm{KE}_n^{(2)} \geq \cdots \geq \sum_{n=1}^{N}\mathrm{KE}_n^{(M)}. \tag{3.27}$$

The underlying successive energy maximisation in POD bears immediate relevance to use the first d (where $1 \leq d \leq M$) POD components in the decomposition (*cf* Equation (3.13)) as a low-order approximation (or equivalently, POD reconstruction) of the flow field for cycle n [64]:

$$\begin{aligned}\mathbf{V}_n = \overline{\mathbf{V}}_{\mathrm{em}} + \mathbf{V}'_n &= \overline{\mathbf{V}}_{\mathrm{em}} + \sum_{m=1}^{M}\left(\boldsymbol{\phi}^{(m)}\cdot C_n^{(m)}\right)\\ &\approx \overline{\mathbf{V}}_{\mathrm{em}} + \sum_{m=1}^{d}\left(\boldsymbol{\phi}^{(m)}\cdot C_n^{(m)}\right),\end{aligned} \tag{3.28}$$

This approximation, from a mathematical point of view, maps the original data set in the real space $\mathbb{R}^M$ onto a d-dimensional ($d \leq M$) linear principal subspace. From a physical point of view, the approximation process discards the low-energy (higher-order) POD components (from the $(d+1)$th to the Mth), and thus extracts the coherent structure (high energetic part) of the turbulent flow field. It should be noted that the ensemble mean must be included in the

calculation so that the reconstructed fields can have a comparable flow speed to the original measured flow field. As such, the ensemble mean can be regarded as the zeroth-order approximation to all the cycles. The choice of the cut-off mode number k is discussed in Chapters 5 and 6.

Although Chen *et al.* [90, 92] recommends performing POD on instantaneous flow fields without subtracting the ensemble mean (i.e., POD on $\{\mathbf{V}_n\}$), it can also be argued that it is better to take the conventional Reynolds decomposition approach and apply POD on the fluctuation flow fields after subtracting the ensemble mean (i.e, POD on $\{\mathbf{V}'_n\}$, as used in this thesis). Since the first POD mode of the instantaneous flow fields is a good estimate of the ensemble average,[(6)] an unnecessary restriction will be added to the strongest fluctuation detected by the second POD mode of $\{\mathbf{V}_n\}$ – that this strongest fluctuation (and all the other modes) must be in a direction which is (almost) orthogonal to the ensemble mean since the POD modes are strictly orthogonal to each other (Equation (3.16)). On the other hand, by subtracting the ensemble mean prior to performing POD, the first POD mode has no restriction in direction and can thus represent the main fluctuation flow structure to the greatest extent. Since the ensemble mean is a linear combination of all the instantaneous flow fields with a common coefficient of $\frac{1}{N}$, the rank of $\{\mathbf{V}'_n\}$ is reduced by one compared to the rank of $\{\mathbf{V}_n\}$. As a result, compared to POD without subtracting the ensemble mean, the total number of POD modes (M) will be reduced by one if the ensemble mean is subtracted before performing POD. Indeed, the differences when using either approach are negligible in the context of this thesis, since a number of POD components are added together to form the low-order approximated flow fields and hence the differences in the individual modes are diluted.

The linear PCA and POD are very similar, but strictly speaking they are not exactly equivalent when conducting the dimensionality reduction [93]. The standard linear PCA aims

[(6)] The RI value, whose definition is given in Equation (3.51), exceeds 0.99 when comparing the the first POD mode of the instantaneous flow fields and the ensemble mean flow field.

at reducing the number of parameters (or features) needed to represent the entire data set, while POD aims at reducing the dimension of the sample space — in the case of the in-cylinder flow fields, the parameters are the velocity components at each location ℓ, and each engine cycle is one sample. Therefore, although their algorithms almost look identical, there still exists a key difference — in the POD algorithm, SVD is applied to the $N \times N$ correlation matrix $\mathbf{G} = (\mathbf{V}')^{\mathsf{T}} \cdot \mathbf{V}'$ (Equation (3.18)), while for linear PCA, SVD is implemented on the $2L \times 2L$ covariance matrix $\mathbf{S} = \mathbf{V}' \cdot (\mathbf{V}')^{\mathsf{T}}$.[7]

Indeed, it can be shown that conducting SVD on the covariance matrix can also generate the same set of POD modes. From Equation (3.15), recall that $\mathbf{Q}$ is a $N \times N$ unitary matrix and both $\mathbf{\Gamma}$ and $\mathbf{\Lambda}$ are $N \times N$ square diagonal matrices where $\mathbf{\Lambda} = \mathbf{\Gamma}^{\odot 2}$, one can derive:

$$\begin{aligned} \mathbf{S} = \mathbf{V}' \cdot \left(\mathbf{V}'\right)^{\mathsf{T}} = \mathbf{P} \cdot \mathbf{\Gamma} \cdot \mathbf{Q}^{\mathsf{T}} \cdot \left(\mathbf{P} \cdot \mathbf{\Gamma} \cdot \mathbf{Q}^{\mathsf{T}}\right)^{\mathsf{T}} \\ \mathbf{S} = \mathbf{P} \cdot \mathbf{\Gamma} \cdot \mathbf{Q}^{\mathsf{T}} \cdot \mathbf{Q} \cdot \mathbf{\Gamma}^{\mathsf{T}} \cdot \mathbf{P}^{\mathsf{T}} \\ \mathbf{S} = \mathbf{P} \cdot \mathbf{\Lambda} \cdot \mathbf{P}^{\mathsf{T}}, \end{aligned} \tag{3.29}$$

where $\mathbf{P}$ is the same $2L \times 2L$ unitary matrix as in Equation (3.15). This equation has the same form as the SVD results of the correlation matrix $\mathbf{G}$ (Equation 3.18), but instead solves the eigenvalue problem of the covariance matrix $\mathbf{S}$. As mentioned before in the discussion of Equation (3.15), the number of POD modes M (rank of the data matrix $\mathbf{V}'$) is restricted by the number of cycles N, and the POD mode matrix $\mathbf{\Phi}$ is the first M columns of the unitary matrix $\mathbf{P}$ since there are only M non-zero values in $\mathbf{\Lambda}$. Therefore, one can also obtain the same set of POD modes by conducting a SVD to the covariance matrix and solve for $\mathbf{P}$, though the computation cost will be higher due to the larger size of the covariance matrix.

With the flow field data arranged as Equation (3.11), the three equivalent algorithms of POD are summarised as follows:

[7] Conventionally a constant $\frac{1}{N}$ is multiplied to both the correlation and the covariance matrices. This constant is omitted in this section since it does not affect the resulting POD modes or coefficients.

Algorithm 1: direct method

1. Solve the eigenvalue problem of $\mathbf{V}'$ as presented in Equation (3.15);
2. Extract the POD modes $\mathbf{\Phi}$ from the unitary matrix $\mathbf{P}$;
3. Compute the POD mode coefficients using $\mathbf{C} = \mathbf{\Phi}^{\mathsf{T}} \cdot \mathbf{V}'$.

Algorithm 2: snapshot method (lowest computational cost)

1. Use Equation (3.17) to compute the correlation matrix $\mathbf{G}$;
2. Solve the eigenvalue problem of $\mathbf{G}$ as presented in Equation (3.18);
3. Use Equation (3.21) to calculate the POD modes;
4. Use Equation (3.19) to find the POD mode coefficients.

Algorithm 3: covariance matrix method (highest computational cost)

1. Use Equation (3.29) to compute the covariance matrix $\mathbf{S}$;
2. Solve the eigenvalue problem of $\mathbf{S}$ as presented in Equation (3.29);
3. Extract the POD modes $\mathbf{\Phi}$ from the unitary matrix $\mathbf{P}$;
4. Compute the POD mode coefficients using $\mathbf{C} = \mathbf{\Phi}^{\mathsf{T}} \cdot \mathbf{V}'$.

These three algorithms obtain the same set of POD modes and mode coefficients. The POD reconstruction (approximation) can then be performed using Equation (3.28).

3.2.2 Kernel principal component analysis

Since POD is a variant of the linear PCA, it can only perform a linear dimensionality reduction operation to the given data set, i.e., mapping the flow data in the real space onto a linear principal subspace. The turbulent flow in the engine cylinder may generate non-linearity in the data set. For instance, as discussed later in Chapter 6, the velocity components measured at the same location in the cylinder at the same engine phase for different cycles generally

follow a Gaussian distribution. As such, a non-linear PCA method is also of interest when filtering (de-noising) the measured in-cylinder flow fields. The kernel PCA (KPCA) offers a generalisation to the standard linear PCA, and enables the non-linear mapping of the input data using any chosen distributions [76, 94]. The following paragraphs briefly discuss the procedure of KPCA, a more detailed derivation can be found in [79].

Given a set of input data $\{\mathbf{x}_r\} \in \mathbb{R}^M$ with P parameters and R observations, the data matrix $\mathbf{x}$ can be arranged as:

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_r & \cdots & \mathbf{x}_R \end{bmatrix}, \tag{3.30}$$

where $\mathbf{x}$ is a $P \times R$ matrix whose rank is M ($M \leq \min\{P, R\}$), each $\mathbf{x}_r$ is a $P \times 1$ column array containing the values of P parameters in the ith observation. The set of data $\mathbf{x}$ can be mapped onto a higher dimensional space $\mathbb{R}^D$ ($D \gg M$), called the feature space, with an arbitrarily large (possibly infinite) number of dimensions, using a non-linear transformation function ψ:

$$\psi: \quad \mathbb{R}^M \xrightarrow{\text{mapping}} \mathbb{R}^D, \quad \mathbf{x}_r \xrightarrow{\text{mapping}} \psi(\mathbf{x}_r). \tag{3.31}$$

The transformation onto the feature space is expected to flatten the non-linear manifold where $\{\mathbf{x}_r\}$ lies. The transformed data, $\psi(\mathbf{x}_r)$, can therefore have linear dependencies in the feature space and then be decomposed by a standard linear PCA. The set of the transformed data $\psi(\mathbf{x})$ can be expressed by a $D \times R$ matrix:

$$\psi(\mathbf{x}) = \begin{bmatrix} \psi(\mathbf{x}_1) & \psi(\mathbf{x}_2) & \cdots & \psi(\mathbf{x}_r) & \cdots & \psi(\mathbf{x}_R) \end{bmatrix}, \tag{3.32}$$

and the covariance matrix $\widetilde{\mathbf{S}}$ of the transformed data is given by:

$$\widetilde{\mathbf{S}} = \frac{1}{N} \cdot \psi(\mathbf{x}) \cdot \psi(\mathbf{x})^{\mathsf{T}} = \frac{1}{N} \sum_{n=1}^{N} \left\{ \psi(\mathbf{x}_r) \cdot \psi(\mathbf{x}_r)^{\mathsf{T}} \right\}, \tag{3.33}$$

where $\widetilde{\mathbf{S}}$ is a $D \times D$ matrix. Analogous to the procedure of the linear PCA, the reconstruction of $\psi(\mathbf{x})$ onto a lower-dimensional space requires solving the eigenvalue problem of $\widetilde{\mathbf{S}}$, which can

be expressed as (suppose there are K non-zero eigenvalues of $\widetilde{\mathbf{S}}$):

$$\widetilde{\mathbf{S}} \cdot \widetilde{\mathbf{\Omega}} = \widetilde{\mathbf{\Omega}} \cdot \widetilde{\mathbf{\Lambda}}, \qquad \widetilde{\mathbf{S}} \cdot \widetilde{\mathbf{\Omega}}_k = \widetilde{\lambda}_k \cdot \widetilde{\mathbf{\Omega}}_k, \tag{3.34}$$

where $\widetilde{\mathbf{\Omega}} = \left[\widetilde{\mathbf{\Omega}}_1 \quad \widetilde{\mathbf{\Omega}}_2 \quad \cdots \quad \widetilde{\mathbf{\Omega}}_k \quad \cdots \quad \widetilde{\mathbf{\Omega}}_K\right]$ is the eigenvector matrix of $\widetilde{\mathbf{S}}$ with a size of $D \times K$, $\widetilde{\mathbf{\Lambda}}$ is a $K \times K$ diagonal matrix whose entries are eigenvalues of $\widetilde{\mathbf{S}}$, the $D \times 1$ column array $\widetilde{\mathbf{\Omega}}_k$ ($k = 1, 2, \cdots, K$) is the kth eigenvector of $\widetilde{\mathbf{S}}$, and $\widetilde{\lambda}_k$ is the kth largest eigenvalue (the kth entry of the diagonal matrix $\widetilde{\mathbf{\Lambda}}$) of $\widetilde{\mathbf{S}}$. Note that the maximum number of non-zero eigenvalues of $\widetilde{\mathbf{S}}$ (K) can exceed the rank of the original data matrix (M) since the size of $\widetilde{\mathbf{S}}$ (D) can be arbitrarily large, i.e., $D \geq K \gg M$. However, as will be discussed later in the discussion of Equation (3.42), the value K is restricted by the number of observations R (i.e., $K \leq R$).

The successful mapping of the original data requires prior knowledge of the non-linear transformation function ψ. However, it is often not obvious how to determine the exact ψ required to align the non-linear data set onto the feature space. In other words, the covariance matrix $\widetilde{\mathbf{S}}$ and its rank K are both unknown. Therefore, the "kernel trick" [76] is used to achieve both the non-linear mapping and the principal components projection in the feature space in one step without having to know the exact expression of ψ. One of the underlying properties in the eigenvalue problem of $\widetilde{\mathbf{S}}$ (Equation (3.34)) is that each eigenvector can be expressed as a linear combination of the transformed data, i.e.:

$$\widetilde{\mathbf{\Omega}}_k = \sum_{r=1}^{R} \left\{\psi\left(\mathbf{x}_r\right) \cdot \alpha_{r,k}\right\}, \tag{3.35}$$

where $\alpha_{r,k}$ is the coefficient for the rth column in the transformed data and the kth eigenvector of the covariance matrix. Therefore, Equation (3.34) can be re-written as:

$$\frac{1}{R}\sum_{i=1}^{R}\left\{\psi\left(\mathbf{x}_i\right) \cdot \psi\left(\mathbf{x}_i\right)^{\mathsf{T}}\right\} \cdot \sum_{j=1}^{R}\left\{\psi\left(\mathbf{x}_j\right) \cdot \alpha_{j,k}\right\} = \widetilde{\lambda}_k \cdot \sum_{i=1}^{R}\left\{\psi\left(\mathbf{x}_i\right) \cdot \alpha_{i,k}\right\}. \tag{3.36}$$

Multiply both sides of Equation (3.36) by $\psi(\mathbf{x}_s)^{\mathsf{T}}$ gives:

$$\psi(\mathbf{x}_s)^{\mathsf{T}} \cdot \frac{1}{R} \sum_{i=1}^{R} \left\{ \psi(\mathbf{x}_i) \cdot \psi(\mathbf{x}_i)^{\mathsf{T}} \right\} \cdot \sum_{j=1}^{R} \{ \psi(\mathbf{x}_j) \cdot \alpha_{j,k} \} = \psi(\mathbf{x}_s)^{\mathsf{T}} \cdot \widetilde{\lambda}_k \cdot \sum_{i=1}^{R} \{ \psi(\mathbf{x}_i) \cdot \alpha_{i,k} \}. \tag{3.37}$$

The kernel trick requires the use of the kernel function, which is defined as:

$$\kappa(\mathbf{x}_i, \mathbf{x}_j) := \psi(\mathbf{x}_i)^{\mathsf{T}} \cdot \psi(\mathbf{x}_j), \tag{3.38}$$

and thus Equation (3.37) can be expressed as:[8]

$$\mathbf{K} \cdot \mathbf{K} \cdot \boldsymbol{\alpha}_k = R \cdot \widetilde{\lambda}_k \cdot \mathbf{K} \cdot \boldsymbol{\alpha}_k \tag{3.39}$$

$$\mathbf{K} \cdot \boldsymbol{\alpha}_k = R \cdot \widetilde{\lambda}_k \cdot \boldsymbol{\alpha}_k, \tag{3.40}$$

where $\mathbf{K}$ is a $R \times R$ symmetrical matrix (called the kernel matrix), whose ith row and jth column element is $\mathbf{K}_{i,j} = \kappa(\mathbf{x}_i, \mathbf{x}_j)$, and $\boldsymbol{\alpha}_k = \begin{bmatrix} \alpha_{1,k} & \alpha_{2,k} & \cdots & \alpha_{r,k} & \cdots & \alpha_{R,k} \end{bmatrix}^{\mathsf{T}}$ is a $R \times 1$ column array containing all the R coefficients for the kth eigenvector ($\widetilde{\boldsymbol{\Omega}}_k$) of the covariance matrix $\widetilde{\mathbf{S}}$ (*cf* Equations (3.34) and (3.35)). The elements in the kernel matrix $\mathbf{K}$ can be directly computed by the input data using different types of kernels. A Gaussian kernel is chosen for this study:

$$\mathbf{K}_{i,j} = \kappa(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(\frac{-\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2} \right), \tag{3.41}$$

where $\kappa(\mathbf{x}_i, \mathbf{x}_j)$ denotes the kernel between a pair of observations $\mathbf{x}_i$ and $\mathbf{x}_j$, $\exp(\cdot)$ denotes the exponential function, σ is a parameter that needs to be chosen based on the data set. The choice of σ is typically based on the distance between the data points, such as the average distance to the nearest neighbour [95] and the median of all the distances between any two data points [96]. In this thesis, σ was chosen to be half of the median of all the distances between any two different

[8] The removal of $\mathbf{K}$ on both sides of Equation (3.39) only excludes the eigenvectors of $\mathbf{K}$ that have zero eigenvalues [79], and hence it does not affect the principal components projection in the following steps.

data points. Further investigations can be made to examine the effect of different choices of σ on each data set.

With the kernel matrix $\mathbf{K}$ given by the input data, each $\boldsymbol{\alpha}_k$ can be solved from Equation (3.40). Note that Equation (3.40) is the eigenvalue problem of the kernel matrix $\mathbf{K}$, which can again be solved by a SVD approach (the same procedure for POD and linear PCA, *cf* Equations (3.15), (3.18) and (3.29)):

$$\mathbf{K} = \widetilde{\mathbf{Q}} \cdot \widetilde{\mathbf{E}} \cdot \widetilde{\mathbf{Q}}^{\mathsf{T}} = \boldsymbol{\alpha} \cdot \left(R \cdot \widetilde{\boldsymbol{\Lambda}} \right) \cdot \boldsymbol{\alpha}^{\mathsf{T}}, \tag{3.42}$$

where $\widetilde{\mathbf{Q}}$ is a $R \times R$ unitary matrix, $\widetilde{\mathbf{E}}$ is a $R \times R$ diagonal matrix containing sorted eigenvalues of $\mathbf{K}$ in a descending order. Suppose there are K non-zero eigenvalues of $\mathbf{K}$ along the main diagonal of the matrix $\widetilde{\mathbf{E}}$ ($\widetilde{\varepsilon}_k$, where $k = 1, 2, \cdots, K$), one can find $\boldsymbol{\alpha} = \begin{bmatrix} \boldsymbol{\alpha}_1 & \boldsymbol{\alpha}_2 & \cdots & \boldsymbol{\alpha}_k & \cdots & \boldsymbol{\alpha}_K \end{bmatrix}$, which has a size of $R \times K$, from the first K columns of the $R \times R$ unitary matrix $\widetilde{\mathbf{Q}}$. It should be noted that the diagonal matrices $\widetilde{\mathbf{E}}$ and $\widetilde{\boldsymbol{\Lambda}}$, which are respectively the eigenvalue matrices of the kernel matrix $\mathbf{K}$ and the covariance matrix $\widetilde{\mathbf{S}}$, have the same numbers of non-zero entries, and the non-zero values differ by a constant factor of R, i.e., $\widetilde{\varepsilon}_k = R \cdot \widetilde{\lambda}_k$. As such, $\mathbf{K}$ and $\widetilde{\mathbf{S}}$ have the same number of non-zero eigenvalues. In other words, the value K (*cf* Equation (3.34)) can be found by computing the rank of the kernel matrix $\mathbf{K}$, and its maximum is R (the size of $\mathbf{K}$).

The column arrays $\boldsymbol{\alpha}_k$ computed by Equation (3.42) are orthonormal, i.e., $\boldsymbol{\alpha}_i^{\mathsf{T}} \cdot \boldsymbol{\alpha}_j = \delta_{ij}$, since they are extracted from a unitary matrix $\widetilde{\mathbf{Q}}$. However, the actual requirement is that the eigenvectors of the covariance matrix $\widetilde{\mathbf{S}}$ are orthonormal. Recall that from Equation (3.35), $\boldsymbol{\alpha}_k$ is the coefficients for the kth eigenvector of the covariance matrix ($\widetilde{\boldsymbol{\Omega}}_k$), the normalisation of $\boldsymbol{\alpha}_k$ can be achieved by:

$$\widetilde{\boldsymbol{\Omega}}_k^{\mathsf{T}} \cdot \widetilde{\boldsymbol{\Omega}}_k = \sum_{i=1}^{R} \sum_{j=1}^{R} \left\{ \psi\left(\mathbf{x}_i\right)^{\mathsf{T}} \cdot \psi\left(\mathbf{x}_j\right) \cdot \alpha_{i,k}^{*} \cdot \alpha_{j,k}^{*} \right\} = 1, \tag{3.43}$$

where the superscript * denote the normalised parameters such that the eigenvectors in the feature space (i.e., $\widetilde{\mathbf{\Omega}}_k$) are orthonormal. Note that although the transformation function ψ is unknown, the term $\psi\left(\mathbf{x}_i\right)^{\mathsf{T}} \cdot \psi\left(\mathbf{x}_j\right)$ is given by the kernel function (*cf* Equations (3.38) and (3.41)). Therefore, Equation (3.43) can be further simplified (using also Equations (3.40) and (3.42)):

$$\begin{aligned}
\widetilde{\mathbf{\Omega}}_k^{\mathsf{T}} \cdot \widetilde{\mathbf{\Omega}}_k = \sum_{i=1}^{R}\sum_{j=1}^{R}\left\{\kappa\left(\mathbf{x}_i, \mathbf{x}_j\right) \cdot \alpha_{i,k}^* \cdot \alpha_{j,k}^*\right\} &= 1 \\
\left(\boldsymbol{\alpha}_k^*\right)^{\mathsf{T}} \cdot \mathbf{K} \cdot \boldsymbol{\alpha}_k^* &= 1 \\
R \cdot \widetilde{\lambda}_k \cdot \left(\boldsymbol{\alpha}_k^*\right)^{\mathsf{T}} \cdot \boldsymbol{\alpha}_k^* &= 1 \\
\widetilde{\varepsilon}_k \cdot \left(\boldsymbol{\alpha}_k^*\right)^{\mathsf{T}} \cdot \boldsymbol{\alpha}_k^* &= \boldsymbol{\alpha}_k^{\mathsf{T}} \cdot \boldsymbol{\alpha}_k \\
\boldsymbol{\alpha}_k^* &= \left(\widetilde{\varepsilon}_k\right)^{-\frac{1}{2}} \cdot \boldsymbol{\alpha}_k, \qquad (3.44)
\end{aligned}$$

where both $\widetilde{\varepsilon}_k$ and $\boldsymbol{\alpha}_k$ are known from the SVD result of the kernel matrix (Equation (3.42)). Replacing $\boldsymbol{\alpha}_k$ with $\boldsymbol{\alpha}_k^*$ in Equation (3.35) ensures the orthonormality of the eigenvectors $\widetilde{\mathbf{\Omega}}_k$ in the feature space.

Despite the eigenvectors in the feature space ($\widetilde{\mathbf{\Omega}}_k$, *cf* Equations (3.34) and (3.35)) cannot be explicitly computed due to the unknown transformation function ψ, the projection of each data point $\psi\left(\mathbf{x}_i\right)$ onto each eigenvectors $\widetilde{\mathbf{\Omega}}_k$ in the feature space can be found by (note that the normalised $\boldsymbol{\alpha}_k^*$ is used hereinafter):

$$\beta_{i,k} = \psi\left(\mathbf{x}_i\right)^{\mathsf{T}} \cdot \widetilde{\mathbf{\Omega}}_k = \sum_{r=1}^{R}\left\{\psi\left(\mathbf{x}_i\right)^{\mathsf{T}} \cdot \psi\left(\mathbf{x}_r\right) \cdot \alpha_{r,k}^*\right\} = \sum_{r=1}^{R}\left\{\kappa\left(\mathbf{x}_i, \mathbf{x}_r\right) \cdot \alpha_{r,k}^*\right\}, \qquad (3.45)$$

or in matrix form: $\boldsymbol{\beta} = \mathbf{K} \cdot \boldsymbol{\alpha}^*$, where $\beta_{i,k}$ (the ith row and kth column element of $\boldsymbol{\beta}$) is the projection of the ith data point onto the kth eigenvector in the feature space.

In the derivations above, an assumption has been made that the transformed data in the feature space is centred such that $\frac{1}{R}\sum_{r=1}^{R}\psi\left(\mathbf{x}_r\right) = 0$. This may not be the case for all the input data, even if the original data were centred in the input space (i.e., $\frac{1}{R}\sum_{r=1}^{R}\mathbf{x}_r = 0$), since the unknown mapping ψ may change the centring property. To generalise the use of KPCA to all

the data (no matter whether they are centred in the feature space or not), Schölkopf *et al.* [76] proposed to modify the kernel matrix $\mathbf{K}$ to the Gram matrix $\widehat{\mathbf{K}}$, which is given by:

$$\widehat{\mathbf{K}} = \mathbf{K} - \mathbf{1}_R \cdot \mathbf{K} - \mathbf{K} \cdot \mathbf{1}_R + \mathbf{1}_R \cdot \mathbf{K} \cdot \mathbf{1}_R, \tag{3.46}$$

where $\mathbf{1}_R$ is a $R \times R$ matrix whose elements all equal to $\frac{1}{R}$. The Gram matrix $\widehat{\mathbf{K}}$ can therefore be computed from the kernel matrix, which is calculated from the input data set (*cf* Equation (3.41)), without having to know the exact transformation function ψ. For a general data set whose mapping in feature space is not necessarily centred, the Gram matrix is used to replace the kernel matrix $\mathbf{K}$ in all the above-mentioned procedures — when obtaining the coefficients $\boldsymbol{\alpha}_k$ (Equation (3.42)), and when evaluating the projection of each data point onto the eigenvectors in the feature space (Equation (3.45)).

A dimensionality reduction process can be performed by taking the first d projections (corresponding to the highest d eigenvalues), which is similar to the procedure used for the linear PCA (*cf* Equation (3.28)):

$$\psi\left(\mathbf{x}_i\right) = \sum_{k=1}^{K} \left(\beta_{i,k} \cdot \widetilde{\boldsymbol{\Omega}}_k\right) \approx \sum_{k=1}^{d} \left(\beta_{i,k} \cdot \widetilde{\boldsymbol{\Omega}}_k\right), \tag{3.47}$$

where the data in the feature space is approximated using $d < K$ dimensions.

It should be noted that, however, the projection process is performed in the feature space with unknown transformation function ψ and eigenvectors $\widetilde{\boldsymbol{\Omega}}_k$, and thus it cannot provide the reconstructed data in the original physical space (called the "pre-images"). Mika *et al.* [77] developed a backward mapping algorithm that finds the pre-images for Gaussian kernels. Suppose the pre-image of the approximated $\psi\left(\mathbf{x}_i\right) \in \mathbb{R}^D$ using the highest d eigenvalues is $\mathbf{z}_i^{(d)} \in \mathbb{R}^M$, the following minimisation problem needs to be solved:

$$\min \left\| \psi\left(\mathbf{z}_i^{(d)}\right) - \sum_{k=1}^{d} \left(\beta_{i,k} \cdot \widetilde{\boldsymbol{\Omega}}_k\right) \right\|^2, \tag{3.48}$$

where $\psi\left(\mathbf{z}_i^{(d)}\right)$ is the mapped $\mathbf{z}_i^{(d)}$ in the feature space. In the case of a Gaussian kernel, this minimisation problem can be simplified to [77]:

$$\mathbf{z}_i^{(d)} = \frac{\sum_{j=1}^{R} \left\{ \eta_j^{(d)} \cdot \exp\left(\frac{-\left\|\mathbf{z}_i - \mathbf{x}_j\right\|^2}{2\sigma^2} \right) \cdot \mathbf{x}_j \right\}}{\sum_{j=1}^{R} \left\{ \eta_j^{(d)} \cdot \exp\left(\frac{-\left\|\mathbf{z}_i - \mathbf{x}_j\right\|^2}{2\sigma^2} \right) \right\}}, \tag{3.49}$$

where $\eta_j^{(d)} = \sum_{k=1}^{d} \left(\beta_{j,k} \cdot \alpha_{j,k}^* \right)$. The value $\eta_j^{(d)}$ for different data points j and different orders of approximation d can be computed from the known $\boldsymbol{\alpha}^*$ (Equation (3.44)) and $\boldsymbol{\beta}$ (Equation (3.45)) matrices. Note that all the parameters in Equation (3.49) are given, except for $\mathbf{z}_i^{(d)}$ itself. The column array $\mathbf{z}_i^{(d)}$ (with a size of $P \times 1$) can be solved by an iterative process:

$$\mathbf{z}_{i,t+1}^{(d)} = \frac{\sum_{j=1}^{R} \left\{ \eta_j^{(d)} \cdot \exp\left(\frac{-\left\|\mathbf{z}_{i,t}^{(d)} - \mathbf{x}_j\right\|^2}{2\sigma^2} \right) \cdot \mathbf{x}_j \right\}}{\sum_{j=1}^{R} \left\{ \eta_j^{(d)} \cdot \exp\left(\frac{-\left\|\mathbf{z}_{i,t}^{(d)} - \mathbf{x}_j\right\|^2}{2\sigma^2} \right) \right\}}, \tag{3.50}$$

where $\mathbf{z}_{i,t}^{(d)}$ and $\mathbf{z}_{i,t+1}^{(d)}$ denote, respectively, the choice of $\mathbf{z}_i^{(d)}$ in time step t and $(t+1)$. The initial choice of $\mathbf{z}_i^{(d)}$ and the stopping criterion need to be chosen according to the data set to ensure a converged result. For the purpose of de-noising, the original data $\mathbf{x}_i$ can be used as the starting point the the iteration [77]. For the flow field data set presented in this thesis, each $\mathbf{z}_i^{(d)}$ consists of P vector magnitudes measured at various cycles or different locations (depends on the arrangement of the data matrix, which will be discussed in detail later in this section). Therefore, the stopping criterion for the data set in this thesis is that the maximum change of vector magnitudes in sequential time steps is smaller than 0.01 m/s. The converged result of $\mathbf{z}_i^{(d)}$ at the chosen dimension d is considered as the dth-order KPCA approximation of the original data $\mathbf{x}_i$. Note that the original data was centred around the ensemble mean, and thus the ensemble

mean needs to be added back to the KPCA approximations (the same procedure as used when evaluating the POD approximations in Equation (3.28)). The ensemble mean is the zeroth-order KPCA approximation for all the cycles.

The algorithm for KPCA with a Guassian kernel can be summarised as follows:

1. Use Equation (3.41) to compute the kernel matrix $\mathbf{K}$ (with a properly chosen σ) from the original data $\mathbf{x}_i$, where $\mathbf{x}_i$ is centred around the ensemble mean;
2. Use Equation (3.46) to convert the kernel matrix $\mathbf{K}$ to the Gram matrix $\widehat{\mathbf{K}}$;
3. Solve the eigenvalue problem of the Gram matrix $\widehat{\mathbf{K}}$ (Equation (3.42)) to obtain $\boldsymbol{\alpha}$ and $\widetilde{\mathbf{E}}$ (note that the kernel matrix $\mathbf{K}$ needs to be replaced by the Gram matrix $\widehat{\mathbf{K}}$);
4. Use Equation (3.44) to normalise $\boldsymbol{\alpha}$ to $\boldsymbol{\alpha}_k^*$;
5. Use Equation (3.45) to obtain projections in the feature space $\boldsymbol{\beta}$ (note that the kernel matrix $\mathbf{K}$ needs to be replaced by the Gram matrix $\widehat{\mathbf{K}}$);
6. Choose a desired dimension d and use Equation (3.50) to obtain $\mathbf{z}_i^{(d)}$ (the dth-order KPCA approximation of the original data $\mathbf{x}_i$);
7. Re-introduce the ensemble mean to $\mathbf{z}_i^{(d)}$.

The paragraphs above present a generalised KPCA algorithm with a Gaussian kernel that can be implemented on any data set. The data set of the fluctuation flow fields can be used in two different ways as the input of the KPCA algorithm:

Approach A

The data set is arranged as a $2L \times N$ matrix (i.e., $\mathbf{V}'$ in Equation (3.11)), each data point $\mathbf{x}_i$ in Equation (3.30) refers to the velocity magnitudes measured at different locations of the same flow field, and has a size of $2L \times 1$ (the same as $\mathbf{V}'_n$). The resulting kernel matrix has a size of $N \times N$.

Approach B

The data set is arranged as a $N \times 2L$ matrix (i.e., the transpose of $\mathbf{V}'$, denote $(\mathbf{V}')^\mathsf{T}$), each data

point $\mathbf{x}_i$ refers to the velocity magnitudes measured at a fixed location within the flow field in different engine cycles, and has a size of $N \times 1$. The resulting kernel matrix has a size of $2L \times 2L$.

As mentioned in the introduction (Section 1.5), KPCA approximations filter out the additive noise (whose type depends on the choice of kernel) from the original data [77]. Therefore, the above-mentioned two data set arrangements have different meanings in terms of the type of noise being filtered. When taking each cycle as a data point (Approach A), KPCA aims to filter out the spatial noise within each flow field. On the other hand, if the velocity magnitudes at a specified grid point are considered as a data point (Approach B), KPCA seeks to filter out the temporal noise at each location among different cycles. Since POD treats each cycle as a data point and the velocity components at a specific location as parameters, the former approach (Approach A) was chosen in this thesis to enable a direct comparison with POD.

3.3 Quantitative metrics for comparison of vector fields

The validation of a flow simulation model requires frequent comparisons between the simulated and measured flow fields at different flow conditions. The averaged or reconstructed flow fields by any of the techniques discussed in the previous sections (Sections 3.1 and 3.2) need to be compared with each other as well. Therefore, it is necessary to develop some metrics for quantitative comparisons of any vector fields, without having to manually perform inspections on numerous individual vector fields. In this section, four different vector field comparison metrics are introduced.

3.3.1 Relevance index

In order to quantify the directional similarity between two vector fields, Liu and Haworth [97] proposed the use of a normalised scalar product of the two vector fields, and termed it the

relevance index (RI):

$$\mathrm{RI} = \frac{\langle \mathbf{A}, \mathbf{B} \rangle}{\|\mathbf{A}\| \cdot \|\mathbf{B}\|}, \tag{3.51}$$

where $\langle \mathbf{A}, \mathbf{B} \rangle$ is the inner product of vector fields $\mathbf{A}$ and $\mathbf{B}$, and $\|\mathbf{A}\|$ represents the $\mathbb{L}^2$ norm of $\mathbf{A}$; each vector field is re-arranged to a $2L \times 1$ column array in the calculation (the same procedure as performed in Equations (3.1)–(3.3)). The RI is a single number that ranges from -1 to $+1$, in which $+1$ means the two vector fields are in perfect alignment.

The use of RI provides opportunities to use a single number to quantify the differences between any two vector fields. It would be useful, on the other hand, to have the spatial information about how well the two flow fields are aligned and where the major differences occur. To achieve this, Chen *et al.* [98] and Zhao *et al.* [14] extended the idea of the RI to point-to-point comparisons, and Zhao *et al.* [14] termed it the local structure index (LSI):

$$(\mathrm{LSI})_\ell = \frac{\langle \mathbf{A}_\ell, \mathbf{B}_\ell \rangle}{\|\mathbf{A}_\ell\| \cdot \|\mathbf{B}_\ell\|}, \tag{3.52}$$

where $(\mathrm{LSI})_\ell$ is the local structure index at location ℓ, $\mathbf{A}_\ell$ is the vector at location ℓ of the vector field $\mathbf{A}$. Instead of a scalar value for RI, the LSI is a scalar field which provides local alignment information for vector pairs at the same spatial location from two vector fields. It should be noted that the spatially-averaged value of an LSI field for two given vector fields is not equivalent to the RI value for the same pair of vector fields — LSI treats each vector pair within the flow fields independently, while the calculation of RI regards the entire flow field as a whole by re-organising the two fields being compared into respective vectors with many dimensions (i.e., $2L$ dimensions for a two-component flow field containing L locations). In other words, LSI is not sensitive to any magnitude differences between the vectors. For instance, if the vectors in two fields are respectively in the same direction but have different magnitudes at each location, LSI produces a unity field, which trivially leads to one for the spatially-averaged value. The RI value, however, accounts for the magnitude differences and is not equal to one unless the two fields are strictly proportional to each other.

In-cylinder flow in a modern GDI engine is often characterised by strong tumble and/or swirl motions to enhance fuel-air mixing, and thus the in-cylinder flow fields usually contain vortices and the flow speeds vary significantly at different locations. As the LSI has no dependence on the magnitudes of the vectors inside each field being compared, the alignment of low magnitude vectors carries the same weight as that of high magnitude vectors. The reversal of the flow direction and rapid flow speed change across a vortex centre cause even very small differences in the vortex centre location to result in very misaligned vectors (high local LSI values). This undesired sensitivity in the low speed regions may lead to misleading conclusions for the validity of simulation results when compared with the experimental data. To enable a point-wise comparison which is sensitive to the alignment of both large and small vector magnitudes, Scott and Willman [99, 100] proposed the use of weightings based on the local vector magnitude compared to the flow field as a whole: the weighted relevance index (WRI), the weighted magnitude index (WMI) and the combined magnitude and relevance index (CMRI), which are discussed in the following sections.

3.3.2 Weighted relevance index

The weighted relevance index incorporates a misalignment penalty based on the LSI values at each spatial location for a pair of flow fields. This penalty is then weighted by the normalised local magnitude of each field which suppresses the penalty for small local velocities. For a given angular difference, this weighting assigns more importance to misaligned high magnitude vectors (a large absolute magnitude difference) than misaligned low magnitude vectors. The WRI value at location ℓ in the vector field is calculated by:

$$(\mathrm{WRI})_\ell = \frac{1-(\mathrm{LSI})_\ell}{2} \times \frac{\|\mathbf{A}_\ell\|}{\mathrm{median}(\|\mathbf{A}_\ell\|)} \times \frac{\|\mathbf{B}_\ell\|}{\mathrm{median}(\|\mathbf{B}_\ell\|)}, \tag{3.53}$$

where $\mathrm{median}(\|\mathbf{A}_\ell\|)$ is the median of the magnitudes for all vectors in $\mathbf{A}$ (including duplicate values, if applicable). The normalisation factor is chosen to be the median of the magnitudes

across the corresponding flow field for robustness against spurious vectors and small pockets of high or low magnitudes determining the weighting factor for the entire field. The WRI produces a field, whose values at each location l are all non-negative (may exceed 1 due to the weighting factor). A lower WRI value means the two fields are aligned better at a certain grid point.[9] The spatially-averaged WRI can also be computed to produce a single value that quantifies the overall directional similarity between two vector fields. A lower spatially-averaged WRI suggests a better overall alignment between the pair of vector fields.[10]

3.3.3 Weighted magnitude index

The weighted magnitude index is a method for quantifying differences in the magnitudes of vectors between two vector fields using a dimensionless metric, which is independent of the alignment of the vectors. The WMI value at location ℓ in the vector field is calculated by:

$$(\mathrm{WMI})_\ell = \left| \frac{\|\mathbf{A}_\ell\| - \|\mathbf{B}_\ell\|}{\mathrm{median}(\|\mathbf{A}_\ell\|, \|\mathbf{B}_\ell\|)} \right|, \tag{3.54}$$

where $|\cdot|$ denotes the absolute value, $\mathrm{median}(\|\mathbf{A}_\ell\|, \|\mathbf{B}_\ell\|)$ is the median of the magnitudes for all vectors in $\mathbf{A}$ and $\mathbf{B}$ (including duplicate values, if applicable). Similarly to the WRI (Equation (3.53)), the values of the WMI are normalised by the median value of the vector magnitudes across both vector fields, in order to provide robustness of the normalisation against spurious vectors and small pockets of high or low magnitudes as before. The WMI is also a scalar field for any two given vector fields, whose values are all non-negative (they may exceed 1 due to the weighting factor). Flow fields with similar local speeds generate low WMI values at the evaluated grid points, and the WMI values at different locations can be spatially

[9] Note that this is different from the LSI (Equation (3.52)) — a lower LSI value means the two fields have a worse alignment at the specified location.

[10] Again, this is different from the RI (Equation (3.51)), where a lower RI value means a poor overall alignment between the two fields.

averaged to provide a single value that quantifies the overall quality of magnitude alignment between the pair of flow fields.

3.3.4 Combined magnitude and relevance index

The WRI and WMI quantify the differences in alignment and magnitude at each location of any two vector fields, respectively. It is also useful to have a single metric which accounts for both differences while retaining the spatial information. One straight-forward option is taking a summation of the WRI and WMI values at each location to produce a new metric. However, the values of WRI and WMI often have very different distributions and ranges for a given set of flow field comparisons, and hence a normalisation process is needed before the summation. The normalisation process is given by:

$$(\mathrm{WRI}_{\mathrm{norm}})_\ell = \frac{(\mathrm{WRI})_\ell - (\mathrm{WRI})_{[0.02]}}{(\mathrm{WRI})_{[0.98]} - (\mathrm{WRI})_{[0.02]}}, \quad (\mathrm{WMI}_{\mathrm{norm}})_\ell = \frac{(\mathrm{WMI})_\ell - (\mathrm{WMI})_{[0.02]}}{(\mathrm{WMI})_{[0.98]} - (\mathrm{WMI})_{[0.02]}}, \quad (3.55)$$

where $(\mathrm{WRI})_{[0.02]}$ represents the 2% quantile of the WRI values (i.e., the threshold value of the WRI below which the best matched 2% of values occur), $(\mathrm{WRI})_{[0.98]}$ represents the 98% quantile of the WRI values (i.e., the threshold value of the WRI above which the worst matched 2% of values occur). This process ensures the central 96% of the WRI and WMI values lie within the range 0 to 1 after the normalisation. For a given set of flow field comparisons, the values of quantiles are constants. Therefore, the grid points with larger WRI or WMI values still have larger $\mathrm{WRI}_{\mathrm{norm}}$ and $\mathrm{WMI}_{\mathrm{norm}}$ values. The combined magnitude and relevance index at location l in the vector field can therefore be calculated by:

$$(\mathrm{CMRI})_\ell = \frac{(\mathrm{WRI}_{\mathrm{norm}})_\ell + (\mathrm{WMI}_{\mathrm{norm}})_\ell}{2}. \quad (3.56)$$

In this way the CMRI values retain sensitivity to both alignment and magnitude differences relevant to the flow fields of interest by ensuring neither the WRI and WMI metric values dominate the contribution to the CMRI.

Chapter 4

Validation of flow model using PIV ensemble mean

In this chapter, the engine operation repeatability and flow structure consistency among multiple PIV measurement planes are firstly discussed. The evolution of PIV ensemble mean flow fields during the intake and compression strokes is illustrated by the data at an example test point. Visual inspections on the PIV and CFD flow fields are then performed on two test points under different engine loads. The CFD flow model is then validated using the PIV ensemble mean flow fields at these two test points. The differences between the measured and simulated flow fields are quantified using the vector field comparison metrics defined in Section 3.3. Several parts of this chapter are adapted from a paper [1] first-authored by the author of this thesis with additional information and more detailed discussions included.

4.1 Engine operation consistency and repeatability

The collection of the multi-plane and multi-test-point data set involves months of work, and tests were conducted on a plane-to-plane basis, in order to minimize the optical alignment work and to maintain consistent location of the light sheet. As a result, experiments at the same test point but in different planes can be separated by weeks or even months. Additionally, the PIV recording at each test points in a specific plane, as mentioned in the previous section, had to be divided into several test runs since the camera memory is not large enough for recording 300 consecutive cycles. Therefore, it is important to check engine operation consistency among measurement planes and the experiment repeatability across test runs.

The pressure transducer installed in the cylinder head (see Appendix A) provides real-time in-cylinder pressure measurements at every crank angle degree, and the in-cylinder pressure serves as a good metric that quantifies the engine operation consistency. The in-cylinder pressure traces for four example test conditions and in four measurement planes are reported in Figure 4.1, where the traces are averaged for 100 or 150 engine cycles in a single test run. The four points differ in loads and compression ratios — the upper row points have a lower load (T1, see the baseline load in Table 2.3) than the bottom row points (T2), and the points on the left column have a lower compression ratio (Piston B, $11:1$) than the points on the right column (Piston C, $12.5:1$). As a result, the peak pressure increases from around 6 bar in the T1 Piston B case (Figure 4.1(a)), to about 13.5 bar at the T2 Piston C point (Figure 4.1(d)). For each condition, the in-cylinder pressure data when conducting PIV tests in different measurement planes are respectively represented in different colours, while different line styles are used to illustrate the pressure traces recorded in multiple test runs for the same PIV measurement plane.

The four test conditions all show a reasonably good experimental repeatability among test runs at each of the four measurement planes (different line styles, same colour). The maximum relative error in peak pressure is less than 1.1% for the two T1 cases, and just exceeds 1.2%

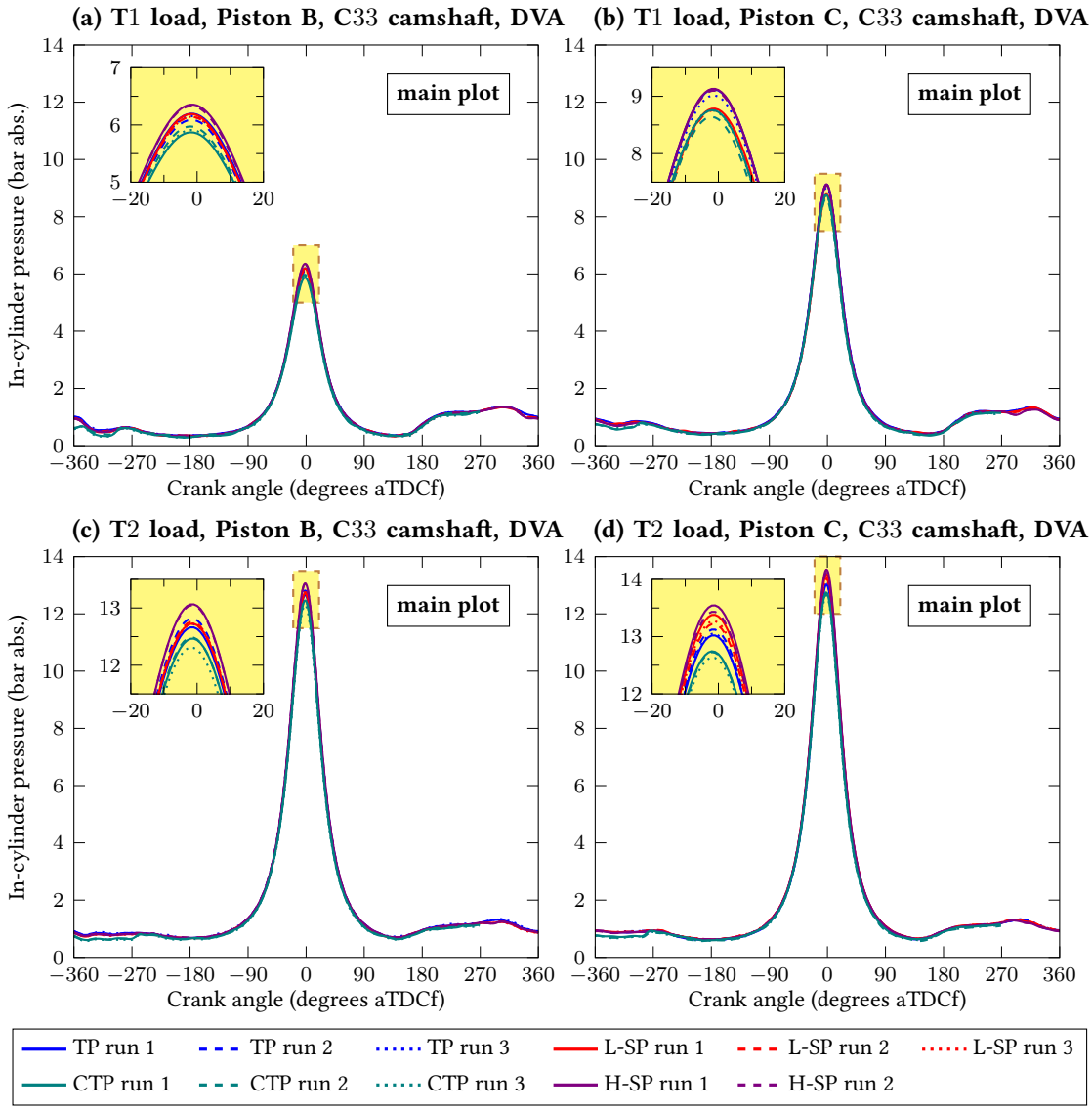


Figure 4.1: In-Cylinder pressure traces under different test conditions, each with a zoomed-in plot at high pressure regions. The zoomed-in region is bounded by a brown dashed box filled with light yellow colour in the corresponding main plot. The four test conditions differ in the engine loads and the piston equipped when the test was conducted. The C33 camshaft was used and the engine was operated with both intake valves open (i.e., DVA) for all four test conditions. The pressure data were recorded when the PIV measurements were taken in different planes: blue lines for the tumble plane (TP), dark green lines for the cross-tumble plane (CTP), red lines for the lower swirl plane (L-SP) and purple lines for the higher swirl plane (H-SP). The pressure data from different test runs are shown in different line styles (solid, dashed and dotted).

for the T2 ones. The consistency over various measurement planes (different colours), despite the greater error (7–8% in all cases), is still acceptable considering that the tests were separated by months. The change of peak pressure may be a combined result from piston ring ageing (which generates variations in sealing performance and blow-by losses) and minor differences in the cylinder volume (and hence the compression ratio) between Perspex and quartz annuli (see Figure A.4) — the TP and L-SP points were using the Perspex annulus, while the H-SP and CTP measurements were using the quartz one.

4.2 Multi-Plane PIV flow fields visualisation and consistency

The in-cylinder flow is highly turbulent and its structure varies from cycle to cycle, hence it is not practical to directly compare individual flow fields that were measured on different planes. Instead, as mentioned in Section 3.1, one can extract the major flow structures that are present in various cycles by applying the Reynolds decomposition [87] to the data set, and obtain the ensemble mean flow field. The Reynolds decomposition averages out the turbulence and cyclic variation in the flow, while retaining the overall flow direction and structure, and hence the resulting ensemble mean enables the flow structure visualisation and comparison among various measurement planes. In this section, the evolution of key in-cylinder flow structures viewed on different measurement planes is discussed, and the consistency in the ensemble means of multi-plane PIV flow fields is examined.

