On the use and interpretation of proper orthogonal decomposition of in-cylinder engine flows

2012 Meas. Sci. Technol. 23 085302

(http://iopscience.iop.org/0957-0233/23/8/085302)

Meas. Sci. Technol. **23** (2012) 085302 (14pp) doi:10.1088/0957-0233/23/8/085302

On the use and interpretation of proper orthogonal decomposition of in-cylinder engine flows

Hao Chen[1], **David L Reuss**[2,3] **and Volker Sick**[1,2]

[1] School of Mechanical Engineering, Shanghai Jiao Tong University, National Engineering Laboratory for Automotive Electronic Control Technology, Shanghai, People's Republic of China
[2] Mechanical Engineering Department, The University of Michigan, Ann Arbor, MI, USA

E-mail: dreuss@umich.edu

Received 13 December 2011, in final form 29 March 2012
Published 28 May 2012
Online at stacks.iop.org/MST/23/085302

Abstract
The proper orthogonal decomposition (POD) has found increasing application for the comparison of measured and computed data as well as the identification of instantaneous and time varying flow structures, particularly cyclic variability in reciprocating internal combustion engines. The patterns observed in the basis functions or modes are sometimes interpreted as coherent structures, though justification of this is not obvious from the mathematical derivations. Similarly, there is no consensus about whether or not the ensemble mean should be subtracted prior to performing POD on a data set. Synthetic flow fields are used here to reveal POD properties otherwise ambiguous in real stochastic flow data. In particular, each POD mode includes elements of all flow structures from all input snapshots and in general, several modes are needed to reconstruct physical flow structures. POD analysis of two experimental in-cylinder engine data is done: one flow condition where every cycle resembles the ensemble-averaged flow pattern, and the other with large cyclic variability such that no cycles resemble the ensemble average. The energy and flow patterns of the POD modes, derived with and without first subtracting the mean, are compared to each other and to the Reynolds decomposed flow to reveal properties of the POD modes.

Keywords: proper orthogonal decomposition, flow analysis, turbulence, engines, cyclic variability, particle image velocimetry

(Some figures may appear in colour only in the online journal)

Introduction

In-cylinder turbulent flow is an important mechanism controlling heat-transfer, fuel–air mixing and rate of combustion in reciprocating internal combustion engines [1]. In-cylinder turbulence is created by the shear flow through the intake valves and large-scale structures formed during the intake stroke that break down and dissipate by the decreasing volume (boundaries) during the compression stroke. Until the availability of high-speed imaging capability in recent years, imaging measurements usually employed phase-dependent sampling, where the velocity from many cycles is ensemble-averaged at fixed crank-angle positions when using particle image velocimetry. Time-dependent studies had been restricted to point measurements using hot-wires or laser Doppler velocimetry. Reynolds-averaged Navier–Stokes, RANS, modeling similarly employs the concept of the average (expected value) and turbulence, which is defined as the fluctuation about the average. In most canonical turbulent flows that were studied, the turbulence is a small perturbation about the mean; thus, the mean shear from the RANS decomposition is a good predictor of the turbulence, which is used to predict the kinetic energy dissipation. In engines the ensemble average may never be present in an individual cycle [2] and

[3] Author to whom any correspondence should be addressed.

the statistical fluctuations about the average are conceptually confounded by cyclic variability of large-scale high-energy coherent structures. Thus the statistical fluctuations may or may not be a good metric for modeling turbulence dissipation. In recent years, the proper orthogonal decomposition (POD) has gained popularity as an analysis tool for evaluating the turbulence of in-cylinder reciprocating engine flows as it provides hope for resolving the discourse between turbulence and cyclic variability. Lumley [14] is generally attributed as being the first to suggest POD for application to atmospheric turbulence as a method to separate 'big eddies' in shear flows, thought to

> '. . . have scales on the order of the average flow, the remainder of the turbulent motion has scales substantially smaller, so that the net effect of this smaller component may be lumped together as an eddy viscosity with considerably more propriety than usual.'