Figure 4.2 shows the isometric view of the PIV ensemble mean flow fields for an example test point at 290° bTDCf (during intake) and 70° bTDCf (during compression) across the lower swirl (x–y plane, where $z = -3.5$ mm), tumble (x–z plane, where $y = -1$ mm) and cross-tumble (y–z plane, where $x = 0$ mm) planes. The engine was equipped with Piston B and C33 camshaft, and operated under the T2 load (see Table 2.3) with both intake valves open (i.e., DVA). The in-cylinder pressure trace for this test point was reported in Figure 4.1(b). The flow fields at the two

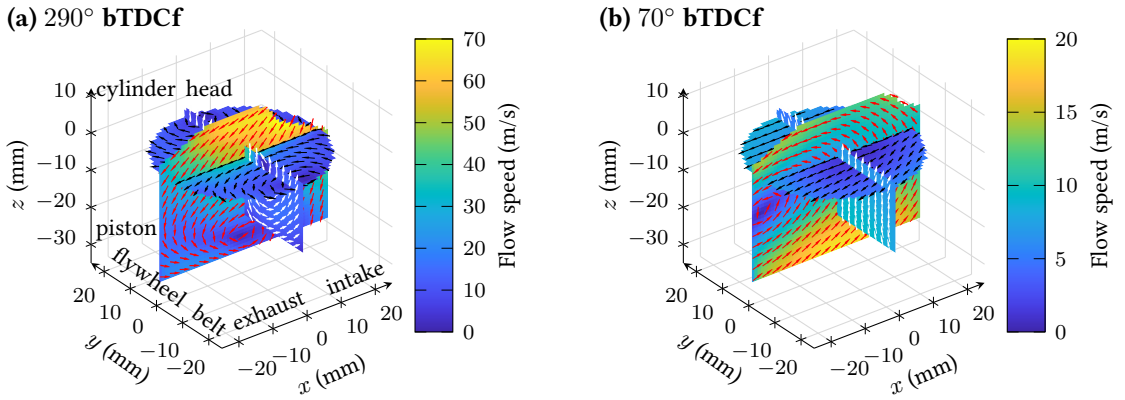


Figure 4.2: Isometric view of two-component in-plane velocities for PIV ensemble mean flow fields for a T2 load test point. The engine was equipped with Piston B and C33 camshaft and operated with both intake valves open (i.e., DVA). The in-plane vectors in each measurement plane are plotted in different colours – red vectors: tumble plane; black vectors: lower swirl plane; white vectors: cross-tumble plane. Note that the colour scales in the two plots are different due to a slower flow motion at 70° bTDCf (during the compression stroke) compared to at 290° bTDCf (during the intake stroke). PIV data in the higher swirl plane (3 mm above the lower swirl plane) are also available at this test point, but are not shown here for a clearer visualisation.

crank angles respectively illustrate two major features of the in-cylinder flow – the intake jet (at 290° bTDCf) and the tumble vortex (at 70° bTDCf). It is explicit that the planar PIV can only visualise a slice of the three-dimensional flow motion, and the following paragraphs discuss how these three-dimensional flow features evolve during the intake and compression strokes from a two-dimensional point of view.

Figures 4.3–4.9 exemplify the evolution of the in-cylinder flow field from early intake (290° bTDCf) to late compression (45° bTDCf) strokes in all four measurement planes for the T2 point. Note that the higher swirl plane data is only available in the late compression stroke (after 90° bTDCf).

During the early intake stroke (Figure 4.3), the strong intake jets from the two inlet valves dominate the flow in all the three available planes. On the tumble plane, the high speed region (colour scale in yellow) is concentrated under the intake valve (top-left corner) as the one inlet jet

overlays with the other, whereas in the cross-tumble and lower swirl planes, the two jets collide and combine into one, and the in-plane flow is quite symmetric with respect to the y axis. High speed regions can be observed in the upper middle of the cross-tumble plane and at the centre of the lower swirl plane, with the surrounding air "circling" around as induced by the intake jets.

The intake jets continue introducing kinetic energy to the in-cylinder flow, and when the intake valve closes (Figure 4.4), the flow induced by the intake jets develops into a so-called "tumble vortex", one of the major features for in-cylinder flow [101]. The vortex rotates with respect to the y axis, and its centre is slightly below the field of view in the tumble plane plot. On the cross-tumble plane, the flow is generally pointing downwards, while the magnitude is very small, meaning that the vortex centre is very close to the measured cross-tumble plane. The in-plane flows on the three planes show very good agreement — the high speed region is near the belt side and under the intake valve; the overall flow direction is from intake to the exhaust at the upper part of the cylinder (see tumble and lower swirl); and flows near the edge of the field of view are moving from the centreline to the sides of the cylinder (see cross-tumble and lower swirl).

As the flow evolves, the tumble vortex centre moves from the intake side to the exhaust side. This process generates a near-zero velocity region as the centre travels through the cross-tumble plane (Figure 4.5), and causes the flow direction to flip from downwards to upwards. At this specific test point, the flow flipping takes about 20 crank angle degrees around 210° bTDCf, and the upper and lower parts of the cross-tumble flow at this crank angle have opposite directions as the transition is happening.

When the piston moves to the bottom dead centre (BDC, Figure 4.6), the centre of the tumble vortex offsets towards the exhaust side, and the flow direction in the cross-tumble plane has fully turned upwards. At this crank angle, the vortex centre is cutting the lower edge of the fields of view on both the tumble and cross-tumble planes, while the centre, from the first glance, is slightly lower in the tumble plane measurements. The flow on the lower swirl plane is still

moving from the intake to the exhaust side, as the vortex centre is far below.

The next key feature is that the tumble vortex will be squeezed by the upward-moving piston motion, and thus its centre moves towards the cylinder head. This process spans over multiple crank angles in the compression stroke, and at 65° bTDCf (Figure 4.7), the tumble vortex centre reaches the lower swirl plane, which again generates an almost stationary flow region and causes the flow direction to flip. The flow in the central region of the lower swirl plane has just turned to the opposite direction (from exhaust to intake) as the vortex centre travelling through this plane, while the flow directions in the surrounding regions are yet to change. On the higher swirl plane, the flow remains moving from the intake to the exhaust side, since the vortex centre is still about 3 mm lower.

The vortex centre finally reaches the higher swirl plane at 55° bTDCf (Figure 4.8), and the flow flipping occurs again. The zero velocity region is now slightly near the exhaust side, and shows a good match with the tumble plane data. The lower swirl plane flow field, on the other hand, has almost changed to uniformly from exhaust to intake, confirming that the vortex centre is now higher than the lower swirl plane.

The flow flipping on the higher swirl plane finishes around 40° bTDCf (Figure 4.9) and at this late-compression phase, the effects of the pop-up on the shaped piston (see Figure A.5) can also be observed. The flow is pushed towards the centreline of the cylinder (see the cross-tumble and lower swirl planes), and accelerates as the piston moves upwards. This finding confirms that one of the original JLR engine design intentions is realised, that the flow under the spark plug is indeed accelerated in order to promote faster combustion.

The intersection of different planes offers an opportunity to quantify the consistency between the PIV experiments at the same test condition, as each pair of planes shares a single component of velocity along the line of intersection. For instance, as shown in Figure 4.2, the tumble plane and the lower swirl plane intersect with each other along a line that is parallel to the x axis, and one can compare the x-direction velocity components measured in these two

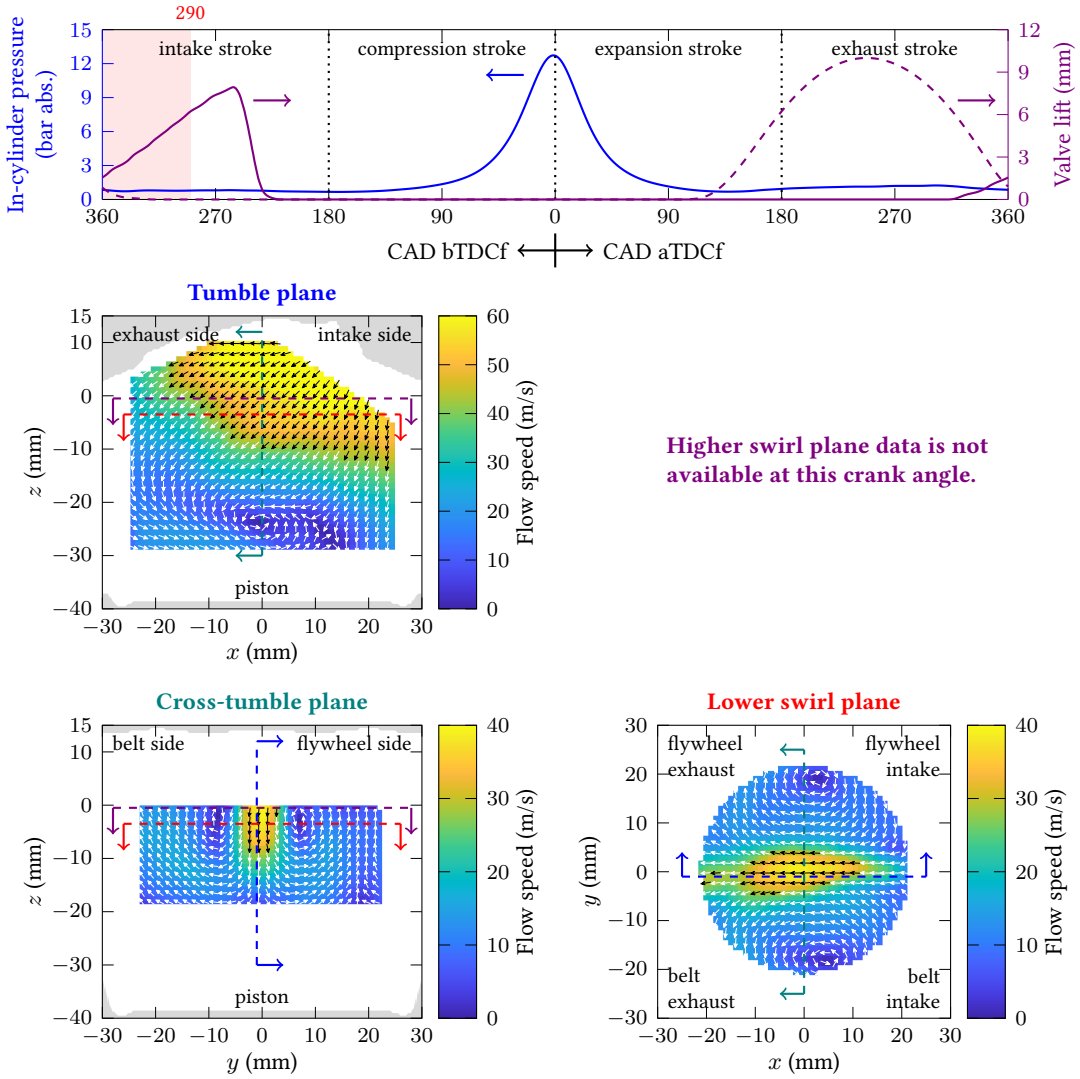


Figure 4.3: Ensemble mean flow fields and engine parameters at 290° bTDCf for the example test point (T2, Piston B, C33, DVA). The top plot shows the engine parameters: in-cylinder pressure trace (blue solid line), intake valve lift (purple solid line) and exhaust valve lift (purple dashed line). The bottom four plots show the PIV ensemble mean flow fields in different measurement planes, while the higher swirl plane data is only available in the late compression stroke (after 90° bTDCf). The coloured dashed lines in each flow field plot mark the corresponding locations of other planes, where the arrows illustrate the viewing direction. Each plane is marked by a different colour: tumble plane (blue), higher swirl plane (purple), cross-tumble plane (dark green) and lower swirl plane (red). Due to the differences in flow speeds, the colour scale may change in different planes (within this figure) or at different crank angles (among Figures 4.3–4.9).

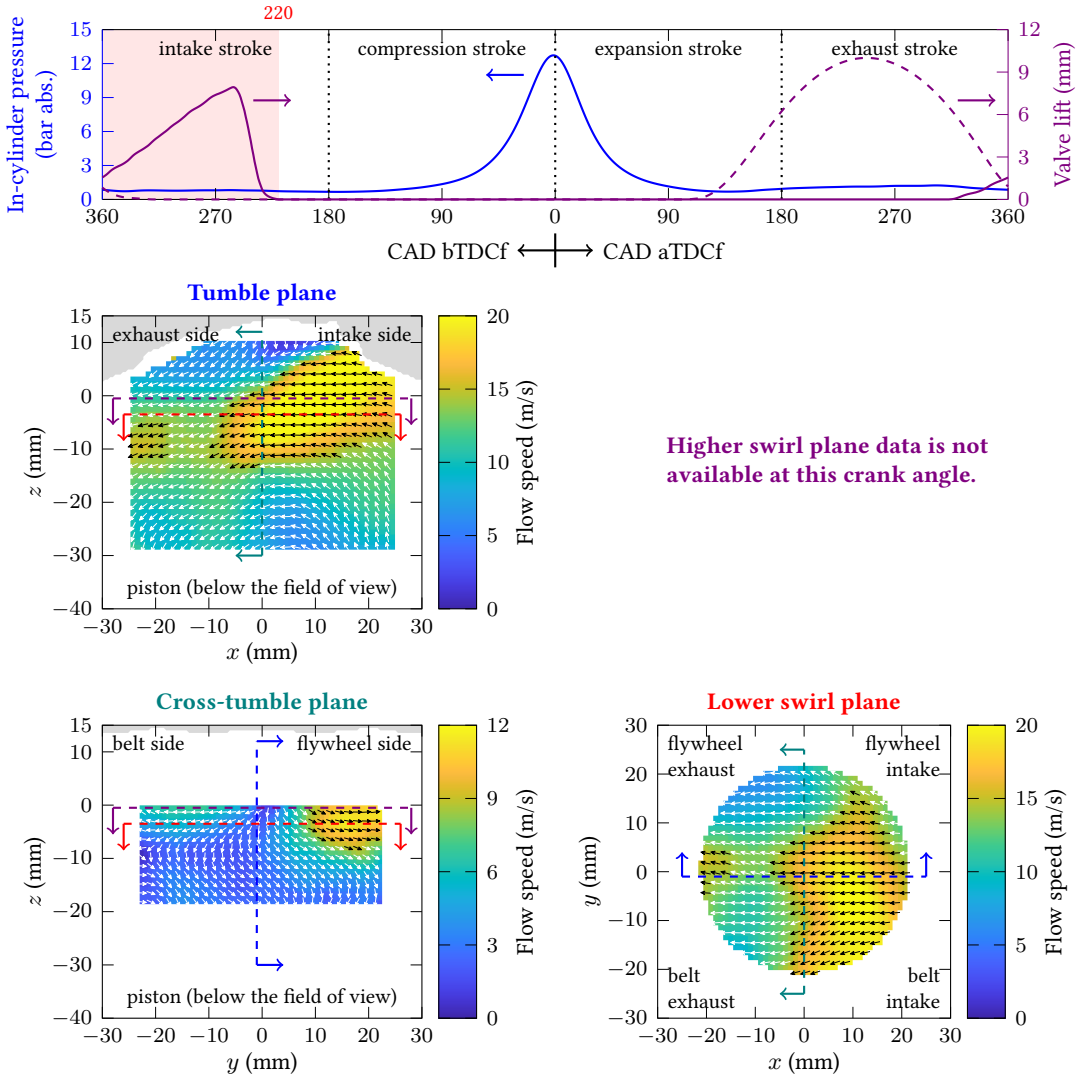


Figure 4.4: Ensemble mean flow fields and engine parameters at 220° bTDCf for the example test point (T2, Piston B, C33, DVA). The top plot shows the engine parameters: in-cylinder pressure trace (blue solid line), intake valve lift (purple solid line) and exhaust valve lift (purple dashed line). The bottom four plots show the PIV ensemble mean flow fields in different measurement planes, while the higher swirl plane data is only available in the late compression stroke (after 90° bTDCf). The coloured dashed lines in each flow field plot mark the corresponding locations of other planes, where the arrows illustrate the viewing direction. Each plane is marked by a different colour: tumble plane (blue), higher swirl plane (purple), cross-tumble plane (dark green) and lower swirl plane (red). Due to the differences in flow speeds, the colour scale may change in different planes (within this figure) or at different crank angles (among Figures 4.3–4.9).

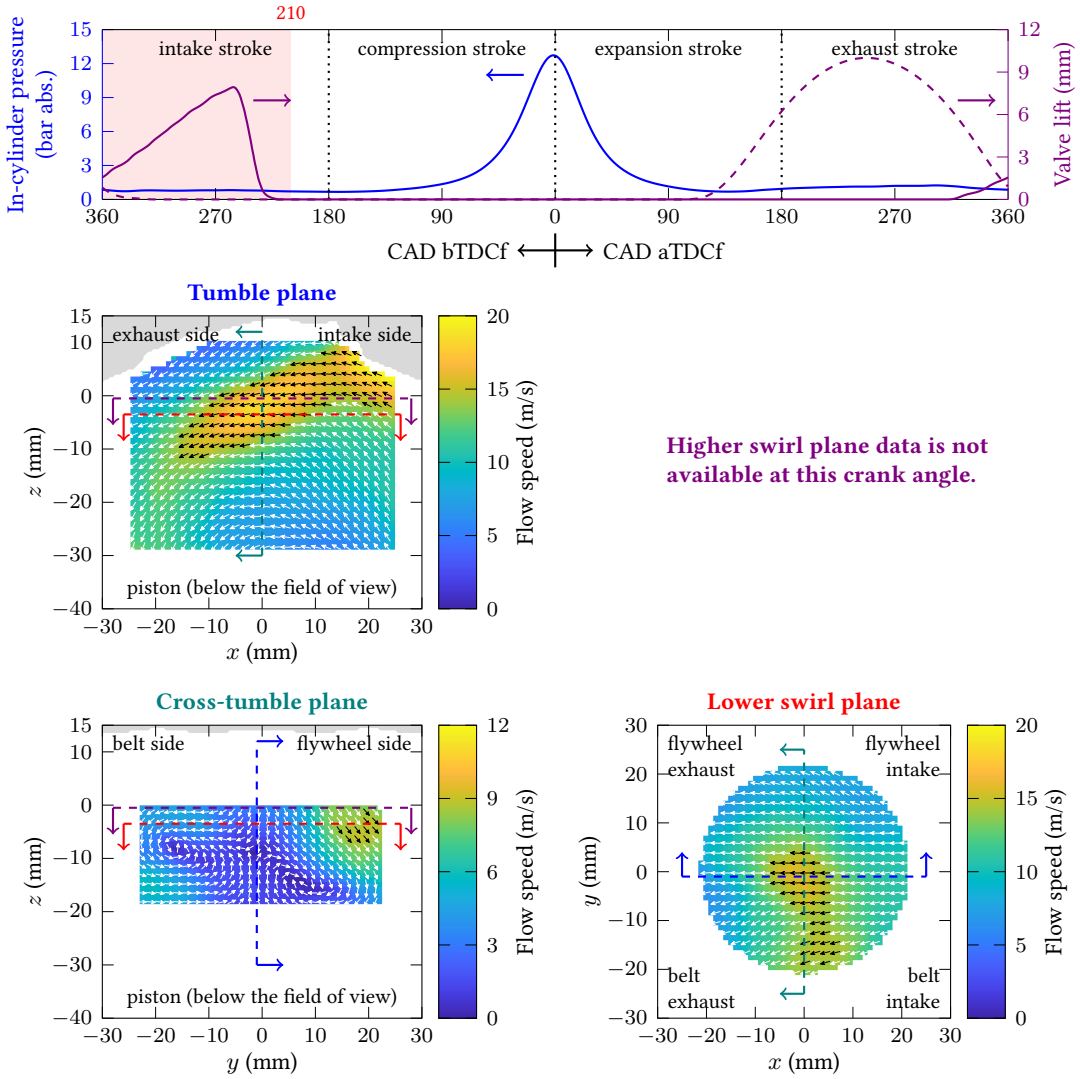


Figure 4.5: Ensemble mean flow fields and engine parameters at 210° bTDCf for the example test point (T2, Piston B, C33, DVA). The top plot shows the engine parameters: in-cylinder pressure trace (blue solid line), intake valve lift (purple solid line) and exhaust valve lift (purple dashed line). The bottom four plots show the PIV ensemble mean flow fields in different measurement planes, while the higher swirl plane data is only available in the late compression stroke (after 90° bTDCf). The coloured dashed lines in each flow field plot mark the corresponding locations of other planes, where the arrows illustrate the viewing direction. Each plane is marked by a different colour: tumble plane (blue), higher swirl plane (purple), cross-tumble plane (dark green) and lower swirl plane (red). Due to the differences in flow speeds, the colour scale may change in different planes (within this figure) or at different crank angles (among Figures 4.3–4.9).

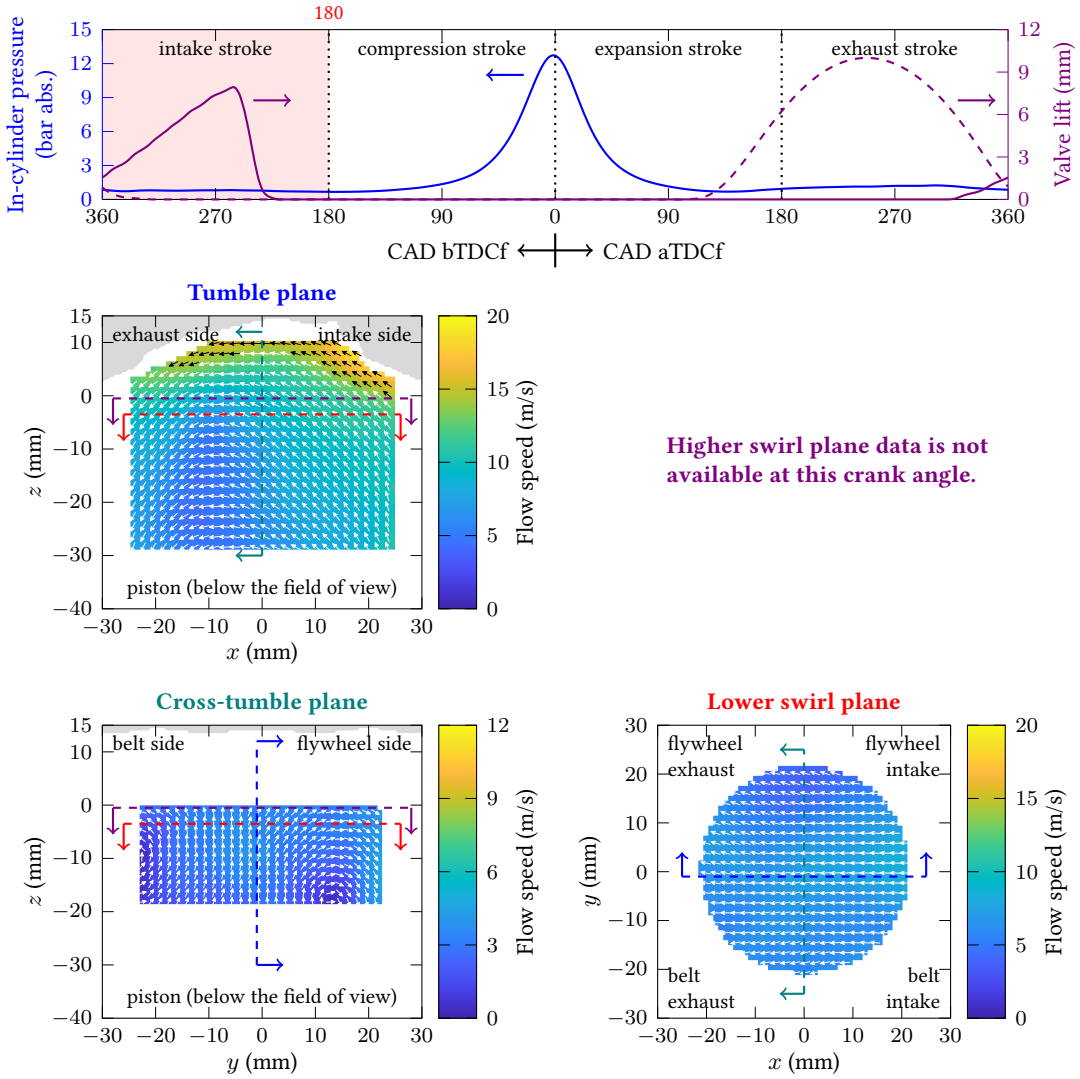


Figure 4.6: Ensemble mean flow fields and engine parameters at 180° bTDCf for the example test point (T2, Piston B, C33, DVA). The top plot shows the engine parameters: in-cylinder pressure trace (blue solid line), intake valve lift (purple solid line) and exhaust valve lift (purple dashed line). The bottom four plots show the PIV ensemble mean flow fields in different measurement planes, while the higher swirl plane data is only available in the late compression stroke (after 90° bTDCf). The coloured dashed lines in each flow field plot mark the corresponding locations of other planes, where the arrows illustrate the viewing direction. Each plane is marked by a different colour: tumble plane (blue), higher swirl plane (purple), cross-tumble plane (dark green) and lower swirl plane (red). Due to the differences in flow speeds, the colour scale may change in different planes (within this figure) or at different crank angles (among Figures 4.3–4.9).

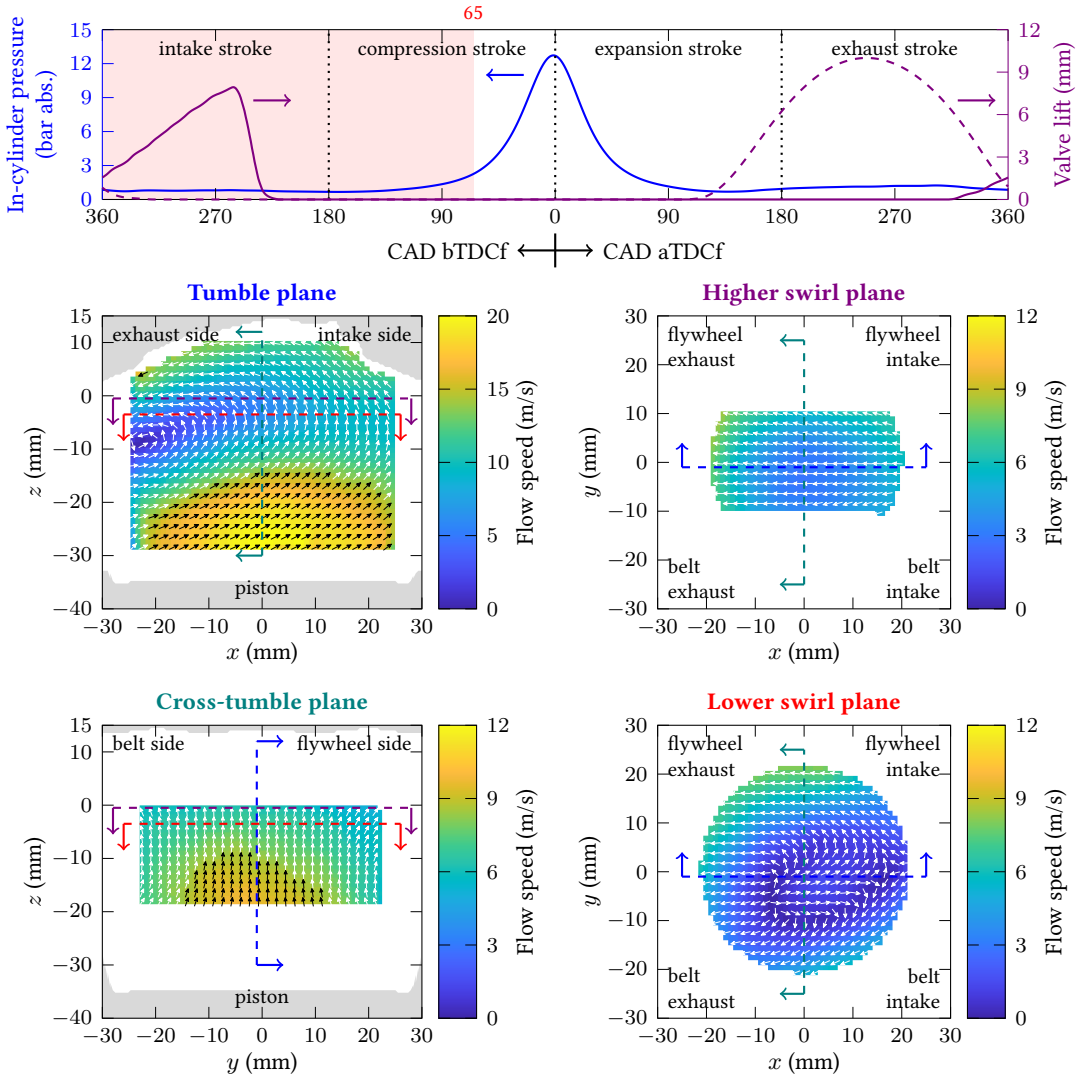


Figure 4.7: Ensemble mean flow fields and engine parameters at 65° bTDCf for the example test point (T2, Piston B, C33, DVA). The top plot shows the engine parameters: in-cylinder pressure trace (blue solid line), intake valve lift (purple solid line) and exhaust valve lift (purple dashed line). The bottom four plots show the PIV ensemble mean flow fields in different measurement planes, while the higher swirl plane data is only available in the late compression stroke (after 90° bTDCf). The coloured dashed lines in each flow field plot mark the corresponding locations of other planes, where the arrows illustrate the viewing direction. Each plane is marked by a different colour: tumble plane (blue), higher swirl plane (purple), cross-tumble plane (dark green) and lower swirl plane (red). Due to the differences in flow speeds, the colour scale may change in different planes (within this figure) or at different crank angles (among Figures 4.3–4.9).

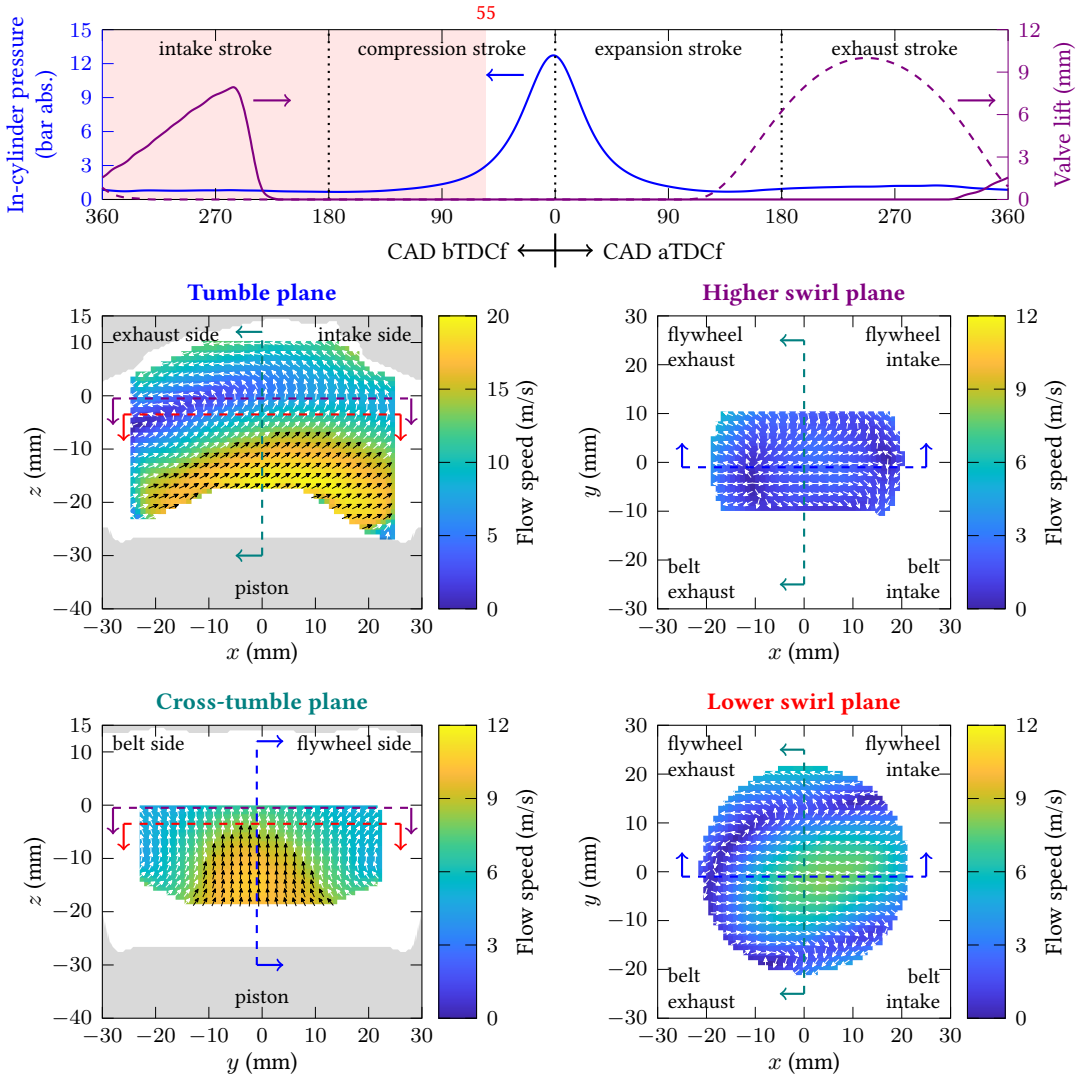


Figure 4.8: Ensemble mean flow fields and engine parameters at 55° bTDCf for the example test point (T2, Piston B, C33, DVA). The top plot shows the engine parameters: in-cylinder pressure trace (blue solid line), intake valve lift (purple solid line) and exhaust valve lift (purple dashed line). The bottom four plots show the PIV ensemble mean flow fields in different measurement planes, while the higher swirl plane data is only available in the late compression stroke (after 90° bTDCf). The coloured dashed lines in each flow field plot mark the corresponding locations of other planes, where the arrows illustrate the viewing direction. Each plane is marked by a different colour: tumble plane (blue), higher swirl plane (purple), cross-tumble plane (dark green) and lower swirl plane (red). Due to the differences in flow speeds, the colour scale may change in different planes (within this figure) or at different crank angles (among Figures 4.3–4.9).

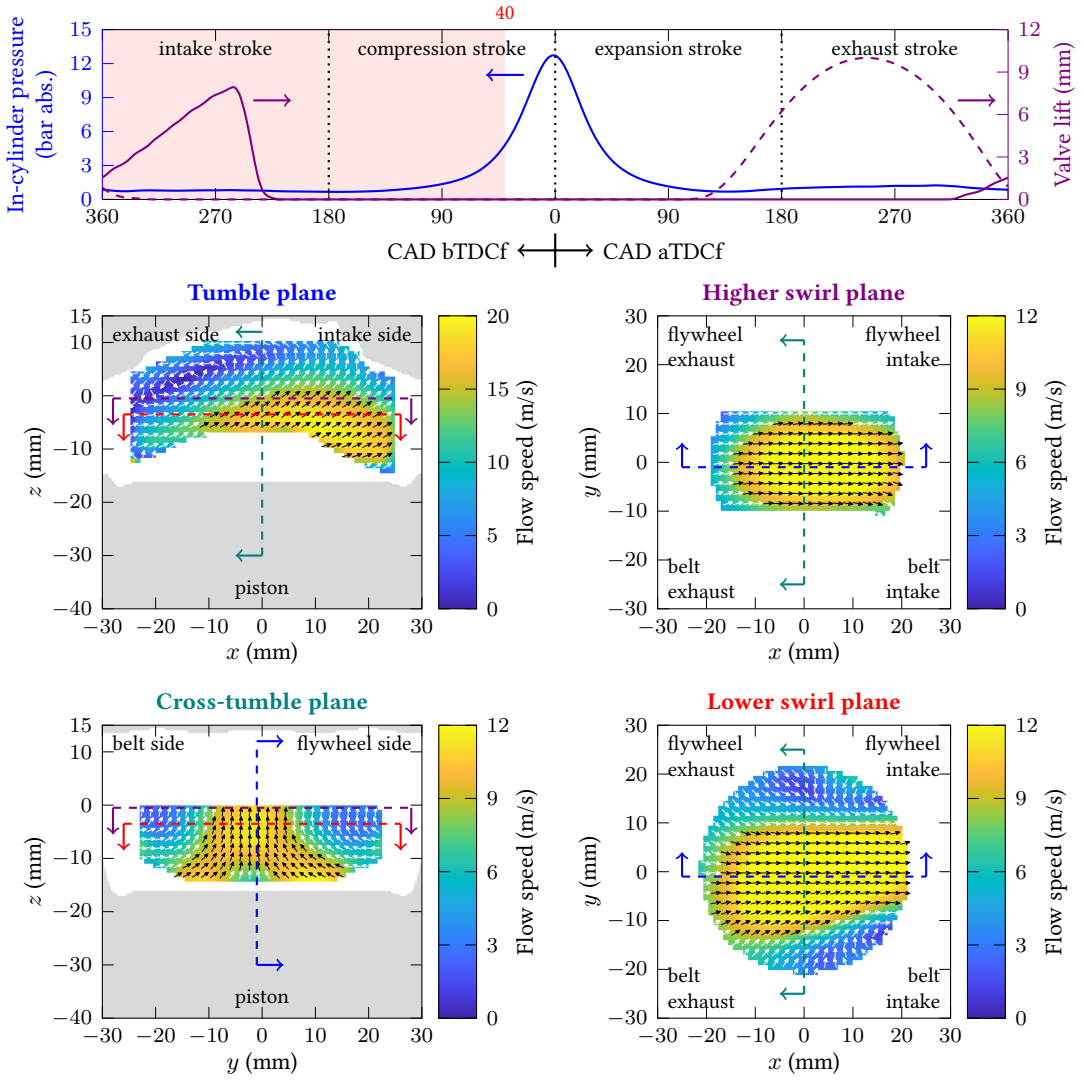


Figure 4.9: Ensemble mean flow fields and engine parameters at 40° bTDCf for the example test point (T2, Piston B, C33, DVA). The top plot shows the engine parameters: in-cylinder pressure trace (blue solid line), intake valve lift (purple solid line) and exhaust valve lift (purple dashed line). The bottom four plots show the PIV ensemble mean flow fields in different measurement planes, while the higher swirl plane data is only available in the late compression stroke (after 90° bTDCf). The coloured dashed lines in each flow field plot mark the corresponding locations of other planes, where the arrows illustrate the viewing direction. Each plane is marked by a different colour: tumble plane (blue), higher swirl plane (purple), cross-tumble plane (dark green) and lower swirl plane (red). Due to the differences in flow speeds, the colour scale may change in different planes (within this figure) or at different crank angles (among Figures 4.3–4.9).

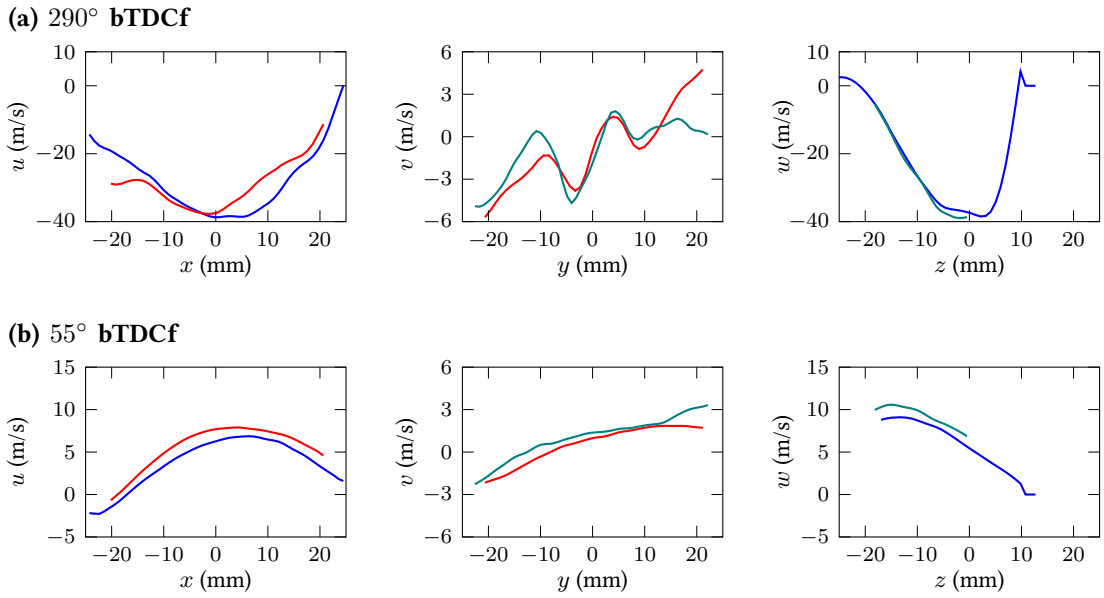


Figure 4.10: Comparison of common-component PIV velocities along the intersections of the three planes: tumble plane (blue), cross-tumble plane (dark green) and lower swirl plane (red).

planes at multiple grid points along the line. The comparison of common-component velocities along the intersections of the three planes are shown in Figure 4.10 for 290° bTDCf (during intake) and 55° bTDCf (during compression). The trend of the velocity profile is well matched across all planes during both intake (top) and compression (bottom) strokes, albeit with small offsets in both position and magnitude when comparing the lower swirl (red) with the other two planes (tumble in blue, cross-tumble in dark green) during the intake stroke.

The multi-plane PIV flow fields show remarkable consistency at this test point despite that the data were recorded with an interval of several months, and such findings, although not presented here for the sake of brevity, can also be observed in the other test points that have data on more than one plane (see Table 2.4). It can be assumed that the major flow structures observed in the ensemble mean flow fields are quite consistent among different measurement planes as long as the engine operation parameters were carefully controlled (see Section 4.1). Therefore, the PIV ensemble means in multiple planes may be used as the validation targets for CFD flow models.

4.3 Visual inspection on PIV and CFD flow fields

Isometric views of the PIV and CFD flow fields across the tumble, cross-tumble and lower swirl planes are presented for six crank angles covering the intake (330°, 300° and 200° bTDCf) and compression (140°, 80° and 50° bTDCf) strokes for two engine loads – the T1 load in Figure 4.11 and the T2 load[(1)] in Figure 4.12, respectively. The T1 load has a lower peak pressure (see Figure 4.1(a)) and a smaller intake air volume flow rate (see Table 2.3) than the T2 load (see Figure 4.1(b)).

On visual inspection there appears to be in most cases a good overall match between the experimental (left columns in Figures 4.11 and 4.12) and simulated flow fields (right columns), with consistency in the structures of the in-cylinder flow across all three planes, for each crank angle and for both test conditions. At 330° bTDCf the intake jet is visible in the tumble plane for both T1 (Figure 4.11(a)–(b)) and T2 (Figure 4.12(a)–(b)) and the directed intake runner, intended to generate high levels of tumble, gives a significant component of the flow in the x-direction from inlet to exhaust which results in the generation of a pair of counter-rotating vortices in the lower swirl plane. At 300° bTDCf the intake jet is well established in all cases. The cross-tumble and lower swirl planes show lower velocities except for a region of horizontal flow across the piston for the T1 PIV tumble plane flow field (Figure 4.11(c)) which is qualitatively not apparent for T2 PIV data (Figure 4.12(c)) or the CFD flow fields at either condition (Figures 4.11(d) and 4.12(d)). The flow fields at later crank angles, 200°, 140° and 80° bTDCf (Figures 4.11(e)–(j) and 4.12(e)–(j)), show the evolution of the main tumble motion as the tumble vortex centre moves from inlet to exhaust over the course of the late intake and early compression strokes. In-plane velocities for the cross-tumble plane are low for these timings due to the dominant flow motion being out of plane. At around 50° bTDCf the rising

[(1)] Data presented and discussed in the previous section (Section 4.2) were also recorded under this test condition (T2, Piston B, C33, DVA).

piston causes the tumble vortex centre to intersect the lower swirl plane for the T1 test point (Figure 4.11(k)–(l)), resulting in low in-plane velocities; the similar behaviour can also be observed for the T2 test point (Figure 4.12), albeit that the tumble vortex centre reaches the lower swirl plane slightly earlier (at around 65° bTDCf, as discussed in Section 4.2).

The RANS simulation has achieved a good qualitative match with the ensemble mean PIV measurements. However, some differences are apparent such as the increased velocity magnitudes within the tumble plane CFD flow fields, the small differences in the tumble vortex centre location between the PIV and CFD flow fields, and the extent of the intake jet for the T1 test condition (Figure 4.11). The PIV ensemble mean for the T2 test condition (Figure 4.12(c)) has lower peak velocities and a broader jet than the CFD jet flows (Figure 4.12(d)); presumably this is a result of oscillations in the jet that are averaged-out when the PIV data is ensemble averaged. The discrepancies between the PIV data and CFD results need to be examined beyond the qualitative visual inspection, which calls for the quantitative comparisons using the metrics introduced in Section 3.3. The WRI, WMI and CMRI metrics will provide greater insight into which flow features are well represented by CFD and which show differences to the experiments and have the potential to inform improvements to the simulation of these in-cylinder flows.

4.4 Investigation using the vector field comparison metrics

The WRI, WMI and CMRI metrics (see Section 3.3) are used to identify key timings within the cycle for the T1 and T2 test conditions discussed in the previous section (Section 4.3) for which differences between the PIV ensemble mean and CFD flow fields are significant in one or more of the tumble, lower swirl or cross-tumble planes. In addition to these two test points,[(2)] two extra

[(2)] The two test points (discussed in Sections 4.3 and 4.4) are: T1, Piston B, C33, DVA and T2, Piston B, C33, DVA. For brevity denote the “T1” and “T2” test points, respectively.

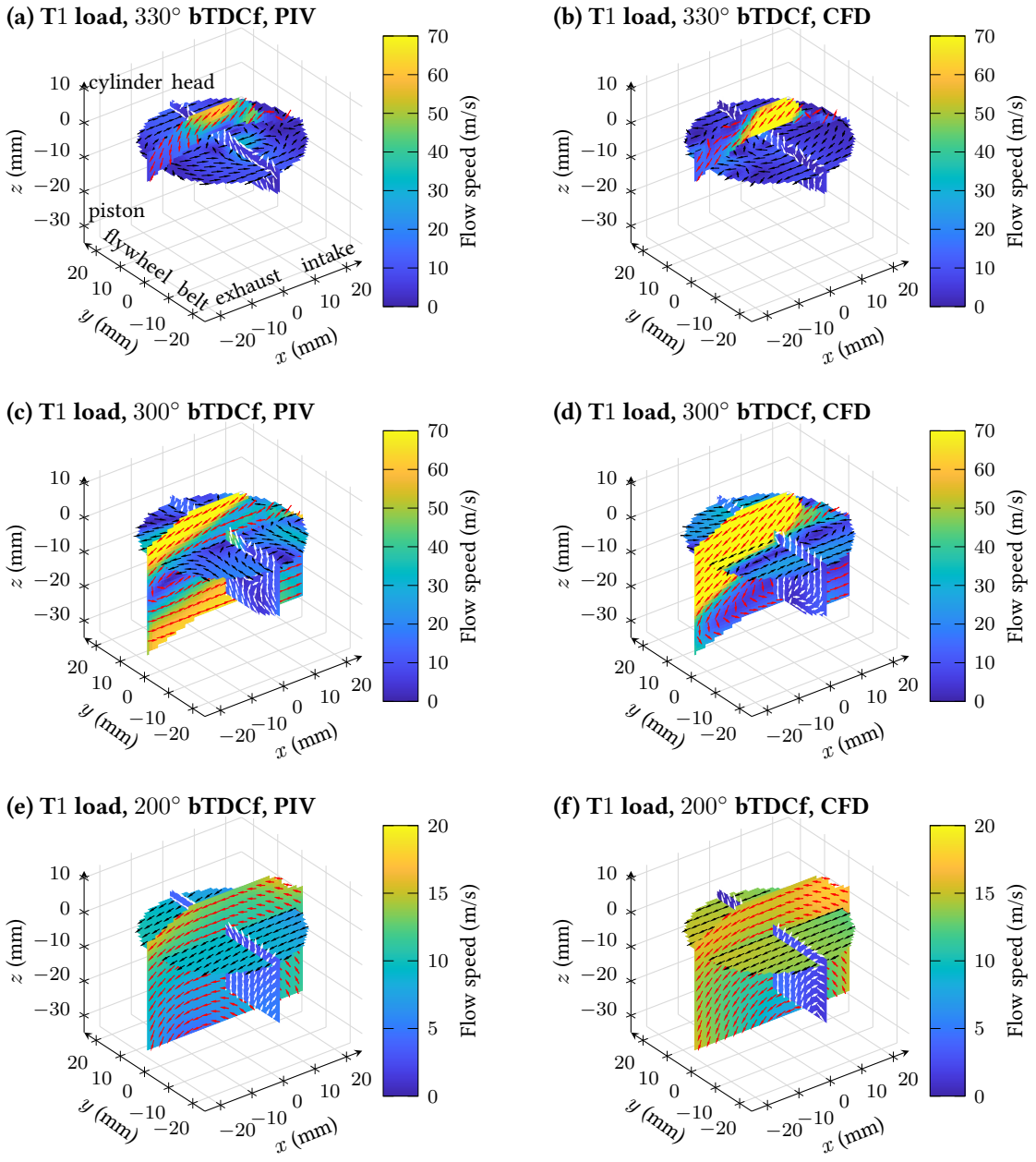


Figure 4.11: Isometric view of two-component in-plane velocities for PIV ensemble mean (left column) and CFD (right column) flow fields at the T1, Piston B, C33, DVA test point. The in-plane vectors in each measurement plane are plotted in different colours – red vectors: tumble plane; black vectors: lower swirl plane; white vectors: cross-tumble plane. Note that the colour scales may differ among the plots for different crank angles due to the change in the flow speed.

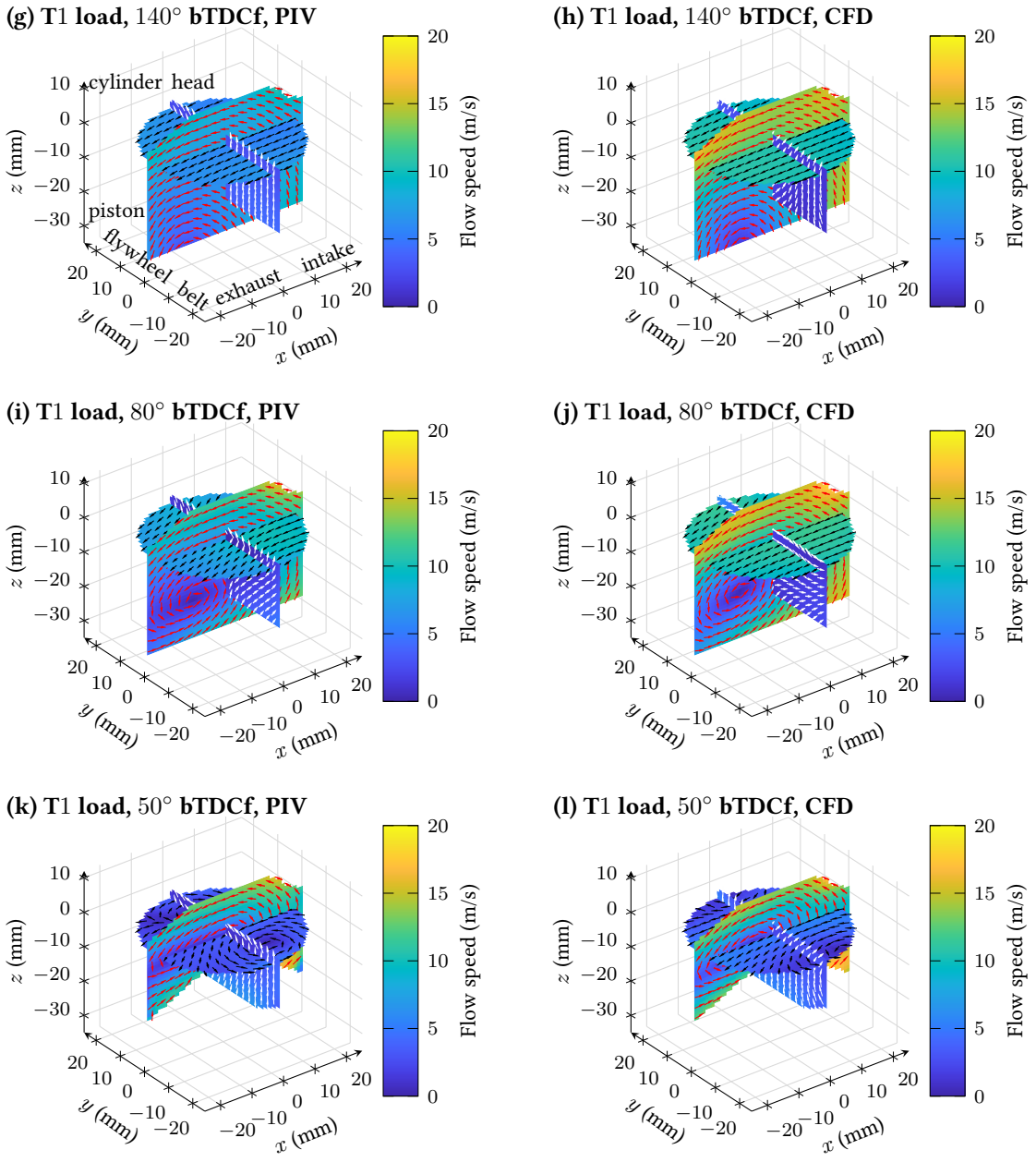


Figure 4.11: Isometric view of two-component in-plane velocities for PIV ensemble mean (left column) and CFD (right column) flow fields at the T1, Piston B, C33, DVA test point. The in-plane vectors in each measurement plane are plotted in different colours – red vectors: tumble plane; black vectors: lower swirl plane; white vectors: cross-tumble plane. Note that the colour scales may differ among the plots for different crank angles due to the change in the flow speed.

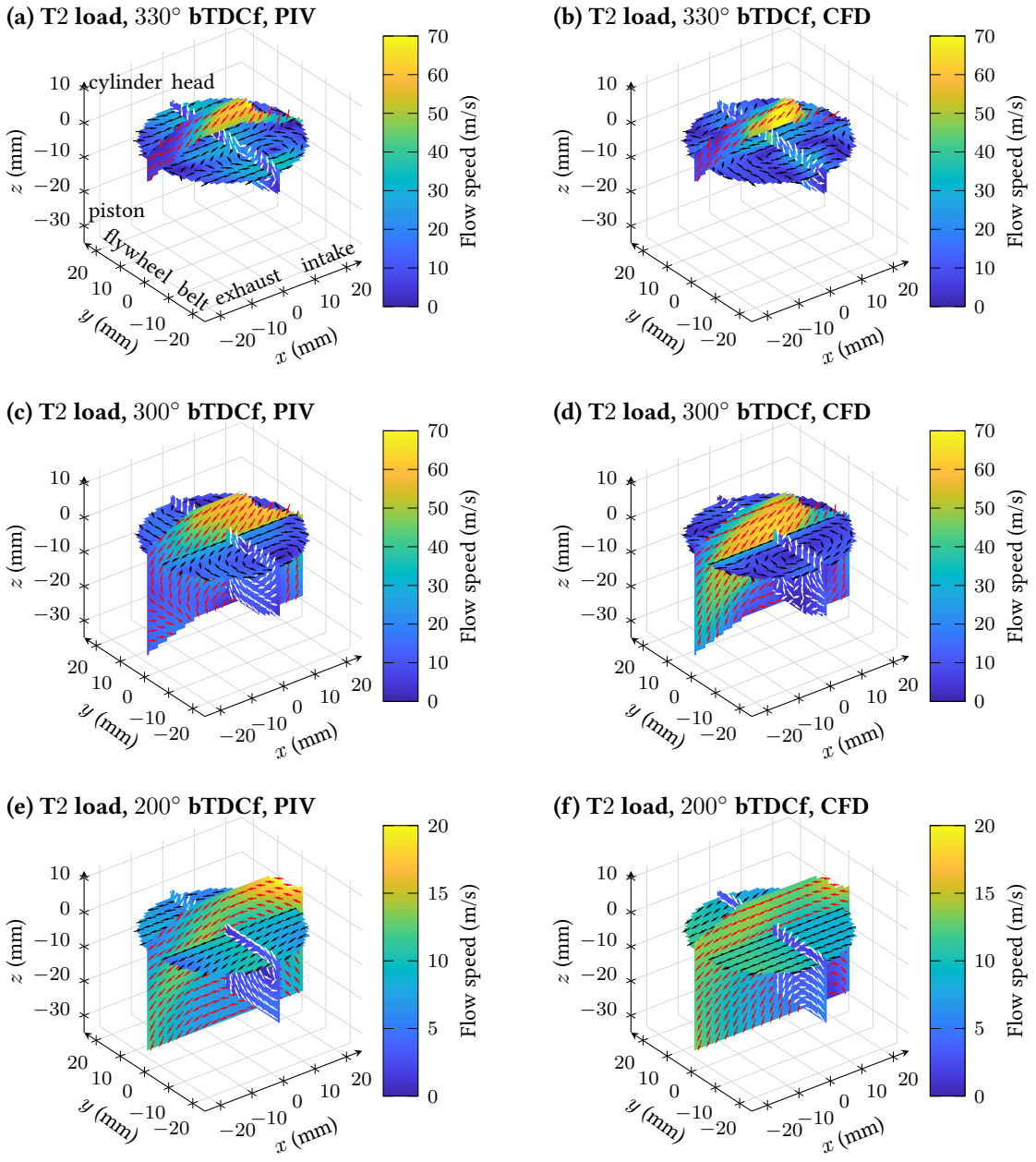


Figure 4.12: Isometric view of two-component in-plane velocities for PIV ensemble mean (left column) and CFD (right column) flow fields at the T2, Piston B, C33, DVA test point. The in-plane vectors in each measurement plane are plotted in different colours – red vectors: tumble plane; black vectors: lower swirl plane; white vectors: cross-tumble plane. Note that the colour scales may differ among the plots for different crank angles due to the change in the flow speed.

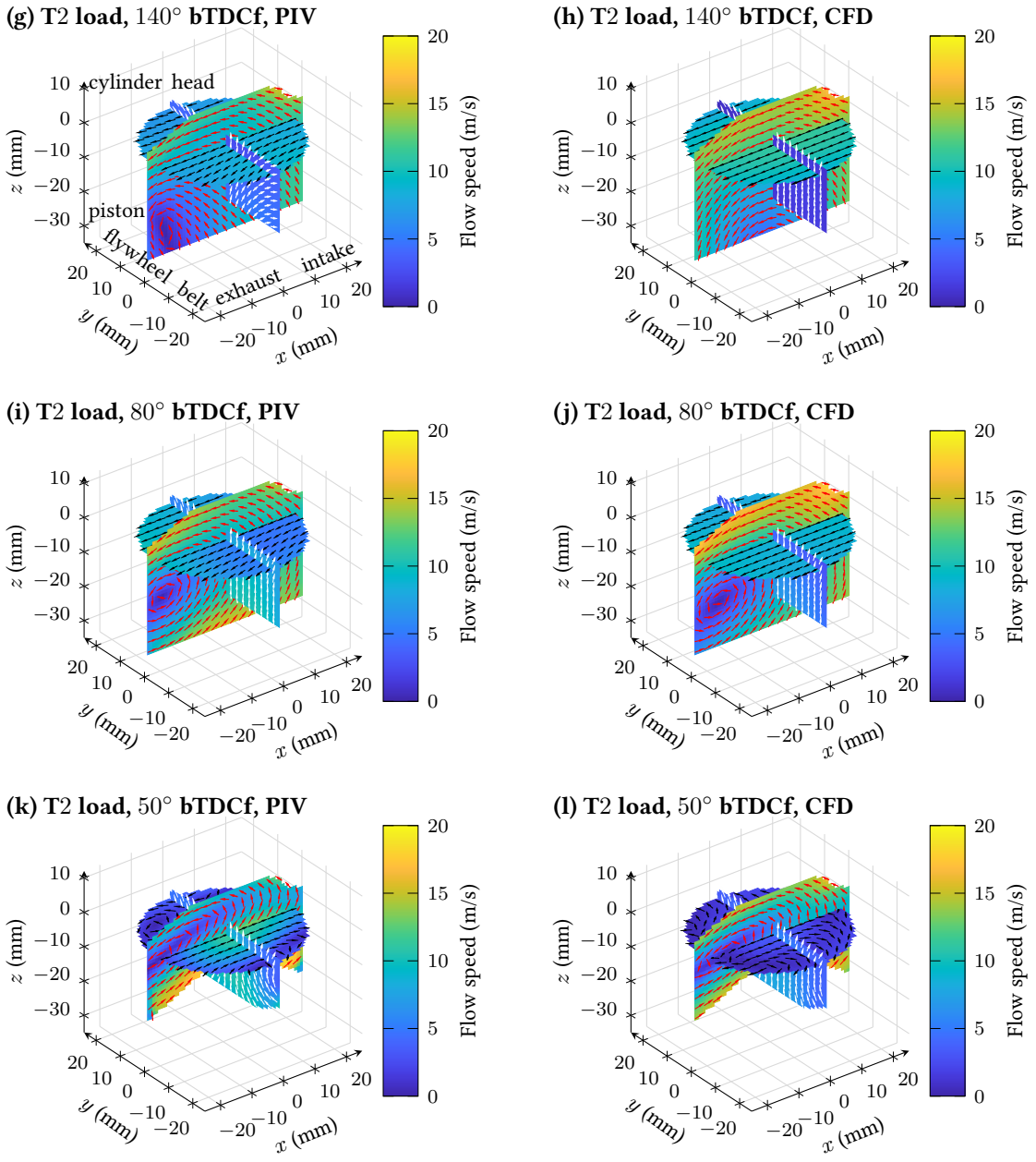


Figure 4.12: Isometric view of two-component in-plane velocities for PIV ensemble mean (left column) and CFD (right column) flow fields at the T2, Piston B, C33, DVA test point. The in-plane vectors in each measurement plane are plotted in different colours – red vectors: tumble plane; black vectors: lower swirl plane; white vectors: cross-tumble plane. Note that the colour scales may differ among the plots for different crank angles due to the change in the flow speed.

test points[3] were used for the normalisation of the WRI and WMI metric values (Equation (3.55)) to produce the CMRI (Equation (3.56)), in order to provide a more comprehensive scale that accounts for the variations in the flow fields for various test conditions. This allows quantitative comparison of differences both between planes and between test points by using the absolute values of the comparison metrics.[4]

4.4.1 Initial investigation using the CMRI

Figure 4.13 presents the spatially-averaged CMRI (hereinafter simply referred as "CMRI") values for the comparison of PIV and CFD flow fields throughout the intake and compression strokes for the T1 and T2 test points. This coarse overview condenses the spatial information within the comparison of PIV to CFD across each plane, for both directional and magnitude differences, into a robust scalar value at each crank angle. The CMRI values for each of the three planes are weighted by one-third and displayed cumulatively at each crank angle, hence the height of the total shaded region represents the overall difference between the PIV and CFD data across all planes. The contribution of the tumble (blue), lower swirl (red) and cross-tumble (dark green) planes to the combined difference in flow fields is given by the height of each regions. The solid purple line (right y axis) in each load shows the mean intake valve lift for all the measured cycles. The CMRI values reported in Figure 4.13 identify four aspects during the intake and compression strokes which merit further investigation into the quality of the match between the CFD and PIV flow fields. These four aspects are discussed in detail in the following paragraphs.

The values of the spatially-averaged CMRI vary at the 61 measured crank angles and in the three measured planes (Figure 4.13). The quality of the match in each plane can be quantified by the arithmetical mean of the CMRI values over the entire crank angle range (61 values for

[3] The two extra test points (used for CMRI calculation only) are: T1, Piston B, bootleg, DVA and T1, Piston C, C33, DVA. Details of the test points were listed in Tables 2.3 and 2.4.

[4] All the values of the WRI and WMI metrics presented in this section are normalised using Equation (3.55).

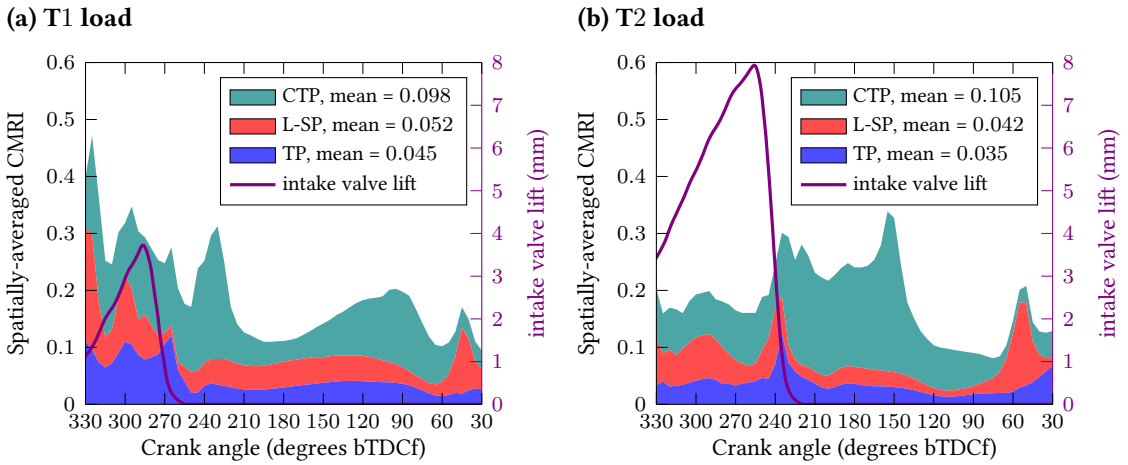


Figure 4.13: Spatially-Averaged CMRI for the T1 and T2 test points, shown as stacked histograms (left y axis) for data in the tumble (blue), lower swirl (red) and cross-tumble (dark green) planes. The solid purple line (right y axis) in each plot shows the mean intake valve lift for all the measured cycles at the listed engine loads. The mean values in each plane reported on the legends refer to the arithmetical mean of the spatially-averaged CMRI values over the entire crank angle range (i.e., the mean of the 61 values from 330° to 30° bTDCf).

each plane), and the mean values are reported on the legends. The cross-tumble plane provides the largest contribution to the differences between the PIV and CFD data for both the T1 (mean CMRI $= 0.098$) and T2 (mean CMRI $= 0.105$) loads. The tumble plane and the lower swirl plane are similarly well matched, but with T2 better matched for both the tumble plane and the lower swirl plane (mean CMRI $= 0.035$ and 0.042) than T1 (mean CMRI $= 0.045$ and 0.052). The mean of all 183 spatially-averaged CMRI values (61 measured crank angles $\times$ three measured planes) at each load suggests that both test points are well matched overall, but T2 provides a slightly better match (overall mean CMRI $= 0.183$) compared to T1 (overall mean CMRI $= 0.195$).

The T2 test point is better matched during the intake stroke (mean CMRI ≈ 0.2 for the crank angles during the intake stroke, Figure 4.13(b)). The CMRI is stably low until a rise during the intake valves closing (IVC) at around 240° bTDCf. The IVC peak in CMRI can be observed in all three planes, where the CMRI values approximately doubles in each plane compared to the

preceding values (at $5°$–$10°$ CA earlier). The peaks in CMRI also occur at a slightly different timing in the three planes ($235°$ bTDCf in TP, $240°$ bTDCf in L-SP and $230°$ bTDCf in CTP), as the flow change due to the intake valve closing may take a different time to reach each plane. On the other hand, the PIV and CFD flow fields at the T1 test point have greater differences during the intake stroke (mean CMRI ≈ 0.3 at the crank angles during the intake stroke, Figure 4.13(a)) — the CMRI values are high until the match improves after $235°$ bTDCf. The lower swirl plane is well-matched by IVC (at around $260°$ bTDCf, CMRI $= 0.018$) but large differences are identified for the tumble (CMRI $= 0.121$) and cross-tumble (CMRI $= 0.137$) planes.

During the development of the tumble flow, the T1 test point has a peak in CMRI (0.313) at around $230°$ bTDCf due to differences in the cross-tumble plane but is well matched (CMRI $<$ 0.14) across all three planes from $210°$ to $150°$ bTDCf (Figure 4.13(a)). The T2 load displays a consistently high total CMRI (≈ 0.25) lasting for approximately $90°$ crank angles from IVC (at around $240°$ bTDCf), mainly arising from differences in the cross-tumble plane (Figure 4.13(b)). The lower swirl plane for T2 (CMRI ≈ 0.025) is better matched than T1 (CMRI ≈ 0.04), while the tumble plane is similarly well matched for both test conditions (both CMRI ≈ 0.03).

During the compression stroke, the T1 test point has a gradual increase in the CMRI values in the cross-tumble plane from $160°$ bTDCf, peaking at $100°$ bTDCf (Figure 4.13(a)). The major contribution to the increase of the cumulative CMRI values arises from the differences in the cross-tumble plane (mean CMRI $= 0.132$ during the compression stroke), as compared to the lower values in the tumble (mean CMRI $= 0.038$) and the lower swirl (mean CMRI $= 0.033$) planes. In contrast, once the high CMRI plateau of the T2 load ends at around $140°$ bTDCf (Figure 4.13(b)), the cumulative CMRI gradually decreases. Excellent matches in the tumble (mean CMRI $= 0.016$) and lower swirl (mean CMRI $= 0.015$) planes can be found throughout the compression stroke of the T2 load, while the differences still mainly exist in the cross-tumble plane (CMRI $= 0.063$). In the late-compression stroke, both loads (T1 at around $45°$ bTDCf, T2 at around $55°$ bTDCf) have a peak in the lower swirl plane CMRI values.

The spatially-averaged CMRI results already give a quantitative overview of where the CFD flow fields matches well with PIV for each test condition. However far more information is available, firstly by further investigation using the WRI and WMI metrics to distinguish between differences in alignment from differences in magnitude, and secondly using the key timings of interest identified for each plane to inform the inspection of the spatially resolved metric fields and the flow fields themselves. The key features of the CMRI values in Figure 4.13 can be assigned as arising from two main characteristics of the in-cylinder flow — the development of the intake jet formed while the inlet valves are open, and the motion of the tumble vortex centre throughout the cylinder for this tumble-dominated flow geometry. Both characteristics are highly three-dimensional flow structures and their dynamics are best studied by consideration of the flow field interactions across all three planes in the following sections.

4.4.2 CMRI peaks during the intake process

The CMRI values of the T1 test point in Figure 4.13(a) show three distinct peaks at 330°, 300° and 265° bTDCf during intake for the tumble plane, and these correspond to significant differences in the simulated and experimental flow fields. Investigation on the flow fields at these key timings (Figure 4.14) reveals these CMRI peaks arise from different features of the intake flow fields. At 330° bTDCf the local flow directions are well matched with low WRI values (Figure 4.14(c)) but there is a large difference in the velocity magnitudes with high WMI values (Figure 4.14(d)) which leads to the first CMRI peak. The high velocity region of the PIV intake jet (Figure 4.14(a)) is located further towards the exhaust while the centre of the intake jet predicted by CFD (Figure 4.14(b)) is closer to the inlet and has a higher velocity than the jet measured by PIV. By 300° bTDCf the intake jet for both PIV and CFD is well established. The CFD jet (Figure 4.14(f)) takes up a much larger region of the upper-exhaust side of the field of view with a clear distinction between the high velocity jet and the slow bulk flow in the central

and inlet side regions of the cylinder. The PIV jet (Figure 4.14(e)) is located much closer to the cylinder head and the overlap between the slow recirculating flow below the PIV jet and the fast lower edge of the CFD jet produces very high WMI values. The PIV flow field also contains a high medium velocity horizontal flow across the piston at around $z = -20$ mm. This cross-flow leads to high WRI values (Figure 4.14(g)) on the exhaust side where it opposes the CFD intake jet direction and high WMI values (Figure 4.14(h)) above the piston surface where the simulated flow is of much lower speed. As the intake valves close, the behaviour of the intake jet differs between the PIV and CFD flow fields. For PIV, the closing valves result in the disappearance of the intake jet at 265° bTDCf (Figure 4.14(i)) when the valve lift is very small and the formation of the bulk tumble flow which persists for the remainder of the cycle. At the same crank angle, in contrast, the small valve lift results in a narrowing of the intake jet and an increase in peak velocity for the CFD flow field (Figure 4.14(j)). This difference in flow speed is captured by the WMI plot in Figure 4.14(k), in which the WMI values are high in the jet region at 265° bTDCf (Figure 4.14(l)). The CFD intake jet disappears once the valves are fully closed at 255° bTDCf (not shown in Figure 4.14 for brevity), regaining a good match with the PIV flow fields in later crank angles (low CMRI, see Figure 4.13(a)).

In contrast to the lower load T1, the higher load T2 has a very good overall match while the intake valve are open (Figure 4.15). For both 330° (Figure 4.15(a)–(d)) and 300° bTDCf (Figure 4.15(e)–(h)) there is an excellent directional match between the PIV and CFD flow fields, with only a minor overestimation of the velocity magnitudes for CFD resulting in slightly raised WMI values (Figure 4.15(h)). The intake jet at T2 (Figure 4.15(e)) is much broader and lower velocity than T1, covering a large portion of the field of view and does not result in the cross-piston flow present in the T1 PIV flow field (Figure 4.14(e)). Once the intake valves have almost fully closed at 235° bTDCf (15% of the peak lift) however, a similar behaviour to that of T1 is observed. The PIV intake jet (Figure 4.15(i)) disappears and the CFD intake jet (Figure 4.15(j)) again narrows and increases in peak velocity magnitude, resulting in high WMI

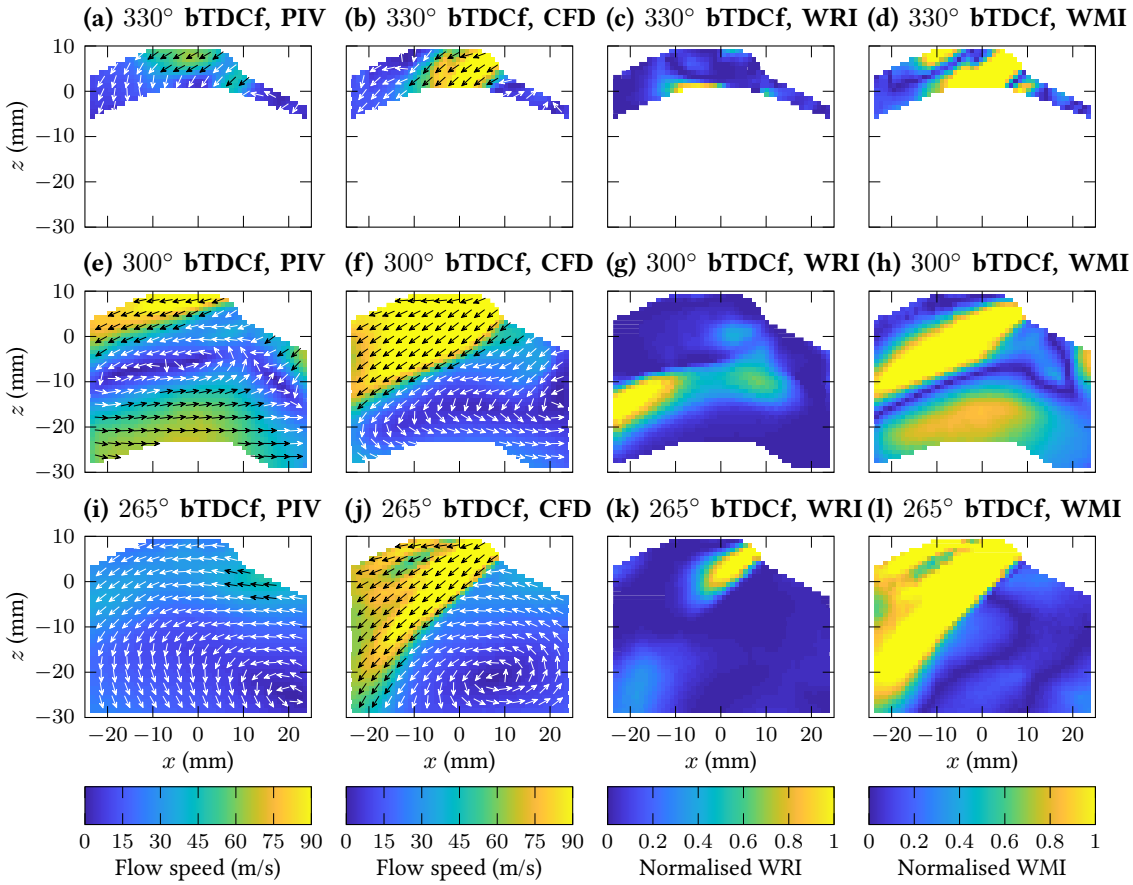


Figure 4.14: PIV ensemble mean (Column 1) and CFD (Column 2) flow fields during the intake process for the T1 test point with local differences in direction and magnitude quantified using the WRI (Column 3) and WMI metrics (Column 4) respectively. Each row displays data from a different crank angle identified as a key timing of interest by a peak in the spatially-averaged CMRI values in Figure 4.13(a).[5]

values (Figure 4.15(l)). Note that the data in the last row (Figure 4.15(i)–(l)) were recorded at 235° bTDCf for the T2 test point, instead of 265° bTDCf for the T1 test point in the previous figure (Figure 4.14(i)–(l)), due to a later intake valve closing (IVC) time at the higher load (T2).

[5] For clarity only every third vector is plotted in each direction, instead of every two vector in other flow field plots in this thesis. This plot style is also adopted in Figures 4.15, 4.16 and 4.18–4.21.

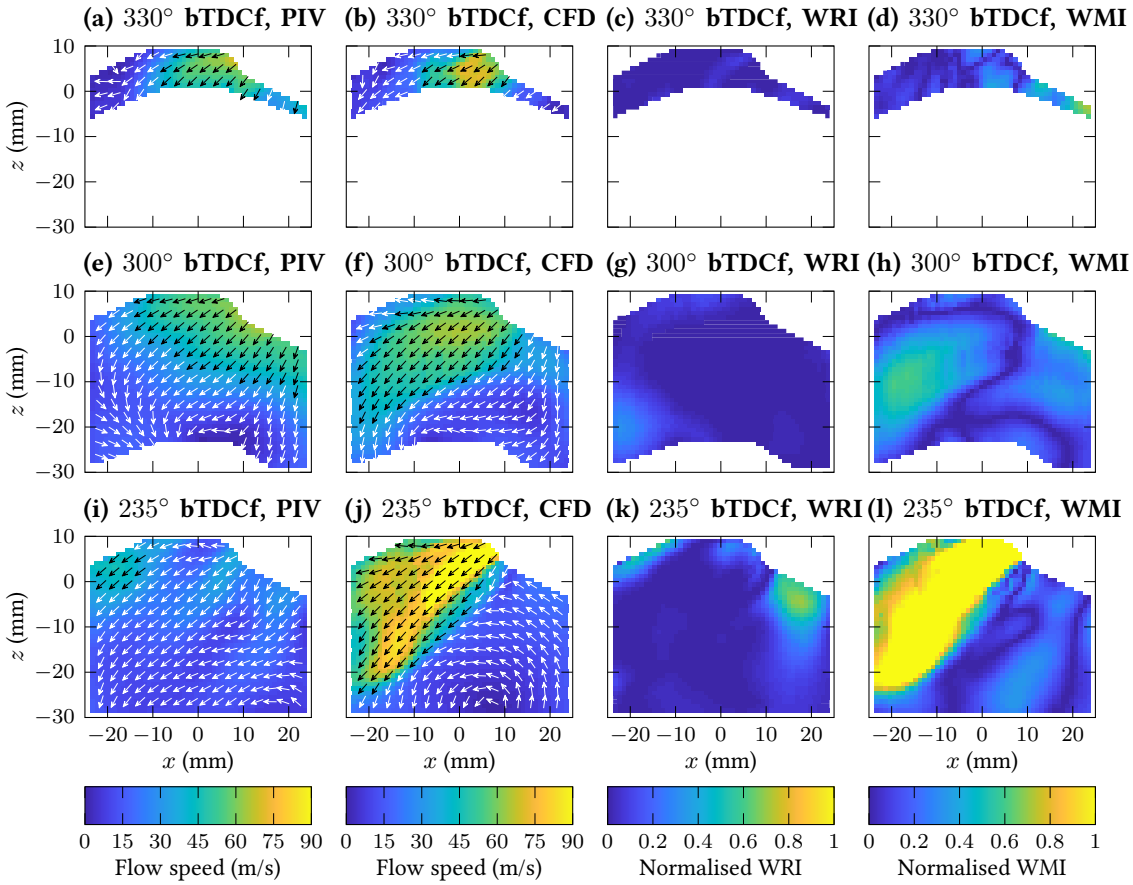


Figure 4.15: PIV ensemble mean (Column 1) and CFD (Column 2) flow fields during the intake process for the T2 test point with local differences in direction and magnitude quantified using the WRI (Column 3) and WMI metrics (Column 4) respectively. Each row displays data from a different crank angle identified as a key timing of interest by a peak in the spatially-averaged CMRI values in Figure 4.13(b).

4.4.3 CMRI peaks due to the tumble vortex motion

The CMRI peak for the T1 test point around 230° bTDCf (Figure 4.13(a)) arises from the tumble vortex centre passing through the cross-tumble measurement plane 10° CA earlier for the PIV measured flow field compared to the CFD simulation (Figure 4.16). The low velocities near the tumble vortex centre first make an appearance in the PIV flow fields at 250° bTDCf (Figure 4.16(a)), disturbing the otherwise downwards flow in the central tumble plane as the

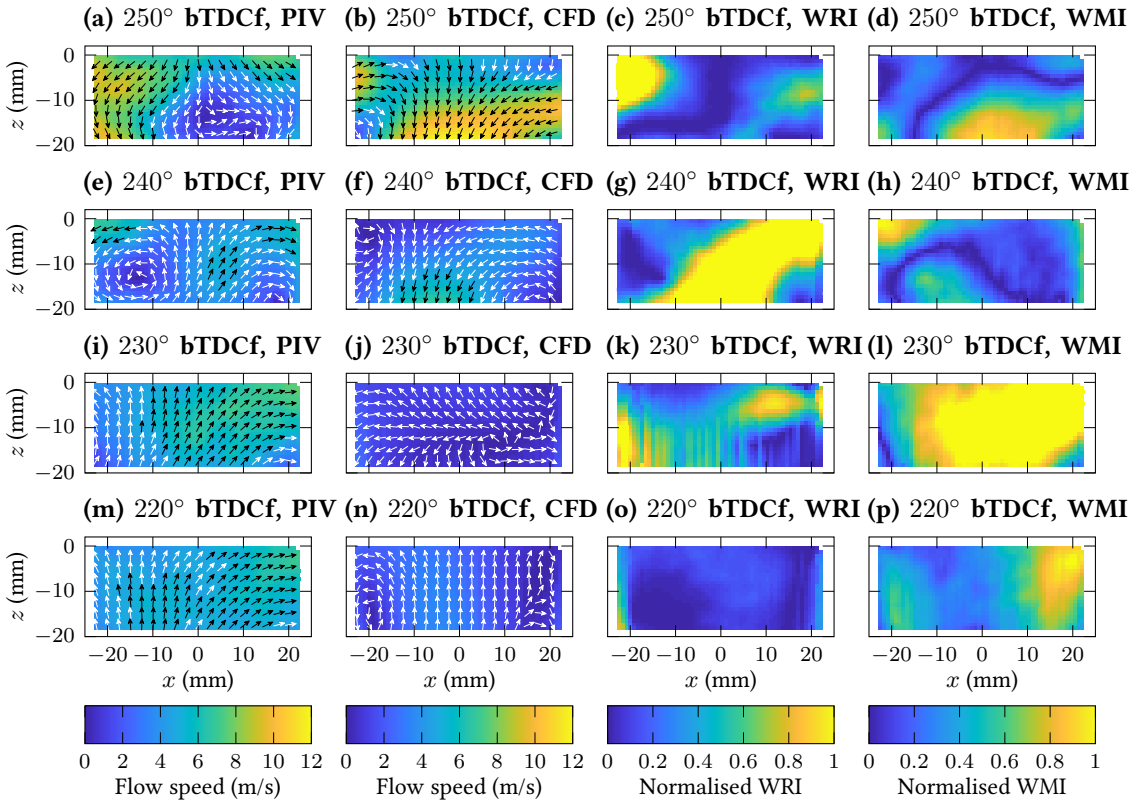


Figure 4.16: Flow direction reversal in the cross-tumble plane for the T1 test point due to the motion of the tumble vortex centre passing through the measurement plane for PIV ensemble mean (Column 1) and CFD (Column 2) flow fields with local differences in direction and magnitude quantified using the WRI (Column 3) and WMI metrics (Column 4), respectively. Note that the flow direction reversal occurs at different crank angles (rows) for PIV ensemble mean and CFD flow fields.

tumble vortex centre is initially formed on the inlet side of the cylinder and the tumble motion is anti-clockwise when viewed in the tumble plane (Figure 4.14(i)). As the tumble vortex centre traverses the cross-tumble plane from the inlet (positive x) to the exhaust (negative x), the flow in the central cross-tumble reverses for PIV at 240° bTDCf (Figure 4.16(e)), resulting in an upwards flow at 230° (Figure 4.16(i)) and 220° bTDCf (Figure 4.16(m)). This vortex crossing occurs 10° CA later for the CFD flow field at 230° bTDCf (Figure 4.16(j)), causing high values of the WRI due to the opposed flows between CFD and PIV at 240° bTDCf (Figure 4.16(g)), followed by high values

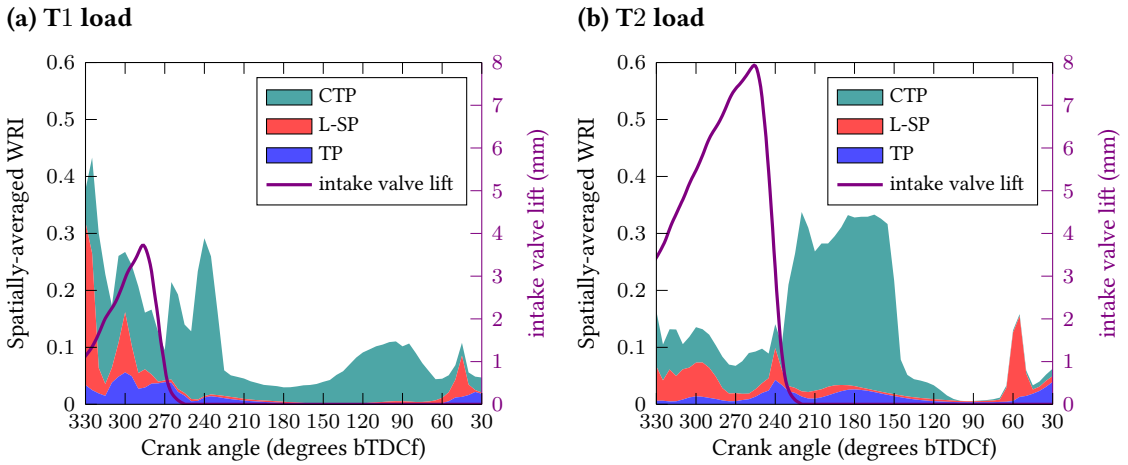


Figure 4.17: Spatially-Averaged WRI for the T1 and T2 test conditions, shown as stacked histograms (shaded areas, left y axis) for data in the tumble (blue), lower swirl (red) and cross-tumble (dark green) planes. The solid purple line (right y axis) in each plot shows the mean intake valve lift for all the measured cycles at the listed engine loads.

of the WMI due to the low velocities within the tumble vortex centre for CFD at 230° bTDCf (Figure 4.16(l)). Once the tumble vortex is found on the exhaust side of the cylinder for both PIV and CFD, there is a good match in both direction and magnitude from 220° to 150° bTDCf as shown by the low CMRI values in Figure 4.13(a).

The extended period of high spatially averaged CMRI values in the cross-tumble plane from from 230° to 140° bTDCf identified in Figure 4.13(b) for the T2 test point may be assigned to a difference primarily in flow direction as high WRI values provide the dominant contribution to the increase in CMRI over these crank angles (Figure 4.17(b)). These directional differences arise from a sequence of flow reversals for both PIV and CFD in the cross-tumble plane as a result of the motion of the tumble vortex centre across the tumble plane during the cycle (Figure 4.18). At 225° bTDCf the intake valves are fully closed and the bulk tumble flow is establishing. The tumble vortex centre for the CFD flow field is at approximately $z = -30$ mm (Figure 4.18(b)), and the tumble motion follows the sloped geometry of the cylinder head giving an upwards flow in the cross-tumble plane ($y - z$ plane, where $x = 0$ mm, Figure 4.18(d)). For the tumble plane

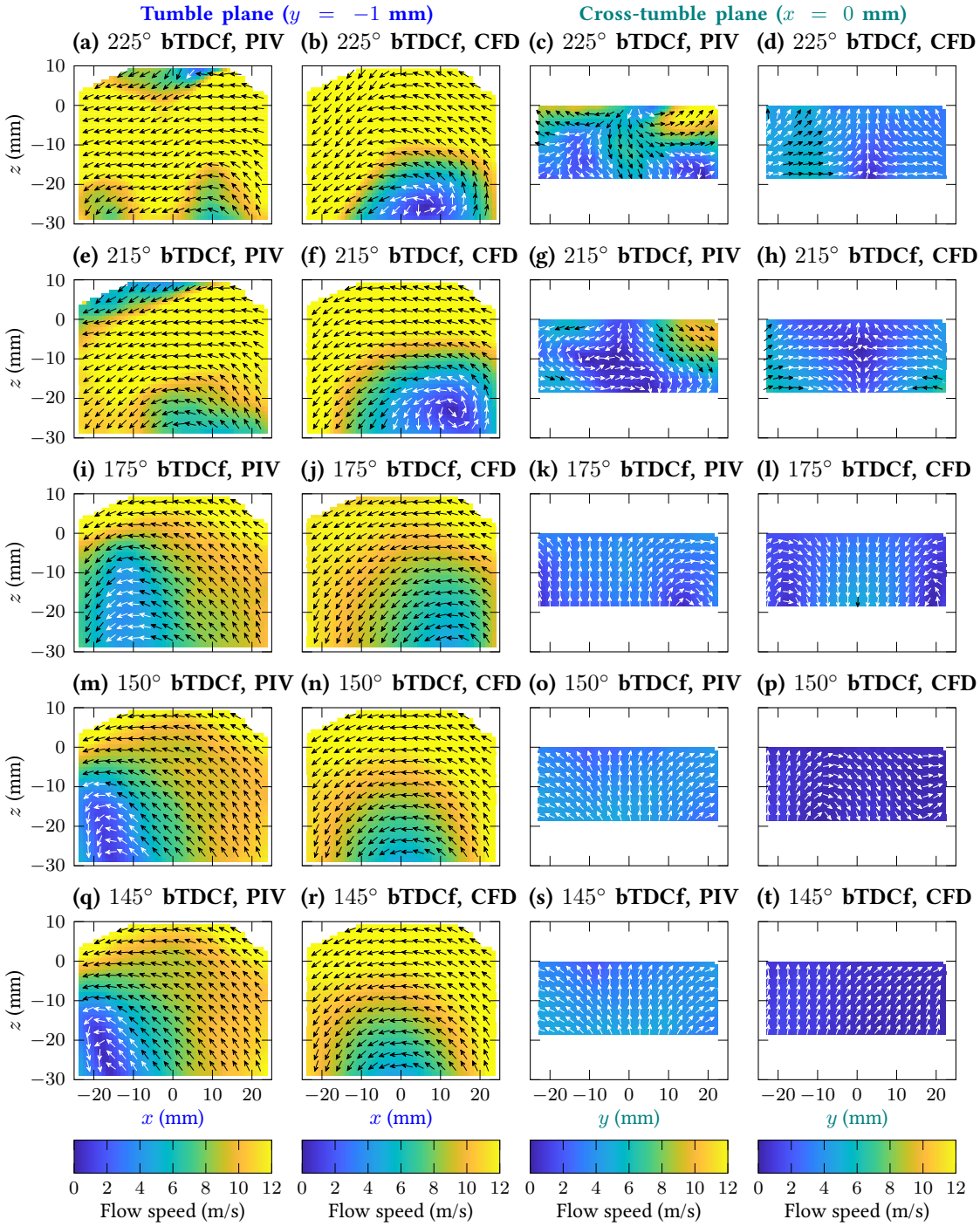


Figure 4.18: Flow direction reversal in the cross-tumble plane (Columns 3 & 4) for the T2 test point while the tumble vortex centre shown in the tumble plane (Columns 1 & 2) crossing from the inlet (positive x) to exhaust (negative x) for PIV (Columns 1 & 3) and CFD (Columns 2 & 4).

($x - z$ plane, where $y = -1$ mm) PIV flow field (Figure 4.18(a)), the tumble vortex centre is located below the field of view and a central region of slow flow near the cylinder head diverts the tumble flow downwards into the centre of the cylinder. This causes the downwards flow in the PIV cross-tumble plane (Figure 4.18(c)) and corresponding high WRI values (Figure 4.17(b)) for the opposed PIV and CFD flow directions (Figure 4.18(c) versus (d)). The presence of this slow flow region is also consistent with a similar observation in an earlier work from our group [99][(6)] using the same engine with a different piston crown geometry.

At 215° bTDCf the tumble vortex centre moves towards the inlet (positive x) for CFD (Figure 4.18(f)) and towards the exhaust (negative x) for PIV (Figure 4.18(e)), resulting in a synchronised switch in both cross-tumble plane flow directions which maintain opposite to each other (Figure 4.18(g) versus (h)).[(7)] These high WRI values (Figure 4.17(b)) persist while the tumble vortex centres of the PIV and CFD flow fields lie on opposite sides of the cylinder (for instance at 175° bTDCf, Figure 4.18(i) versus (j)). It is not until at 150°–145° bTDCf when the CFD tumble vortex centre traverses the cross-tumble plane to the exhaust side of the cylinder (Figure 4.18(n) and (r)) and reverses the CFD flow direction in the cross-tumble plane (Figure 4.18(r)) to be in line with the PIV flow direction (Figure 4.18(q)), that the WRI values improve (i.e., decrease). After 145° bTDCf both PIV and CFD have an upwards flow direction in the cross-tumble plane (Figure 4.18(q) and (r)) since the tumble vortex centre is on the exhaust side (negative x, Figure 4.18(s) and (t)), and this good match (low WRI value) is maintained for the rest of the compression stroke (Figure 4.17(b)).

Towards the end of the compression stroke (after 60° bTDCf), the flow fields in the tumble and cross-tumble planes are well matched for both T1 and T2 with spatially-averaged CMRI values

[(6)] This paper was published before the author of this thesis joined the group, hence it only included the data that were measured in the tumble plane. The author of the thesis conducted tests for the lower swirl plane, the higher swirl plane and the cross-tumble plane.

[(7)] The cross-tumble plane flow direction switch for the PIV ensemble mean flow field was discussed in detail in Section 4.2 and illustrated by Figures 4.4 and 4.5 using data at the same T2 test point.

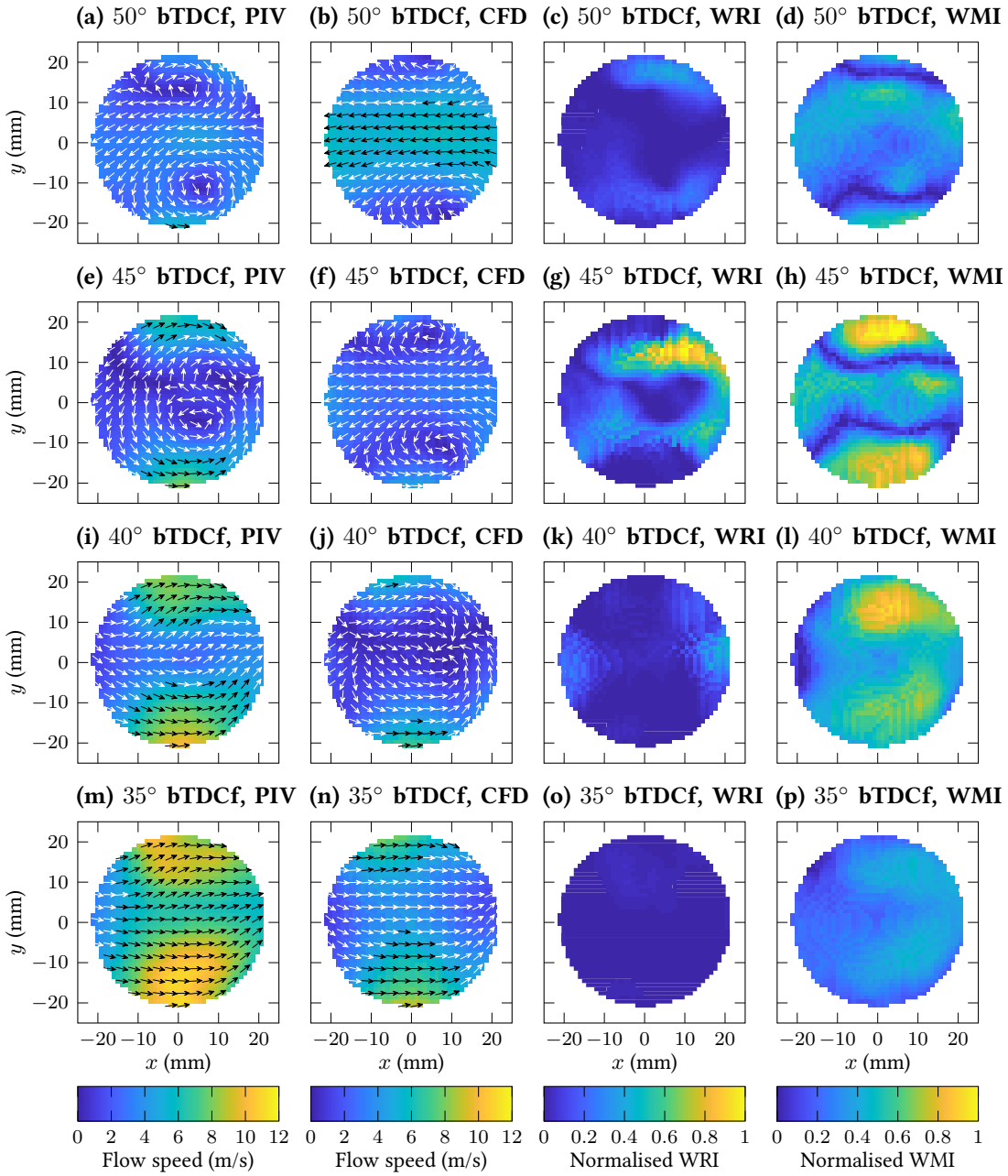


Figure 4.19: Flow direction reversal in the lower swirl plane for the T1 test point while the tumble vortex centre passing through the measured plane for PIV ensemble mean (Column 1) and CFD (Column 2) flow fields with local differences in direction and magnitude quantified using the WRI (Column 3) and WMI metrics (Column 4), respectively. Note that the flow direction reversal occurs at different crank angles (rows) for the PIV ensemble mean and CFD flow fields.

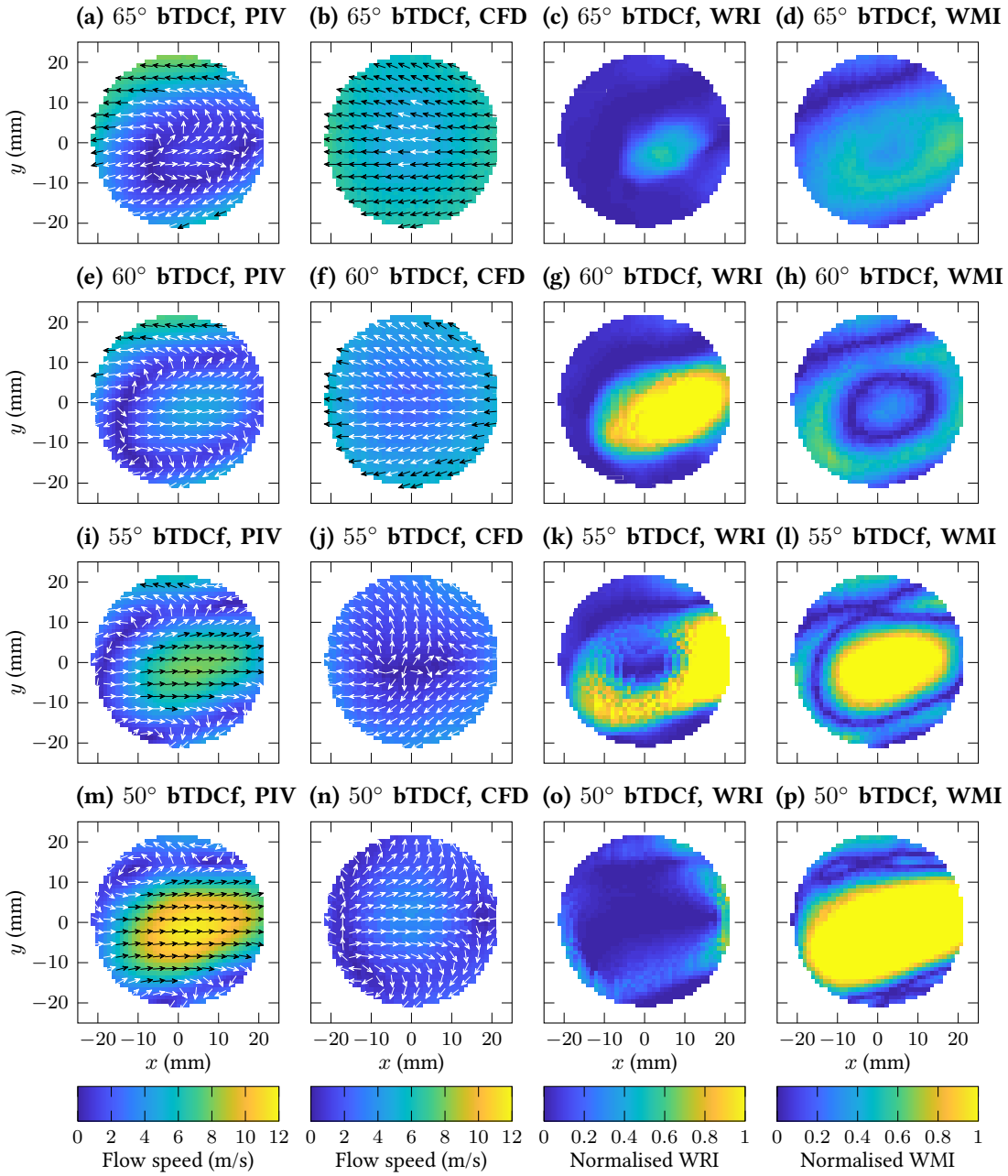


Figure 4.20: Flow direction reversal in the lower swirl plane for the T2 test point while the tumble vortex centre passing through the measured plane for PIV ensemble mean (Column 1) and CFD (Column 2) flow fields with local differences in direction and magnitude quantified using the WRI (Column 3) and WMI metrics (Column 4), respectively. Note that the flow direction reversal occurs at different crank angles (rows) for the PIV ensemble mean and CFD flow fields.

below 0.05 (blue and dark green shaded areas in Figure 4.13). The lower swirl plane CMRI values (red shaded area), however, display a peak at $45°$ bTDCf for T1 (Figure 4.13(a)) and at $55°$ bTDCf for T2 (Figure 4.13(b)), and hence further investigations on the flow fields are needed.

Inspection of the lower swirl plane flow fields identifies a delayed flow reversal for CFD compared to PIV as the tumble vortex centre is carried through the swirl measurement plane by the rising piston compressing the in-cylinder flow (Figure 4.19). Initially the tumble vortex centre for both PIV and CFD lies below the swirl measurement plane (Figures 4.11(c)–(j) and 4.12(c)–(j)). The lower swirl plane flow at $z = -3.5$ mm is therefore from inlet to exhaust for the majority of the intake and compression strokes. For the T1 load, at $50°$ bTDCf a pair of counter-rotating vortices appear in the PIV lower swirl plane flow field at approximately $y = \pm 10$ mm (Figure 4.19(a)). Similar vortices also appear in the CFD flow field, albeit $5°$ CA later at $45°$ bTDCf (Figure 4.19(f)), at which time the PIV flow field (Figure 4.19(e)) has started to reverse its direction from exhaust to inlet, hence the high WRI (Figure 4.19(g)) and WMI (Figure 4.19(h)) indices. The PIV tumble vortex centre passed through the swirl measurement plane from $45°$–$40°$ bTDCf (Figure 4.19(e) and (i)), leaving the PIV to measure the horizontal components of velocity in the lower half of the tumbling flow (Figure 4.19(i)). At $40°$–$35°$ bTDCf (Figure 4.19(j) and (n)) the CFD flow direction also reverses, recovering the good match in the lower swirl plane found throughout the majority of the cycle (low spatially-averaged CMRI values in Figure 4.13(a) and low spatially-averaged WRI values in Figure 4.17(a)).

For the T2 load, the flow field also reverses first for PIV at $65°$ bTDCf (Figure 4.20(a))[8] but with a $10°$ CA delay before the CFD flow field reverses at $55°$ bTDCf (Figure 4.20(j)). Instead of a pair of counter-rotating vortices observed at the T1 test point (Figure 4.19(a)), the flow reverses in the central region of the lower swirl plane at T2 (Figure 4.20(a)), surrounded by an annulus of low in-plane velocities. This may be linked to the offset in the T2 (Figure 4.12(k)–(i)) tumble

[8] The flow direction reversal for the PIV ensemble mean flow field was discussed in detail in Section 4.2 and illustrated by Figures 4.7 and 4.8 using data at the same T2 test point.

vortex centre location being approximately 20 mm towards the exhaust compared to the central location for T1 (Figure 4.11(k)–(i)).

4.4.4 Asymmetric CFD flow during compression in the cross-tumble plane

Aside from the timings of flow reversals discussed in the previous section (Section 4.4.3), the flow fields in the cross-tumble plane between the IVC (at around 260° bTDCf for T1 and 240° bTDCf for T2) and 160° bTDCf display nearly uniform vertical flow directions for both the T1 (Figure 4.16) and T2 (Figure 4.18) test points. For the T2 test point, once the final flow reversal occurs at 150° bTDCf for CFD (Figure 4.18(p)), the CMRI values show a good match across all three planes until 70° bTDCf (Figure 4.13(b)), with a vertical flow direction in the cross-tumble plane (Figure 4.18). On the other hand, after 160° bTDCf the cross-tumble plane CMRI values for T1 increase to peak at 100° bTDCf (dark green shaded area in Figure 4.13(a)) as a result of increased WRI values (Figure 4.17(a)), as opposed to the excellent match between CFD and PIV for T2 with near-zero WRI values (Figure 4.17(b)) over the same region in the cross-tumble plane. The increased WRI values for T1 are due to the appearance of asymmetry in the CFD flow field which is not observed in the PIV flow field (Figure 4.21).

At 160° bTDCf the CFD cross-tumble flow field for T1 begins to differ from the vertical PIV flow field, as suggested by the increasing WRI values in Figure 4.17(a). By 130° bTDCf, in comparison to the previously well-matched PIV flow field (Figure 4.21(a)), the CFD velocities on the right of the cross-tumble plane increase in magnitude and rotate towards the left of the cross-tumble plane where the CFD velocity magnitude is reduced (Figure 4.21(b)). At 100° bTDCf the CFD flow field (Figure 4.21(d)) has become horizontal across the central 20 mm of the cross-tumble plane, producing high WRI values (Figure 4.17(a)) when compared to the PIV flow field which remains vertical (Figure 4.21(e)). This difference in flow characteristic persists until 60° bTDCf when the rising piston causes the region of horizontal flow in CFD to

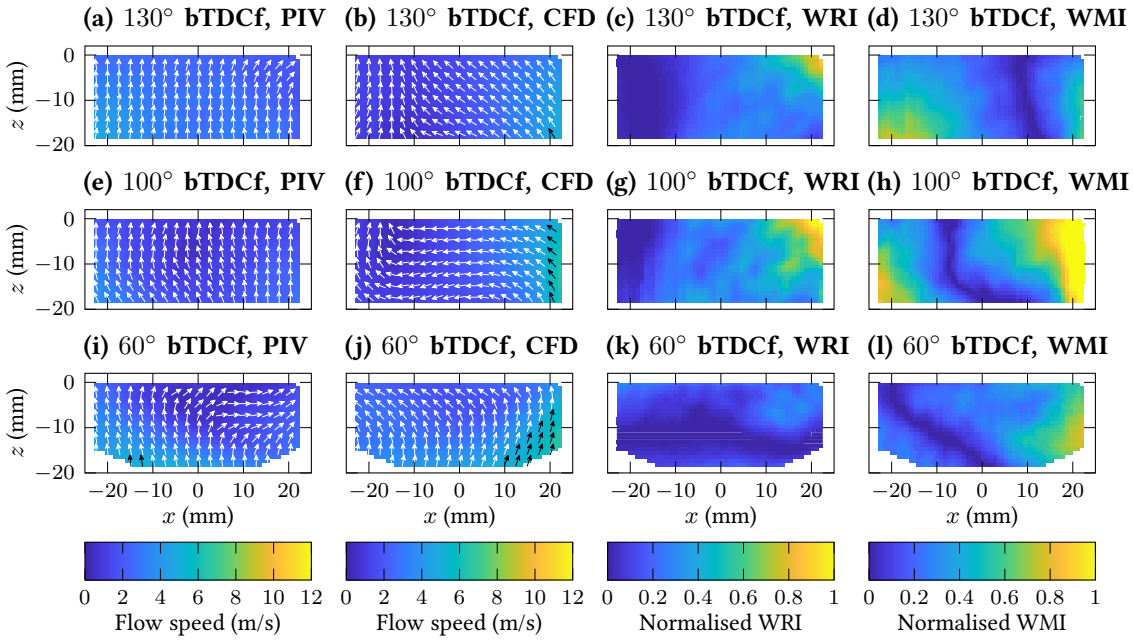


Figure 4.21: Asymmetric mid-compression cross-tumble flow field feature for the T1 test point shown in the PIV ensemble mean (Column 1) and CFD (Column 2) flow fields with local differences in direction and magnitude quantified using the WRI (Column 3) and WMI metrics (Column 4), respectively. The asymmetric flow feature is only observed in the simulated flow fields (Column 2).

leave the cross-tumble field of view, to be replaced by vertical flow near the piston which is well matched to the PIV flow field (Figure 4.21(i) versus (j)). This phenomenon is not expected because the engine geometry is symmetrical to the mid-valve tumble plane ($y = 0$ plane), and hence the simulation ought to generate symmetrical flows on both sides of the cylinder when viewed in the central cross-tumble plane ($x = 0$ plane). Further discussion is being made with the JLR CFD team to investigate the reason, and the current hypothesis is that the CFD mesh may not be perfectly symmetrical to the mid-valve tumble plane.

4.5 Chapter summary

In this chapter, high-speed PIV measurements of the in-plane velocities across the tumble, cross-tumble and lower swirl planes of the optically accessible engine (Appendix A) have been quantitatively compared to the RANS simulations of the flow fields for motored operation under two different engine loads. The flow fields measured in various planes show high consistency, which enable the comparison with the RANS data. Spatially-resolved differences in alignment and magnitude between the experimental and simulated in-cylinder flow fields have been quantified both independently using the weighted relevance index (WRI) and weighted magnitude index (WMI), as well as in combination using the combined magnitude and relevance index (CMRI). Spatially-averaged CMRI values for each plane provide a rapid and quantitative method for evaluating the quality of the match between the PIV and CFD flow fields both between planes and between test conditions. Overall the higher load test condition (T2) provided the best match between simulation and experiment. Application of the CMRI, WRI and WMI metrics also enabled the identification of key timings within the intake and compression strokes, specific to each test point, for further investigation into differences between the PIV and CFD flow fields. In general, a very good match was found between the PIV and CFD flow fields across all three planes and for both test conditions. Differences that were identified include:

- The RANS simulation accurately captures the intake flow characteristics for T2 but predicts higher velocities within the intake jet for the lower load test point (T1). For both test points the increased jet velocities predicted by CFD during inlet valve closing were not observed for PIV.

- A series of flow direction reversals in the cross-tumble and swirl planes arise due to the development of the bulk tumble motion and the subsequent motion of the tumble vortex

centre during the cycle. Differences in the location of the tumble vortex centre between the PIV and CFD flow fields resulted in delays of $5°$–$10°$ CA in the in-plane flow reversals for CFD compared to PIV.

- A low velocity asymmetric flow in the cross-tumble plane was predicted by CFD during compression for T1 but not observed in the PIV measurements which showed a uniform vertical flow for the corresponding crank angles.

The ensemble mean PIV fields, each derived from 300 cycles, are compared to the RANS simulated flow fields. As such, the effect of cycle-to-cycle fluctuations and turbulence present in each individual PIV cycle is averaged out, resulting in mean velocities which are less extreme than those found in an individual cycle. Indeed, where differences in magnitude are present, the RANS velocity magnitudes for T1 and T2 are typically higher than the corresponding mean PIV magnitudes. A representative single measured cycle after dimensionality reduction may provide a more appropriate comparison to the RANS flow fields, and this idea is discussed in the next chapters (Chapters 5 and 6).

Chapter 5

Validation of flow model using POD reconstructed PIV data

In this chapter, the RANS simulated flow at a specific crank angle during the intake stroke in the cross-tumble plane is chosen to exemplify its validation process against averaged or filtered PIV data that were computed using different approaches. A set of metrics is then developed to quantify the intake jet profile. Various methods are tested, including the phase-averaged ensemble mean, the modified mean flow field based on a speed average, and the individual cycles approximated by the low-order POD reconstruction, among which the low-order POD reconstructed flow fields give a fairer and more robust comparison with the simulated data. During the reconstruction process, a stability test is also developed to determine the number of modes needed in the low-order POD reconstruction. The majority of contents in this chapter are adapted from a paper [2] first-authored by the author of this thesis with additional information and more detailed discussions included.

5.1 Motivation

The phase-averaged PIV ensemble mean flow fields serve as good validation targets for the RANS flow model in most cases, and their differences can be quantified by the vector field comparison metrics (Section 3.3) as discussed in the previous chapter (Chapter 4). During the intake stroke, however, engine flow fields measured by the PIV technique show the flow to have high turbulence and cycle-to-cycle variations that include intake jet flapping [102] — the direction of the jet flows from each of the two intake valves changes in different snapshots, and the resulting merged jet from the two intake streams can swing to various directions [103]. When viewed from the intake side of the engine (Figure 5.1), the intake streams from the two intake valves collide near the symmetric centreline plane of the cylinder ($y = 0$), resulting in a roughly downward-moving central intake jet in the central cross-tumble plane ($y - z$ plane, where $x = 0$ mm). Despite the engine being geometrically symmetric with respect to the central plane ($y = 0$), the intake jet flapping may cause the two intake streams to have different strengths and thus the central intake jet can swing from left to right (Figure 5.1(a)–(c)), or translate to one side of the engine (Figure 5.1(a) versus (d)) in different engine cycles even if the measurements were taken at the same crank angle. The ensemble averaging of the experimental data (Figure 5.1(e)), however, smooths out the fields, cancelling horizontal components of the flow and causing its velocity magnitude to be underestimated. Therefore, if the ensemble mean is used to validate the RANS modelled flow field (Figure 5.1(f)) in the presence of the intake jet flapping, the validation would be heavily skewed and may even result in a false conclusion.

The mismatch between the averaged PIV flow fields and the RANS simulation lies within their fundamental differences in averaging mechanisms — PIV takes multiple measured vector fields and performs a pure mathematical average to generate an artificial field, while RANS adopts averaged physical quantities from the PIV experiment into a physics-based model that is

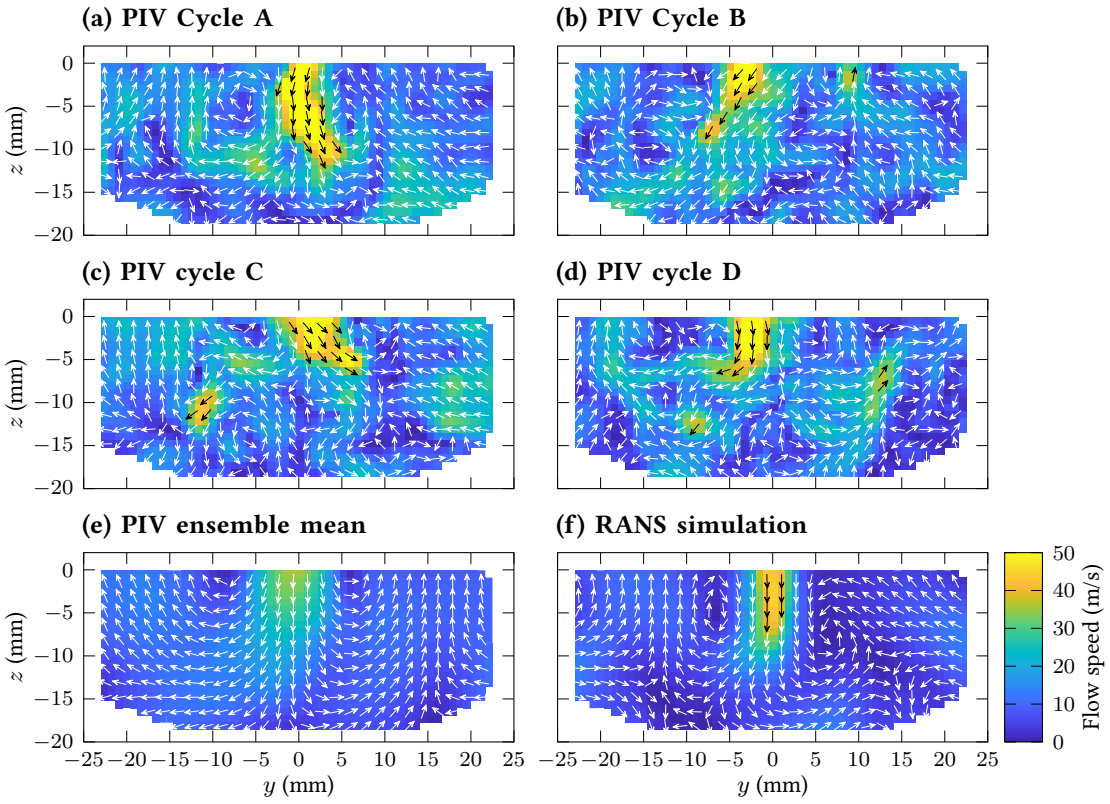


Figure 5.1: Flow fields in the cross-tumble plane at 300° bTDCf (during the intake stroke when intake valves are open). Data were taken at the T2, Piston B, C33, DVA test point.[(1)]

governed by thermo-fluid equations. An alternative method to the ensemble mean is therefore required to filter the turbulence and cycle-to-cycle variation within the flow fields, and to extract their major organised fluid motions (known as "coherent structures") to enable a fair validation for the RANS flow predictions. The proper orthogonal decomposition (POD) technique introduced in Section 1.5 and Section 3.2.1 provides a tool that can reduce the flow complexity and extract the dominant structure of the in-cylinder turbulent flow measured in

[(1)] It is worth reiterating that, as mentioned in a footnote for Figure 2.7, the vectors are in unit length and only illustrate the flow direction, while the magnitude of the vector at each location are shown by the colour map. For clarity only every second vector is plotted in each direction, while the false colour background uses the actual spatial resolution from the PIV experiment. The colour (black/white) of each vector is chosen to contrast with the false colour background. This plot style is adopted in this thesis for all the plots containing vectors, unless otherwise specified.

each cycle (at the same crank angle). After the dimensionality reduction by POD, the POD-reconstructed/approximated flow fields (*cf* Equation (3.28)) are used to validate the simulation results.

5.2 Quantification of intake jet profile

Once the coherent structures of the flow fields from individual cycles have been extracted, comparison to RANS requires quantification of the intake jet characteristics. Point-by-point comparison metrics for vector fields, such as the WRI, WMI and CMRI (Section 3.3) developed by Scott and Willman [99, 100] to compare RANS simulated data to PIV measurements, can quantify the similarity of pairs of fields and identify regions of interest. However, point-wise comparisons cannot accommodate more in-depth or subjective concepts of similarity such as a coherent structure being displaced or distorted between two fields. In the individual flow fields in the cross-tumble plane, the intake jet typically maintains a coherent structure but shows significant cycle-to-cycle variation in its shape, direction and location (Figure 5.1(a)–(d)). As such, any straightforward point-by-point comparison of velocities (whether provided by the WRI, WMI and CMRI metrics or another methods) would not provide the most appropriate way of quantifying the similarity of the jet structure among these fluctuating fields or between the measurements and the simulation. In the context of strong cyclic fluctuations of a coherent structure, the term "similarity" is better defined by key characteristics of the intake jet and the range of values these quantities take, rather than by a point-by-point comparison of the vector fields. Therefore rather than a point-wise comparison, quantities such as the intake jet angle, penetration length and width need to be defined and evaluated for each flow field. A zoomed-in plot for the RANS simulated flow field at 300° bTDCf (Figure 5.1(f)) is presented in Figure 5.2(a) near the intake jet region to exemplify the procedures of finding the jet angle, the jet penetration length and the nominal jet width.

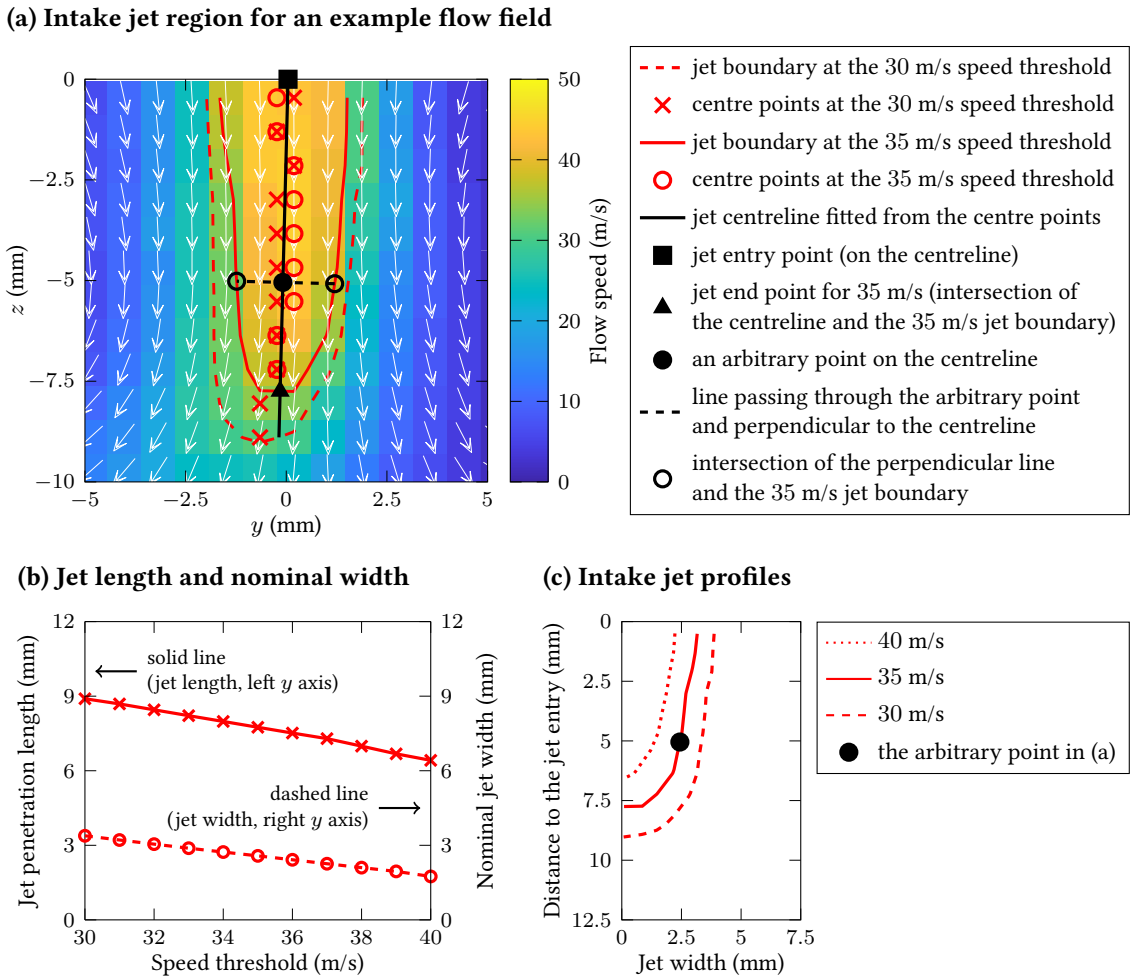


Figure 5.2: Intake jet profile definitions illustrated by an example flow field (the RANS simulated flow field at 300° bTDCf). (a) Intake jet region and notations of various definitions used to compute the jet profile metrics (labelled in the legend): the jet penetration length is given by the distance between the jet entry (black square) and end points at a given speed threshold (black triangle for 35 m/s); the jet width is a function of both the length along the jet and the speed threshold — for an arbitrary point on the jet centreline (black filled dot), the jet width at 35 m/s is given by the distance between two intersection points on the 35 m/s jet boundary (black hollow circles); For clarity, in this figure only, all the vectors are plotted in white regardless of their magnitudes, and the vector spatial resolution is the same as the actual PIV experiment. (b) Jet penetration length (red solid line with crosses, left y axis) and nominal jet width (red dashed line with hollow circles, right y axis) at different speed thresholds. (c) Intake jet profiles (width as a function of length) at different speed thresholds; The values for the arbitrary point in (a) are also marked.

Table 5.1: Robustness of the jet angle calculation using different combinations of speed thresholds

Speed thresholds (m/s)			Jet angle (degrees)
Min.	Max.	Increment	
25	40	1.0	3.91
25	35	1.0	4.09
30	45	1.0	1.09
30	40	1.0	1.45
30	40	0.1	1.43

5.2.1 Jet angle

The high-speed intake jet was first distinguished from the low-speed background by choosing appropriate speed thresholds. For instance, the red contour lines in Figure 5.2(a) illustrate the jet boundary using the 30 and 35 m/s speed thresholds, respectively. The jet region is therefore bounded by the top edge of the PIV field of view ($z = 0$, at the fire-deck) and the jet boundary at the given speed threshold. For each speed threshold, the centre points on each row (constant z values) of its jet region were then obtained (for example, red crosses for 30 m/s and red hollow circles for 35 m/s in Figure 5.2(a)). In order to have a unique and robust jet centreline (black solid line) for a given flow field, the centre points were found for a sequence of threshold values as defined in Table 5.1, in terms of the range of the threshold values and their increment. The centreline of the jet was defined by performing a linear fit to all the centre points, no matter whether or not they were obtained using the same speed threshold (i.e., consider the red crosses and the red hollow circles as equivalent in this step). This approach puts more weight on the higher-speed regions — the centre points at higher speed regions will be accounted for more times than the points in lower speed regions during the line fit process — and thus avoids the jet centreline being skewed towards the tip of the jet, of which direction fluctuates drastically

among the low speed thresholds. The angle between the jet centreline and the jet propagation direction (in this case the negative z direction) is the jet angle.

A robustness check (Table 5.1) for the jet angle of the example flow field (Figure 5.2(a)) shows that the evaluated jet angle is stable within a few degrees for different combinations of speed thresholds. All the jet angles presented in the later sections in this chapter were obtained by using the speed threshold combination which has a minimum speed threshold of 30 m/s, at 1 m/s increment, to a maximum of 40 m/s (highlighted in yellow in Table 5.1). All the speed thresholds within the highlighted combination will be used to evaluate the jet penetration lengths and nominal jet widths for any given flow field, in order to provide an overall comparison of jet profiles between the experimental and simulated data.

5.2.2 Jet penetration length

Once the jet centreline is determined, the jet penetration length for each flow field can be found by calculating the largest distance among its intersections with the jet region at a certain speed threshold. For instance, the distance between the jet entry point (black square in Figure 5.2(a)) and the intersection of the centreline and the jet boundary (black triangle) gives a jet penetration length of 7.7 mm at the 35 m/s threshold. The jet penetration length decreases at higher speed thresholds (red solid line with crosses in Figure 5.2(b), see left y axis). The slope of the line also shows the velocity gradient along the jet propagation direction — a more negative (steeper) slope means the velocity gradient is smaller, since the distance between two constant speed lines is larger.

5.2.3 Nominal jet width

Given an arbitrary point on the jet centreline inside the jet region (black solid dot in Figure 5.2(a)), a unique line (black dashed line) can be drawn that passes through the point and

perpendicular to the jet centreline. When this perpendicular line intersects with the contour of each given speed threshold at two point locations on opposite sides of the jet centreline (black hollow circles for 35 m/s in Figure 5.2(a)), the jet width corresponding to that point on the centreline is defined as the distance between these two intersections. If there is only one intersection detected (occurs when the arbitrary point on the jet centreline is too close to the entry of the jet), the jet width for that arbitrary point is not defined. Figure 5.2(c) illustrates the jet width as a function of length along the jet (distance away from the jet entry, black hollow square in Figure 5.2(a)) at three different speed thresholds. Due to the convergent shape, the jet width generally decreases from the entry (top) to the tip of the jet (bottom). The jet is wider for lower speed thresholds for the same jet length, which can also be intuitively observed in Figure 5.2(a). Figure 5.2(c) also shows that the arbitrary point in Figure 5.2(a) is about 5 mm away from the jet entry, and the jet width at this location is 2.4 mm for the 35 m/s speed threshold.

In order to have a single quantity for the jet width, the median of the jet width at all lengths along the jet for a certain speed threshold is selected to be the nominal jet width (red dashed line with hollow circles in Figure 5.2(b), see right y axis). The median was chosen against the mean to provide a more robust result that would otherwise be skewed by the highly-curved jet tip. Similar to the jet length (red solid line), the nominal jet width is also larger at lower speed thresholds. The slope of the nominal jet width versus speed thresholds illustrates the velocity gradient perpendicular to the jet propagation direction — again, a more negative (steeper) slope leads to a smaller velocity gradient. These jet quantification metrics are used to evaluate the similarity of the flow fields in the presence of a flapping intake jet in later sections of this chapter.

5.3 Comparison between RANS results and averaged PIV data

Before using the POD technique to filter the PIV multi-cycle data and validate the RANS flow model, it is worth comparing the RANS results with the conventional averaged PIV data first to investigate why the PIV ensemble mean and other averaged flow fields fail to be an appropriate validation target when the flapping intake jet is present in the field of view. In this section, the PIV ensemble mean and speed-based average flow fields are respectively compared to the CFD simulated flow field.

5.3.1 Ensemble mean

The ensemble mean flow field at 300° bTDCf (Figure 5.1(e)) is presented again in Figure 5.3(b) with a zoomed-in plot near the jet region. The flow speed inside the central jet region is notably slower than the RANS simulation (Figure 5.3(a), also plotted in Figures 5.1(f) and 5.2(a)). Again, if the simulation results were validated against the ensemble mean, one would easily conclude that the model over-estimated the jet speed. However, as visualised before in Figures 5.1(a)–(d), the actual jet speed in individual PIV cycles is much higher and closer to the RANS prediction.

Careful examination of the velocity data measured on an example single point ($y = 0.19, z = -2.15$ mm) inside the jet region for 300 cycles (Figure 5.4) suggest that large cyclic variation exists in the horizontal velocity components, for which the histogram (Figure 5.4(a)) has two peaks with comparable heights – both near 25 m/s (absolute values) but with opposite signs. Such a bimodal distribution is a classic signature of a flapping jet. The ensemble mean (Equation (3.5)) averages out the horizontal velocity components, resulting in an almost zero horizontal speed ($\overline{v}_{\text{em}} = -0.037$ m/s). Its magnitude (Equation (3.6)) is therefore much smaller than the jet speed of individual cycles.

Viewed from another perspective, each of the two velocity components at a certain grid point can be regarded as a random variable, of which a histogram (Figure 5.4(a) for horizontal

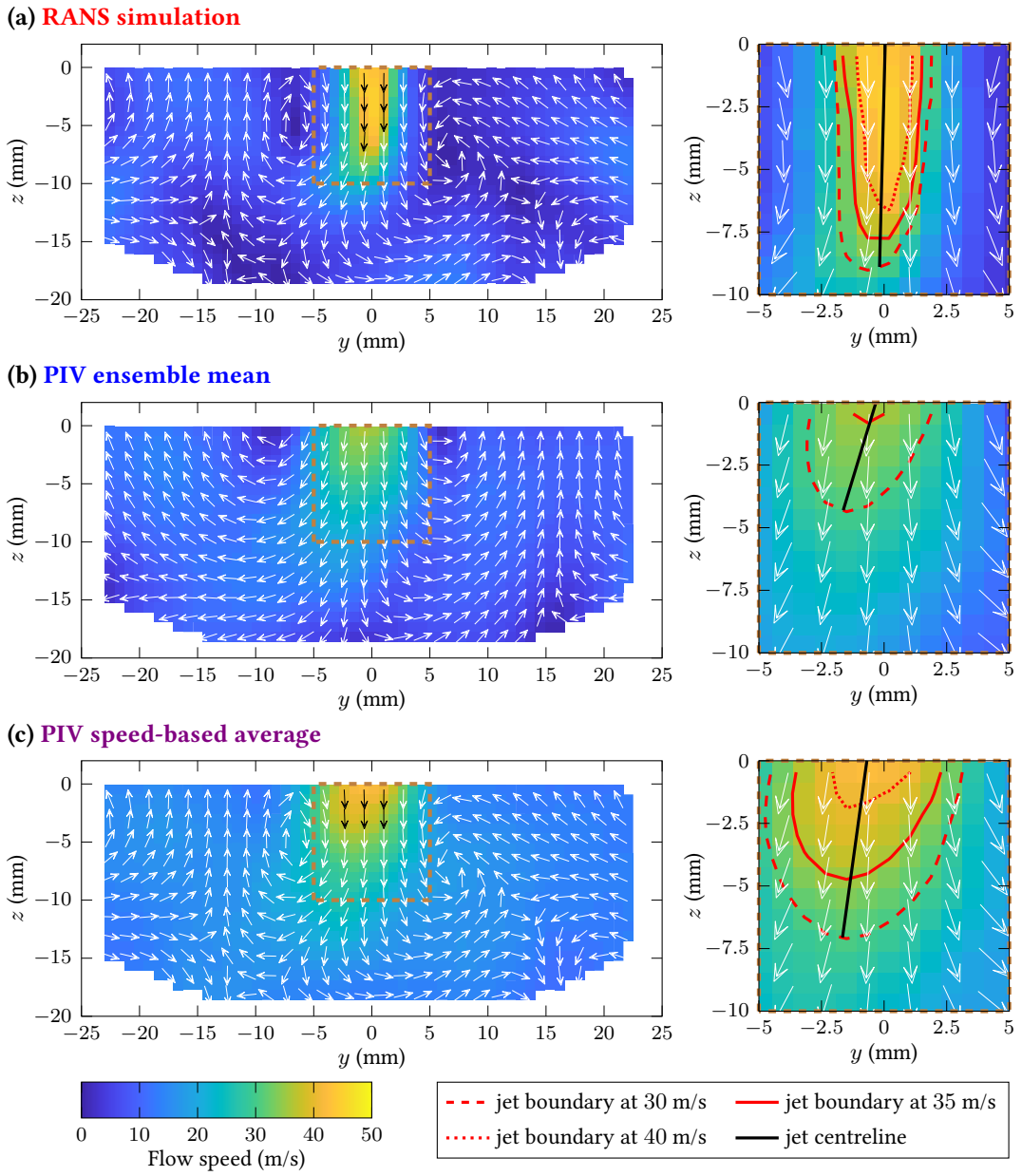


Figure 5.3: RANS simulated flow field versus the averaged PIV flow fields at 300° bTDCf. The zoomed-in region (brown dashed box) of each figures in the left column is plotted in the right column. The jet boundaries at various speed thresholds and the jet centreline are superposed on the zoomed-in plots (right column). Note that the 40 m/s jet region (red dotted line) for the PIV ensemble mean is undefined due to low flow speeds. For clarity, in the zoomed-in plots in this figure only, all the vectors are plotted in white regardless of their magnitudes.

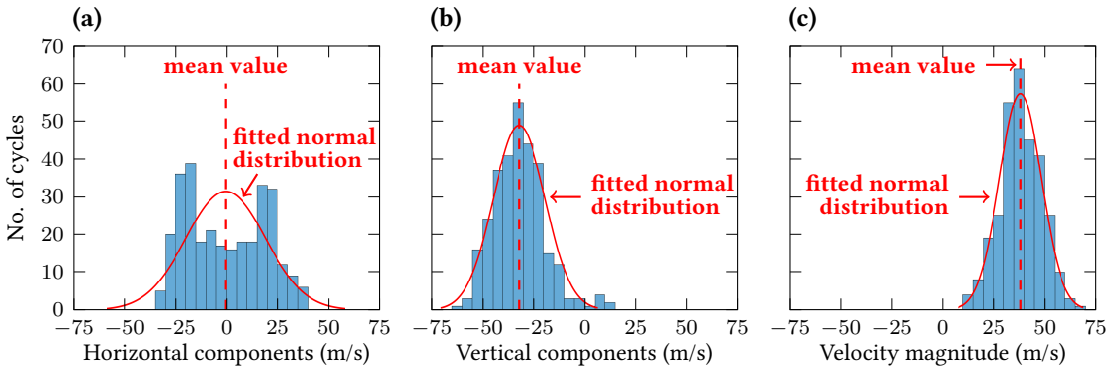


Figure 5.4: Speed distributions at a single point inside the jet region ($y = 0.19, z = -2.15$ mm) for 300 cycles at 300° bTDCf, illustrated by histograms of (a) 300 horizontal velocity components, (b) 300 vertical velocity components and (c) 300 velocity magnitudes (i.e., flow speeds).

components, (b) for vertical ones) can be fitted by a normal distribution (red solid lines), and the validity of the fit can be examined by the chi-squared goodness-of-fit test [104]. The horizontal components give a poor match ($\chi^2\,(7, N = 300) = 102$, $p \approx 0 < 0.05$), resulting from the swinging central jet that originated from the collision of the two intake streams; the vertical components generally follow the normal distribution ($\chi^2\,(5, N = 300) = 7.6, p = 0.18 > 0.05$) and have negative values, showing that the majority of cycles have a downward central jet at that location. Consequently, using the vertical component of the ensemble mean to approximate the overall flow behaviour along the z axis would be appropriate since it is equivalent to using the mean of a normal distribution to estimate the ensemble. Nevertheless, the same analogy cannot be applied for the horizontal component as the normal distribution assumption is no longer valid.

Quantifying the jet profile (Figure 5.5(a)) using the approaches defined in the previous section (Section 5.2) also confirms that the ensemble mean has a much shorter jet penetration length (blue solid line with crosses) at any chosen speed threshold compared to the RANS model (red solid line with crosses). At higher speed thresholds (over 35 m/s), the jet profile (Figure 5.5(b)) for the ensemble mean cannot even be defined since it significantly under-represents the jet speeds present within individual cycles. As such, the ensemble mean would not be a suitable

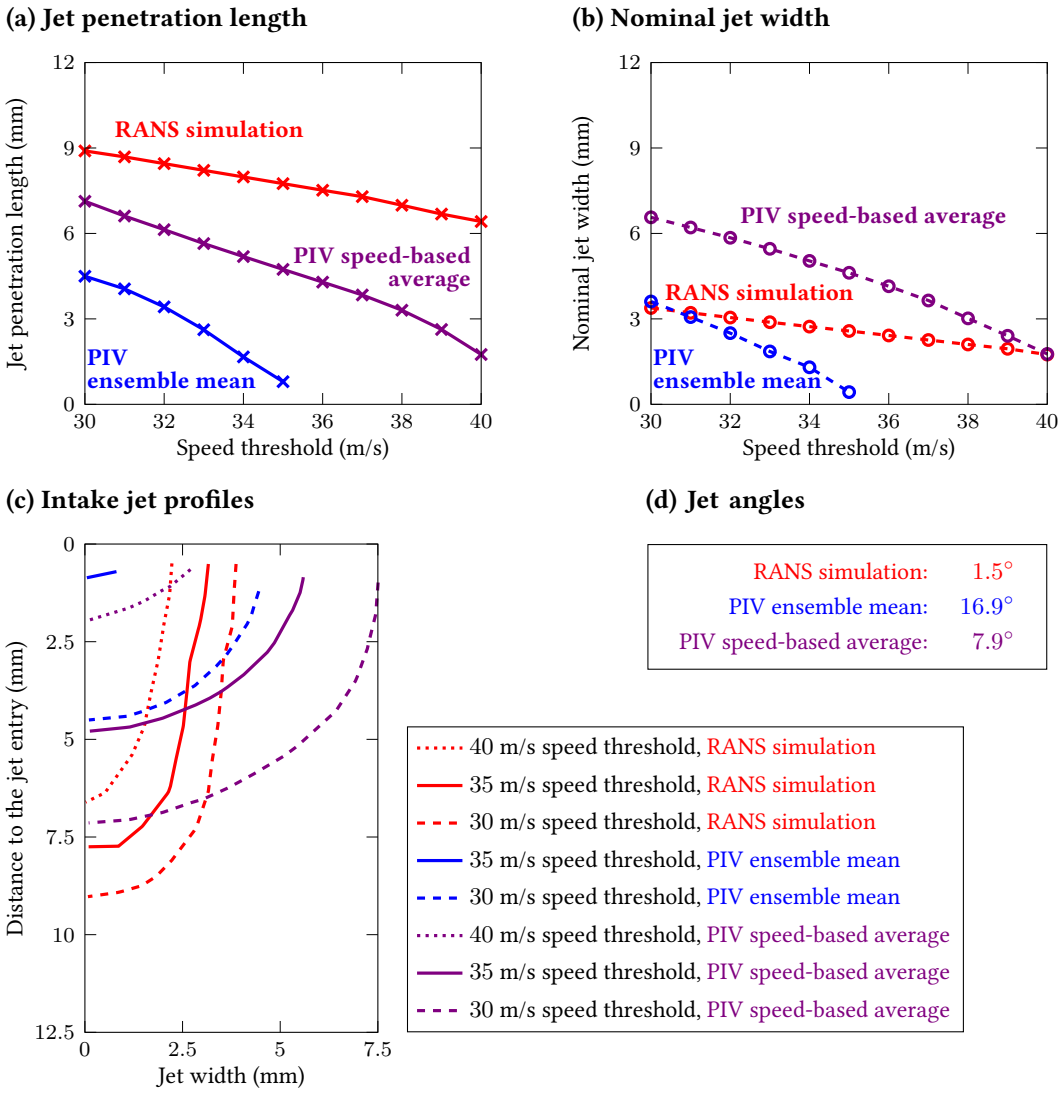


Figure 5.5: Intake jet quantification metrics for the RANS simulated and averaged PIV flow fields at 300° bTDCf. (a) Jet penetration lengths and (b) nominal jet widths for RANS simulated (red), PIV ensemble mean (blue) and PIV speed-based averaged (purple) flow fields as functions of the speed threshold. (c) Intake jet profiles (width as a function of length) at the 40 m/s (dotted lines), 35 m/s (solid lines) and 30 m/s (dashed lines) speed thresholds; note that the maximum flow speed in the PIV ensemble mean is lower than 40 m/s, so the jet profile for the PIV ensemble mean is not defined at the 40 m/s speed threshold. (d) Jet angles for different types of flow fields.

representation of the intake flow fields in the cross-tumble plane when there is a jet in the field of view with significant direction changes among different cycles.

5.3.2 Speed-Based averaging

The cancellation between horizontal components with opposing signs during calculation of the ensemble mean (Figure 5.4(a)) inspires an alternative approach – averaging the magnitude of vectors (speed) at each grid point. The speed of each vector is always non-negative, and therefore such so-called speed-based averaging avoids cancellation by negative values. The histogram of the velocity magnitudes (Figure 5.4(c)) at the previous example point inside the jet region matches with the normal distribution at 5% significance level ($\chi^2\,(5, N = 300) = 10.47, p = 0.06$), and therefore its mean can be used as an approximation of the ensemble behaviour.

Figure 5.3(c) illustrates the speed-averaged flow field, of which flow direction is adopted from the ensemble mean (Figure 5.3(b)), while the colour map is replaced by the mean of magnitudes ($\overline{M}_{\text{sa}}$, Equation (3.8)). The flow speed inside the jet region is significantly higher than the ensemble mean, and is much closer to the actual jet speed measured in individual cycles Figures 5.1(a)–(d), as well as the RANS modelled data.

The jet profile in the PIV speed-averaged field, however, is quite different from the one in the RANS model. The speed-averaged jet (purple lines in Figure 5.5(a)) was found to have a smaller penetration length (solid lines with crosses) and a larger nominal width (dashed lines with hollow circles) than the RANS modelled jet (red lines). Both (purple) lines are steeper than the corresponding (red) lines for the RANS data, showing that the velocity gradients both along and perpendicular to the jet propagation direction are smaller, which is in line with the smoother jet profile observed in Figure 5.3 (c) compared with the RANS simulation (Figure 5.3 (a)). Additionally, as illustrated in Figure 5.5(b), its curvature (width as a function of length) is not similar to the RANS model. The RANS modelled jet (red lines) is “slim” and has a “test tube-like”

curvature — the entry is not very wide, and the width almost remains constant until the tip of the jet, at which the width rapidly reduces to zero. The speed-averaged jet (purple lines), on the other hand, shows a "bowl-like" curvature — it is much wider near the entry, and gradually shrinks as the length increases — as a result from the averaging of the swinging intake jet.

It should be noted that despite the speed-averaging method resolving the issue of under-representation of the intake jet speed, the background flow speed outside of the jet region is also increased compared to the ensemble mean. This is a direct result of the speed-averaging method since the directional fluctuations (whether cycle-to-cycle variations in the central jet structure or turbulent fluctuations across the flow field) no longer cause the flow speed cancellation, and any local high speed spots within individual flow fields, such as the one near ($y = -10, z = -10$ mm) in Figure 5.1(c), will contribute towards the mean speed.

In addition to the flow speed and jet profile, some other flow features are also smoothed out by the two above-mentioned averaging methods, which makes the RANS validation questionable if they were used as validation targets. For instance, the intake jet applies a shear stress on the relatively slower in-cylinder flow, and creates so-called "shear vortices" in the instantaneous flow fields (one of which is centred near $y = 6, z = -6$ mm in Figure 5.1(a)). This flow feature is predicted by the RANS simulation (Figure 5.3 (a)), but is smeared out in both averaging methods (Figure 5.3 (b) and (c)) since the vortex centres in individual cycles may have different locations. The flow directions at the top-left and top-right corners of the PIV field of view also differ appreciably in direction (y components) between the RANS and the averaged PIV data using both methods. Given these disadvantages in using the averaged flow fields, they would not always be suitable choices as RANS model validation targets, especially when the highly-fluctuating intake jets are present.

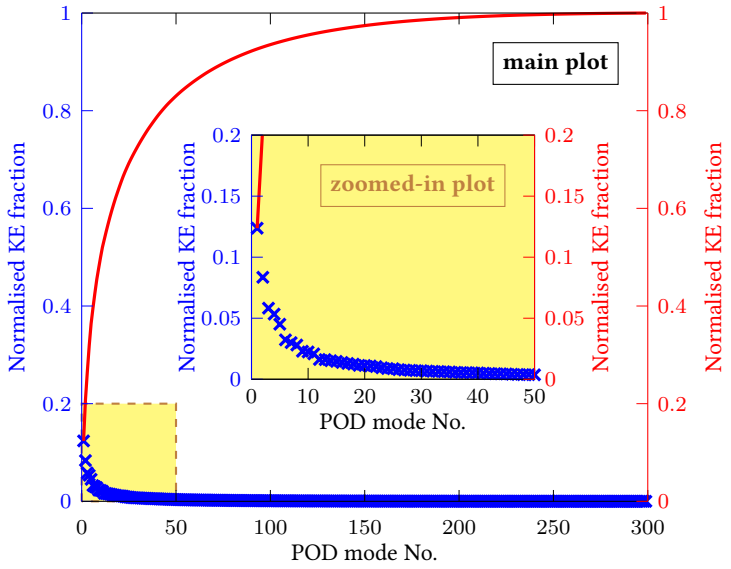


Figure 5.6: POD kinetic energy spectra at 300° bTDCf with a zoomed-in plot at lower-order modes. The zoomed-in region is bounded by a brown dashed box filled with light yellow colour in the main plot.

5.4 Comparison between RANS results and low-order POD reconstructed PIV data

Proper orthogonal decomposition has been used widely to filter turbulent flow data and extract coherent structures from individual flow fields (Section 1.5). This section aims to examine the possibility of validating a RANS simulated flow using approximated individual cycles by low-order POD reconstruction.

5.4.1 Low-Order POD approximation of PIV data

Given the fluctuation flow fields at 300° bTDCf for 300 engine cycles, POD (Equation (3.13)) decomposes them into 299 modes,[(2)] and also partitions their kinetic energy

[(2)] The number of POD modes is smaller than the number of cycles since the ensemble mean, which is a linear combination of all the cycles, was subtracted prior to conducting POD (Equation (3.10)).

content in a successively maximised manner (Equation (3.27)). As a result, the amount of kinetic energy distributed into each mode (orange crosses in Figure 5.6) decreases as the mode number increases. Most of the fluctuation kinetic energy is concentrated in lower order modes — the first ten modes (see cumulative sum in Figure 5.6) collected over 50% of the total kinetic energy, while less than 7% is in the last 200 modes (Mode 100 to the last). Therefore, it would be reasonable to approximate and extract the key features of the flow data in each cycle by reconstructing a field using the sum of several low order POD terms (Equation (3.28)). One key question remaining to be answered would be, how many POD modes should be included in the reconstructed data to represent the coherent flow structure, without introducing small-scale turbulence that may bias its comparison with the RANS modelled data. The answer requires detailed observations to the approximated flow fields, which are discussed in the following paragraphs.

The POD reconstructed fields using different numbers of POD modes (Equation (3.28)) for two arbitrarily selected engine cycles, Cycle A (Figure 5.1(a)) and Cycle B (Figure 5.1(b)), are shown in Figure 5.7. Each flow field has a 35 m/s contour highlighted in red marking the intake jet region and a brown box (3×3 mm) illustrating the vortex location of the original flow fields (Figure 5.7(i), bottom row) generated by the shearing flow from the central intake jet. It should be noted that the apparent flow patterns in individual POD modes (which for brevity are not shown here) do not represent actual flow features but only the relative strength of the fluctuations in different regions of the measured plane. The coherent flow structures in the instantaneous flow fields (Figure 5.7(i)) are instead preserved in the approximated flow fields (Figure 5.7(a)–(h)) once sufficiently large numbers of low-order POD components are added to the ensemble mean (*cf* Equation (3.28)) [64, 92]. The focus of this section lies only on the low-order approximated flow fields and thus discussions of the individual POD mode shapes are not presented.

The jet profiles are quantified by the method previously discussed in Section 5.2, and the results can be referred in Figure 5.8 — the jet angle (Figure 5.8(a)), jet penetration length

(Figure 5.8(b)), and nominal jet width (Figure 5.8(c)) are each plotted as a function of the number of POD modes used in the approximation for Cycle A (dark green lines) and Cycle B (orange lines). Note that the jet angle (Figure 5.8(a)) is not a function of the speed threshold, while two speed thresholds around 35 m/s (solid line with crosses for lower speed threshold at 34 m/s, dashed line with circles for higher speed threshold at 36 m/s) were chosen in order to show the robustness of the other two metrics (Figure 5.8(b) and (c)) versus the choice of speed threshold. The jet length and nominal width for the 35 m/s speed threshold, though not shown for a clearer visualisation, would lie between the solid and the dashed lines of each cycle.

For both cycles, at a first glance, the approximated flow fields (Figure 5.7(a)–(h)) become closer to the original flow fields (Figure 5.7(i)) as more terms are included (from top to bottom), while differences still remain between the two cycles. For Cycle A (left column), adding the first two POD terms (Figure 5.7(b) and (c)) barely changes the intake jet profile from the ensemble mean (Figure 5.7(a), which is identical for both Cycles A and B, and was shown in Figure 5.1(e) and Figure 5.3(b) before) — the intake jet speed is still under-estimated; the jet length is less than 2 mm (Figure 5.8(b)) since the 35 m/s contour (red solid line) can only be seen at the top of the field of view. After adding the third term (Figure 5.7(d)), however, the jet profile changes significantly — despite the jet speed still being lower than the original snapshot (Figure 5.7(i)); their jet lengths are now comparable. The error in the shear vortex location, marked by the brown boxes in Figure 5.7,[(3)] is also within 2 mm. However, the jet angle of the third term is positive (to the left, dark green solid line in Figure 5.8(a)), while the actual jet angle from the original snapshot ought to be negative (to the right, see Figure 5.7(i)). The reconstructed jet profile fluctuates strongly between the third- and the 11th-order of approximations (Figure 5.7(d)–(f)). For instance, at the fifth-order (Figure 5.7(e)), the contour

[(3)] These boxes refer to the location of a shear vortex induced by the intake jet in the corresponding original flow fields (Figure 5.7(i)), and are kept at the same locations for all the plots of the same cycle (presented in the same column).

illustrates a shorter and narrower jet region, and the values in the metrics also change rapidly from one approximation to the next (dark green lines in Figure 5.8). It should be noted that such apparently significant changes may, to a large extent, result from the slight magnitude changes around 35 m/s in the approximation, since the flow feature shown by the colour map and vectors is almost unaltered. On the other hand, the change does show that the flow speed inside the jet region is still fluctuating when more modes are included. After accounting for the 11th term (Figure 5.7(f)), the jet profile becomes more stable and evolves more gradually at higher-order approximations — only small changes can be observed visually from the 11th- to the 23rd-order approximations (Figure 5.7(f)–(h)), and the values of metrics (dark green lines in Figure 5.8) also converge to constants. For instance, the jet length is fluctuating between $10.6 - 12.5$ mm (slightly more than two vector spacings) from the 11th- to the 50th-order approximations. Therefore, Cycle A needs 11 POD modes to represent its coherent structure.

Cycle B (right column of Figure 5.7), on the other hand, shows another trend in its low-order POD reconstructed flows. The sudden flow feature change by adding one mode (Mode 3 for Cycle A) still exists but appears earlier — the first-order approximation (Figure 5.7(b)) differs greatly from the ensemble mean (Figure 5.7(a)) and illustrates a visible jet region with around 5.5 mm jet length and 3.5 mm nominal width at 35 m/s (orange lines in Figure 5.8(b) and (c)), as distinct from the zeroth-order (ensemble mean), which again almost has no jet region defined under this speed threshold. The ”evolution” of low-order approximations after the first-order (Figure 5.7(c)–(h)) appears to be very gradual, and the values of all three metrics (orange lines in Figure 5.8) change by small amounts. However, it is not until adding the 18th-order term (Figure 5.7(g)) to the approximation that the shear vortex starts to occur near the region of the original flow field (highlighted by the brown box). The jet penetration length also has a sudden increase (Figure 5.8(b)) before the 18th-order. Small differences in the jet profile can still be observed if more terms are added after the 18th-order (see, for instance, the 23rd-order approximation, Figure 5.7(h)), yet the change is considered to

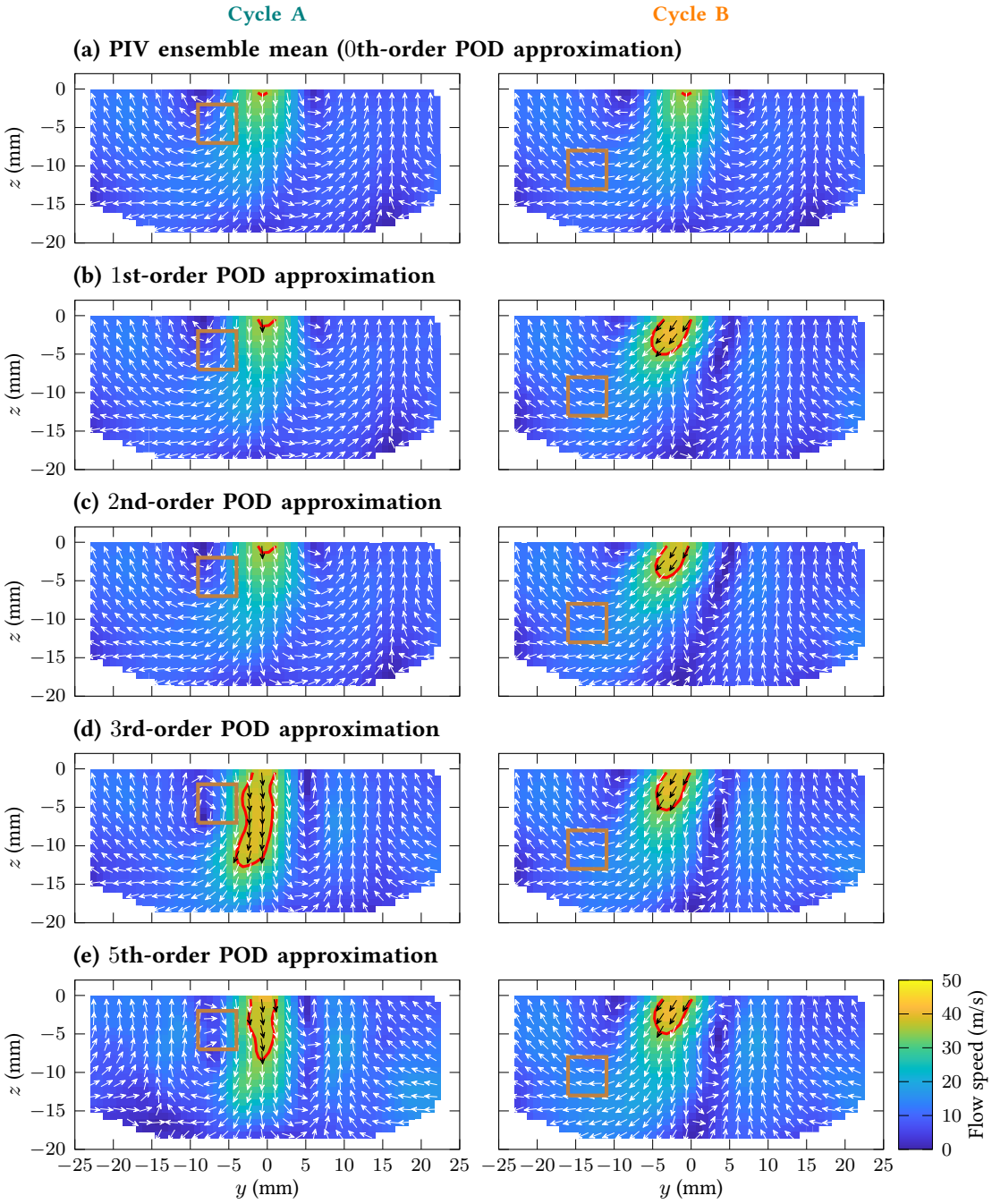


Figure 5.7: Approximated flow fields by the low-order POD reconstruction. Please refer to the detailed figure caption on the next page.

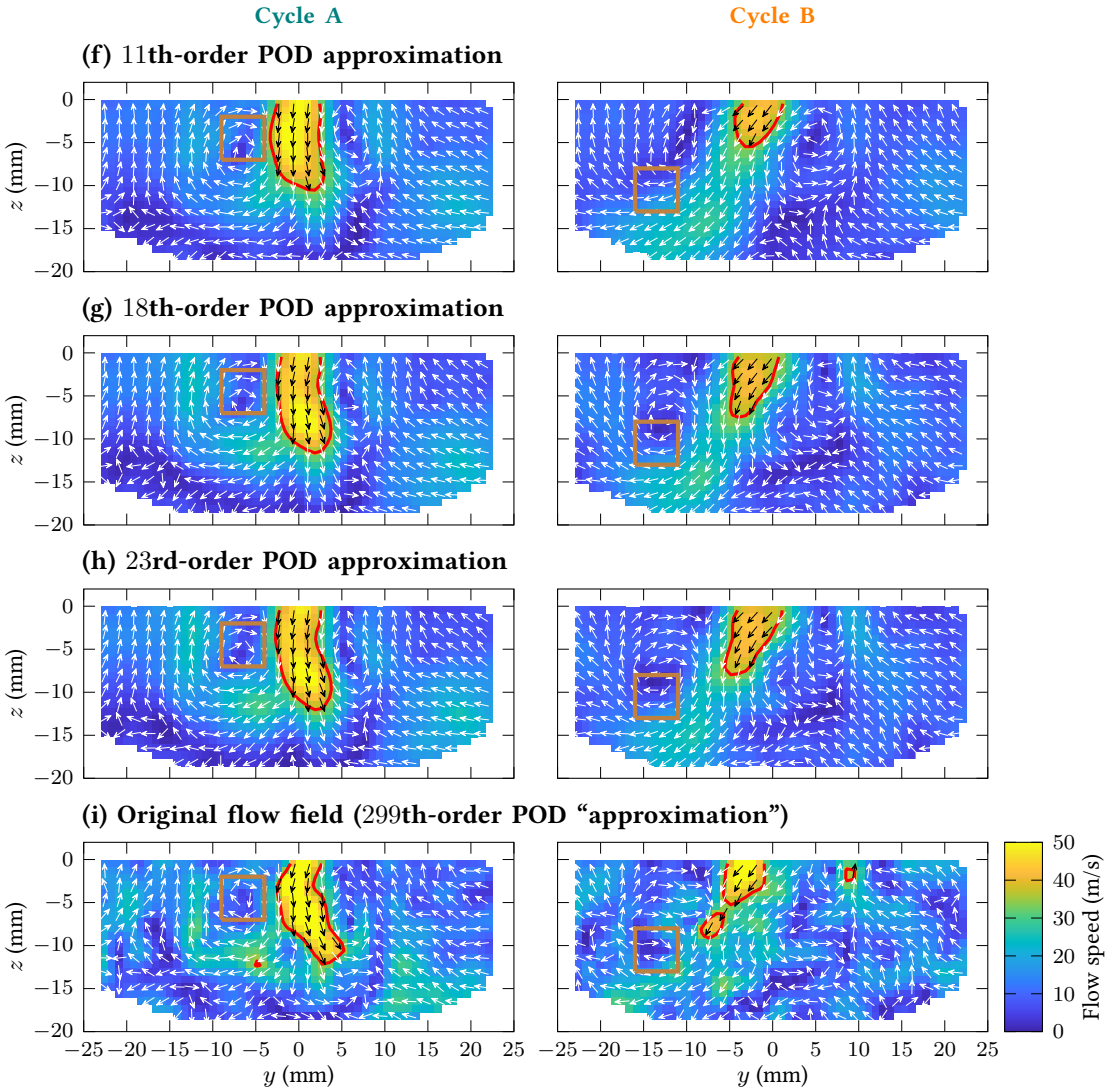


Figure 5.7: Approximated flow fields by the low-order POD reconstruction for Cycle A (left column, whose original flow field was also shown in Figure 5.1(a)) and Cycle B (right column, *cf* Figure 5.1(b)) at 300° bTDCf using different numbers of POD modes (i.e., orders). The brown boxes show the location of a shear vortex induced by the intake jet in the corresponding original flow fields, and the position of the brown boxes are kept the same for plots in the same column. The red contour lines illustrate the jet region at the 35 m/s speed threshold. Note that the ensemble mean (first row on the previous page) is the zeroth-order of approximation for both cycles.

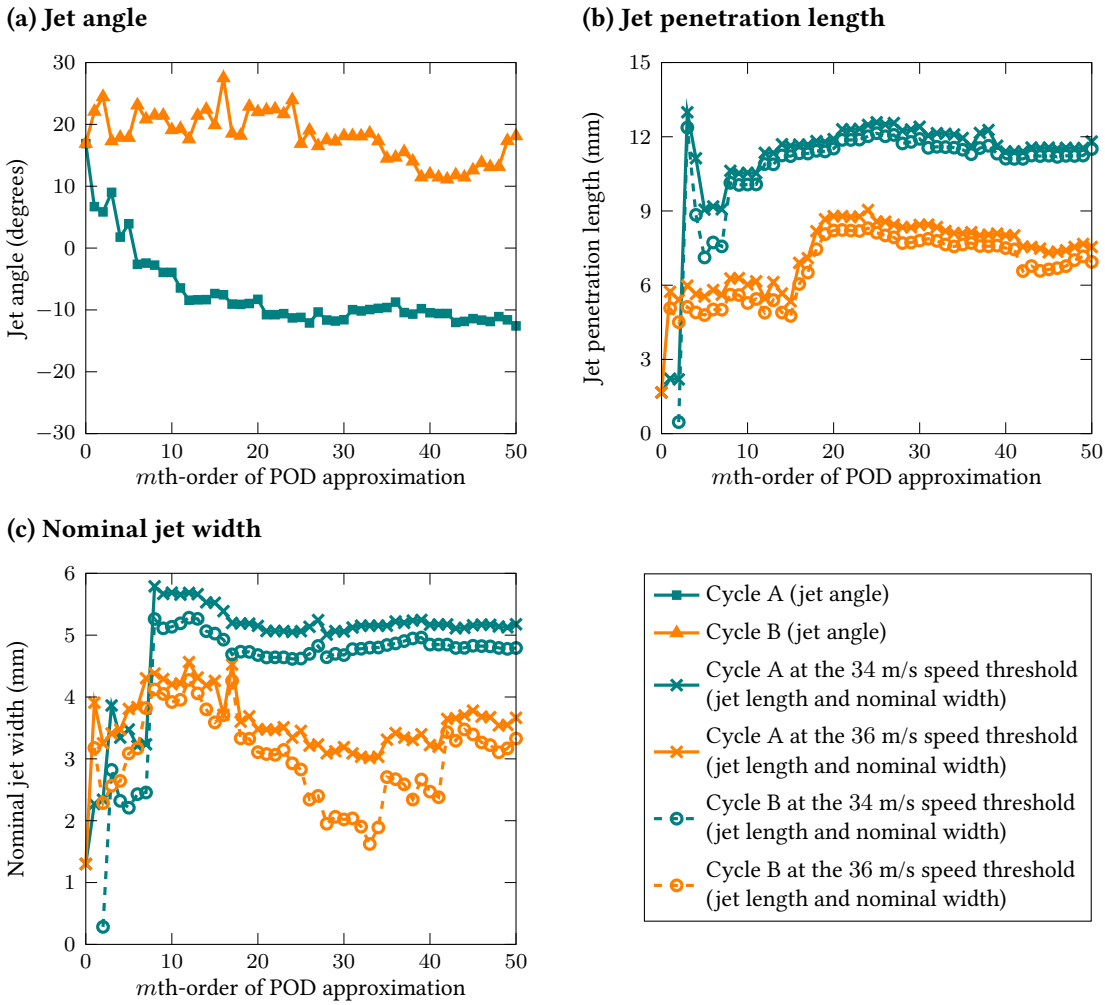


Figure 5.8: Intake jet quantification metrics for different orders of POD approximations at 300° bTDCf for Cycle A (dark green) and Cycle B (orange). Note that the jet angle is not a function of speed thresholds, while the jet penetration length and nominal width change with respect to the speed thredholds.

be negligible since most of the flow features (quantified in Figure 5.8) are already stable. As a result, 18 modes are needed to provide a coherent structure of Cycle B.

The reasons that different numbers of POD modes are needed for each cycle lie not only with cycle-to-cycle variation in the engine (i.e., in the measured PIV fields), but in the mechanism of

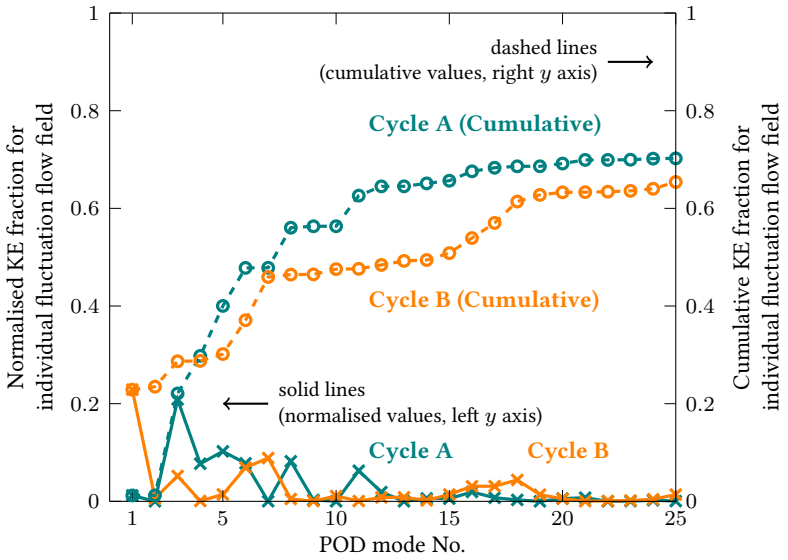


Figure 5.9: Kinetic energy fraction collected by POD modes for Cycle A (dark green lines) and Cycle B (orange lines) at 300° bTDCf. The solid lines show the normalised KE fraction distributed into every POD mode for each cycle (left y axis), and the dashed lines show the values cumulated from the first POD mode (right y axis).

the POD calculation as well. POD computes the modes by successively maximising the total amount of kinetic energy from all the 300 fluctuation flow fields, so that the kinetic energy collected by lower order modes is always no less than the higher modes (Equation (3.27)), yet the kinetic energy content of each cycle does not necessarily distribute into the modes in the successive maximisation manner. For instance, as shown in Figure 5.9, less than 2% of the kinetic energy of the fluctuation part of Cycle A is distributed into the first two modes (dark green lines), while Mode 3 alone collects more than 20%, causing the sudden flow feature change mentioned previously from the second- to the third-order of approximations (left column of Figure 5.7(c) and (d); see also the metrics as dark green lines in Figure 5.8). A similar observation can also be made to the fluctuation velocity field of Cycle B (orange lines in Figure 5.9) — over 22% of its kinetic energy is in Mode 1, and the fraction does not exceed 9% in any of the remaining single modes. Therefore, one can expect the most noticeable change in flow structure would

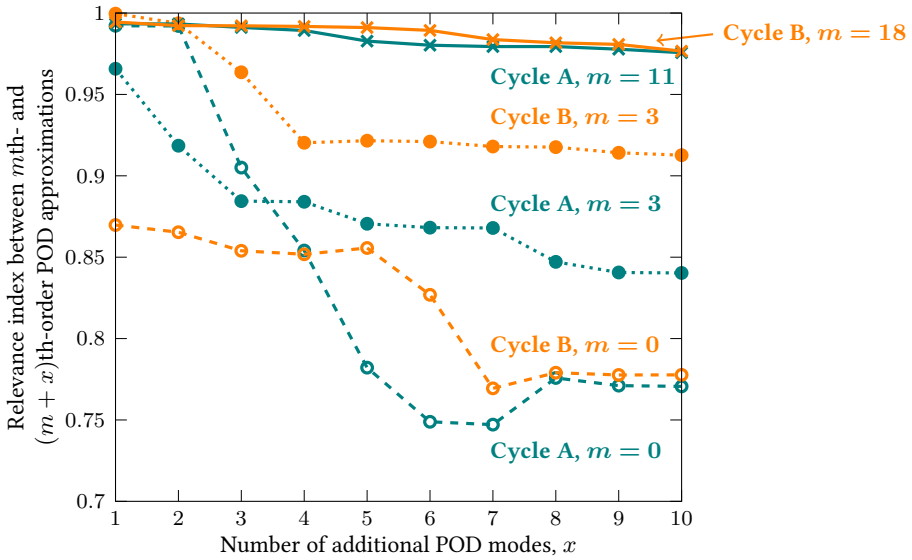


Figure 5.10: Relevance index between the mth- and $(m+x)$th-orders of POD approximations at 300° bTDCf for Cycle A (dark green lines) and Cycle B (orange lines). Each line with the same colour uses a different baseline mode number m.

occur between the zeroth- (ensemble mean) and the first-order of approximations (right column of Figure 5.7(a) and (b), see also orange lines in Figure 5.8).

Additionally, as the kinetic energy of the fluctuation part of Cycle A rapidly accumulates (63% of total, dark green dashed line in Figure 5.9) in the first 11 modes, the 11th-order POD approximation would have sufficient flow features to approximate Cycle A. The kinetic energy accumulation of the fluctuation part of Cycle B (orange dashed line), in contrast, is slower — although the first mode already has 22% of the kinetic energy and the percentage quickly increases to 46% in Mode 7, the cumulative kinetic energy fraction undergoes a very flat region with an additional seven modes only introducing another 3%, until it ramps up from 49% (Mode 14) to 60% (Mode 18). As a result, the change from different orders of approximation would be more gradual (right column of Figure 5.7), and Cycle B requires more modes in its reconstruction.

Other than the three intake jet profile quantification metrics, the "stability" of each

approximated field can also be evaluated by adding a number of extra components to a baseline approximation and calculating the directional similarity (quantified by the relevance index, *cf* Equation (3.51)) before and after. The relevance indices between the mth-order (baseline) and $(m + x)$th-order approximations are computed and plotted in Figure 5.10 for Cycle A (dark green lines) and Cycle B (orange lines). When the baseline is set to be the ensemble mean ($m = 0$), adding one ($x = 1$) or two ($x = 2$) extra mode(s) does not change the flow structure for Cycle A (dark green dashed line with hollow circles, both RI > 0.99), while including the third component ($x = 3$) causes the RI to drop below 0.91, and the value continues to decrease as more components are added (increasing x). The values match with the previous findings by observing the approximated fields (left column of Figure 5.7) and the corresponding intake jet quantification metrics (Figure 5.8). Moving the baseline approximation to higher modes, for instance $m = 3$, one can further observe the effect of adding more modes to the third-order reconstructed flow of Cycle A (dark green dotted line with filled circles). The change in flow features is smaller (higher RI), while the RI value still drops to 0.85 after including eight more terms ($x = 8$, i.e., the comparison between the third-order and the 11th-order of POD approximations), showing that three modes are still not sufficient for the flow feature to stabilise. The calculation continues for higher orders of baseline approximations, until at the 11th-order ($m = 11$, dark green solid line with crosses), the RI values are higher than 0.97 regardless of the number of extra components ($x = 1, 2, \cdots, 10$), and thus the approximation of Cycle A stabilises at Mode 11.

A similar procedure can also be implemented to Cycle B (orange dashed line with hollow circles) — the RI is below 0.87 after introducing the first term to the ensemble mean ($m = 0, x = 1$) and reduces even further for higher orders ($x > 1$), which agrees with the appearance of the intake jet in the first-order POD approximation (right column of Figure 5.7) and the gradual change of the intake jet profile (orange lines in Figure 5.8) after the first mode. The flow feature is more but not sufficiently stable at $m = 3$ — the minimum RI for all x is higher than 0.91, yet

a sudden drop of RI can be observed from $x = 2$ (fifth-order approximation) to $x = 4$ (seventh-order). At $m = 18$, all the RI values finally stay above 0.97, showing that the 18th-order POD reconstructed data for Cycle B is stable.

Zhuang and Hung [67] took a similar approach when constructing the coherent structure in their quadruple POD analysis, except that they chose to set $x = 1$ for their 100-cycle data set. For a large data set with 300 snapshots (299 computed POD modes) that include highly fluctuating intake jets, a larger x value ought to be used in order to avoid the flow fields being apparently stable at lower order modes, since the flow features are more distributed in different modes and a single extra component may not be enough to change the flow structure. For instance, Cycle B would be falsely considered to be stable at its third-order approximation ($m = 3$, orange dotted line in Figure 5.10) if only one extra mode ($x = 1$) is considered, since the RI is larger than 0.99. The actual decrease of RI that occurs after adding the Mode 7 component ($m = 3$, $x = 4$), on the other hand, would not be observed.

5.4.2 Stability test of low-order POD approximations

As exemplified by the previous paragraphs, determining the number of modes needed for a stable low-order reconstructed flow field of a specific cycle requires comprehensive evaluation for each order of approximation, which can be time-consuming to implement on a 300-cycle data set. Therefore, a stability test is proposed in order to automate this procedure. An mth-order approximation is assumed to reach a stable flow structure if, comparing to the corresponding $(m + x)$th-order approximations ($x = 1, 2, \cdots, 10$), any four out of the five following criteria based on intake jet characteristics (Criteria 1–3), kinetic energy content (Criterion 4) and overall vector-wise flow field direction (Criterion 5) are met:

1. the jet angle fluctuates within ± 5 degrees;

2. the change of the jet penetration length is smaller than 1.68 mm (twice the PIV vector spacing) at the 35 m/s speed threshold;

3. the change of the nominal jet width is smaller than 0.84 mm (i.e., PIV vector spacing) at the 35 m/s speed threshold;

4. the difference in kinetic energy is smaller than 10% of the total kinetic energy collected by all POD components;

5. the relevance indices are higher than 0.95 with any of the choices of x.

Among all the order of approximations for the same cycle which passed the above stability test, the one with the lowest order (i.e., uses least number of components) is chosen to be the approximated flow field of that cycle. The threshold of the nominal jet width is smaller than the one of the jet penetration length to compensate for their differences in absolute value ranges. The tests were conducted for a maximum of ten extra modes (denote x), assuming that the flow feature is unlikely to suffer a sudden change if it is already stable for ten extra components. It should be noted that the result from this stability test depends inevitably on choosing appropriate thresholds subject to the data set, yet this is considered to be reasonable since a similar subjective decision has to made when manually examining the flow fields.

For all the 300 cycles, the number of lower-order POD modes needed for each cycle is shown in Figure 5.11. Over half of the cycles require $10 - 20$ POD components (*cf* Equation (3.13)) to reach a stable flow feature (blue filled histogram), and over 80% of them need fewer than 20 components (blue solid line with crosses for the cumulative number of cycles), if four out of the five previously mentioned criteria are met. Take Cycle A (left column of Figure 5.7) as an example, the respective minimum number of modes needed to meet each criterion is: 11 for jet angle (Figure 5.8(a)), 4 for jet penetration length (Figure 5.8(b)), 8 for nominal jet width (Figure 5.8(c)), 7 for kinetic energy fraction (Figure 5.9) and 11 for the relevance index (Figure 5.10). Therefore, the

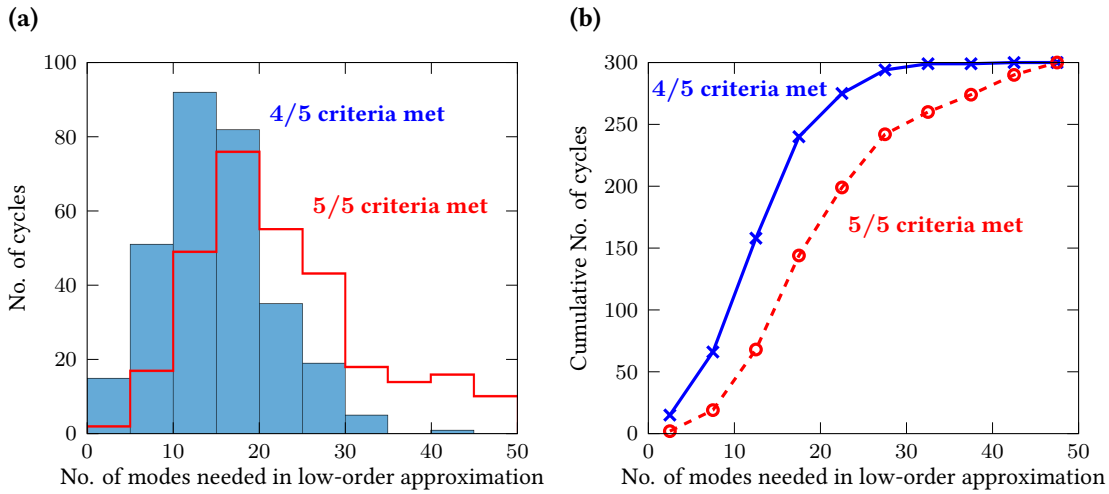


Figure 5.11: Number of POD modes needed in the low-order approximation for each of the 300 cycles at 300° bTDCf identified by different numbers of stability test criteria. Blue filled histogram in (a) and blue solid line in (b): four out of the five stability test criteria are met; Red unfilled histogram in (a) and red dashed line in (b): all the five stability test criteria are met. The lines in (b) show the cumulative values in (a).

stability test suggests that the flow feature of Cycle A stabilises at the 11th-order approximation.

The required modes increase when more constraints are put onto the stability test. For instance, if the test needs to meet all the five criteria instead of four, the majority of cycles would need about $15 - 25$ modes (red unfilled histogram in Figure 5.11), and the 80% cumulative line (red dashed line with hollow circles) shifts to $25 - 30$ modes. However, this may cause the stability test to over-constrain the results since the threshold for each criterion was used globally for all cycles but not all of them are always suitable for every cycle. Hence the previously proposed stability test only required meeting four out of the five criteria.

One way to examine whether the choices of thresholds are appropriate would be to see which criterion is most likely to be the "extra" one — the one which was not accounted for in the stability test. Among all the 300 cycles, 77 of them have the jet angle as the unsatisfied criterion; the jet penetration length, nominal jet width and kinetic energy fraction each have 33, 37 and 32 cycles respectively; the relevance index criterion was not used in 82 cycles; the remaining 39 cycles

have "tied" criteria, meaning that multiple criteria give the same order of approximation and the extra one will not have any effect even if it was considered. The results shows that the jet angle and relevance index thresholds may be too restrictive, resulting in the higher mode number when including the fifth criterion. Efforts may be made in the future to find a method that can choose the thresholds properly, such that most of the cycles have tied criteria and give a consistent number of modes, or make the five criteria equally restrictive (i.e., being the extra criterion for the same number of cycles). Additionally, the first three criteria (jet angle, penetration length and nominal width) were selected based on the specific flow features present on the measured plane during the intake stroke of this work. These criteria may not be applicable to other flow situations measured at a different crank angle or on another PIV plane. On the other hand, the other two criteria (kinetic energy difference and relevance index) are independent of the flow features, and thus can be applied to any flow field with appropriate selection of threshold values. It is recommended that the stability criteria should include both the flow features of interest and flow feature-independent metrics in other applications.

5.4.3 RANS versus low-order POD reconstructed PIV flow fields

The approximated flow field using low-order POD reconstruction extracts the key features and filters the small-scale turbulence and noise from the original PIV measurements, providing an alternative validation target for the RANS simulated data which were modelled based on averaged boundary conditions from the PIV experiments. Their similarity in terms of the intake jet profile can be again quantified by the metrics developed in an earlier section (Section 5.2), for which results are shown in Figure 5.12. Based on the stability test discussed in the previous section, Cycle A uses its 11th-order approximation, while Cycle B is reconstructed by 18 POD components. For the readers' convenience, the flow fields of the RANS simulation (Figure 5.1(f) and Figure 5.3(a)), the PIV ensemble mean (Figure 5.1(e) and Figure 5.3(b)), and

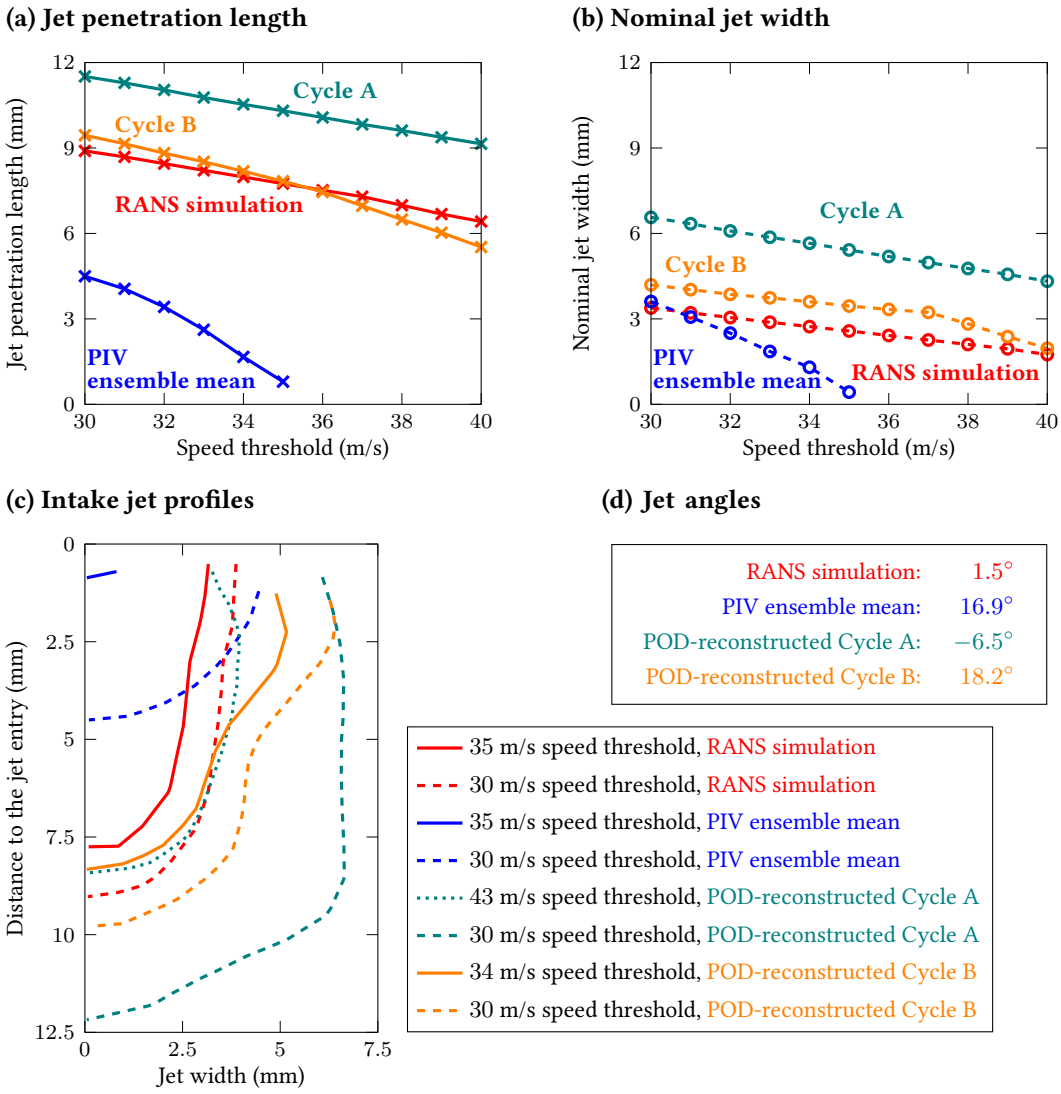


Figure 5.12: Intake jet quantification metrics for the low-order POD approximated flow fields compared to the RANS simulation and PIV ensemble mean at 300° bTDCf. (a) Jet penetration lengths and (b) nominal jet widths for the RANS simulation (red), the PIV ensemble mean (blue), the POD-reconstructed Cycle A (dark green) and the POD-reconstructed Cycle B (orange) as functions of speed thresholds. (c) Intake jet profiles (width as a function of length) at different speed thresholds, where the data for the POD-reconstructed Cycle A are plotted at 43 m/s (dark green dotted line) and 30 m/s (dashed line), while the data for the POD-reconstructed Cycle B are plotted at 34 m/s (orange solid line) and 30 m/s (dashed line). (d) Jet angles for different types of flow fields.

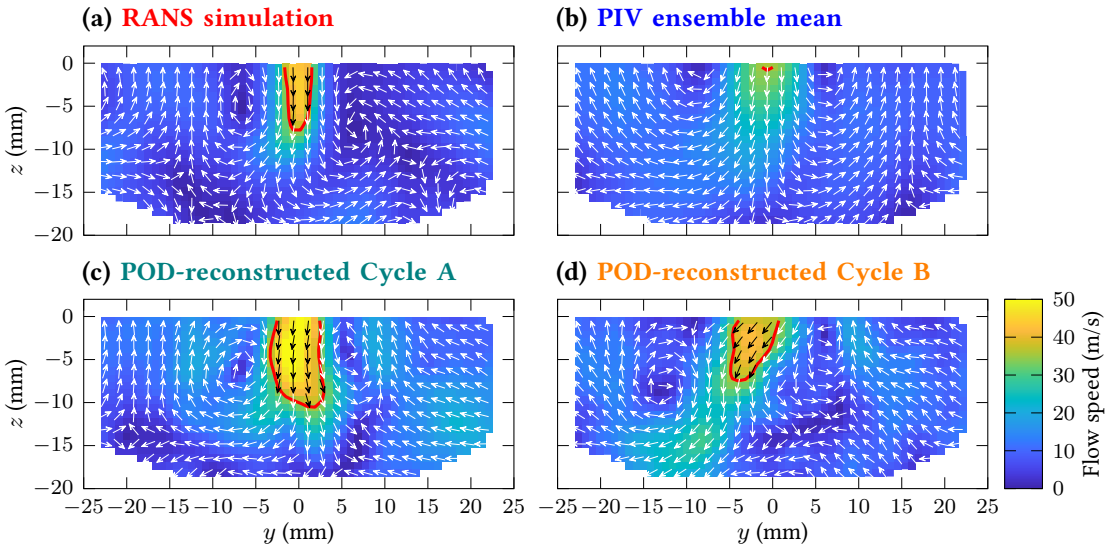


Figure 5.13: RANS simulated flow field versus the PIV ensemble mean and low-order POD reconstructed flow fields at 300° bTDCf. The red contour lines illustrate the jet region at the 35 m/s speed threshold for each flow field. These flow fields have been plotted in earlier figures: (a) in Figure 5.1(f) and Figure 5.3(a); (b) in Figure 5.1(e), Figure 5.3(b) and Figure 5.7(a); (c) in Figure 5.7(f), left column; (d) in Figure 5.7(g), right column.

the low-order POD reconstructed Cycle A (left column of Figure 5.7(f)) and Cycle B (right column of Figure 5.7(g)) are plotted again in Figure 5.13.

The jet profiles of the low-order approximations of both cycles and the RANS model are comparable in many aspects. For Cycle A's approximation, despite its jet penetration length (dark green solid line with crosses in Figure 5.12(a)) at each speed threshold being larger than the corresponding value of the RANS model (red solid line with crosses), the slopes of the two lines are close, meaning that they have similar velocity gradients along the jet penetration direction. Similar observations can also be made for the comparison in the nominal jet widths as a function of the speed threshold (dark green and red dashed lines with hollow circles in Figure 5.12(b)), showing that the perpendicular velocity gradients of the approximated Cycle A and the RANS simulation are close to each other as well. As a result, the central intake jet of the Cycle A's approximation (dark green solid line in Figure 5.12(c); flow field shown in Figure 5.13(c)) has

similar curvature with the one of the RANS simulation (red solid line in Figure 5.12(c); flow field shown in Figure 5.13(a)) — their width almost keeps constant in the main jet body, and the width rapidly shrinks to zero at the tip of the jet. However, one can also observe that the jet in the Cycle A's approximation is much wider and longer at the same speed threshold (30 m/s).

Given the comparable velocity gradients and jet curvature, the relative characteristics of the RANS and low-order POD approximated PIV intake jets are well matched. By subsequently increasing the speed threshold of Cycle A's approximation to find the best match versus the RANS simulated data, the scalar jet speed difference between the measurement and the simulation may be quantified. The best match was found at the 43 m/s speed threshold (dark green dashed line in Figure 5.12(c)), at which the absolute values of the jet length and width of the Cycle A's approximated flow field are in line with the RANS simulated data, in addition to the existing consistency in the intake jet curvature. Therefore, the speed difference inside the jet region is about 13 m/s. Also, the jet angle calculation (mentioned previously in "Quantification of Intake Jet Profile") shows that the jet angle difference between the approximated Cycle A and the RANS flow field is around 8° (labelled in Figure 5.12(a) and (b)).

The low-order POD reconstructed flow field of Cycle B (Figure 5.13(d)) has almost the same jet penetration length (orange solid line with crosses in Figure 5.12(a)) and slightly larger nominal jet width (orange dashed line with hollow circles in Figure 5.12(b)) compared with the RANS simulated data (red lines in Figure 5.12(a) and (b)). The differences in the slopes of the corresponding lines are also small, showing that the velocity gradients are similar. Consequently, the central jet of the Cycle B's approximation (orange solid line in Figure 5.12(c)) has a profile that is close to the one of the RANS model (red solid line), and they have comparable length and width. It should be noted that, however, Cycle B's approximation has a much wider jet at the entry compared to the RANS simulation. The jet speed difference between the approximated Cycle B and the RANS simulation is about 4 m/s, which can again be estimated by increasing the speed threshold of Cycle B to 34 m/s (orange dashed line in

Figure 5.12(c)) and comparing to the RANS simulation (at 30 m/s). Despite these similarities, the jet angle (labelled in Figure 5.12(a)) of the reconstructed Cycle B deviates about 17° from that of the RANS simulation, which can also be observed by comparing their flow fields (Figure 5.13(d) versus (a)).

Viewing the entire 300-cycle dataset (Table 5.2 and Figure 5.14), the low-order POD reconstruction method provides a much fairer comparison to the RANS simulated data than using the other two averaging methods. The RANS model shows acceptable consistency with the coherent flow structures of the experimental data obtained by the low-order POD approximation, despite that, it cannot provide information of the in-cylinder flow field cyclic variation. The jet regions for the RANS simulation and the low-order POD reconstructed flow fields can be defined at comparable jet speeds (detectable jet speeds in Table 5.2), albeit the jet speed is slightly lower in the RANS model. The jet length for the RANS model is close to the 300-cycle median of the low-order POD approximations (values reported in Table 5.2, see also a direct comparison in Figure 5.14(b)). While the RANS simulation predicts a smaller nominal width than the 300-cycle median, its value is still within the cyclic variation range (bounded by the 10% and 90% quantiles) of the experimental data (Figure 5.14(c)). The central intake jet profile of the RANS simulation and its velocity gradients are also in line with the corresponding properties for both example cycles (Figure 5.12). The mismatches with individual cycles, such as jet speed with Cycle A and jet angle with Cycle B, can easily be caused by the strong cycle-to-cycle variation in the flapping intake jet. For instance, the differences between the 10% and 90% quantiles of the jet angles for the low-order POD reconstructed flow fields are about 38° in Table 5.2, and the jet angles for all the 300 reconstructed flow fields span a range of 112° in Figure 5.14(a), as a result of the uneven flow collision from the two intake valves.

The RANS simulated flow field (Figure 5.13(a)) also captures several smaller-scale flow structures within the approximated PIV fields (Figure 5.13(c) for Cycle A and Figure 5.13(d) for Cycle B). For instance, the two shear vortices induced by the intake jet are centred respectively

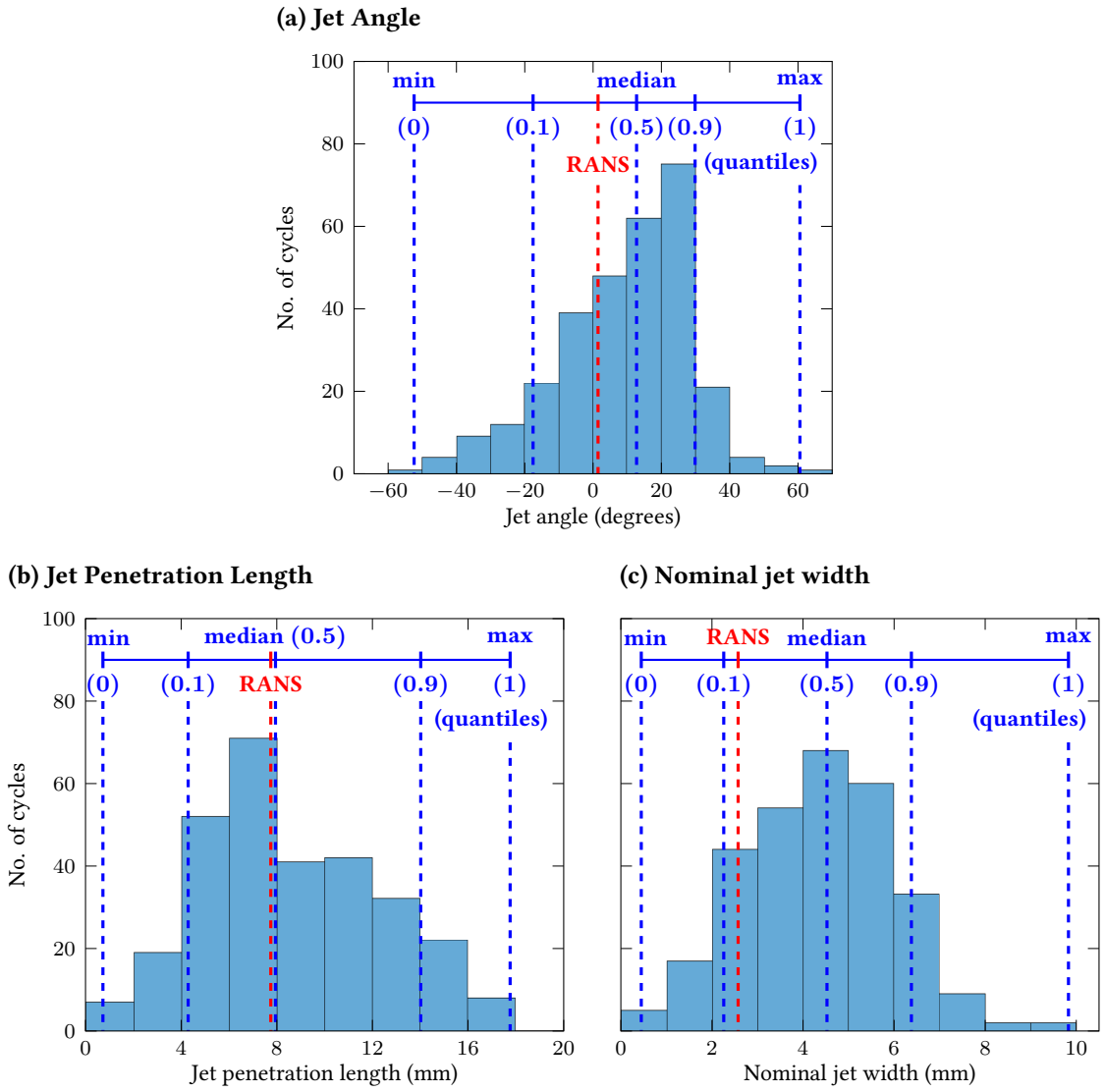


Figure 5.14: Intake jet quantification metrics for the low-order POD reconstructed cycles (blue, 300 cycles in total) versus the RANS results (red) at 300° bTDCf. The jet penetration length and nominal jet width were evaluated at the 35 m/s speed threshold. The numbers in parentheses refer to the quantiles of values in each histogram. These statistical values are also reported in Table 5.2.

Table 5.2: RANS model validation results using different methods for all the 300 cycles at 300° bTDCf

Comparison metrics		RANS simulation	PIV averaging / filtering methods		
			Ensemble mean	Speed-Based average	Low-Order POD reconstruction
Maximum detectable speed threshold (m/s)[a]		42	35	42	44
Jet angle (degrees)		1.45	16.9	7.85	12.7 ± 19.5, $(-17.6, +29.9)$[b]
Jet penetration length (mm)	at 30 m/s	8.89	4.50	7.13	10.5 ± 3.86, $(5.63, 15.7)$[b]
	at 35 m/s	7.75	0.80	4.74	7.95 ± 3.69, $(4.29, 14.0)$[b]
Nominal jet width (mm)	at 30 m/s	3.38	3.61	6.57	5.70 ± 1.92, $(3.41, 7.55)$[b]
	at 35 m/s	2.58	0.43	4.62	4.53 ± 1.62, $(2.26, 6.38)$[b]

[a] The jet region can be defined using the speed thresholds equal to or smaller than the specified values, and therefore the jet quantification metrics can be calculated. A smaller value means the maximum flow speed within the flow field is smaller.
[b] Showing the median $\pm$ one standard deviation, and the (10%, 90%) quantiles of the results for 300 cycles, respectively.

at ($y = -6.56, z = -5.52$ mm) and ($y = +6.94, z = -7.21$ mm) in the simulated flow field, which only deviates from the shear vortex centres in the Cycle A's approximation at ($y = -7.40, z = -6.36$ mm) and ($y = +6.09, z = -6.36$ mm) by one vector spacing (0.84 mm) in both y and z directions. The RANS-predicted flow directions in the lower-speed regions are also within the range of cyclic variations for the POD reconstructed fields. At the top-left corner of the PIV field of view, the RANS predicts a flow direction that is in line with the Cycle A's approximation (inclined towards the positive y direction), while slightly differing from the Cycle B's approximation (inclined towards the negative y direction). The flow directions for the RANS field and the approximations of both cycles are aligned near the top-right corner of the PIV field of view.

If the RANS model was validated only against the ensemble mean, it would be falsely considered to be a poor model that significantly over-estimated the flow speed (much higher detectable jet speeds in Table 5.2). Other biased conclusions would also be drawn — for instance, the model had a much longer intake jet (in Table 5.2 for two example speed thresholds, see other speed thresholds in Figure 5.12(a)), and even appeared to generate an incorrect jet profile (Figure 5.12(c)). Indeed, it was the ensemble mean (Figure 5.13(b)) itself that failed to be a representative flow field of the 300 cycles — the jet length and nominal width for both speed thresholds are out of the corresponding ranges of the low-order POD reconstructions (Table 5.2), and the jet profile is significantly different from the original cycles (Figure 5.1(a)–(d)) or their approximations (Figure 5.13(c) and (d)). The speed-based averaging method generates closer values of the jet lengths and nominal widths to the RANS model and the low-order POD reconstructions (Table 5.2), but at the expense of an unavoidable background speed increase discussed in an earlier section (Section 5.3.2).

Additionally, the low-order POD reconstructions retain the variations of the jet quantification metrics in different measured cycles, and provide ranges of values that the simulation results should aim for. For instance, the 10% and 90% quantiles in Table 5.2 can be

used as the validation lower and upper limits of the jet length and width. It should be noted that using the experimental median values as validation targets can sometimes be too restrictive — the RANS model only predicts a reference cycle without considering cyclic variations and hence the model can be considered accurate if the metrics fall within the corresponding experimental cyclic variation ranges (as illustrated in Figure 5.14(b) and (c)). On the other hand, the jet angles obtained from experiments vary significantly (Table 5.2 and Figure 5.14(a)) due to the flapping intake jet, making the resulting validation range overly large and easy for the RANS model to achieve. Therefore, caution is needed when interpreting the jet angle predicted by RANS models even if the value falls within the validation range (i.e., experimental cyclic variation range).

5.5 Chapter summary

In this chapter, the modelled flow fields were validated against the experimental measurements under highly-fluctuating flow conditions such as the engine intake process. Metrics were developed to quantify the profile of the central intake jet generated by the collision of the two intake air streams (from both valves), which allowed a detailed comparison between different flow fields containing the central intake jet in the cross-tumble plane. The RANS model simulates the flow fields by using the averaged boundary conditions from the PIV experiment, and it only generates one flow field at any fixed crank angle. Several methods were investigated to average the multi-cycle measured PIV flow fields to enable the validation. The conventional phase-averaged ensemble mean was first examined, yet its central intake jet was heavily affected by the horizontal velocity component cancellation — the jet speed was found to be largely under-represented compared to individual PIV measurements and the jet profile was greatly smoothed, such that it lacked key flow features and was not a representative flow field of the ensemble during the intake process. The speed-based averaging method resolved

the issue of speed under-representation, but at the cost of introducing an unavoidable flow speed increase outside of the jet region, and the smoothing problem still remained.

These findings led to the necessity of using non-averaged individual PIV measured fields as RANS validation targets. These highly fluctuating flow fields must be filtered to allow a fair comparison with the non-fluctuating RANS simulation. Proper orthogonal decomposition (POD) was therefore implemented on the PIV data set, and the flow fields were reconstructed by using several low-order POD modes to extract the coherent flow structures and discard the smaller-scale turbulence and measurement noise. Efforts were made to find an appropriate cut-off mode number for the POD reconstruction in order to provide a robust approximation to the original measured flow fields, and a novel stability test was proposed to systemise the procedure. The low-order POD reconstructed flow fields retained the major flow features of the central intake jet, and enabled a fair validation of the RANS simulated data. The results showed that the current RANS model provides a sufficiently accurate simulation in terms of the maximum penetration length, the nominal width, and the velocity gradients of the highly fluctuating central intake jet, though the jet speed is slightly under-estimated.

The methods developed here can also be implemented on other data sets when the ensemble mean fails to be a good representative of any highly fluctuated flow fields. Future focuses may involve finding best-matched PIV cycles versus the RANS simulation to examine whether the similarity in the intake jet profile during the early intake stroke (300° bTDCf) leads to similarity in later events such as mixture preparation after direct gasoline injection or flame propagation during the combustion, as well as modifying the RANS model based on the validation results.

Chapter 6

Comparison of POD and KPCA methods in flow model validation

In this chapter, two different manifold reduction techniques, the proper orthogonal decomposition (POD) technique (also known as the linear principal component analysis, linear PCA in statistics) and its non-linear extension, the kernel principal component analysis (KPCA) method, are implemented for analysing in-cylinder vector fields measured by the particle image velocimetry (PIV) technique and for validating the flow simulation model. The contents of this chapter refer to a follow-up research from a paper [3] co-first-authored by the author of this thesis. Some parts of the chapters (Section 6.1–6.4) are adapted from the paper with additional information and more detailed discussions included, while a different KPCA approach was used in Section 6.5.[(1)]

[(1)] Approach A (discussed in Section 3.2.2) is used in this thesis in order to provide a direct comparison between the KPCA and the POD method, while Approach B was applied in the paper. The same data set is used in this thesis and in the paper.

6.1 Motivation

In the previous chapter (Chapter 5), the proper orthogonal decomposition (POD) technique was applied to the PIV flow fields, in order to avoid the flow speed underestimation due to the component cancellation embedded in the computation of the PIV phase-averaged ensemble mean, and to account for the cycle-to-cycle variation in the measured flow data. The low-order POD reconstructed PIV flow fields, after filtering out the smaller-scale turbulent fluctuations and noise in the experiment, enable a fairer validation of the simulated flow data and provide the ranges of each parameter for the flow model to validate against (Section 5.4.3). However, the results of the low-order POD reconstruction inevitably depend on the choice of the cut-off mode number (i.e., the numbers of POD modes to be included in the low-order reconstruction). Despite the efforts many researchers (Section 1.5) and the author's stability criteria in Section 5.4.2, criticism still exists on the lack of objectivity when determining the cut-off mode number [60]. Therefore, it is necessary to explore alternative PIV data analysis techniques that allow a cleaner separation between the coherent flow motion (composed by the lower-order modes) and smaller-scale turbulent fluctuations (included in the higher-order modes), in order to provide a less subjective validation target to flow simulation models.

As discussed in Section 1.5, the difficulty in finding the appropriate cut-off POD mode number mostly results from the non-linearity in the input data. The kernel principle component analysis (KPCA) technique, as a generalisation of POD,[(2)] provides an opportunity to map the data into a chosen non-linear space, and may lead to a less ambiguous cut-off choice. In the following sections, both KPCA and POD techniques are respectively implemented on a set of tumble plane data at the T2 test load,[(3)] with the goals of comparing their performances on the coherent structures extraction for PIV flow fields and the validation of flow simulation models.

[(2)] POD is known as the linear PCA in statistics.

[(3)] Data were taken in the tumble plane at the T2, Piston B, C33, DVA test point. Data for the same test point but in different measurement planes were also used in Chapters 4 and 5.

6.2 RANS versus PIV ensemble mean and individual cycles

Before implementing manifold reduction techniques such as POD and KPCA to the PIV measured flow fields, it is worth investigating the conventional option of directly comparing the RANS simulation with the PIV ensemble mean. The RANS simulation (Figure 6.1(a)), the PIV ensemble mean (Figure 6.1(b)) and the individual PIV flow fields (two example cycles[(4)] in Figure 6.1(c) and (d)) are obtained at 280° bTDCf during the intake stroke for the T2 test point. The PIV field of view is blocked by the open intake valve at the top-right corner of Figure 6.1(b)–(d). Therefore, the comparison between the simulated and measured flow fields are taken only at the locations where the PIV vectors are available. Qualitatively a good match of the intake flow structure can be seen between the simulation and the PIV ensemble mean, while the single cycle flow fields show greater degrees of smaller-scale fluctuations. The flow speed in the PIV ensemble mean, however, is noticeably smaller compared to the RANS simulation and the presented individual cycles. Also, the intake jet penetration length in the ensemble mean field is much shorter. This is a result from the velocity component cancellation among different cycles when computing the ensemble mean, which was discussed previously in Section 5.3.1.[(5)]

The differences between flow fields can be quantified by the vector field comparison metrics which were introduced in Section 3.3. Figure 6.2 shows the relevance index (*cf* Equation (3.51)) versus the RANS simulation for the PIV ensemble mean (blue line) and all the 300 individual cycles (red crosses). Despite its lower flow speed, the ensemble mean has a higher RI compared to any of the individual cycles, suggesting a better match between the PIV ensemble mean and

[(4)] The two example cycles were randomly selected and recorded consecutively in a test run. These two cycles are not the same cycles as presented in the published paper [3], in order to ease the concern of "cherry-picking" and show the methods that are being presented in this chapter works for any cycle.

[(5)] The values of velocity components at a single grid point vary significantly from cycle to cycle. The ensemble mean, due to its natural averaging process (*cf* Equation (3.5)), can be skewed by the low values in some cycles, resulting in a much smaller flow speed in the averaged flow field. An extreme case of this phenomenon was presented by the horizontal velocity component cancellation in Section 5.3.1.

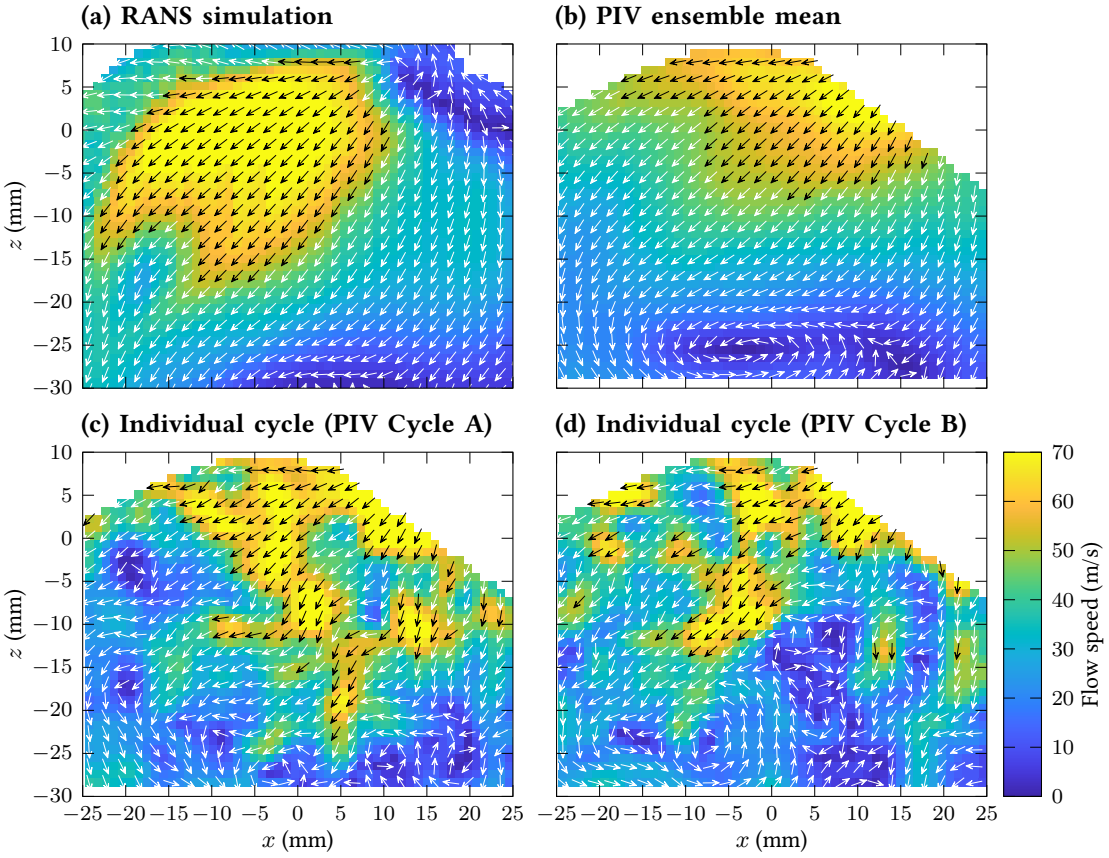


Figure 6.1: Flow fields in the tumble plane at 280° bTDCf (during the intake stroke when intake valves are open). Data were taken at the T2, Piston B, C33, DVA test point.[6]

the RANS simulation. It should be noted that the differences between the non-filtered individual cycles and the RANS simulated flow field are exaggerated due to the existence of the local small-scale fluctuations in the measured flow fields, while in the ensemble mean such fluctuations are cancelled out by the averaging process.

In addition to a single scalar value from the relevance index, one can obtain the spatially-

[6] It is worth reiterating that, as mentioned in a footnote for Figure 2.7, the vectors are in unit length and only illustrate the flow direction, while the magnitude of the vector at each location are shown by the colour map. For clarity only every second vector is plotted in each direction, while the false colour background uses the actual spatial resolution from the PIV experiment. The colour (black/white) of each vector is chosen to contrast with the false colour background. This plot style is adopted in this thesis for all the plots containing vectors, unless otherwise specified.

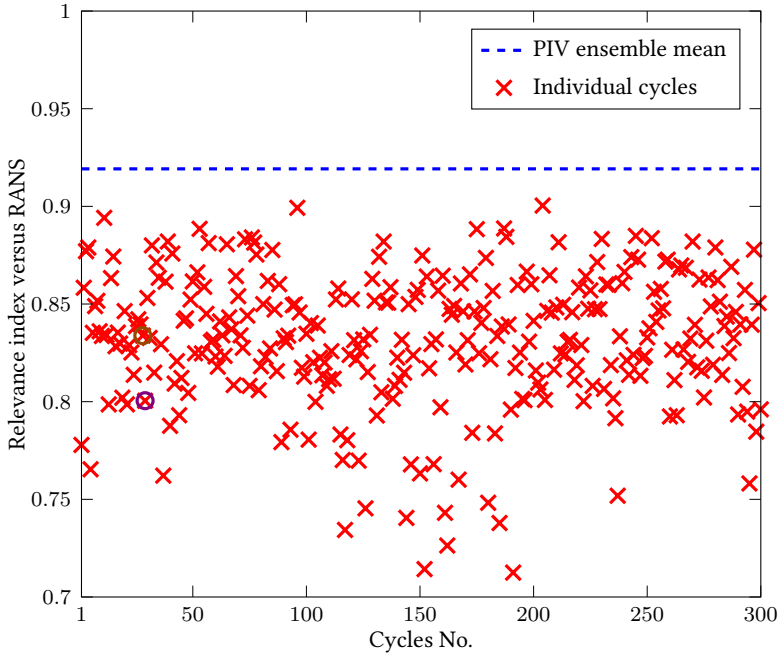


Figure 6.2: Relevance index (RI) versus the RANS data for the PIV ensemble mean (blue dashed line) and all the 300 individual cycles (red crosses) at 280° bTDCf. The values for the two example cycles whose flow fields were shown in Figure 6.1 are additionally marked by the brown circle (Cycle A, Figure 6.1(c)) and the purple circle (Cycle B, Figure 6.1(d)), respectively.

resolved quality-of-match information in the flow direction and in the flow speed (i.e., velocity magnitude) independently by the weighted relevance index (*cf* Equation (3.53)) and the weighted magnitude index (*cf* Equation (3.54)). Figure 6.3 shows the WRI and WMI fields when comparing the PIV ensemble mean and the two example cycles respectively to the RANS simulation.[7] The simulated flow field and the ensemble mean have an excellent directional match (Figure 6.3(a)). Cycle A also shows a similar good directional match with the RANS data (Figure 6.3(c)), while flow direction differences are found near the tumble vortex centre for Cycle B (Figure 6.3(e)) — a large region of high values are observed at the bottom part of the WRI field, likely caused by the fact that the RANS simulation is not able to capture small eddies in the individual cycles.

[7] Note that a lower value of WRI and WMI indicates a better match between the two vector fields, while a higher RI value means a better match.

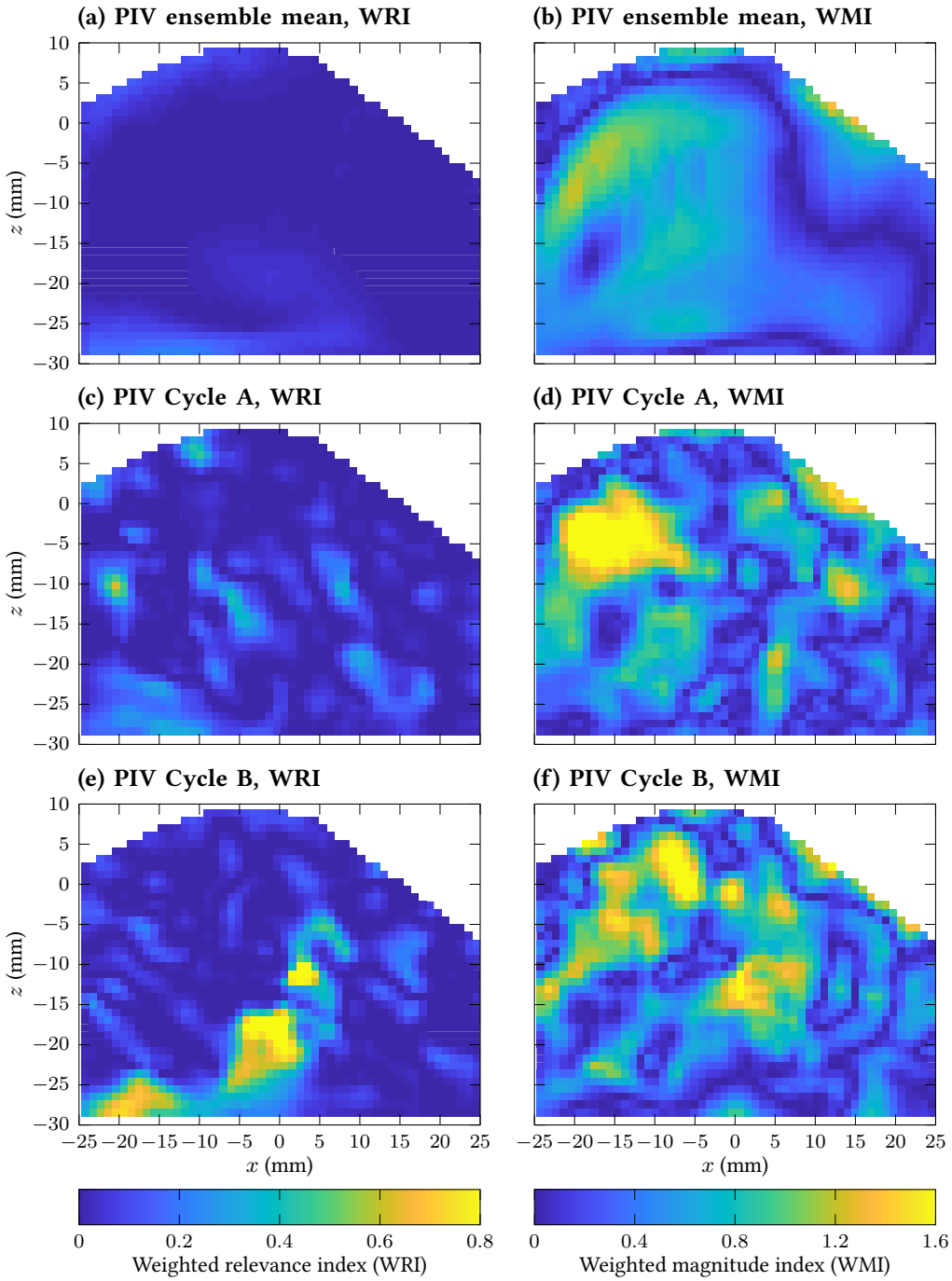


Figure 6.3: Weighted relevance index (WRI, left column) and weighted magnitude index (WMI, right column) fields versus the RANS data for the PIV ensemble mean (top row) and two example cycles (middle row for Cycle A, bottom row for Cycle B) at 280° bTDCf.

The WMI field for the PIV ensemble mean (Figure 6.3(b)), on the other hand, has larger values across the whole field compared to its WRI field (Figure 6.3(a)). This suggests that compared to the differences in the flow direction, the flow speeds in the ensemble mean field deviate to a greater extent from the simulation data — the flow speed differences are particularly larger near the edge of the intake jet (at around $x = -20, z = -10$ mm), at which the flow directions vary significantly from cycle to cycle and hence the flow speed underestimation due to the component cancellation in the ensemble mean becomes more substantial. Local high WMI values in the flow speeds can also be observed in individual cycles (Figure 6.3(d) and (f)), and the locations at which high values occur vary in different cycles.

The spatial information from the WRI and WMI fields can be condensed back to single values to provide a more intuitive comparison by taking the spatial average of each field. Figure 6.4 shows the spatially-averaged WRI and WMI for each individual PIV cycle (red crosses) and the PIV ensemble mean (the intersection of two blue dashed lines, also marked by the blue square) compared with the RANS simulation. In line with the previous finding using the relevance index (Figure 6.2), the PIV ensemble mean has the a better directional match with the RANS simulation compared to any of the individual cycles (lowest WRI in Figure 6.4). However, a number of individual cycles are found to have better agreements with the simulated flow field in terms of the flow speed (lower WMI) compared to the ensemble mean.

Despite there exists some individual cycles that can have a reasonably good similarity with the RANS simulation (almost as good as the ensemble mean in terms of the flow direction and may even better in terms of the flow speed, see Figure 6.4), it is worth noting that the measured flow fields include small-scale turbulent features and experimental noise, while the RANS only provides an estimation of the bulk flow behaviour based on averaged thermo-fluid equations and does not contain these fluctuations. Additionally, the RANS data may contain artifacts of modelling assumptions. Therefore, it is questionable to compare the individual cycles directly to the RANS simulated flow field in the first place. These findings bring the necessity of using some

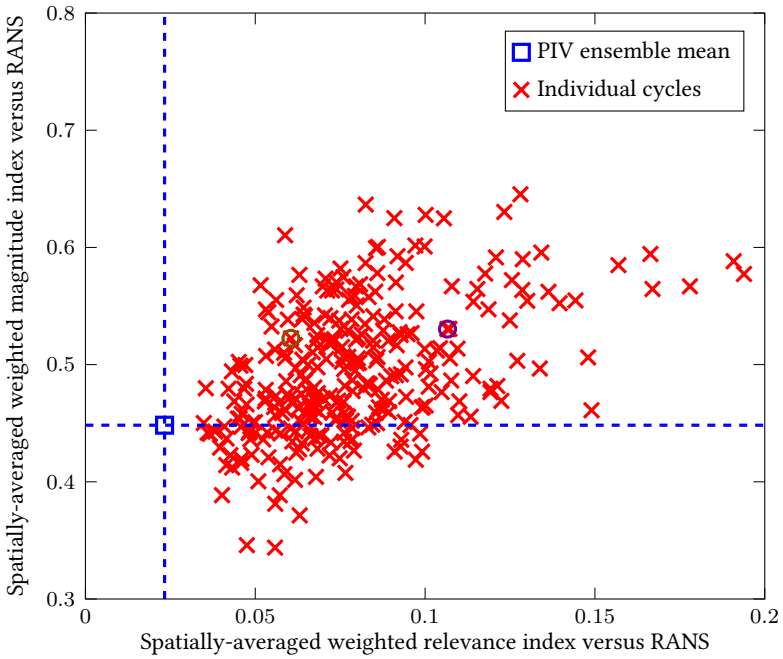


Figure 6.4: Spatially-Averaged weighted relevance index (WRI) and weighted magnitude index (WMI) versus the RANS data for the PIV ensemble mean (the intersection of the blue dashed lines, also marked by the blue square) and all the 300 individual cycles (red crosses) at 280° bTDCf. The values for the two example cycles are additionally marked by the brown circle (Cycle A) and the purple circle (Cycle B), respectively.

data-processing techniques to filter the instantaneous flow fields, in order to determine whether the flow structures and the flow speeds are reliably captured by the RANS simulation.

6.3 POD (linear PCA) analysis

Proper orthogonal decomposition (POD) has been proposed as an approach to make objective quantitative comparisons between the experimental and simulated data that goes beyond the ensemble averaging. POD can provide insights into in-cylinder flow dynamics and can be used to quantify the cycle-to-cycle variation (CCV) of the in-cylinder flow (Section 1.5). In the previous chapter (Chapter 5), it has been demonstrated that the low-order POD

reconstructed flow fields can be used for the validation of a RANS simulation in the cross-tumble plane even when a flapping intake jet exists in the field of view. In this section, the POD technique is implemented again on the data in the tumble plane, but with different flow comparison metrics and an alternative cut-off mode determination algorithm. Since the flow fields in the tumble plane is more organised than in the cross-tumble plane, it is now appropriate to implement the point-wise vector field comparison metrics (WRI and WMI, Section 3.3) to quantify the flow structure similarities and changes for a pair of flow fields in the entire tumble plane field of view, in replacement of the customised flow comparison metrics (Section 5.2) that focuses on the flapping intake jet in the middle of the cross-tumble plane field of view.[8]

Given the mathematical process after centring the data (Equation (3.10)), the proper orthogonal decomposition (*cf* Equation (3.13)) applied to the 300-cycles data set generates 299 POD modes with a decreasing amount of kinetic energy captured by each successive mode (*cf* Equation (3.27)). Different criteria are available in literature to identify the cut-off mode number for different flow motions (Section 1.5). In order to decide the number of modes needed to reconstruct the flow fields, the quadruple POD method [66, 67] is applied here[9] — the turbulent flow field is decomposed into four separate classes: mean, coherent, transition and turbulent flow motion. The coherent flow motion contains large-scale vortices and plays a critical role in the cyclic variations. The transition part is identified as a passage in the energy cascade between the large-scale vortices in the coherent turbulence and the smallest ones in the incoherent background turbulence. The turbulent part represents the small-scale structures

[8] It is worth reiterating that, as mentioned in Section 5.2, the point-wise comparisons cannot accommodate more in-depth or subjective concepts of similarity such as a coherent structure being displaced or distorted between two fields in the cross-tumble plane data set. This is why the intake jet profile was quantified by key characteristics of the intake jet such as jet angle, penetration length and width in Section 5.2.

[9] The mode stability criteria proposed by the author and introduced in Section 5.4.2 can also be implemented here with necessary modifications. The quadruple POD method is chosen because it is a more widely adopted algorithm and suits the need of comparing a general POD analysis procedure with the KPCA algorithm in Section 6.5.

in the flow field, which are assumed to be homogeneous and isotropic.

In the quadruple POD approach, Qin *et al.* [66] and Zhuang and Hung [67] proposed the use of the relevance index (RI, *cf* Equation (3.51)) among the reconstructed flow fields as the tool to determine the proper cut-off mode — the number of POD modes required in the coherent part is determined by the RI between the fluctuation flow structure reconstructed by the first m modes and the flow structure using one additional mode (i.e., using the first $(m+1)$ modes).[(10)] If the addition of a POD mode to the flow reconstruction does not modify the resulting field significantly, in other words, if the RI before and after adding a POD mode is close to 1, the contribution of that POD mode is negligible, and the previous mode is identified as the cut-off mode for the coherent part of the turbulence. It should be noted that, for a given RI threshold, the cut-off mode number for each cycle may be different due to the strong CCV in the flow fields and the nature of successive maximisation of the sum of the sample variances during the POD computation process.[(11)] Instead of using different cut-off mode numbers for each cycle (as used in Chapter 5), Qin *et al.* [66] chose to take the average RI of all cycles, and set a cycle-averaged $\text{RI} = 0.95$ threshold as the cut-off between the coherent and the transition flow motions.

Figure 6.5 shows the averaged RI over 300 cycles (blue dots) with an increasing number of modes at 280° bTDCf. A cycle-averaged RI over 0.95 is achieved when comparing the flow fields that include the first eight POD modes and the first nine POD modes in the calculation, and hence the cut-off mode number is eight. It should be noted that, however, a wide range of RI values is observed for each chosen mode over 300 cycles, as shown by the 10% and 90% quantiles (labelled by the edges of the error bars). Indeed, there are over 30% cycles whose RI value has not reach 0.95 using the first eight modes, and for those cycles, their coherent structures may

[(10)] Note that here the "fluctuation flow structure" is reconstructed by the first several modes and excludes the ensemble mean, as opposed to the "low-order POD apporximation" which includes the ensemble mean (*cf* Equation (3.28)).

[(11)] POD successively maximises the sum of the sample variances in all cycles, not the variance for each cycle (*cf* Equation (3.26)). Therefore, the extent of the flow features collected up to a certain numbers of POD modes may vary from cycle to cycle.

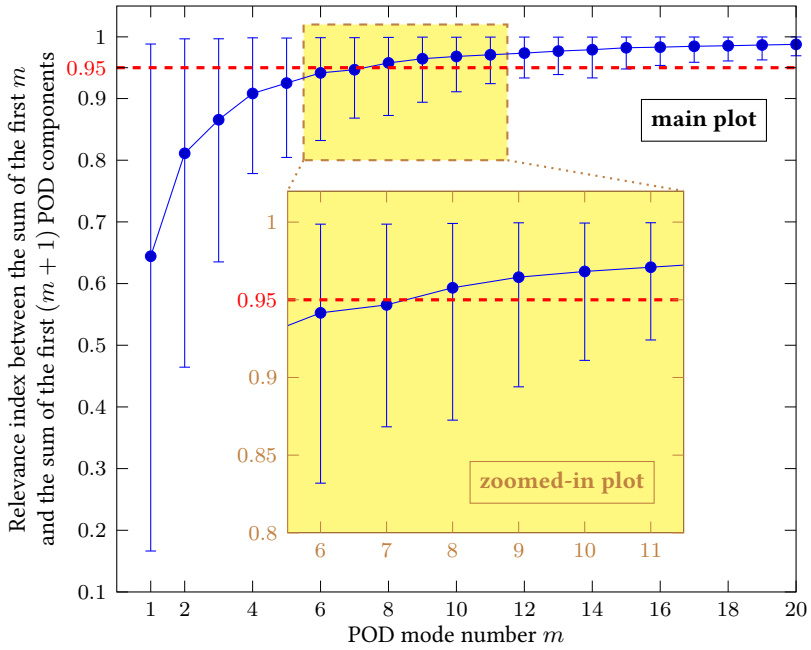


Figure 6.5: Cut-Off POD mode number identification for the PIV flow fields measured at 280° bT-DCf using the quadruple POD method. The relevance index (RI) values between the sum of the first m and the sum of the first $(m+1)$ POD components are plotted for each POD mode number m, and a cutoff threshold of cycle-averaged RI $= 0.95$ (red dashed line) is chosen according to [66, 67]. Note that the RI values vary in different cycles. Blue dots: the 300-cycle average; error bars: the 10% and 90% quantiles of 300 values at a specific mode number. The zoomed-in region is bounded by the brown dashed lines filled with light yellow colour in the main plot.

not yet be converged. The disadvantage caused by this phenomenon will be discussed later in this section.

Figure 6.6 shows the POD-reconstructed flow fields (ensemble mean plus the first m POD components, *cf* Equation (3.28)) for the two example cycles (left column: Cycle A and right column: Cycle B); the same numbers of POD modes are included in the flow fields on each row. The first-order POD-reconstructed flow fields (Figure 6.6(b)) qualitatively shows a high degree of similarity with the ensemble-averaged experimental flow field (Figure 6.6(a)). The cycle-to-cycle variation of the flow field can be seen with the addition of increasing numbers of POD modes (Figure 6.6(c)–(e)), until when all 299 modes are included in the reconstruction, the resulting flow

field becomes identical to the original single cycle flow field (Figure 6.6(f)). It is worth noting that, despite that the quadruple POD identifies eight modes are sufficient for a "stable" coherent structure, extra POD modes included in the reconstruction process still contribute to a change in the flow field. For instance, small flow structure changes can be found between the eighth (Figure 6.6(d)) and the 13th (Figure 6.6(e)) orders of approximations for both cycles.

A quantitative comparison between the POD-reconstructed flow field using eight modes and the RANS simulated flow field is performed using the WRI and WMI metrics. For both cycles (Figure 6.7(a) and (c)), when compared to the RANS simulation, the eighth-order POD approximations have a lower WRI across the whole field than the corresponding original cycles (Figure 6.3(c) and (e)), indicating that the POD-approximated flow fields are closer to the RANS flow field in terms of the flow direction. On the other hand, since POD successively maximises the kinetic energy within the data set (Equation (3.27))), the POD approximations tend to retain the flow speed from the original cycles. As a result, their WMI fields versus the RANS data (Figure 6.7(b) and (d)) remain at a similar level to the original cycles (Figure 6.3(d) and (f)), albeit that the WMI fields for the POD approximations are smeared out more uniformly since the small-scale turbulence within the original cycles are filtered out by the reconstruction process. Such comparisons can be made for all the 300 cycles, providing not only the main flow field information that can be obtained from the PIV ensemble mean, but also the cyclic variation information in the coherent flow field structures that goes beyond the scope of the PIV ensemble mean.

Although the reconstructed flow field using eight modes has shown both qualitatively and quantitatively a better validation target for the RANS flow field, one would argue that the results may be affected by the choice of the cut-off threshold. The POD reconstructed flow fields in Figure 6.6 shows that the inclusion of any additional modes in the POD analysis may produce a distinct change in the reconstructed flow fields, and may further affect the validation results. To examine the influence of extra modes on the validation results, the WRI and WMI

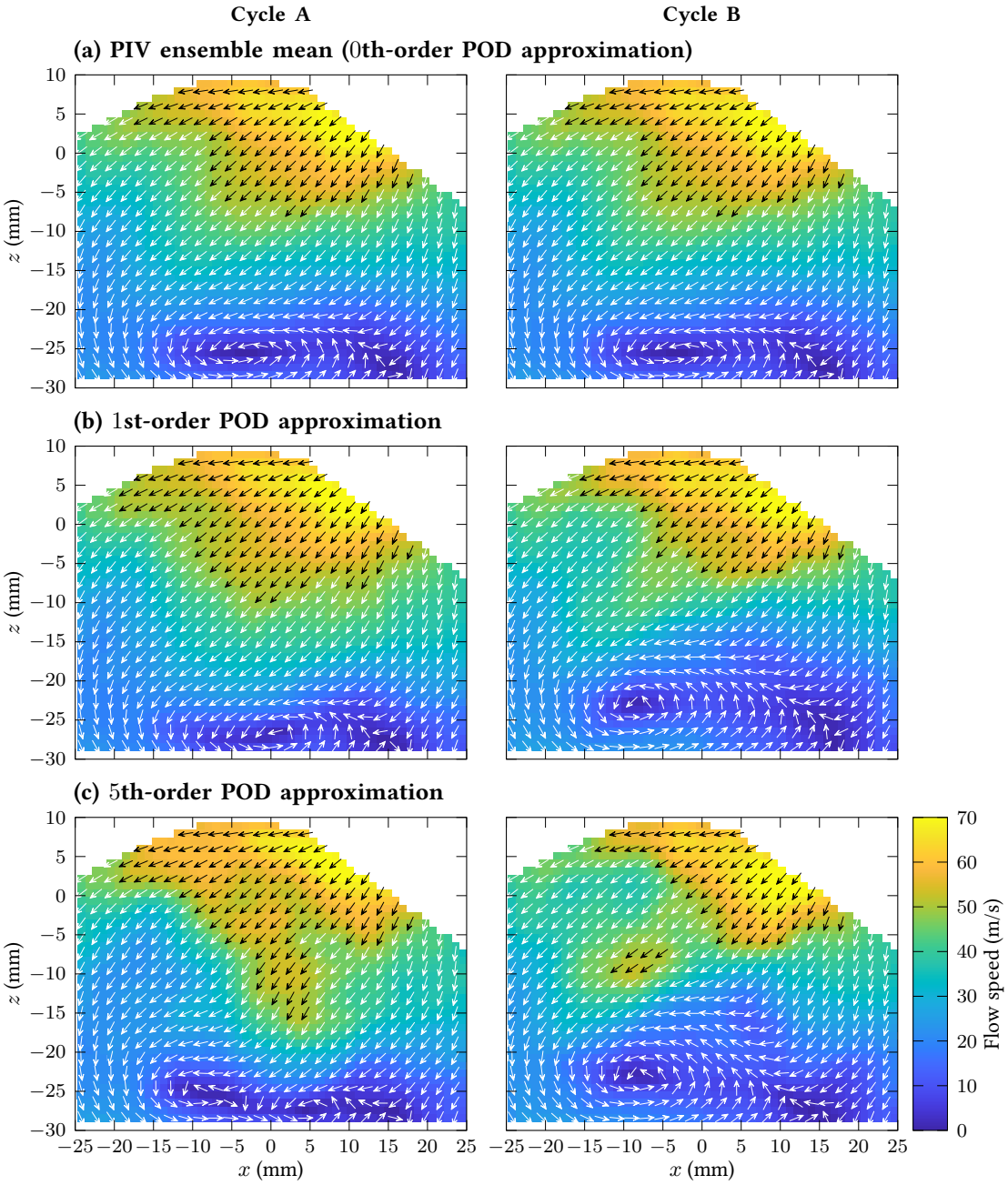


Figure 6.6: POD-Based reconstruction for two example cycles (left column for Cycle A, right column for Cycle B) at 280° bTDCf. Different numbers of modes are included in the reconstruction for each row.

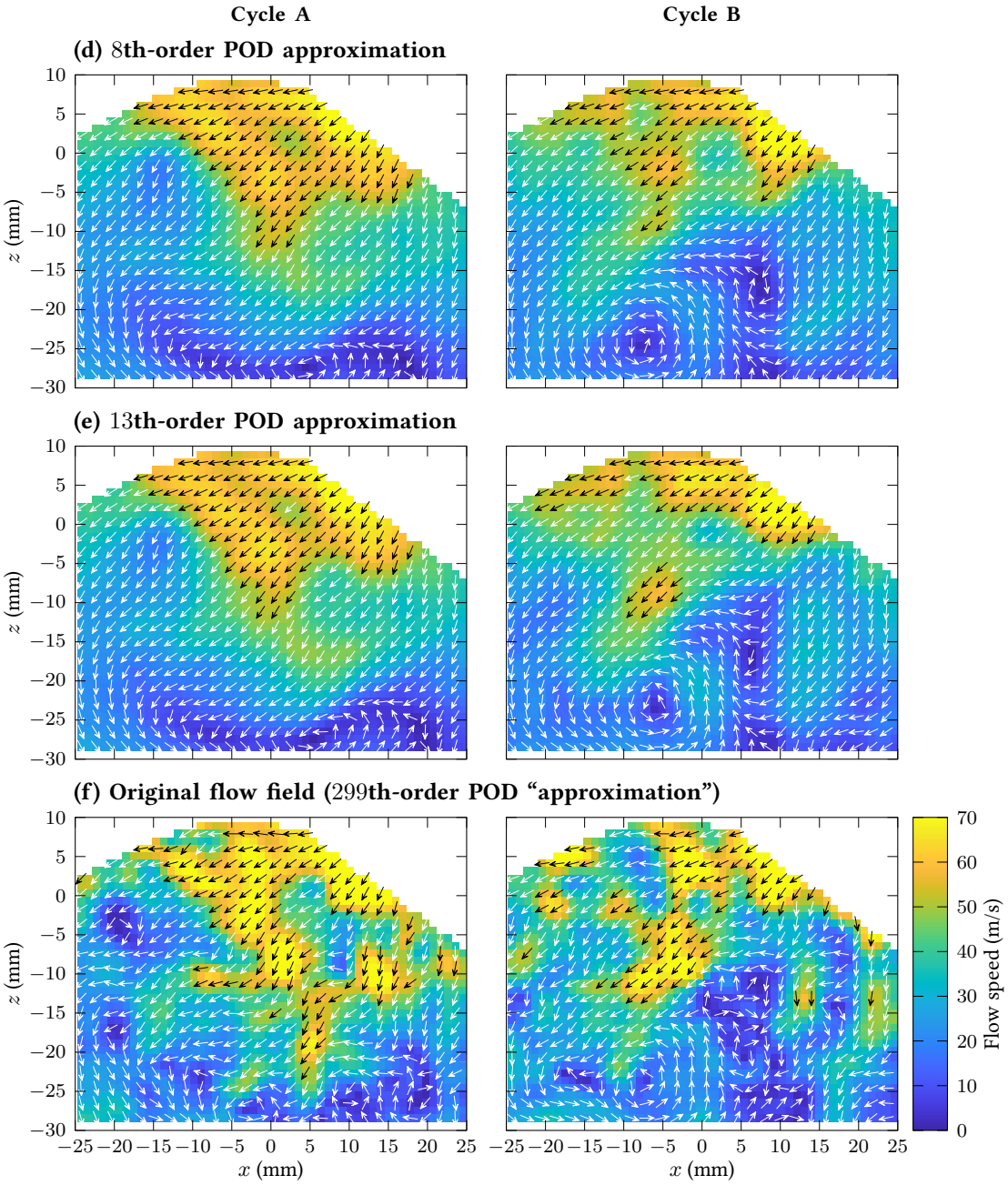


Figure 6.6: POD-Based reconstruction for two example cycles (left column for Cycle A, right column for Cycle B) at 280° bTDCf. Different numbers of modes are included in the reconstruction for each row.

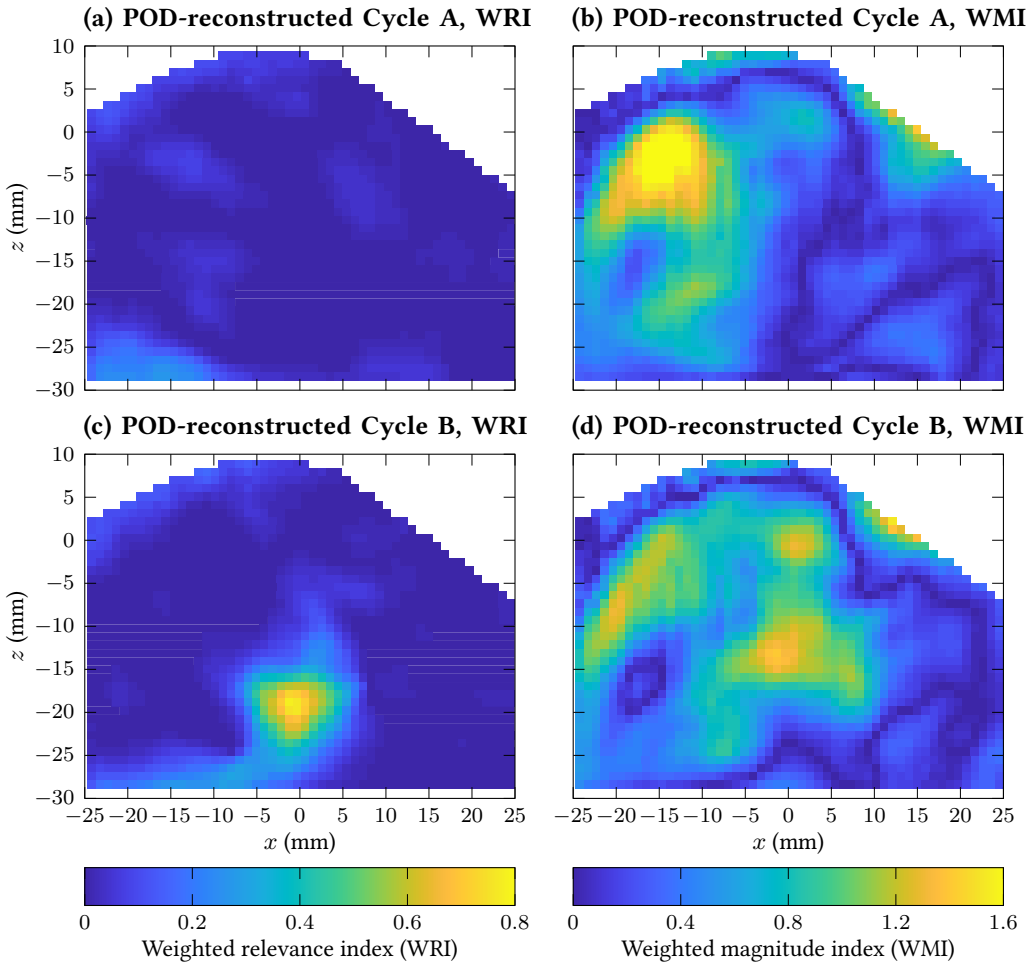


Figure 6.7: Weighted relevance index (WRI, left column) and weighted magnitude index (WMI, right column) fields versus the RANS data for two POD-reconstructed cycles (top row for Cycle A, bottom row for Cycle B) at 280° bTDCf.

fields in Figure 6.7 are again spatially averaged to provide a single value (respectively for WRI and WMI) that quantifies the overall quality of match in terms of flow direction and speed between POD-reconstructed flow fields and the RANS simulated flow field. Figure 6.8 illustrates the spatially-averaged WRI (top plot) and WMI (bottom plot) for POD-reconstructed Cycle A (brown crosses) and Cycle B (purple circles) using different numbers of dimensions

(i.e., POD modes). Generally speaking, after including at least eight POD modes,[12] the values for both metrics and for both cycles show an increasing trend, resulting from the fact that more random fluctuations are included in the POD approximations at higher orders. On the other hand, the increasing trend after the eighth-order approximations is less obvious in the spatially-averaged WMI values — it is only observed in Cycle A (brown crosses in the bottom plot), while the values for Cycle B (purple circles) fluctuates around the value for the eighth-order approximation.

When using the two metrics to quantify the quality of match between the POD approximations and the RANS data, one would prefer to have constant (or at least close to constant) values that do not depend on the choice of numbers of dimensions included in the POD reconstruction, so that the validation results after the comparison can be more robust and objective. However, the spatially-averaged WRI values may change by a significant amount as more dimensions are included in the POD reconstruction. For instance, the spatially-averaged WRI values for Cycle B (purple circles in the top plot of Figure 6.8) increases by 26% from 0.065 at the eighth-order approximation to 0.082 at the 13th-order approximation (after adding five more modes). If Cycle B were used to compare with the CFD data, one might draw different conclusions on the quality of match in flow directions when using different numbers of POD modes to reconstruct this cycle. Indeed, this ambiguity in determining the actual levels of match versus RANS occurs not only in a single cycle (in this example, Cycle B), but in many other cycles within the data set as well.

Figure 6.9 illustrates the changes in the spatially-averaged WRI (top plot) and WMI (bottom plot) values when adding more modes into the POD reconstruction compared to the values computed using the corresponding eighth-order POD approximations for all the 300 cycles in the data set. The changes are shown in % from the baseline (i.e., the value for the corresponding

[12] The cut-off mode number suggested by the quadruple POD method was eight. A vertical dashed blue line is plotted at the eighth POD mode in Figure 6.8 to guide the readers.

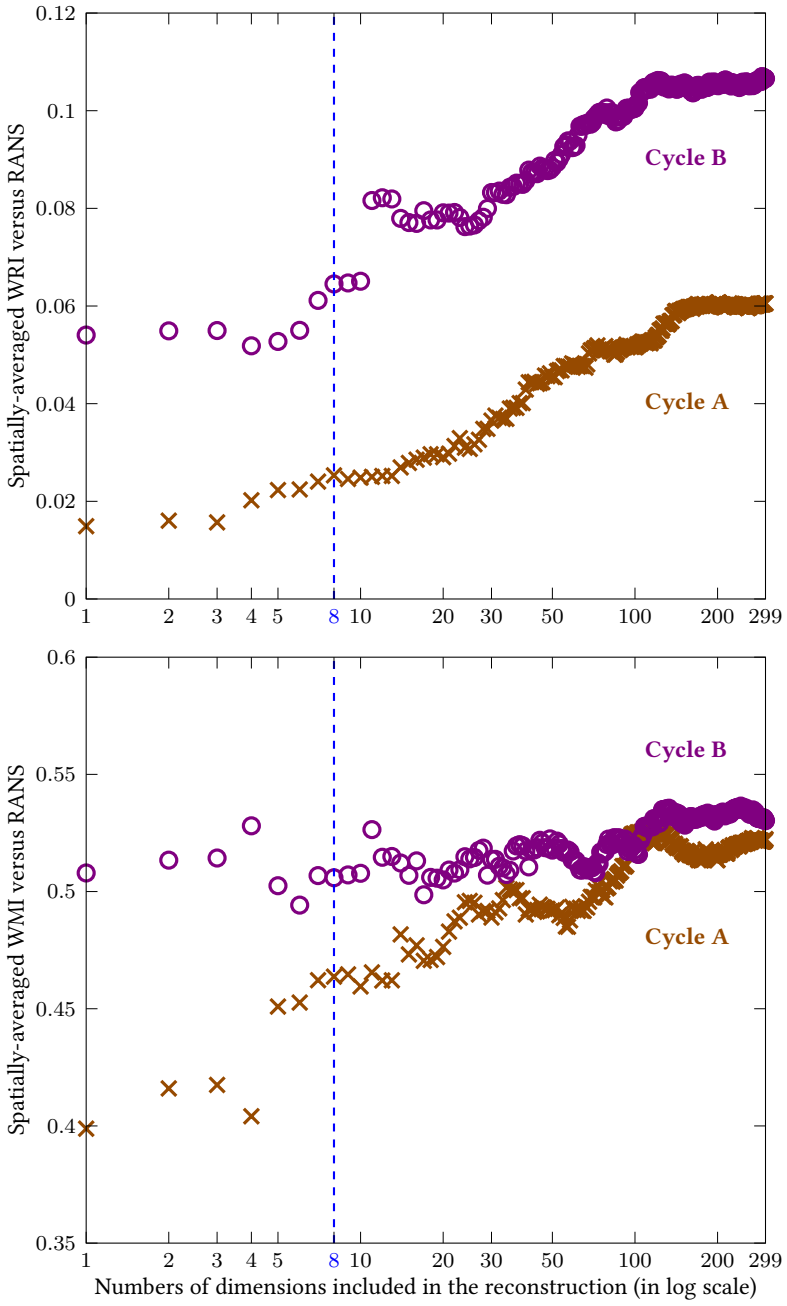


Figure 6.8: Spatially-Averaged WRI (top plot) and WMI (bottom plot) versus the RANS data for the POD-reconstructed flow fields of Cycle A (brown crosses) and Cycle B (purple circles) with varying numbers of dimensions at 280° bTDCf. Eight POD modes (blue dashed line) was chosen as the cut-off mode number based on the result shown in Figure 6.5.

eighth-order POD approximations), whose values are computed by:

$$\%\text{change} = \frac{\text{metric using } (x+8) \text{ dimensions} - \text{metric using } 8 \text{ dimensions}}{\text{metric using } 8 \text{ dimensions}} \times 100\%. \quad (6.1)$$

For instance, the above-mentioned 26% increase in the spatially-averaged WRI metric for Cycle B after including five extra dimensions are illustrated by a purple circle at the coordinate $(5, 26\%)$ in the top plot. The dark green dots refer to the median of the values for 300 cycles and the error bars are bounded by the 20% (lower) and 80% (upper) quantiles for the 300 values at each number of extra modes. Both metrics show an increase trend as the inclusive of more dimensions, while the increases in the spatially-averaged WRI (top plot) are particularly noticeable — the 300-cycle median value rises by 13% with five extra modes, and for over 20% cycles, their values at the 43th-order approximations (i.e., 35 extra modes) are doubled (or even more than doubled) compared to the values for the eighth-order approximations. Due to the unstable values of WRI and WMI, it can be concluded that choosing a proper number of modes is critical for the POD reconstructions when they are used to validate the RANS data.

Similar results in the literature also indicate the same challenge in the choice of the cut-off mode [74, 105], especially when there exist multiple criteria to determine the number of modes required for the approximation. Additionally, as mentioned earlier in this section, although the quadruple POD method suggests using a certain cut-off mode number identified by the cycle-averaged RI threshold, the actual numbers of POD modes needed to reach the same RI value for each cycle are different, which creates another layer of uncertainty when using any selection criteria. The variations of the cut-off POD mode number among different cycles call for the implementation of an alternative data dimensionality technique that can lead to a less ambiguous cut-off between the coherent flow structure that needs to be included while validating the RANS data, and the small-scale turbulence that needs to be excluded so that not to bias the validation. The analysis in the following section focuses on the inherent structure hidden in the data that can inspire the use of a more suitable flow field filtering method.

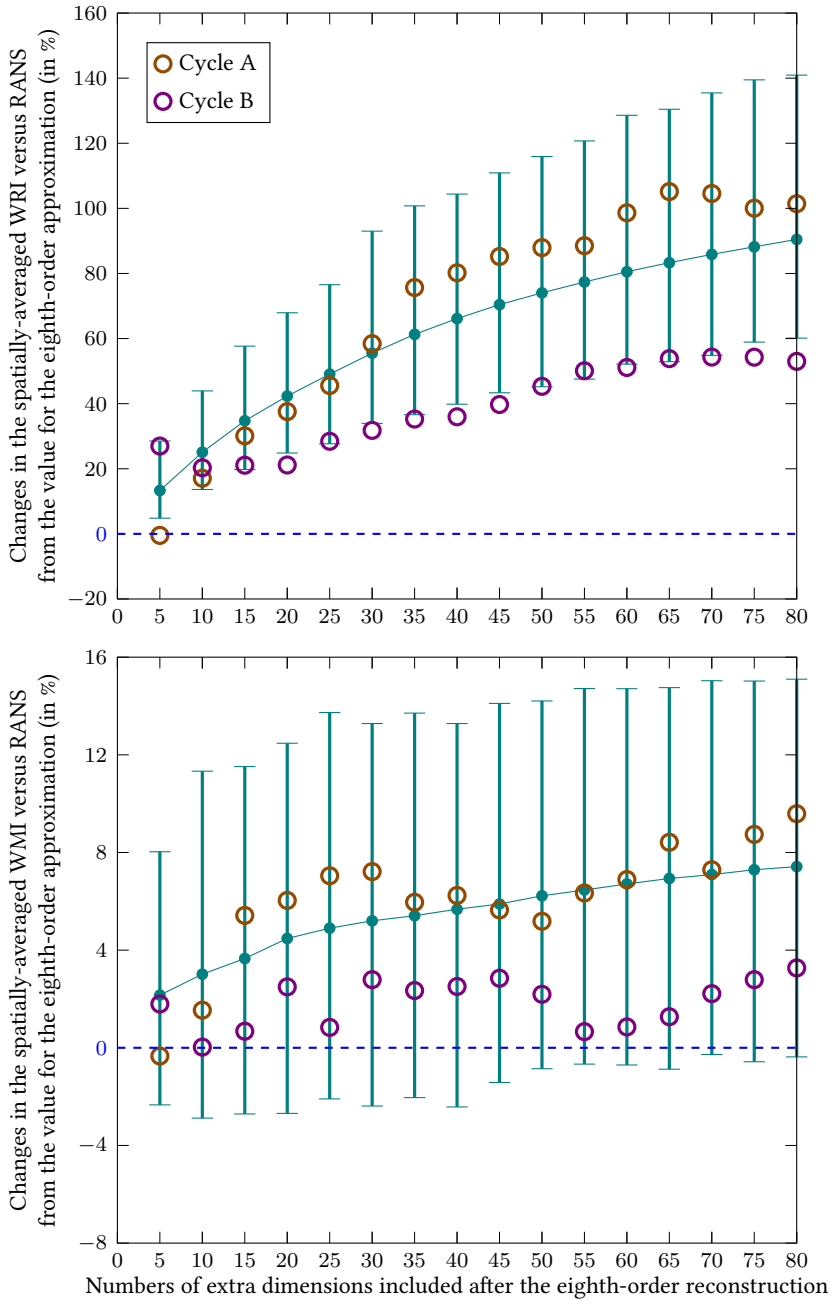


Figure 6.9: Changes in the spatially-averaged WRI (top plot) and WMI (bottom plot) values when extra dimensions are included after the eighth-order POD reconstruction. Dark green dots: the 300-cycle median; error bars: the 20% and 80% quantiles of 300 values at a specific number of extra dimensions. The changes are shown in % from the baseline at the eighth-order POD approximation.

6.4 Statistical analysis at individual grid points

Recall that the POD technique considers the velocity components at each grid point to be a random variable (Section 1.5). In order to investigate alternative flow field decomposition methods other than the POD technique, it is useful to begin from examining the statistical distribution of these random variables. In statistics, the chi-squared (χ^2) test for normality [106] is used to determine whether or not a random variable follows a normal distribution. For this study, the null hypothesis is that the vectors on a specific location in the flow field (300 samples in total) are sampled from a normal distribution with a mean and a variance estimated from the 300 samples. The significance level (α), which is the probability of falsely rejecting the null hypothesis when it is actually true (known as the Type I error, or a "falsely positive" conclusion in statistics), is chosen as 5% when implementing the hypothesis test. In order to account for the directional aspect of the velocity, the vector is decomposed into two components: the horizontal component u and the vertical component w. For each component at any location in the flow field, a separate chi-squared test for normality is performed. Figure 6.10 presents the statistical test result for the entire flow field over all the 300 cycles at 280° bTDCf. The majority (over 58%) of the flow field presented a normal distribution for both velocity components (failed to reject the null hypothesis in both tests, marked in dark blue in Figure 6.10).[13] However, small regions in the intake stream and at the bottom edge of the field of view (near the tumble vortex centre) saw scalars rejecting the null hypothesis of being sampled from a normal distribution at the 5% significance level: among the 1870 grid points in Figure 6.10, there are about 11% grid points rejected the null hypothesis in the test of u (marked in light blue), about 21% grid points rejected the null hypothesis in the test of w (marked in green), and about 10% grid points rejected the null hypothesis in both tests (marked

[13] More strictly speaking, in statistics, failing to reject the null hypothesis means "there is no sufficient evidence showing that the samples are not from a normal distribution".

in yellow). The deviation from a Gaussian distribution is presumably caused by a stronger CCV in the intake stream, and the displacement of vortex centres among different cycles.

To further examine the Gaussian properties of the flow field, the distributions of the velocity components at six different locations (marked A–F in Figure 6.10) are analysed in more detail. Figure 6.11 shows the distribution of both horizontal and vertical velocity components of four locations in the bulk flow region ((a)–(d)) and two locations in the intake stream ((e) and (f)), where a fitted normal distribution (red solid line) is superposed onto each histogram. The p values reported in each sub-figure indicate whether the null hypothesis is rejected — if the p value is greater than the significance level (5%, or 0.05), the null hypothesis cannot be rejected, in other words, the distribution is close to a normal distribution. The histograms show, despite there exists cycle-to-cycle variation in all locations, the extent of CCV varies among different regions. The CCV is stronger in the intake stream ((e) and (f)) compared to in the bulk flow ((a)–(d)), and hence the p values are much lower, meaning that the rejection of the null hypothesis (that the samples are from a normal distribution) is more statistically significant in the intake stream. On the other hand, for most of the locations in the bulk flow (as exemplified by Figure 6.11(a)–(d)), even though some of the locations have failed the hypothesis criteria ($p < 0.05$), it can be seen that qualitatively their deviations from a normal distribution are not significant. It is concluded that from the statistical study, for this flow condition, a non-linear correlation exists in the flow field — the vectors at the majority of the grid points follow (or are close to following) a Gaussian distribution. Therefore, mapping the data using a Gaussian distribution onto a higher dimensional space could help flatten the non-linearity that has been exhibited, and potentially give a less ambiguous cut-off threshold in the PCA-based algorithms.

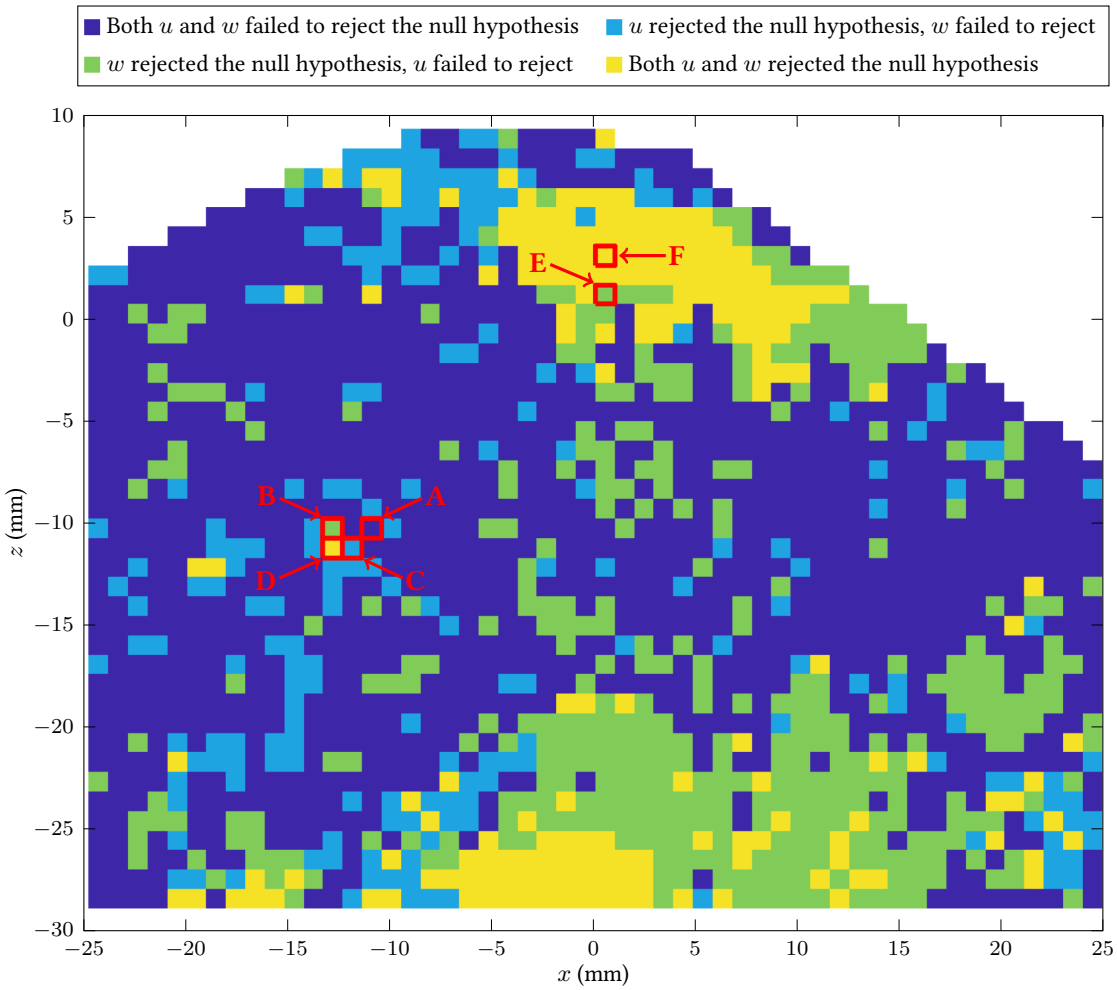


Figure 6.10: Chi-Squared test for normality to each location in the PIV flow fields measured at 280° bTDCf. The null hypothesis is that the vectors on a specific location in the flow field (300 samples in total) are sampled from a normal distribution with a mean and a variance estimated from the samples. The significance level of the test is chosen as 0.05. The horizontal (u) and vertical (w) components are respectively tested. The colours at each grid point indicate whether the null hypothesis is rejected at the chosen significance level, as labelled in the legend. If the p value for the test is smaller than the chosen significance level (0.05), the null hypothesis is rejected. The red boxes and their labels mark six locations from A–F, at which the single-point velocity component distributions and the p values for both velocity components are reported in Figure 6.11.

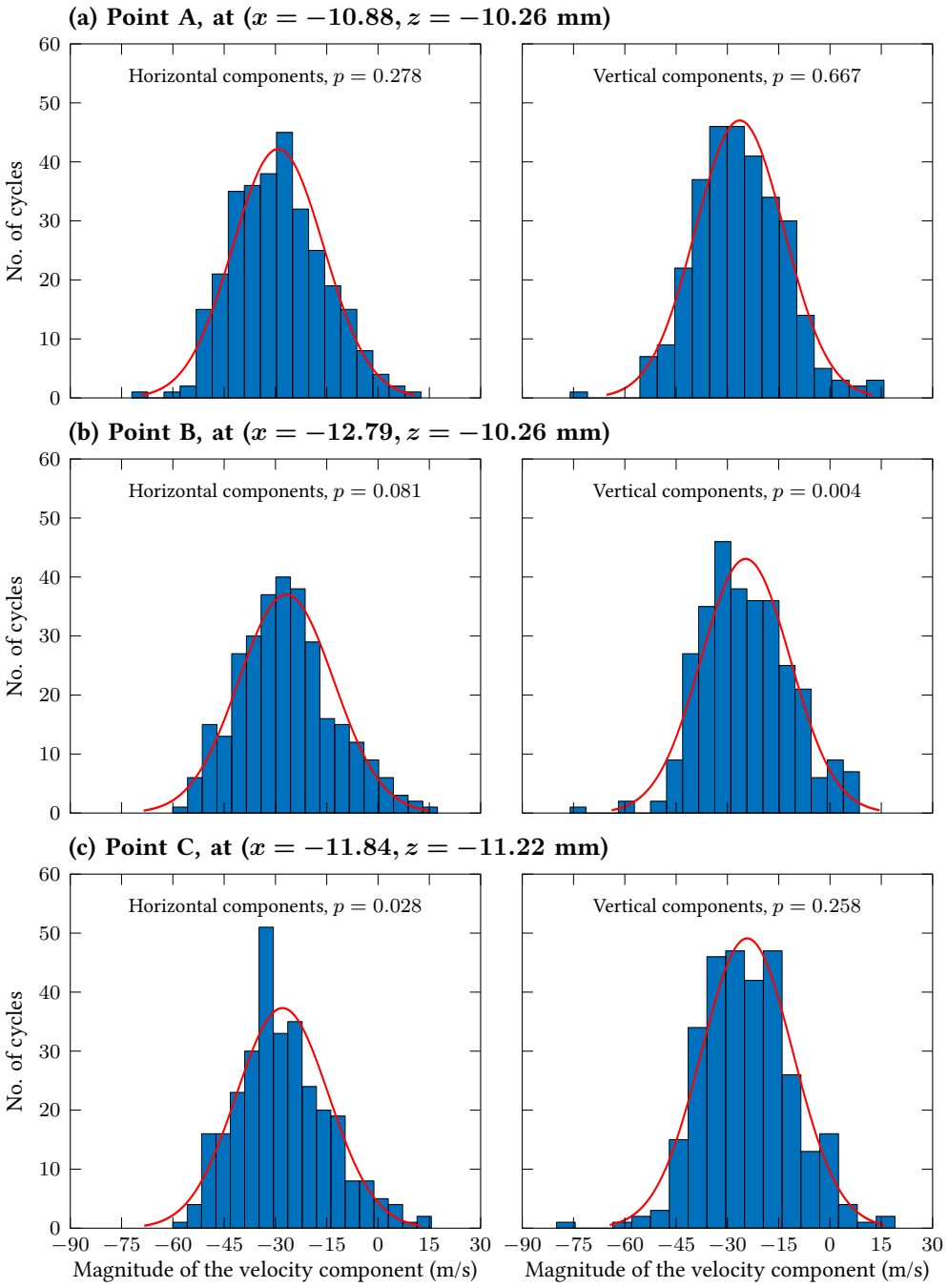


Figure 6.11: Single-Point velocity components distributions at 280° bTDCf. Left column: horizontal components, right column: vertical components. Each row show data at different locations (marked by the red boxes in Figure 6.10). The red lines show a normal distribution fit to each histogram. The p values from the chi-squared test for normality are also reported.

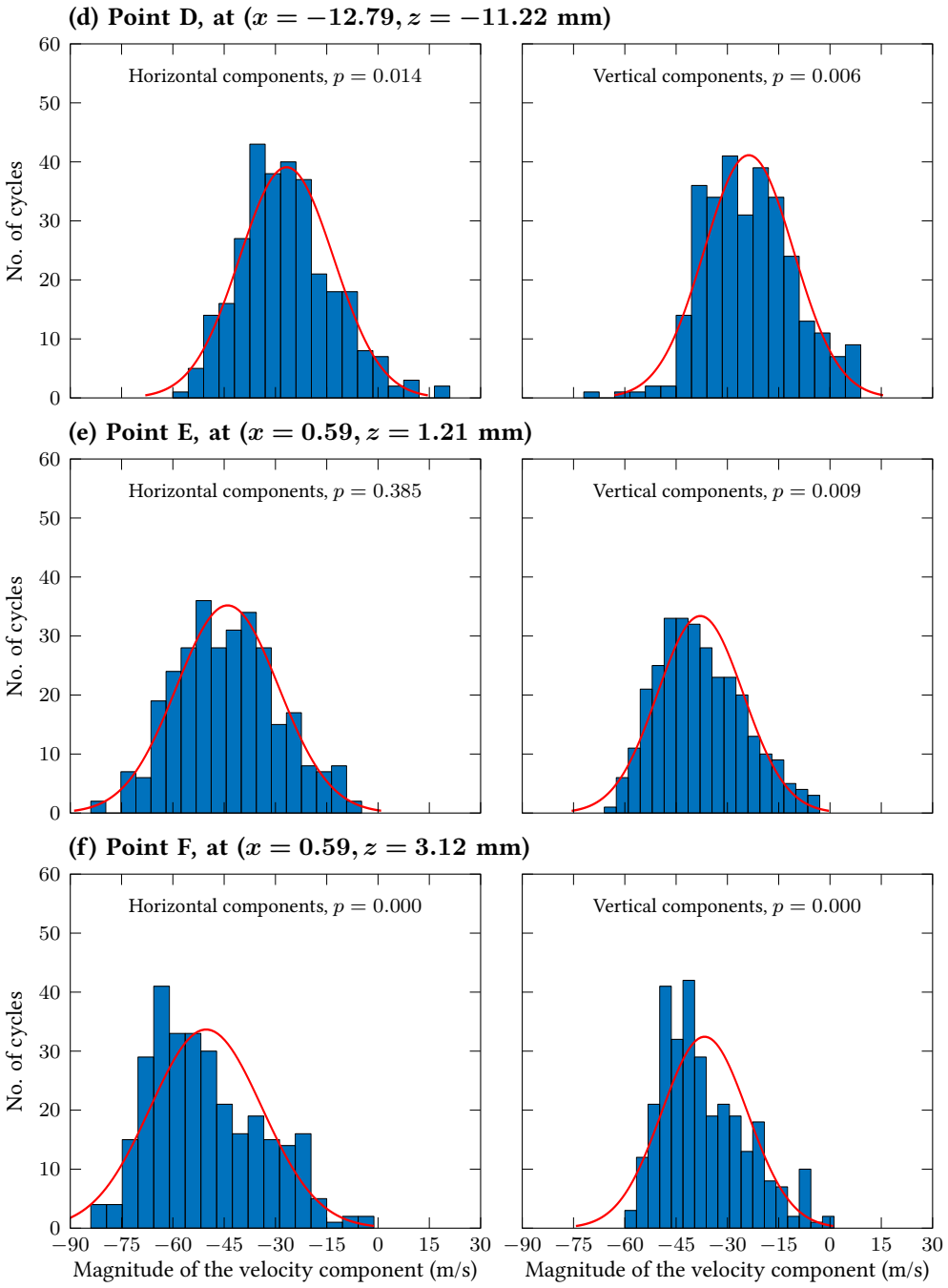


Figure 6.11: Single-Point velocity components distributions at 280° bTDCf. Left column: horizontal components, right column: vertical components. Each row show data at different locations (marked by the red boxes in Figure 6.10). The red lines show a normal distribution fit to each histogram. The p values from the chi-squared test for normality are also reported.

6.5 Application of kernel PCA (KPCA)

From the statistical analysis of the flow field (Section 6.4), there exist strong Gaussian properties in the velocity components. Therefore, when considering PCA-based manifold reduction techniques, a non-linear option may be more suitable than the linear version (i.e., POD). As discussed in the literature review (Section 1.5), the kernel PCA (KPCA) method offers an opportunity to map the data into a selected non-linear space, where a Gaussian kernel is chosen for this study. Following the algorithm introduced in Section 3.2.2, KPCA was implemented on the same data set as used for the POD technique in Section 6.3. The following paragraphs discuss the results of the KPCA method.

For a fair comparison with the POD results, eight KPCA dimensions were also used as the baseline when computing the WRI and WMI fields versus the RANS data. The same procedure used in Figure 6.9 was applied again to the KPCA approximations, in order to compute the spatially-averaged WRI and WMI values for the KPCA-reconstructed flows using different numbers of extra dimensions, and to further examine the "stability" of both metrics when more dimensions are included in the reconstruction. Figure 6.12 shows the resulting values, which are computed using Equation (6.1) and plotted as the change in % of each metric versus the corresponding baseline (eighth-order) reconstruction — the POD results are compared to the eighth-order POD reconstruction, and the KPCA results are compared to the eighth-order KPCA reconstruction. The data for the KPCA reconstructions are plotted in orange, while the values for the POD reconstructions (the same data previously reported in Figure 6.9) are shown again in dark green for an easier direct comparison between the two reconstruction methods. The dark green dots and orange squares illustrate, respectively, the 300-cycle median values for each method, and the error bars show the range of the central 60% values (bounded by the 20% and 80% quantiles).

A noticeable improvement can be found in the spatially-averaged WRI values (top plot)

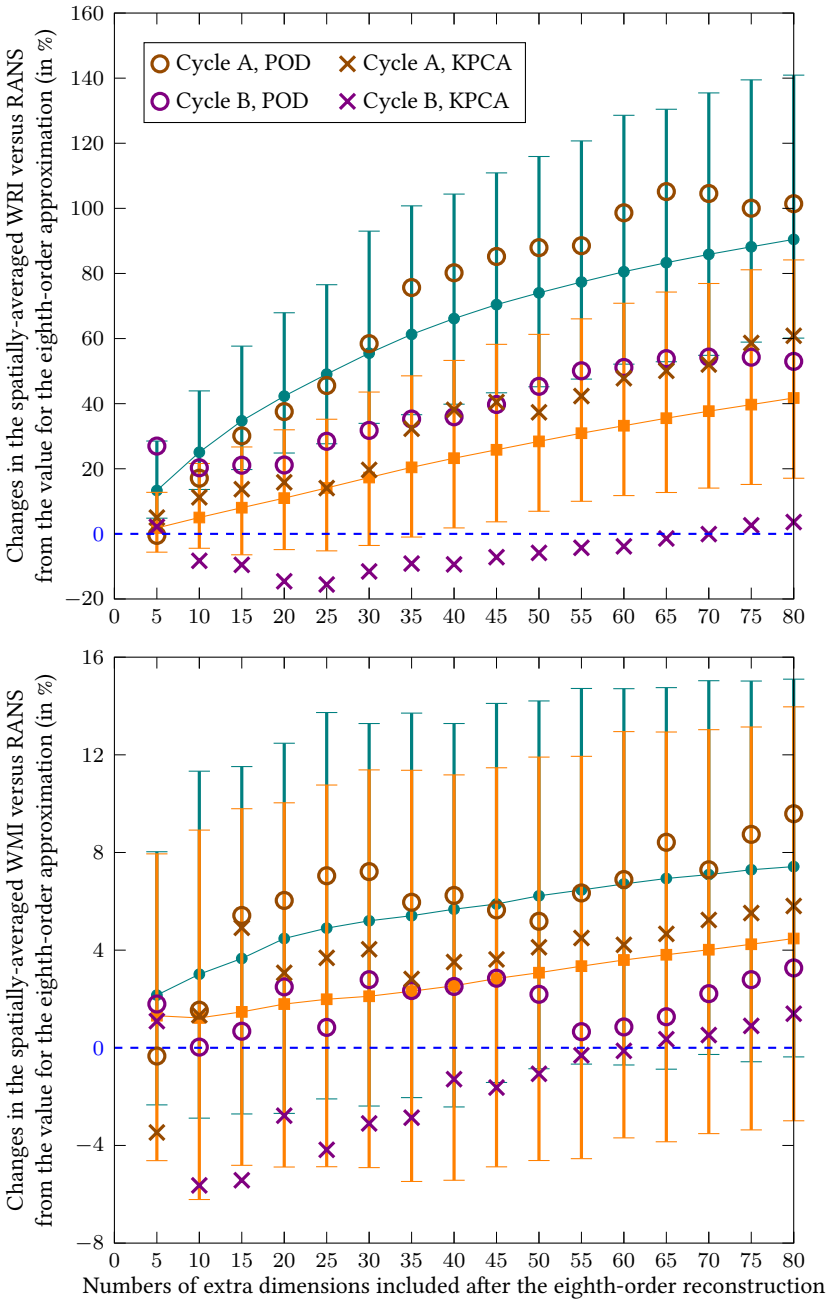


Figure 6.12: Changes in the spatially-averaged WRI (top plot) and WMI (bottom plot) values when extra dimensions are included after the eighth-order POD (dark green) and KPCA (orange) reconstruction. Dots/squares: the 300-cycle median; error bars: the 20% and 80% quantiles of 300 values at a specific number of extra dimensions. The changes are shown in % from the baseline at the eighth-order approximation for each reconstruction method. The same values for the POD method were previously reported in Figure 6.9.

when using the KPCA method compared to the POD method — the changes in the metric values are much smaller (closer to zero, highlighted using a blue dashed line) at any number of extra dimensions. For instance, the 26% increase (previously discussed in Section 6.3) in the Cycle B's WRI value with five extra dimensions using the POD method (the purple circle at the coordinate $(5, 26\%)$) drops to 1.8% when the KPCA approach is implemented (the purple cross at the coordinate $(5, 1.8\%)$). Indeed, the Cycle B's WRI values suffer a much smaller deviation from the baseline values at all the dimensions for the KPCA method (purple crosses) than the POD method (purple circles). A similar observation can also be made for Cycle A (brown crosses for KPCA, brown circles for POD). In other words, the quality of match versus the RANS data in terms of the flow direction is more stable when using different numbers of KPCA dimensions. As such, if the KPCA reconstructions were used to validate the RANS data and to examine whether the flow direction is matched well between them, the conclusions would be more robust than the findings made from the POD reconstructions.

On the other hand, the improvement on the spatially-averaged WMI (bottom plot of Figure 6.12) for the KPCA method is less obvious. Both the POD method (dark green) and the KPCA approach (orange) show a reasonably small deviation from the baseline value, while the stability in the values for the KPCA approach is slightly higher than that of POD (i.e., closer to the zero change). For individual cycles, Cycle A has a small improvement in its KPCA approximations (brown crosses) than its POD reconstructions (brown circles), while the changes for Cycle B using both methods (purple circles for POD, purple crosses for KPCA) are comparable and both close to zero. In conclusion, the robustness in terms of the flow speed difference with the RANS data are equally good for both the POD and KPCA reconstructions when using various numbers of dimensions.

Figure 6.13 illustrates the KPCA-reconstructed flow fields for the two example cycles (left column: Cycle A and right column: Cycle B), of which the same numbers of dimensions are included in the flow fields on each row. The KPCA-reconstructed flow fields behave in a similar

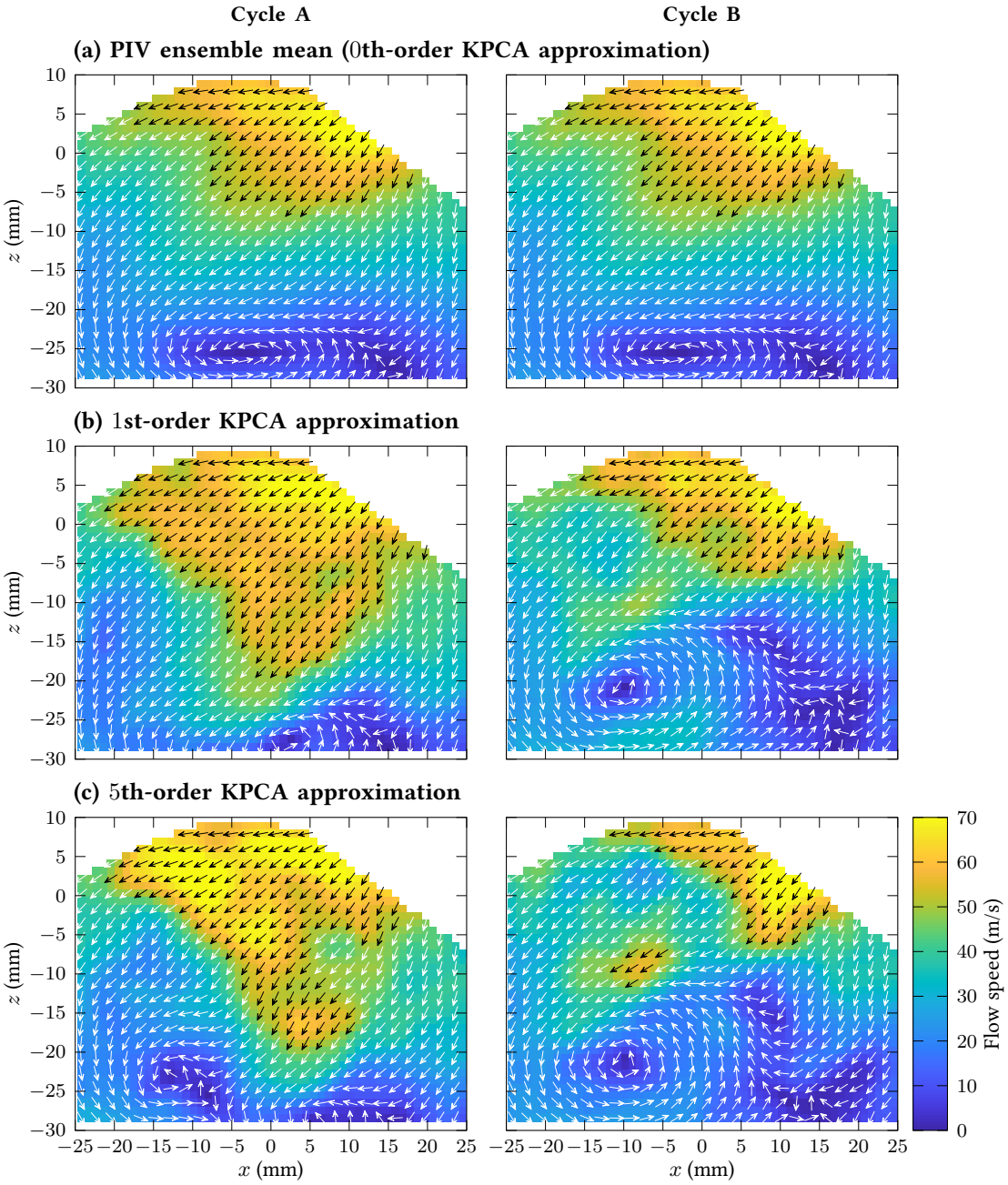


Figure 6.13: KPCA-Based reconstruction for two example cycles (left column for Cycle A, right column for Cycle B) at 280° bTDCf. Different numbers of dimensions are included in the reconstruction for each row.

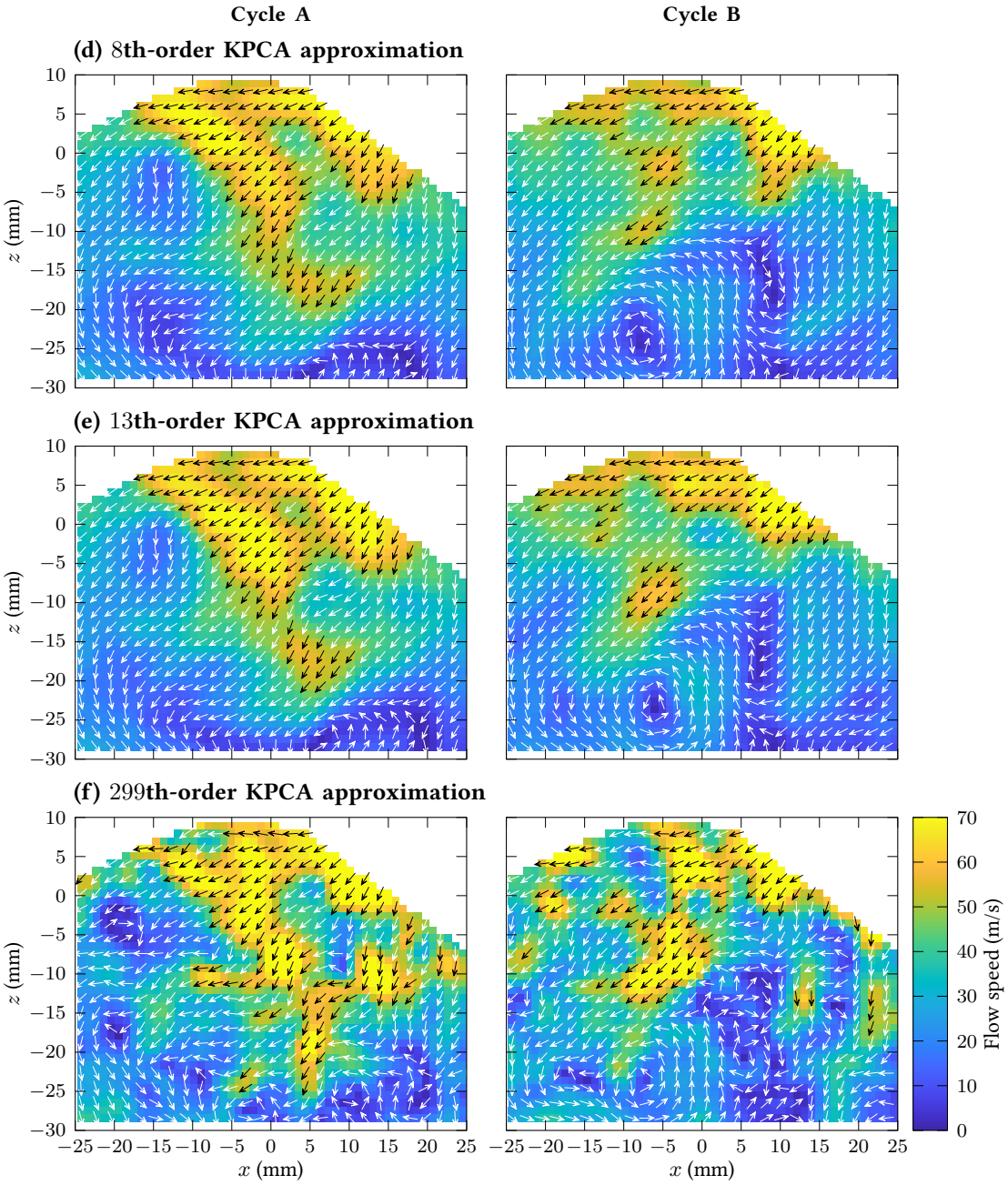


Figure 6.13: KPCA-Based reconstruction for two example cycles (left column for Cycle A, right column for Cycle B) at 280° bTDCf. Different numbers of dimensions are included in the reconstruction for each row.

way to the POD-reconstructed flow fields (shown in Figure 6.6). The first-order KPCA approximations in Figure 6.13(b) are very close to the PIV ensemble mean (Figure 6.13(a)), while the flow fields start to deviate more from the ensemble mean as more dimensions are included in the flow field reconstruction (Figure 6.13(c)–(f)). When using all 299 dimensions, the KPCA approximations resemble the original cycles with negligible differences caused by the numerical rounding error during the reconstruction process. At higher orders, the KPCA approximations also have small changes in the flow structures as had when using the POD method. For instance, the tumble vortex centre (near $x = -8, z = -25$ mm) has a small offset ($\approx$ 2 mm) between the eighth-order KPCA approximated Cycle B (right column of Figure 6.13(d)) and its 13th-order KPCA approximation (right column of Figure 6.13(e)). However, as suggested in Figure 6.12, such a small difference has minimal influences on the spatially-averaged WRI and WMI values.

The reconstruction of the flow field using the KPCA method has shown an advantage in terms of the convergence of the spatially-averaged WRI and WMI after the inclusion of at least eight dimensions. Therefore, one may also use eight dimensions for KPCA reconstructions and compare the reconstructed flow fields with the RANS data. Additionally, it is also possible to make a direct comparison between the POD and the KPCA methods to examine their effectiveness in maintaining a stable coherent flow structure when the same number of dimensions (in this case, eight dimensions) is used for the reconstruction. Recall that the WRI and WMI metrics can also provide spatial information, the two metrics can also be used to check the level of agreement with the RANS data across the entire flow field, and the results can be compared against the WRI and WMI fields for the individual cycles (Figure 6.3(c)–(f)) and the POD reconstructions (Figure 6.7). Figure 6.14 shows the WRI (left column) and WMI (right column) fields for the KPCA reconstructed flow field of Cycle A (top row) and Cycle B (bottom row). The WRI values for the KPCA-reconstructed Cycle A (Figure 6.14(a)) are low across the whole field, suggesting an excellent directional match to the RANS data. For Cycle B,

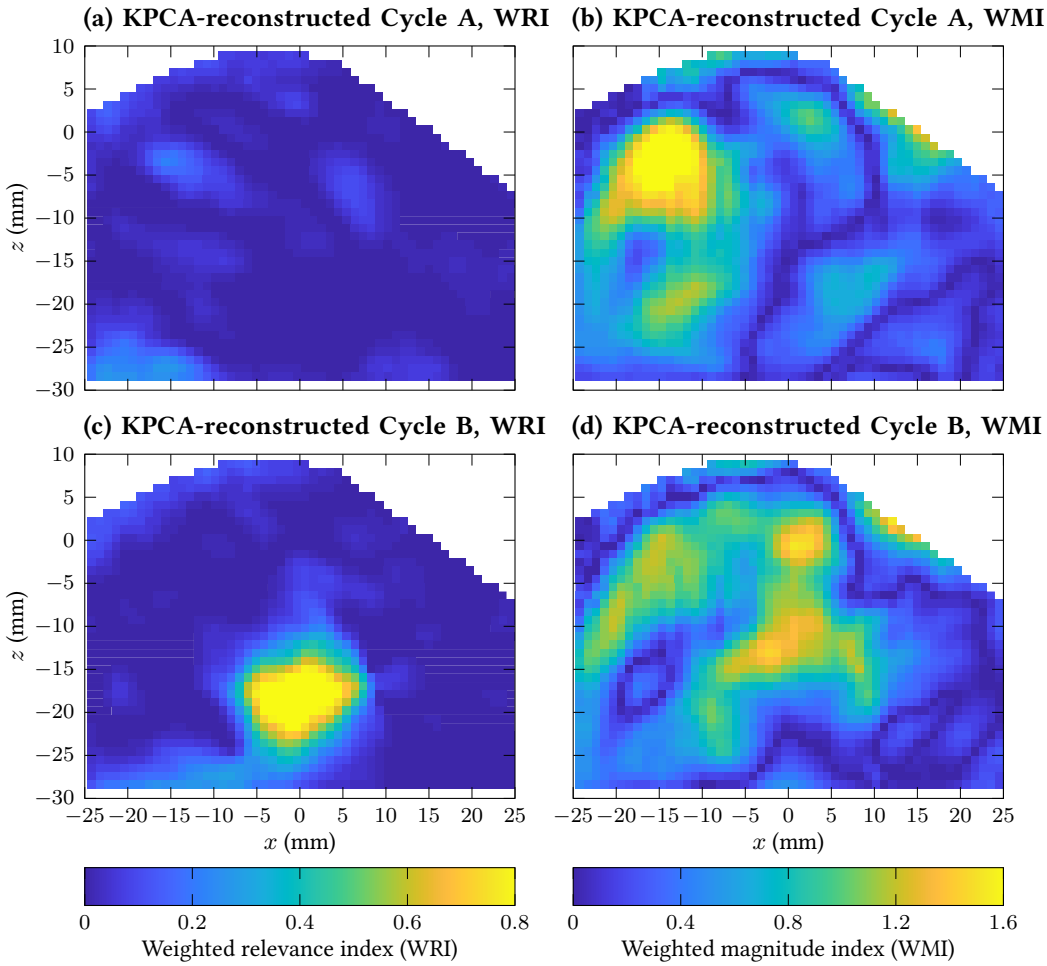


Figure 6.14: Weighted relevance index (WRI, left column) and weighted magnitude index (WMI, right column) fields versus the RANS data for two KPCA-reconstructed cycles (top row for Cycle A, bottom row for Cycle B) at 280° bTDCf.

higher WRI values are observed in its KPCA reconstruction (Figure 6.14(c)) at the bottom part of the field of view, resulting from the disagreement between the tumble vortex centre locations in the measured and the simulated flow fields. The WRI values across the whole field of view for the KPCA approximations (Figure 6.14(a) and (c)) are significantly smaller than the values for the corresponding individual cycles (Figure 6.3(c) and (e)), showing that the KPCA approximations match better with the RANS data than the original cycles in terms of the flow

direction. For instance in Cycle A, the high values at several small pockets within its original flow field (Figure 6.3(c)) do not occur in its KPCA reconstruction (Figure 6.14(a)), since the KPCA reconstruction process filtered out the noise and small-scale fluctuations that may result in a mismatch when comparing with a smooth RANS flow field. On the other hand, since the KPCA approximations retain the flow speed in the original cycles, the WMI values for the KPCA approximations (Figure 6.14(b) and (d)) and for the corresponding individual cycles (Figure 6.3(d) and (f)) are roughly at the same magnitude. It can also be observed in Figure 6.14 that both the WMI and WRI fields for the KPCA approximations are smoother than the ones for the individual cycles (Figure 6.3(c)–(f)) due to the filtering process being embedded in the KPCA algorithm. Similar findings were also made when investigating the POD approximations using the WRI and WMI fields (Figure 6.7). In fact, when comparing the WRI and WMI fields for both reconstruction methods (Figure 6.7 for POD, Figure 6.14 for KPCA), despite there are some local differences, minimal distinction can be found in their overall levels of match with the RANS data. It is worth reiterating that, however, the flow direction change among different orders of KPCA reconstructions are smaller than the change among the corresponding POD approximations. Therefore, the conclusions that are drawn based on the WRI and WMI fields are more robust when using the KPCA method.

Following the same approach as applied in Figure 6.4, one can use the spatially-averaged WRI and WMI values for all the 300 cycles (Figure 6.15) to quantitatively compare the level of match with the RANS simulation among various PIV data processing methods. The PIV data were processed by the four different methods introduced in this chapter: the PIV ensemble mean (a single value, labelled by the intersection of two blue dashed lines and the blue square), the individual cycles (300 values, marked by the red crosses), the POD-reconstructed flow fields (300 values, marked by the dark green circles), and the KPCA-reconstructed flow fields (300 values, marked by the orange circles). Both reconstructed flow fields include eight dimensions, which is a value determined by the quadruple POD analysis in Section 6.3). The two dimensionality

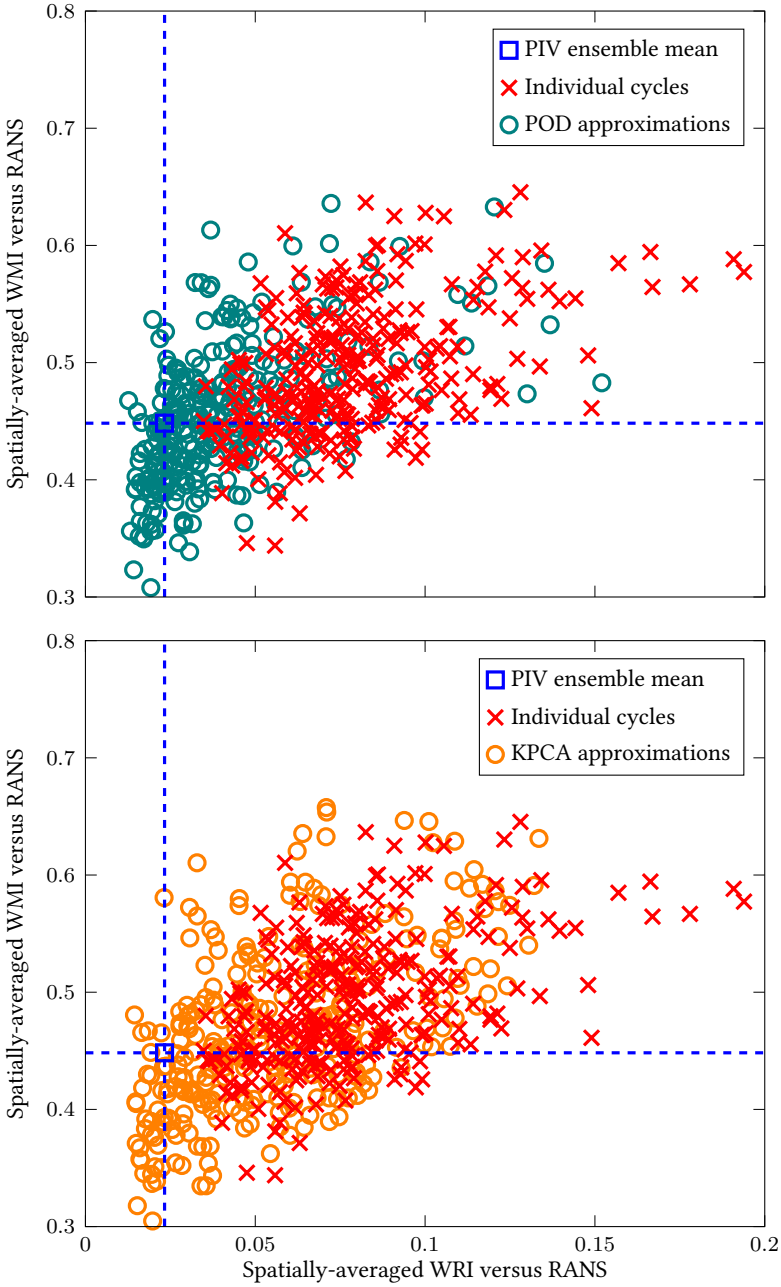


Figure 6.15: Spatially-Averaged WRI (top plot) and WMI (bottom plot) versus the RANS data for four different validation targets: the PIV ensemble mean (the intersection of the blue dashed lines, also marked by the blue square), all the 300 individual cycles (red crosses), the POD reconstructed flow field (dark green circles, top figure only) and the KPCA reconstructed flow field (orange circles, bottom figure only) at 280° bTDCf.

reduction techniques (POD and KPCA) incorporate the CCV information that cannot be obtained from using the PIV ensemble mean. The reconstructed flow fields from both techniques, when comparing to the RANS data, lead to much smaller spatially-averaged WRI and WMI values than the values for the individual cycles, indicating that the RANS predictions match better with the reconstructed (or "filtered") flow fields. Limited differences are found over all cycles when comparing these two methods, which again show that they are equally suitable in the validation of the RANS simulation. However, as highlighted earlier in this section, when the experimental scalars show Gaussian (or near-Gaussian) properties, the KPCA technique has a significant strength in the stability of the WRI and WMI values when using different numbers of dimensions, which may lead to a more robust conclusion of the RANS model validation.

6.6 Chapter summary

In this chapter, the accuracy of conventional methods in validating the RANS simulated flow field via either ensemble averaged or individual PIV flow fields measured in the tumble plane of the optically accessible engine (Appendix A) are firstly examined. The data set contains flow fields measured at a fixed crank angle for 300 cycles. The results suggest that both methods have limitations when validating the simulation results. While the PIV ensemble mean generally shows a good agreement for the main flow structure with the RANS simulation, the cycle-to-cycle variation (CCV) in the flow fields (a key feature of in-cylinder flows) is excluded during the validation. Additionally, the PIV ensemble mean has a lower overall flow speed than is observed in individual PIV images, which may lead to a biased conclusion for the validation. On the other hand, using the individual cycles directly as the validation targets for RANS data exaggerates the differences between the measured and simulated data; the process itself is also questionable since the measured flow fields are highly turbulent, while the RANS simulation only provides an estimation of the bulk flow behaviour based on averaged thermo-fluid equations.

Two manifold reduction methods were then proposed to help make a more objective comparison between the Reynolds-averaged Navier-Stokes (RANS) simulations and the PIV measurements when the cycle-to-cycle variations (CCV) are present in the flow fields. Both qualitative and quantitative studies are given to show the capability and the improvement of each method in capturing the main flow field features compared to the single cycle experimental flow fields, and both techniques additionally capture the CCV information compared to the PIV ensemble mean. Also, the reconstructed PIV data from both methods show a similar quality of match with the RANS data from the current flow simulation model, while there are some differences remain for these two methods.

The number of modes needed for the POD reconstruction was determined using the quadruple POD analysis, which is based on the relevance index reaching 0.95 before and after adding an additional mode. The POD reconstructed flow field shows an improvement in terms of the quality of match with the RANS simulation, as quantified by the weighted relevance index and the weighted magnitude index. Although the POD method shows improvements both qualitatively and quantitatively from using the individual cycles, a challenge exists when finding an non-ambiguous method to determine the number of modes needed in the POD reconstruction. The inherent numerical process in the POD method makes it critical to choose appropriate criteria for the cut-off mode number, as the reconstructed flow field varies significantly with a different number of modes. As a result, the conclusions of the RANS validation were found to depend strongly on numbers of modes included in the POD reconstruction. The current study used a previously developed quadruple POD analysis where the separation of the main flow field is based on the cycle-averaged relevance index. However, a varying RI was also found for different cycles which makes objectively choosing a fixed number of modes for flow field reconstruction even more challenging, and an appropriate

choice usually requires a detailed investigation on the flow feature of interest.[14]

To investigate possible less ambiguous ways to choose the number of dimensions needed for the PCA-based reconstructed flow fields, a statistical analysis is performed on the velocity components at each grid point. For the tumble plane at the investigated condition during the intake process, a Gaussian distribution over 300 cycles was found across the majority of the flow field. Therefore, the kernel principal component analysis (KPCA), a non-linear extension of the conventional linear PCA method, was used as a post-processing tool to reconstruct the in-cylinder flow field. With the implementation of a Gaussian kernel, KPCA allows flattening of the non-linear Gaussian correlation between the cycles in a higher dimensional manifold, and hence a linear PCA can be used to reconstruct the flow field in the non-linear space. Smaller changes were observed in the KPCA reconstructed flow fields comparing to the POD method when varying the numbers of dimensions, which leads to a more robust validation result for the RANS model. A qualitative study also indicates that the KPCA is also capable of capturing the main flow field for each cycle while incorporating the CCV. Further quantitative study also shows that without using dedicated criteria to choose the number of dimensions, the KPCA reconstructed flow field can have comparable accuracy to the POD reconstructed flow fields. The findings in this chapter imply that, when the flow fields can be represented by known distributions (such as a Gaussian distribution), the KPCA method with an appropriate kernel is seen to provide more stable values of vector field comparison metrics, and hence in this case it is a promising candidate for PIV data analysis in both extracting the coherent structures within the flow field and in validating the RANS simulations. Additionally, since the flow features do not alter significantly for the KPCA reconstructed flow fields at different numbers of dimensions, it is possible to choose a universal cut-off dimension for multiple

[14] As exemplified before in Section 5.4.2, a stability test based on specific flow feature of interest is recommended, and in this case the numerical values in the test are chosen by the user who defines a level of match needed for the validation.

engine cycles. Further studies will focus on flow fields exhibiting other statistical distributions such as a bimodal distribution in the cross-tumble plane (as shown in Figure 5.4(a)). The methods introduced in this chapter are also applicable for multi-cycle simulation data (such as the large eddy simulation, LES), and for other types of turbulent flow data that need de-noising and reconstruction.

Chapter 7

Conclusion, summary and discussion

In this chapter, the key findings in this thesis are summarised with some further discussions. The originalities and novelties of the work presented in previous chapters are firstly discussed, together with some learning points from various topics. Suggestions are also made on the validation procedure and on the possible extensions of the current work. A list of publications that are related to this thesis is also provided.

7.1 Highlights of originalities, novelties and learning points

The main objective of this thesis is to validate in-cylinder flow simulation models using experimental data. This involves measuring the in-cylinder flow fields and developing methods to compare the experimental and the simulated data. The original and novel aspects of the work are listed below; some learning points are also highlighted.

In-cylinder flow data set for multiple test points and measurement planes

A high-speed planar particle image velocimetry (PIV) technique was applied to measure the in-cylinder flow fields of a near-production optically accessible engine under different test conditions and in multiple planes in order to provide a comprehensive data set for the comparison with the simulated data. The engine was operated at a total of eleven test conditions (see Table 2.4) that are of industrial interest specified by Jaguar Land Rover. The changes among the test points include engine operation loads (see Table 2.3), camshaft profiles, numbers of intake valve actuated, and compression ratios (varied by using different pistons) — many of which require using different hardware configurations for the test runs. The flow conditions also vary significantly for each test point, hence the PIV seeding density needs to be adjusted according to the actual images at the desired test point. Four different planes (see Figures 4.2–4.9) were selected to provide representative flow fields within the cylinder for each test point. The measurement in each plane requires a separate optical layout (see Section 2.2) — each takes weeks to design and to set up. Changing the measurement plane involved designing the laser beam and light sheet formation layouts, aligning and calibrating the PIV camera, and providing necessary guarding to ensure laser safety. During this process, efforts were made to avoid background scatter from various directions caused by the reflective surfaces within the cylinder, by means of carefully designing the light sheet entry route and imaging direction, as well as painting the metal surfaces in black. Additionally, for a specified

test condition and measurement plane, the flow speed may vary significantly among crank angles, hence the separation time between the pulses from each of the two laser cavities has to change from crank angle to crank angle to maintain a suitable particle moving distance (see Section 2.1.3). This involves many iterations (i.e., a trial-and-error process) in experiments since the actual flow speed is unknown until it is calculated via PIV cross-correlation. It should also be noted that, although it seems that the data collection time is short — 300 cycles were recorded at each test point in each measurement plane, which corresponds to less than a minute of operation, or more precisely, 24 seconds at 1500 rpm for the T1 and T2 loads — on average it takes about five days to collect 300 cycles of data at a specified test point (excluding the time needed when changing optical configurations). It is found during the actual tests that a stable engine operation may be hard to achieve, especially when the engine load is low (hence the throttle is harder to control). The optical access (provided by the transparent cylinder liner and the piston window) can be easily blocked by various oils (such as the engine lubricants and the PIV seeding oil), which makes a test run invalid. The cylinder liner therefore requires frequent cleaning between tests and each cleaning process involves an engine reassembly.

High fidelity PIV data in an optically accessible engine with production engine pistons

The optically accessible engine was equipped with shaped crown pistons which are used in actual production engines (see Appendix A). The shaped crown pistons, compared to the flat-top pistons used in conventional research optical engine, has a reduced optical access and introduces more highly-reflective surfaces, and leads to a more challenging optical alignment environment. In addition to careful alignment of the optics and design of the experimental setup (see Section 2.2), a cycle filter was proposed in Section 2.4 to effectively remove the background scatter, resulting in high fidelity data with low vector removal rate (less than 0.1% of all vectors) and minimal uncertainty (less than 5% of local flow speeds).

Consistency and repeatability of PIV experiments separated by months

The optics and camera were aligned to measure the flow fields in a specified plane for all the test points before being re-organised for the measurements in another plane, so that the measurement plane can be fixed among multiple test points and the time needed for the realignment is kept to a minimal. This means the measurements at the same test point but in different planes have to be separated by months. Despite the difficulty, it can be seen in Section 4.1 that the same test point can still be reached with acceptable consistency and repeatability in terms of the in-cylinder pressure traces. In Section 4.2, it is further observed that the flow structures in the ensemble mean PIV flow fleids match reasonably well among different measurement planes.

Comparison of PIV and RANS flow fields in multiple planes

In the previous works by former group members [55, 99, 100], who also achieved high-quality experimental data and developed the vector field comparison metrics, the validation of RANS-simulated in-cylinder flow fields was limited to the usage of the PIV ensemble mean on a single plane (the tumble plane). The work in this thesis (see Chapter 4) extends the comparisons to multiple measurement planes. It is illustrated in Sections 4.2–4.4 that the evolution of the three-dimensional key flow structures among different crank angles in the engine cylinder can be clearly viewed when observing the in-plane flow from the orthogonally-oriented planes (tumble, swirl and cross-tumble plane). For instance, the motion of the tumble vortex induced by the intake flow generates different flow directions in different measurement planes, and causes an in-plane flow direction reversal when its centre passes through the measurement plane. A detailed comparison using the vector field comparison metrics (see Section 3.3) to the multi-plane data set in Section 4.4 gave a clearer and more comprehensive view on how better the simulation model matches with the experiments, and the discrepancies between the experimental and simulated

data were detected at a variety of crank angles. The simulation was shown to have a better match with the experiment at a higher engine load, and there exists an offset between the locations of the tumble vortex centre in the measured versus the predicted flow fields, as illustrated by a delay in the in-plane flow-reversals for CFD compared to PIV.

Examination of methods beyond the conventional approach of the PIV ensemble mean for RANS model validation

The experimental PIV data needs to be averaged (or processed) to enable its comparison with the RANS-simulated data — this is because the simulation only provides a single realisation (i.e.,flow field) of the bulk flow behaviour at a specified crank angle, while the experiment measures the instantaneous flow fields in multiple engine cycles with smaller-scale turbulence and experimental noises that are either not resolved by or not included in the RANS model. Conventionally, the averaging process refers to the "ensemble-averaging", i.e., taking the arithmetical mean of the velocity vectors measured at the same location and at the same crank angle for multiple cycles, and a so-called "ensemble mean" flow field of the specified crank angle (see Equation (3.5)) is generated from the averaging process. This approach is widely used (for instance, [26, 99, 100]) for the comparison between PIV and RANS in-cylinder flow fields, and was also applied in Chapter 4 when investigating the overall level of match between the measured flow fields and the RANS model. Despite the ensemble mean providing an acceptable representation of the multi-cycle flow fields in many cases, it was discussed in Section 5.1 that blindly using the ensemble mean to validate the RANS flow simulation model can sometimes cause misleading conclusions. When there exists a flow feature that shows strong directional variations among different cycles, for instance in the case of a swinging intake jet (see Figure 5.1), the ensemble mean will smear out that feature and under-represent the actual flow speed in individual cycles. As a result, should the RANS model be directly compared to the ensemble mean, a false conclusion may be drawn that the model is

over-estimating the flow speed — whereas in reality it is the ensemble mean itself which is not representing the actual flow speed appropriately. Therefore, PIV data processing methods beyond the conventional scope of ensemble mean are needed in order to provide fairer validation targets for the RANS model — two of which, namely proper orthogonal decomposition (POD) and its generalised form, kernel principal component analysis (KPCA), were examined in Chapters 5 and 6, and their algorithms were provided in Section 3.2. Both methods were applied to filter the individual cycles to enable the comparison with the RANS-simulated flow field. The "reconstructed" (i.e., filtered) flow fields were found to retain the coherent flow structure within each measured flow fields, while effectively excluding the smaller-scale turbulence structures and noises in experiments which may bias the comparison (again, these structures are not included in the RANS data), and hence they were used as alternative validation targets for the RANS model. Unlike the PIV ensemble mean which is a single averaged field, the reconstructed flow fields contains the cyclic variation information and hence there are multiple (the same number as the number of cycles) fields that the RANS model can validate against at a specified crank angle. A RANS flow field is considered to be a sufficiently accurate prediction of the coherent flow behaviour if its key flow features are similar to any of the reconstructed experimental flow fields, or the deviation from the reconstructed flow fields lies within the range of the cyclic variation (see Section 5.4.3). The level of similarity can be quantified by the methods summarised in the following paragraph.

Metrics to quantify the level of similarity between flow fields

The RANS flow model validation process requires frequent comparisons of similarity between the experimental and simulated flow fields. Two sets of metrics were used to serve this purpose in this thesis. The first set (introduced in Section 3.3), namely the weighted relevance index (WRI), the weighted magnitude index (WMI) and the combined magnitude and relevance index (CMRI), were developed by previous group members [55, 99, 100]. These

metrics extended the concept of the standard relevance index used by Liu *et al.* [97] (see Equation (3.51)) from a field-to-field comparison (which generates a number for two given flow fields) to a point-by-point comparison (which generates a spatially-resolved field), hence it is now possible to know where the discrepancies occur in the flow fields. The flow speed and flow direction comparisons are also separated in two different metrics in order to better quantify the discrepancy in these two key aspects. As exemplified by Section 4.4, these metrics effectively identified the crank angle and the location at which the simulated flow fields do not match with the measured ones from a large data set (two test points, each have 300 measured cycles and 61 measured crank angles). The values will be high at a certain location in the WRI, WMI and CMRI fields if there exists a large discrepancy between the two given flow fields.

On the other hand, as discussed in Section 5.2, these point-by-point comparison metrics may not be appropriate to use when the coherent structure of interest (in this case a flapping intake jet resulting from the collision of two intake streams from both valves) shows significant cycle-to-cycle variation in terms of the location. Therefore, another set of metrics (introduced in Section 5.2) based on the actual flow features were proposed by the author of this thesis. The intake jet was firstly quantified based on its angle, penetration length and width in each flow field, and the values of these "jet profile quantification metrics" were then compared between the experimental and simulated flow fields. The accurate simulation is expected to have similar values of these metrics to the ones of the experiments.

Convergence of the POD-reconstructed flow fields and the stability test

The implementation of the POD-reconstructed flow fields in Chapter 5 successfully avoided the possible bias in the flow speed due to the averaging process while computing the PIV ensemble mean. As illustrated by Figure 5.7, however, the flow features in the POD-reconstructed flow fields depend on the number of POD modes included during the reconstruction process, and hence the cut-off mode number needs to be carefully chosen so

that a stable flow structure can be achieved for the simulation model validation.

In Section 5.4.2, a stability test was introduced to check whether a POD-reconstructed flow field has reached a stable flow structure. The stability criteria were developed based both on the flow features of interest and on parameters that are independent of the flow structures such as the amount of the kinetic energy content collected by the reconstructed flow fields. A cut-off mode number is then selected for each measured flow field, and the resulting "low-order POD approximated flow fields", when being used to compare with the RANS data, can generate reliable validation conclusions.

It was also pointed out in Section 5.4.2 and in a later section (Section 6.1), however, the choice of the stability criteria is not free from subjective views. Although the flow feature parameters are selected based on the actual level of similarity needed (i.e., by choosing a required accuracy of the parameters, see Section 5.4.2), it can still be argued that, when applying the POD reconstruction technique, a clear separation of the coherent flow structure from the smaller-scale turbulence is hard to achieve and requires careful examinations on many individual flow fields to ensure the selected parameters meet with the actual requirement. As discussed in Section 1.5, this difficulty in choosing an appropriate cut-off POD mode number has been demonstrated by other researchers both in the scope of in-cylinder flow analysis [66, 67, 72, 73], as well as in broader research areas such as in turbulent flow data processing [59, 68, 74] and in statistics [70]. Therefore, in Chapter 6, the KPCA method was implemented to examine whether a more objective choice of the cut-off mode number (more precisely, in KPCA it is called the "cut-off dimension") can be achieved, and whether this issue can be bypassed.

Implementation of KPCA in engine in-cylinder flow analysis

The KPCA method (whose algorithm was provided in Section 3.2.2) is a generalised POD (or in statistics, linear PCA) process which is usually used for a non-linear manifold in statistics, but unlike the POD method, it has not yet gained attention in the engine in-cylinder flow analysis

community. As the in-cylinder flow is highly turbulent, it may have some non-linear correlations between the flow fields, and hence a non-linear PCA method such as the KPCA method is worth attempting for the PIV data filtering process.

The successful implementation of the KPCA technique requires some knowledge about the statistical distribution of the data so that an appropriate kernel can be selected to effectively filter the original data and to provide reconstructed data with a reduced dimensionality. As such, a statistical analysis was first performed in Section 6.4, and for the majority of the locations in the flow field, the horizontal and vertical velocity components at each grid point for different cycles at the same crank angle were found to respectively follow a Gaussian distribution. A Gaussian kernel was then used on the data set in Section 6.5, and the resulting KPCA-reconstructed flow fields, when compared to the POD-reconstructed ones, show a more stable trend in the validation metrics (spatially-averaged WRI and WMI) as a function of the number of dimensions used (see Figure 6.15). In other words, when applying KPCA to the flow fields at the given flow condition (intake flow measured in the tumble plane), the values of the validation metrics, and hence the conclusion of whether the flow simulation model is sufficiently accurate, no longer depends greatly on the choice of the cut-off dimension — using any large enough cut-off dimension number (in this case above eight) in the KPCA reconstruction process would lead to the same validation conclusion. It is recommended, on the other hand, not to choose an extremely large cut-off dimension number (for instance in this case not higher than 88) in order to avoid unnecessary inclusion of smaller-scale turbulence structures in the reconstructed flow fields.

It should be noted that, although the KPCA method has the potential to bypass the ambiguity in choosing a specific cut-off dimension number (because the range of choice is very wide), it requires specific knowledge about the statistical distribution of the data in order to choose the corresponding kernel. If the statistical distribution is unknown or not universal for all the grid points in a given flow field, the KPCA may not show a better converged trend in the validation

metrics compared to the linear PCA (i.e., POD). The computational cost for the KPCA method is also much higher — for the given 300-cycle data set, it took about six seconds for the POD analysis, but the KPCA algorithm needs more than four hours on the same computer. [(1)]

Suggestions on the procedure of the flow model validation

As mentioned in the previous paragraphs, each of the three methods (ensemble mean, POD and KPCA) introduced in this thesis has its own benefits and disadvantages when being used as the validation target of the RANS model — these findings are summarised in Table 7.1. A recommended flow model validation procedure is therefore proposed as follows:

1. Determine key flow features of interest that the flow model needs to match with, such as the bulk flow motion discussed in Chapters 4 and 6, and the intake jet profile discussed in Chapter 5;
2. According to the flow features of interest, choose between the overall flow field comparison metrics (see Section 3.3) and a set of metrics based on the flow features (see Section 5.2) as the validation metrics;
3. Compute the ensemble mean and compare firstly with the individual measured flow fields to check whether the actual flow speed and key flow feature is correctly represented by the ensemble mean;
    (1) If the ensemble mean provides a good representative of the coherent flow structure, it can be used as the validation target of the simulation model;

[(1)] The time corresponds to the computation of the KPCA-reconstructed flow fields using all dimensions from the first-order to the 299th-order approximations. The majority (over 99%) of the computational cost results from the backward mapping algorithm which requires iteration (see Equation (3.50)). Provided that the number of dimensions is known (for instance by choosing a large enough number), it should take about 50 seconds to perform the entire KPCA analysis at the desired number of dimensions. In contrast, there is no backward mapping process in the POD algorithm, so every order of POD-reconstructed flow field can be computed within a very short period of time through simple matrix multiplications (see Equation (3.28)).

(2) If the ensemble mean fails to provide a good representation of the coherent flow structure, apply the POD reconstruction process and proceed with Step 4;

4. Check whether a clear cut-off POD mode number can be easily obtained by plotting the validation metrics specified in Step 2 as a function of mode number (as exemplified by Figures 5.8 and 6.15);

   (1) If the validation standard does not alter significantly with the cut-off POD mode number (as exemplified by Figure 5.8), apply a stability criteria (see Section 5.4.2) and use the resulting POD-reconstructed flow fields as validation targets for the simulation model;

   (2) If there exists ambiguities in the determination of the cut-off POD mode number (as exemplified by Figure 6.15), attempt the KPCA method in Step 5;

5. Examine the statistical distribution of the data and see if an appropriate kernel can be chosen based on the distribution;

   (1) If yes, apply the KPCA method and use the resulting KPCA-reconstructed flow fields as validation targets for the simulation model;

   (2) If not, it would be better to use the POD-reconstructed flow fields with cautions that the validation conclusions may be affected by the subjective choice of cut-off POD mode number.

The above procedure took the limited time for an industrial engine design loop into account, tried to avoid high computational cost if possible, flagged out the possible biases in the validation process, and focused on getting a robust validation conclusion with minimal subjective inputs. It should be noted that, on the other hand, the validation of a flow simulation model inevitably involves some human decisions. For instance, the desired thresholds for the validation metrics in Step 2 have to be chosen by the user to determine whether the model is "accurate enough"; the validation methods need also be selected based on the actual flow features and the budget.

Table 7.1: Comparisons of each validation target

Validation target	Advantages	Disadvantages
Ensemble mean	Easy to implement Quick check for many test points	Possible flow speed under-representation No cyclic variation information
POD reconstruction	Relatively low computational cost Contains cyclic variation information Retains coherent flow structures Represents actual flow speeds	Requires careful choice of the cut-off mode number
KPCA reconstruction	Less ambiguity in choosing the cut-off dimension number Contains cyclic variation information Retains coherent flow structures Represents actual flow speeds	High computational cost Requires knowledge of data statistical distribution

7.2 Future work

The work in the thesis can be extended in many different ways:

Validation of the flow simulation model using the entire data set

A large in-cylinder flow data set has been collected in the same optically accessible engine (in nine different engine configurations, four different measurement planes, for 61 crank angles ranging from the early intake stroke to the late compression stroke, and each has 300 cycles), and only a small portion of the data were presented in the thesis to illustrate the possible validation methods. An immediate future work involves applying these methods to the entire data set in order to provide a more complete validation to the current flow simulation model.

Modification of the current flow simulation model

The validation conclusion suggests that the current flow simulation model matches better with the experiments at a higher engine load. Therefore, adjustments need to be made in the flow simulation model so that it can accurately predict the flow behaviour at different test points. Once a new flow simulation model is built by the Jaguar Land Rover CFD team, it needs to be validated again with the experimental data.

Extension to the multi-cycle simulations

The current validation process is limited to a RANS model due to the consideration in the computational cost for industrial use. The RANS model, no matter how accurate it is, cannot provide the cyclic variation information. On the opposite, the multi-cycle experimental data presented in the thesis and the POD analysis in Table 5.2 demonstrated that there exists strong cyclic variation within the engine, which is one of the most important features of in-cylinder flow. In order to capture the cyclic variation, alternative simulation approaches, such as the large-eddy

simulation (LES), must be used. The methods developed in the thesis can also be implemented to validate the LES data. For instance, both the PIV and the LES data can be respectively filtered by the POD or the KPCA method, and the resulting reconstructed flow fields can be compared against each other. Another option is to conduct the POD analysis on a combined data set that contains both the PIV and the LES flow fields, and examine whether the LES data can be identified as outliers — if the LES data is actually a good prediction, they should not be detected as outliers. The identification of outliers using a POD-based algorithm has been demonstrated in a previous publication [107] first-authored by the author of this thesis.

Application of the KPCA method using a different kernel

The KPCA method was implemented with a Gaussian kernel in Chapter 6, and when there exist Gaussian correlations among the data points, it showed promising benefits in reducing the ambiguities in choosing the cut-off dimension number. However, the effectiveness of KPCA in other flow conditions has not yet been examined. As exemplified in Figure 5.4, the velocity data can show a bimodal distribution in the intake jet region. Therefore, another possible future work is to find an appropriate kernel that can be used for this flow condition (and later for other flow conditions as well).

7.3 Publications relating to this thesis

The following publications are related to the thesis:

[1] L. Shen, C. Willman, R. Stone, et al., "Multi-Plane PIV Measurements in a Gasoline Direct Injection Engine", SAE Int. J. Adv. & Curr. Prac. in Mobility **3**, 233–239 (2021).

[2] L. Shen, C. Willman, R. Stone, et al., "On the Use of Particle Image Velocimetry (PIV) Data for the Validation of Reynolds Averaged Navier-Stokes (RANS) Simulations during the

Intake Process of a Spark Ignition Direct Injection (SIDI) Engine", International Journal of Engine Research **23**, 1061–1081 (2021).

[3] X. Fang, L. Shen, C. Willman, et al., "Manifold reduction techniques for the comparison of crank angle-resolved particle image velocimetry (PIV) data and Reynolds-averaged Navier-Stokes (RANS) simulations in a spark ignition direct injection (SIDI) engine", International Journal of Engine Research **23**, 1275–1294 (2021).

References

[1] L. Shen, C. Willman, R. Stone, et al., "Multi-Plane PIV Measurements in a Gasoline Direct Injection Engine", SAE Int. J. Adv. & Curr. Prac. in Mobility **3**, 233–239 (2021).

[2] L. Shen, C. Willman, R. Stone, et al., "On the Use of Particle Image Velocimetry (PIV) Data for the Validation of Reynolds Averaged Navier-Stokes (RANS) Simulations during the Intake Process of a Spark Ignition Direct Injection (SIDI) Engine", International Journal of Engine Research **23**, 1061–1081 (2021).

[3] X. Fang, L. Shen, C. Willman, et al., "Manifold reduction techniques for the comparison of crank angle-resolved particle image velocimetry (PIV) data and Reynolds-averaged Navier-Stokes (RANS) simulations in a spark ignition direct injection (SIDI) engine", International Journal of Engine Research **23**, 1275–1294 (2021).

[4] IEA, *Global EV Outlook 2019* (International Energy Agency, Paris, 2019).

[5] N. O. Kapustin and D. A. Grushevenko, "Long-term electric vehicles outlook and their potential impact on electric grid", Energy Policy **137**, 111103 (2020).

[6] T. R. Hawkins, B. Singh, G. Majeau-Bettez, et al., "Comparative Environmental Life Cycle Assessment of Conventional and Electric Vehicles", Journal of Industrial Ecology **17**, 53–64 (2013).

[7] V. R. J. H. Timmers and P. A. J. Achten, “Chapter 12 - Non-Exhaust PM Emissions from Battery Electric Vehicles”, in *Non-Exhaust Emissions*, edited by F. Amato (Elsevier, London, 2018), pp. 261–287.

[8] K. Senecal and F. Leach, *Racing Toward Zero: The Untold Story of Driving Green* (SAE International, Warrendale, PA, 2021).

[9] N. C. Onat, S. Gumus, M. Kucukvar, et al., “Application of the TOPSIS and intuitionistic fuzzy set approaches for ranking the life cycle sustainability performance of alternative vehicle technologies”, Sustainable Production and Consumption **6**, 12–25 (2016).

[10] N. C. Onat, M. Kucukvar, and S. Afshar, “Eco-efficiency of electric vehicles in the United States: A life cycle assessment based principal component analysis”, Journal of Cleaner Production **212**, 515–526 (2019).

[11] G. Kalghatgi, “Is it really the end of internal combustion engines and petroleum in transport?”, Applied Energy **225**, 965–974 (2018).

[12] A. C. Alkidas, “Combustion advancements in gasoline engines”, Energy Conversion and Management **48**, 2751–2761 (2007).

[13] R. Stone, *Introduction to Internal Combustion Engines*, 4th ed. (Red Globe Press, London, 2012).

[14] F. Zhao, M. Liu, P. Ge, et al., “Multi-plane time-resolved Particle Image Velocimetry (PIV) flow field measurements in an optical Spark-Ignition Direct-Injection (SIDI) engine for Large-Eddy Simulation (LES) model validations”, Oil & Gas Science and Technology - Rev. IFP Energies nouvelles **74** (2019).

[15] V. Sick, M. C. Drake, and T. D. Fansler, “High-speed imaging for direct-injection gasoline engine research and development”, Experiments in Fluids **49**, 937–947 (2010).

[16] V. Sick, “High speed imaging in fundamental and applied combustion research”, Proceedings of the Combustion Institute **34**, 3509–3530 (2013).

[17] D. Veynante and L. Vervisch, “Turbulent combustion modeling”, Progress in Energy and Combustion Science **28**, 193–266 (2002).

[18] V. K. Krastev, A. D’Adamo, F. Berni, et al., “Validation of a zonal hybrid URANS/LES turbulence modeling method for multi-cycle engine flow simulation”, International Journal of Engine Research **21**, 632–648 (2020).

[19] A. A. Raut and J. M. Mallikarjuna, “Effects of direct water injection and injector configurations on performance and emission characteristics of a gasoline direct injection engine: A computational fluid dynamics analysis”, International Journal of Engine Research **21**, 1520–1540 (2020).

[20] M. Theile, M. Reißig, E. Hassel, et al., “Numerical analysis of the influence of early fuel injection on charge motion in a direct injection spark ignition engine using scale-resolving simulations”, International Journal of Engine Research **21**, 664–682 (2020).

[21] I. Ko, F. Rulli, S. Fontanesi, et al., “Methodology for the large-eddy simulation and particle image velocimetry analysis of large-scale flow structures on TCC-III engine under motored condition”, International Journal of Engine Research **22**, 2709–2731 (2021).

[22] H. Zhao, *Laser diagnostics and optical measurement techiques in internal combustion engines* (SAE International, Warrendale, PA, 2012).

[23] T. Léonard, L. Y. M. Gicquel, N. Gourdain, et al., “Steady/Unsteady Reynolds-Averaged Navier–Stokes and Large Eddy Simulations of a Turbine Blade at High Subsonic Outlet Mach Number”, Journal of Turbomachinery **137**, 041001 (2015).

[24] V. Yakhot, S. A. Orszag, S. Thangam, et al., “Development of turbulence models for shear flows by a double expansion technique”, Physics of Fluids A **4**, 1510–1520 (1992).

[25] W. A. Gardner, A. Napolitano, and L. Paura, “Cyclostationarity: Half a century of research”, Signal Processing **86**, 639–697 (2006).

[26] X. Yang, S. Gupta, T.-W. Kuo, et al., “RANS and Large Eddy Simulation of Internal Combustion Engine Flows—A Comparative Study”, Journal of Engineering for Gas Turbines and Power **136**, 051507 (2014).

[27] N. Wu, W. K. Bushe, and M. H. Davy, “On the experimental validation of combustion simulations in turbulent non-premixed jets”, Combustion Theory and Modelling **14**, 855–874 (2010).

[28] O. Stein, B. Böhm, A. Dreizler, et al., “Highly-resolved LES and PIV Analysis of Isothermal Turbulent Opposed Jets for Combustion Applications”, Flow, Turbulence and Combustion **87**, 425–447 (2011).

[29] X. Fang, R. Ismail, N. Sekularac, et al., “On the Prediction of Spray A End of Injection Phenomenon Using Conditional Source-Term Estimation”, SAE Technical Paper, 2020-01–0779 (2020).

[30] X. Fang, N. Sekularac, and M. Davy, “Parametric Studies of a Novel Combustion Modelling Approach for Low Temperature Diesel Spray Simulation”, in Proceedings of the asme 2020 internal combustion engine division fall technical conference (2020), ICEF2020–2924.

[31] P. C. Miles, “The history and evolution of optically accessible research engines and their impact on our understanding of engine combustion”, in Proceedings of the asme 2014 internal combustion engine division fall technical conference, Vol. 2, edited by A. Dybbs and B. Ghorashi (2014), ICEF2014–5701.

[32] L. Bryant, “On velocity gradients in PIV interrogation”, Scientific American **216**, 102–112 (1967).

[33] C. A. Amann, “A perspective of reciprocating-engine diagnostics without lasers”, Progress in Energy and Combustion Science **9**, 239–267 (1983).

[34] D. Clerk, “Cylinder actions in gas and gasoline engines”, in *SAE transactions*, Vol. 16 (Society of Automotive Engineers, Inc., New York City, 1921), pp. 33–92.

[35] H. R. Ricardo, “Notes on the design of high-speed internal combustion engines”, Proceedings of the Institution of Automobile Engineers **20**, 488–504 (1926).

[36] H. Glyde, “Experiments to determine velocities of flame propagation in a side valve petrol engine”, Journal of the Institution of Petroleum Technologists **16**, 756–776 (1930).

[37] G. M. Rassweiler and L. Withrow, “High-speed motion pictures of engine flames”, Industrial and Engineering Chemistry **28**, 672–677 (1936).

[38] F. W. Bowditch, “A new tool for combustion research a quartz piston engine”, SAE Technical Paper, 610002 (1961).

[39] J. Wolfrum, “Lasers in combustion: from basic theory to practical devices”, in *Symposium (international) on combustion*, Vol. 27 (Elsevier, Amsterdam, 1998), pp. 1–41.

[40] B. Williams, M. Edwards, R. Stone, et al., “High precision in-cylinder gas thermometry using Laser Induced Gratings: Quantitative measurement of evaporative cooling with gasoline/alcohol blends in a GDI optical engine”, Combustion and Flame **161**, 270–279 (2014).

[41] B. Williams, M. Edwards, R. Stone, et al., “In-Cylinder Temperature Measurements Using Laser Induced Grating Spectroscopy and Two-Colour PLIF”, SAE International Journal of Engines **10**, 2191–2201 (2017).

[42] C. Willman, L. Shen, J. Camm, et al., "Cycle-to-Cycle Variation Analysis of Two-Colour PLIF Temperature Measurements Calibrated with Laser Induced Grating Spectroscopy in a Firing GDI Engine", SAE Int. J. Adv. & Curr. Prac. in Mobility **1**, 1404–1419 (2019).

[43] M. Raffel, C. E. Willert, S. T. Wereley, et al., *Particle image velocimetry: a practical guide* (Springer, Berlin, 2007).

[44] R. J. Adrian and J. Westerweel, *Particle image velocimetry* (Cambridge University Press, Cambridge, 2011).

[45] M. Bitter, S. Scharnowski, R. Hain, et al., "High-repetition-rate PIV investigations on a generic rocket model in sub- and supersonic flows", Experiments in Fluids **50**, 1019–1030 (2011).

[46] K. Kikuchi and O. Mochizuki, "High-repetition-rate PIV investigations on a generic rocket model in sub- and supersonic flows", Measurement Science and Technology **22**, 064002 (2011).

[47] T. Justham, S. Jarvis, A. Clarke, et al., "Simultaneous study of intake and in-cylinder IC engine flow fields to provide an insight into intake induced cyclic variations", Journal of Physics: Conference Series **45**, 146–153 (2006).

[48] H. Zhuang, D. L. S. Hung, J. Yang, et al., "Investigation of Swirl Ratio Impact on In-Cylinder Flow in an SIDI Optical Engine", Journal of Engineering for Gas Turbines and Power **138**, 081505 (2016).

[49] S. Buhl, F. Gleiss, M. Köhler, et al., "A Combined Numerical and Experimental Study of the 3D Tumble Structure and Piston Boundary Layer Development During the Intake Stroke of a Gasoline Engine", Flow, Turbulence and Combustion **98**, 579–600 (2017).

[50] F. Zhao, P. Ge, H. Zhuang, et al., "Analysis of Crank Angle-Resolved Vortex Characteristics Under High Swirl Condition in a Spark-Ignition Direct-Injection Engine", Journal of Engineering for Gas Turbines and Power **140** (2018).

[51] D. Butcher, A. Spencer, and R. Chen, "Influence of asymmetric valve strategy on large-scale and turbulent in-cylinder flows", International Journal of Engine Research **19**, 631–642 (2018).

[52] R. Stiehl, J. Bode, J. Schorr, et al., "Influence of intake geometry variations on in-cylinder flow and flow-spray interactions in a stratified direct-injection spark-ignition engine captured by time-resolved particle image velocimetry", International Journal of Engine Research **17**, 983–997 (2016).

[53] C. Jainski, L. Lu, A. Dreizler, et al., "High-speed micro particle image velocimetry studies of boundary-layer flows in a direct-injection engine", International Journal of Engine Research **14**, 247–259 (2013).

[54] J. Bode, J. Schorr, C. Krüger, et al., "Influence of the in-cylinder flow on cycle-to-cycle variations in lean combustion DISI engines measured by high-speed scanning-PIV", Proceedings of the Combustion Institute **37**, 4929–4936 (2019).

[55] B. Scott, "Flow and combustion in direct injection spark ignition engines", PhD thesis (University of Oxford, 2019).

[56] J. Westerweel, "Fundamentals of digital particle image velocimetry", Measurement Science and Technology **8**, 1379–1392 (1997).

[57] J. Westerweel, "On velocity gradients in PIV interrogation", Experiments in Fluids **44**, 831–842 (2008).

[58] J. L. Lumley, "The structure of inhomogeneous turbulent flows", in *Atmospheric turbulence and radio wave propagation*, edited by A. Yaglom and A. Tatarsky (Nauka, Moscow, 1967), pp. 166–178.

[59] B. P. Epps and E. M. Krivitzky, "Singular value decomposition of noisy data: noise filtering", Experiments in Fluids **60**, 1–23 (2019).

[60] F. Rulli, S. Fontanesi, A. d'Adamo, et al., "A critical review of flow field analysis methods involving proper orthogonal decomposition and quadruple proper orthogonal decomposition for internal combustion engines", International Journal of Engine Research **22**, 222–242 (2021).

[61] P. S. Abraham, K. Liu, D. C. Haworth, et al., "Evaluating large-eddy simulation (LES) and high-speed particle image velocimetry (PIV) with phase-invariant proper orthogonal decomposition (POD)", Oil & Gas Science and Technology - Rev. IFP Energies nouvelles **69**, 41–59 (2014).

[62] M. Voisine, L. Thomas, J. Borée, et al., "Spatio-temporal structure and cycle to cycle variations of an in-cylinder tumbling flow", Experiments in Fluids **50**, 1393–1407 (2011).

[63] M. Fogleman, J. L. Lumley, D. Rempfer, et al., "Application of the proper orthogonal decomposition to datasets of internal combustion engine flows", Journal of Turbulence **5**, N23 (2004).

[64] L. Shen, K.-Y. Teh, P. Ge, et al., "Temporal Evolution Analysis of In-Cylinder Flow by Means of Proper Orthogonal Decomposition", International Journal of Engine Research **22**, 1714–1730 (2020).

[65] S. Roudnitzky, P. Druault, and P. Guibert, "Proper orthogonal decomposition of in-cylinder engine flow into mean component, coherent structures and random gaussian fluctuations", Journal of Turbulence **7**, N70 (2006).

[66] W. Qin, M. Xie, M. Jia, et al., "Large eddy simulation and proper orthogonal decomposition analysis of turbulent flows in a direct injection spark ignition engine: Cyclic variation and effect of valve lift", Science China Technological Sciences **57**, 489–504 (2014).

[67] H. Zhuang and D. L.S. Hung, "Characterization of the effect of intake air swirl motion on time-resolved in-cylinder flow field using quadruple proper orthogonal decomposition", Energy Conversion and Management **108**, 366–376 (2016).

[68] B. P. Epps and A. H. Techet, "An error threshold criterion for singular value decomposition modes extracted from PIV data", Experiments in Fluids **48**, 355–367 (2010).

[69] X. Ma, G. E. Karniadakis, H. Park, et al., "DPIV-driven flow simulation: a new computational paradigm", Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences **459**, 547–565 (2003).

[70] S. Bradde and W. Bialek, "PCA meets RG", Journal of Statistical Physics **167**, 462–475 (2017).

[71] K. Liu, D. Haworth, X. Yang, et al., "Large-eddy Simulation of Motored Flow in a Two-valve Piston Engine: POD Analysis and Cycle-to-cycle Variations", Flow, Turbulence and Combustion **91**, 373–403 (2013).

[72] S. Buhl, F. Hartmann, and C. Hasse, "Identification of large-scale structure fluctuations in engines using POD-based conditional averaging", Oil & Gas Science and Technology - Rev. IFP Energies nouvelles **71**, 1 (2016).

[73] D. Butcher and A. Spencer, "Cross-correlation of POD spatial modes for the separation of stochastic turbulence and coherent structures", Fluids **4**, 134 (2019).

[74] B. P. Epps and E. M. Krivitzky, "Singular value decomposition of noisy data: mode corruption", Experiments in Fluids **60**, 1–30 (2019).

[75] J. Shlens, "A tutorial on principal component analysis", ArXiv preprint **1404.1100** (2014).

[76] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear Component Analysis as a Kernel Eigenvalue Problem", Neural Computation **10**, 1299–1319 (1998).

[77] S. Mika, B. Schölkopf, A. Smola, et al., "Kernel PCA and de-noising in feature spaces", in Advances in neural information processing systems 11 - proceedings of the 1998 conference, nips 1998, Advances in Neural Information Processing Systems (1999), pp. 536–542.

[78] H. Hoffmann, "Kernel PCA for novelty detection", Pattern Recognition **40**, 863–874 (2007).

[79] C. M. Bishop, *Pattern recognition and machine learning (information science and statistics)* (Springer-Verlag, Berlin, Heidelberg, 2006).

[80] H. Mirgolbabaei and T. Echekki, "Nonlinear reduction of combustion composition space with kernel principal component analysis", Combustion and Flame **161**, 118–126 (2014).

[81] M. R. Malik, P. Obando Vega, A. Coussement, et al., "Combustion modeling using Principal Component Analysis: A posteriori validation on Sandia flames D, E and F", Proceedings of the Combustion Institute **38**, 2635–2643 (2021).

[82] P. A. Franken, A. E. Hill, C. W. Peters, et al., "Generation of optical harmonics", Physical Review Letters **7**, 118–119 (1961).

[83] M. P. Wernet, "Particle displacement tracking applied to air flows", in Laser anemometry-advances and applications, Vol. 1, edited by A. Dybbs and B. Ghorashi (Aug. 1991), pp. 327–336.

[84] B. Lecordier, M. Mouqallid, S. Vottier, et al., "CCD recording method for cross-correlation PIV development in unstationary high speed flow", Experiments in Fluids **17**, 205–208 (1994).

[85] LaVision GmbH, *DaVis 8.4 Software Product Manual* (LaVision, Göttingen, 2016).

[86] V. Yakhot and S. A. Orszag, “Renormalization group analysis of turbulence. I. Basic theory”, Journal of Scientific Computing **1**, 3–51 (1986).

[87] S. B. Pope, *Turbulent flows* (Cambridge University Press, Cambridge, 2000).

[88] A. Chatterjee, “An introduction to the proper orthogonal decomposition”, Flow, Turbulence and Combustion **78**, 808–817 (2000).

[89] Y. C. Liang, H. P. Lee, S. P. Lim, et al., “Proper orthogonal decomposition and its applications - Part I: Theory”, Journal of Sound and Vibration **252**, 527–544 (2002).

[90] H. Chen, D. L. Reuss, D. L. S. Hung, et al., “A practical guide for using proper orthogonal decomposition in engine research”, International Journal of Engine Research **14**, 307–319 (2012).

[91] L. Sirovich, “Turbulence and the dynamics of coherent structures. I. Coherent structures”, Quarterly of Applied Mathematics **45**, 561–571 (1987).

[92] H. Chen, D. L. Reuss, and V. Sick, “On the use and interpretation of proper orthogonal decomposition of in-cylinder engine flows”, Measurement Science and Technology **23** (2012).

[93] A. García-González, A. Huerta, S. Zlotnik, et al., “A kernel Principal Component Analysis (kPCA) digest with a new backward mapping (pre-image reconstruction) strategy”, ArXiv preprint **2001.01958** (2020).

[94] B. Schölkopf, A. Smola, and K. R. Müller, “Kernel principal component analysis”, Lecture Notes in Computer Science **1327**, 583–588 (1997).

[95] P. Arias, G. Randall, and G. Sapiro, “Connecting the out-of-sample and pre-image problems in Kernel methods”, in Proceedings of the ieee computer society conference on computer vision and pattern recognition, CVPR.2007.383038 (2007).

[96] N. Thorstensen, F. Segonne, and R. Keriven, "Normalization and preimage problem in gaussian kernel PCA", Proceedings - International Conference on Image Processing, ICIP, 741–744 (2008).

[97] K. Liu and D. C. Haworth, "Development and assessment of POD for analysis of turbulent flow in piston engines", SAE Technical Paper, 2011-01–0830 (2011).

[98] H. Chen, H. Zhuang, D. L. Reuss, et al., "Influence of Early and Late Fuel Injection on Air Flow Structure and Kinetic Energy in an Optical SIDI Engine", SAE Technical Paper, 2018-01–0205 (2018).

[99] B. Scott, C. Willman, R. Stone, et al., "Novel metrics for validation of PIV and CFD in IC engines", SAE Technical Paper, 2019-01–0716 (2019).

[100] C. Willman, B. Scott, R. Stone, et al., "Quantitative metrics for comparison of in-cylinder velocity fields using particle image velocimetry", Experiments in Fluids **61**, 1–16 (2020).

[101] C. Arcoumanis, Z. Hu, C. Vafidis, et al., "Tumbling motion: a mechanism for turbulence enhancement in spark-ignition engines", SAE Technical Paper, 900060 (1990).

[102] F. Hartmann, S. Buhl, F. Gleiss, et al., "Spatially Resolved Experimental and Numerical Investigation of the Flow through the Intake Port of an Internal Combustion Engine", Oil & Gas Science and Technology - Rev. IFP Energies nouvelles **71** (2016).

[103] P. Ge and D. L. S. Hung, "Impingement-Induced Stability Analysis of Intake Manifold Air Jets Using High Speed PIV and POD", International Journal of Automotive Technology **22**, 403–414 (2021).

[104] J. L. Devore, *Probability and Statistics for Engineering and the Sciences*, 8th ed. (Brooks/Cole, Boston, 2010).

[105] B. Enaux, V. Granet, O. Vermorel, et al., “Large eddy simulation of a motored single-cylinder piston engine: numerical strategies and validation”, Flow, Turbulence and Combustion **86**, 153–177 (2011).

[106] J. Devore, *Probability and Statistics for Engineering and the Sciences*, 9th ed. (Cengage Learning, 2015).

[107] L. Shen, K.-Y. Teh, P. Ge, et al., “Detecting Outliers in Crank Angle Resolved Engine Flow Field Datasets for Proper Orthogonal Decomposition Analysis”, SAE Technical Paper, 2017-01–0612 (2017).

Appendix A

Single cylinder optically accessible engine and its components

The data presented in this thesis is obtained on a single cylinder optical engine designed by Jaguar Land Rover Limited (JLR), and the combustion system is based on its Ingenium petrol engine family (Engine ID: AJ200). The differences between the optical and the baseline thermal engines are kept to a minimum for better examining the "real-world" engine behaviour, yet necessary modifications have to be made in order to allow sufficient optical access. For instance, the bore is increased from 83 mm in the thermal engine to 85 mm in the optical engine, to compensate for the extra space occupied by the optical components. Table A.1 lists the specifications of this optical engine. The author of this thesis performed daily testing and maintenance work of the engine, but was not involved in building up the engine.[(1)] The optical system alignment of the PIV experiments based on the setup discussed in Section 2.2 and the PIV tests were conducted by the author of this thesis. The following paragraphs provides a

[(1)] The engine was built before the author of this thesis joined in the research group. Prof. Richard Stone was in charge of assembling the engine, with help from Dr. Christopher Willman and Dr. Blane Scott. Their efforts are much appreciated.

Table A.1: Optical engine specifications

Parameters	Descriptions
Maximum Engine Speed	2000 rpm
Bore	85.0 mm
Stroke	90.3 mm
Displacement Volume	512 cc
Compression Ratio[a]	12.5 : 1

[a] Designed value when using production (shaped, non-flat) piston. In the optical engine, the actual compression ratio is usually smaller due to sealing and blow-by losses.

brief description of the engine setup, and for a detailed discussion, the readers may refer to Scott's thesis [1].

Figure A.1 shows a schematic of the engine test cell. A direct-current dynamometer with a Control Techniques Four-Quadrant controller is used to provide speed or torque control to the engine. The engine phase (crank angle) is recorded by a crank encoder mounted on the crankshaft, which by default provides 360 pulses per resolution (i.e., one flag at every crank angle degree, hereinafter denoted CAD). The signals are sent to both the Engine Timing Control System (ETCS) and the Data Acquisition (DAQ) units for synchronization. The ETCS, based on an in-house LabView program [2], controls the fuel injection and ignition timings, and sends transistor-transistor logic (TTL) pulses to a programmable timing unit (PTU, from LaVision), which then triggers the laser and/or camera as needed. The engine performance is logged by two National Instruments (NI) data acquisition cards with another customized LabView program [2]. The high-rate NI 6070E card has a maximum sample rate of 1 MS/s, and records high-frequency data such as intake valve lift and in-cylinder pressure at every crank angle degree. The low-frequency parameters are logged by NI PCI-6024E (maximum sample rate: 2.6 kS/s) for every cycle.

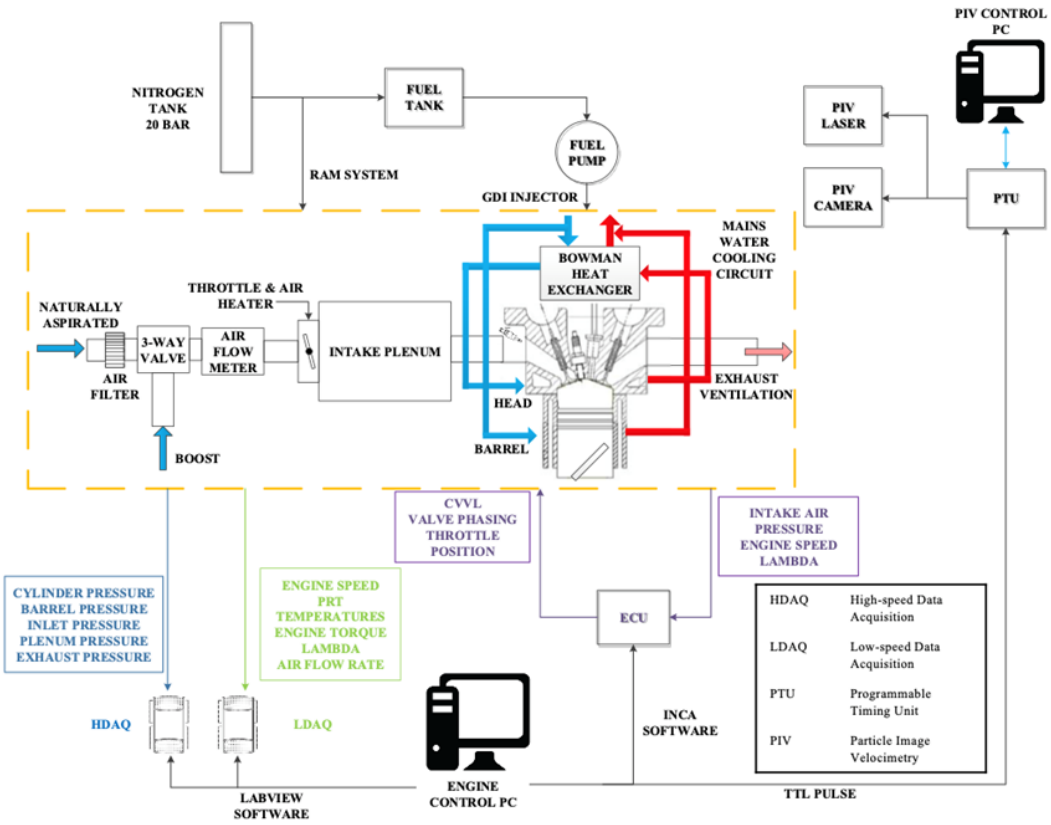


Figure A.1: A schematic of the single-cylinder optically accessible engine test cell. This figure was kindly provided by Dr. Blane Scott [1].

The engine setup is illustrated in Figure A.2 with two views: a 45° view from the exhaust-belt-side (left) and a side view from the belt (right). The cylinder liner is separated into two parts: the upper part is made in quartz (denoted "annulus") and is transparent to ultraviolet and visible light, providing optical access from the side of the engine; the lower liner is in metal (denoted "barrel"), which allows cooling water to flow around and thus, by keeping the piston rings cool, reduce frictional heating between the moving piston and the inner surface of the liner. The metal liner also enables the pressure measurement inside the barrel with a pressure transducer (Kistler 4075A5).

The cylinder head (Figure A.3) has two intake and two exhaust valves. A continuously

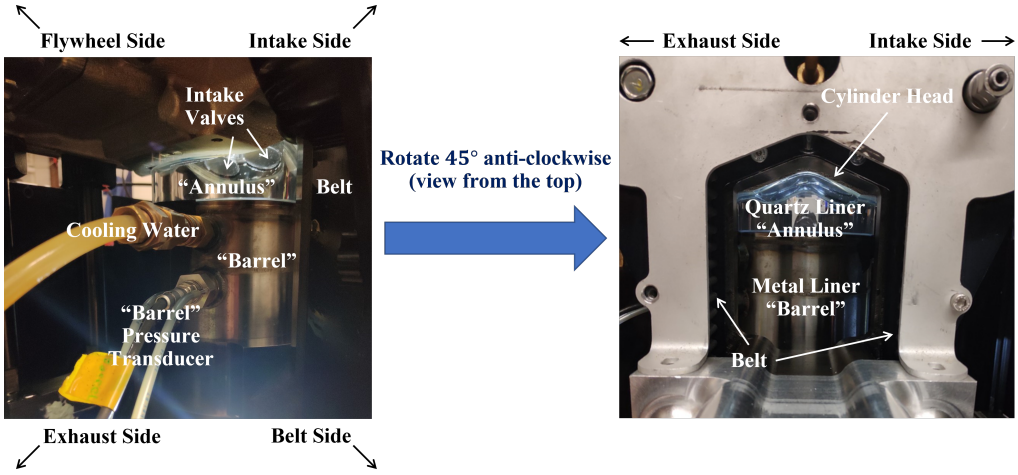


Figure A.2: A schematic of the single-cylinder optically accessible engine. Left: 45° view from the exhaust-belt-side; Right: side view from the belt-side.

variable valve lift (CVVL) system installed between the intake camshaft and valves enables flexibility in intake valve lifts by retarding valve opening and/or preceding valve closing. The real-time intake valve lift is measured using a Sensitec valve train-measurement system (VTMS.DSE.02) through pre-machined grooves on the inlet valve stems. The exhaust valves are connected to the exhaust camshaft by conventional roller finger followers (RFF) system, and exhaust maximum opening point (EMOP) can therefore be controlled. The exhaust valve lift is calculated by the known valve profile for a given EMOP. Pressure transducers (Kistler 4075A10) record the pressures inside the inlet and exhaust runners.

The fuel is pumped by an external system up to 300 bar in pressure (measured by Kistler 4065B0500), and the injector is vertically-mounted on the pent roof (Figure A.3). The spark plug is placed near the exhaust side. A pressure transducer (Kistler 6043A60) is also installed on the pent roof to measure real-time in-cylinder pressure. Fire decks (marked by the white brick-board shade) are placed between each intake or exhaust valves, and served as zero points along the stroke direction. The piston top dead centre (TDC) is about 1 mm below the fire decks, and the actual distance depends on the type of piston used (see Figure A.5).

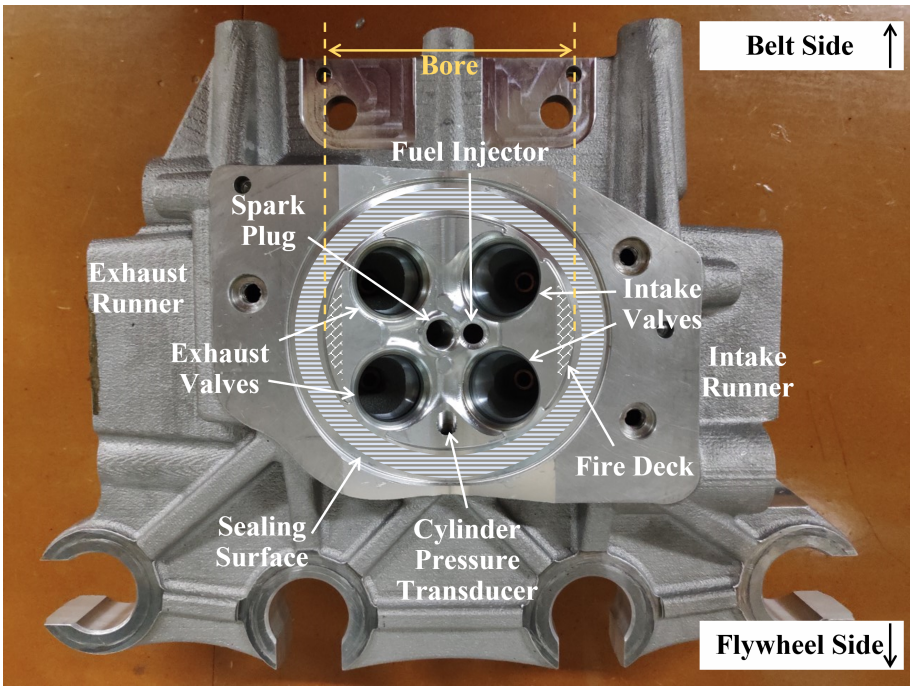


Figure A.3: Bottom view of the cylinder head. Valves and other instruments are not installed. The white striped shade marks the sealing surface for the annulus (cylinder liner), and the white brick-board shade marks the location of the fire decks.

The centre of the cylinder head is surrounded by a curved surface (marked by the white striped shade in Figure A.3). The quartz annulus (right of Figure A.4) with corresponding curvature on its upper surface, which then forms a tight seal between the engine head and the two-component cylinder liner, and maximizes the optical access near the head. In actual practise, the quartz annulus was found to be quite fragile, and thus a Perspex annulus (left of Figure A.4) was also used as an alternative when the in-cylinder temperature is low enough. PTFE[(2)] gaskets (Garlock GYLON® 3504) are compressed between the components to provide an enhanced sealing performance.

The lower part of the engine follows a conventional Bowditch [3] optical engine design (left-side of Figure A.5). A shaped piston with quartz window insert is screwed to the slotted tubular

[(2)] PTFE stands for polytetrafluoroethylene.

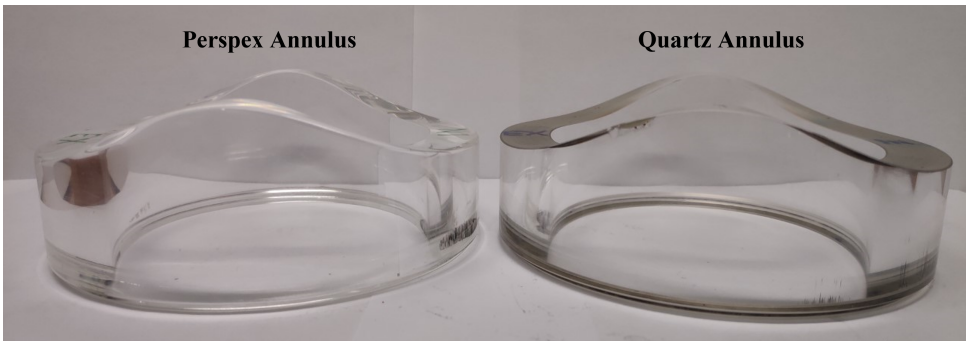


Figure A.4: Annulus with curved top surface. The curvatures of the top surfaces match with the sealing surface of the cylinder head. Left: made in Perspex; Right: made in quartz.

extension, and the combination is bolted down to the bottom end piston to which the crankshaft is linked. Two spacers are used to ensure the correct spacing and alignment with the pent roof, respectively. A $45°$ mirror is placed beneath the piston window, which gives the bottom view of the cylinder head despite the upper piston moving. A metal retaining ring is placed under the piston window to hold the window in place.

Pistons with shaped crowns are provided by JLR for testing engine behaviour under different compression ratios (top-right of Figure A.5). Piston C has the same crown as the production piston, and can give a $12.5:1$ compression ratio, while Piston B has a lower compression ratio ($11:1$) since its crown is flatter (and thus the clearance volume is larger). The shaped crown limits the diameter of the piston window, and thus reduces the field of view, to 46 mm compared to using a flat piston (60 mm), yet preserves the similarity to the production thermal engine to the greatest extent possible. Such a compromise is considered to be acceptable as the purpose of this experiment is to examine the in-cylinder flow field of a production engine.

Despite the similarities to the production piston, the optical piston used in our experiment still has several necessary modifications, chief among them is that the piston rings are positioned much lower (see low-ring versus high-ring pack Piston C in Figure A.5). The rings are lowered by

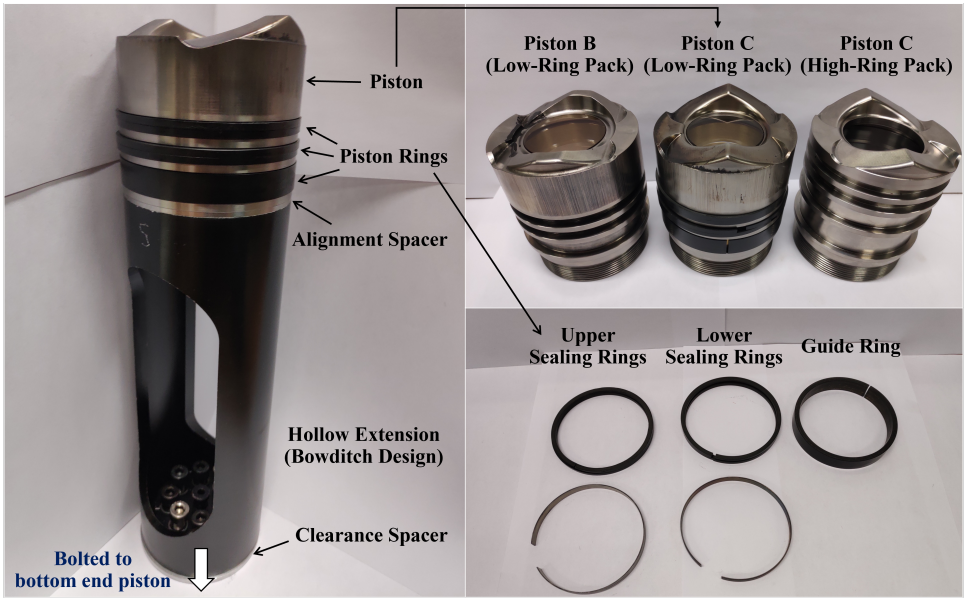


Figure A.5: Optically accessible engine piston assembly. Left: Bowditch design with hollow extension rod and low-ring pack piston C; Top-Right: pistons with shaped crown; Bottom-Right: piston rings.

the height of the quartz annulus to avoid direct contact between the piston rings and the quartz annulus, and thus protect the annulus from frictional heating and scratching. However, it also introduces extra volume in the cylinder and reduces the effective compression ratio. Additionally, the piston rings (bottom-right of Figure A.5 are made of polyamide-imides (Torlon) to replace the metal ones in thermal engines since no lubrication is supplied. Three sets of piston rings are used: the upper two sets (each with two Torlon rings and a metal spring ring) provide sealing between the piston and the metal barrel, while the lowest one guides the piston translation motion. The quartz window insert also has a different heat transfer coefficient compared to the metal piston, resulting in deviations in the near-surface temperature. The bulk flow motion is of interest in this report is assumed not to be significantly affected by these changes, and in any case, the simulation model which the PIV data validates (detailed in Section 2.5) is based on the optical engine setup instead of the thermal one.

References for Appendices

[1] B. Scott, “Flow and combustion in direct injection spark ignition engines”, PhD thesis (University of Oxford, 2019).

[2] X. Wang, “Instantaneous in-cylinder heat transfer and combustion analysis in spark ignition engines”, PhD thesis (University of Oxford, 2008).

[3] F. W. Bowditch, “A new tool for combustion research a quartz piston engine”, SAE Technical Paper, 610002 (1961).