Conceptually, this is applicable to in-cylinder engine flows, where large-scale, energetic structures associated with cycle-to-cycle variability require an objective means to be separated from the dissipating portions of the flow. POD has been used in engine flow studies for about 10 years for a range of topics as the following examples illustrate. Baby *et al* [3] utilized POD to analyze PIV data taken in an optical engine to show that POD can separate 'true turbulence' (in fact small scale turbulent fluctuations) from cycle-to-cycle variations. Druault *et al* [4] employed POD to determine flow information between two consecutive PIV flow images. Roudnitzky *et al* [5] implemented the POD technique to PIV measurements obtained in the tumble plane of spark ignition engine flow. The in-cylinder flow was decomposed into an average part, a coherent part and a random Gaussian fluctuations part. Fogleman *et al* [6] introduced a novel approach called phase-invariant POD on both computational fluid dynamics (CFD) and PIV data. The phase-invariant POD modes were desirable to provide a suitable basis for low-dimensional models, which would describe the breakdown process of tumble. Kapitza *et al* [7] utilized POD to investigate the role of the intake port flow on in-cylinder flow fluctuations. Vosine *et al* [8] also used POD to investigate in-cylinder flow structures and their cycle-to-cycle fluctuations. POD has also been used as an objective means for comparing PIV results and data from large eddy simulations [9, 10].

The interpretation and application of POD can be perplexing. The patterns observed in the modes are sometimes interpreted as extraction of coherent structures, though justification of this is not obvious from the mathematical derivations. For low-order estimates of the flows, Holmes *et al* [11] suggest that the number of low-order modes should capture 90% of the total energy and neglect no modes with more than 1% of the total energy; this is a pragmatic truncation of the spectrum but nonetheless an ad hoc definition. Often discussed, but by no means obvious, is whether POD should be performed with or without first subtracting the average; as noted by Chatterjee [12] this does not affect the calculation, but only the interpretation. It is commonly stated that the first mode is equivalent to the ensemble mean.

Furthermore, it is a common practice to quantify the equivalency of POD results of different sample sets and low-order estimates of different velocity distributions using the energy spectra and cumulative energy distributions. However, equivalency of energy spectra is of reduced value if the basis functions ('flow' patterns) are not equivalent. The relevance index introduced by Liu and Haworth [9] is better suited to quantify equivalency of flow patterns through a correlation of the projection of one flow pattern onto another. Then, PDFs of the POD coefficients can be employed to reveal the cycle-to-cycle variability of modes.

Here, as also used by Kapitza *et al* [7], POD is first applied to synthetic velocity snapshots to answer questions about fundamental properties of POD not discernable from the complexity of real turbulent flows. The intent is to gain physical understanding by inputting velocity fields with known characteristics and then observing the characteristics of the resulting POD. The POD is then applied to two engine data sets, which represent two extremes of in-cylinder flow [2]. One extreme is the undirected flow using an undirected port and standard valve, where the ensemble-average flow structure is never observed in individual cycles. The other is a directed flow produced by a shrouded valve, which creates a highly repeatable high-swirl flow and the ensemble-average flow pattern is apparent and dominant in every cycle. In production-engine design, a directed flow is formed by either directed ports (as in 4-valve pent-roof heads and Diesel engine ports) or by valve shrouding (the upstream side of the valve or in the head-casting around the valve) in an attempt to create a repeatable combination of swirl and/or tumble flow. Though these two flows here are not flows identical to those in modern engines, they do represent two extremes likely to occur in an engine, namely an ensemble-average flow that is either highly directed versus undirected by the intake ports or valves.

The purpose of this study is to show (1) what physical properties of the coherent structures are captured by the POD, (2) how these properties are captured and distributed among the modes and (3) provide the relationship between the POD of the original velocity, V, the POD after subtracting the ensemble average, V–$\langle V \rangle$ and the RANS decomposition.

1. POD definitions

The POD technique decomposes the original velocity or scalar field into a sum of weighted, linear, basis functions or modes. It is considered an empirical basis because the functions are computed from the structures in the original data fields rather than prescribing them *a priori* as in Fourier decomposition, which is an alternative linear basis. The creation of the basis functions is a statistical correlation method that forces the basis functions to be normalized and orthogonal. It is natural to apply the POD to velocity, since the optimization of the basis in L^2 space separates and ranks the modes according to energy [11]. Applied to scalar fields, the square of the intensity has no physical meaning, but the basis functions and coefficients are created just the same. As with velocity distributions, the scalar POD is useful for extracting pattern features from scalar intensity [13]. The POD can be carried out using either the

classical method [14] or the equivalent method of snapshots [15]. The method of snapshots is used here, as in most engine applications, because there are usually far more measured data (measurement or computational grid points) compared to the sample size, and thus the method of snapshots is far more computationally efficient [16]. Here, only the mathematical descriptions necessary to define the nomenclature are repeated as comprehensive descriptions are available in the literature [11, 12, 16].

The input data for the POD consist of K two-dimensional velocity fields (snapshots) sampled in time, $V^{(k)} = (u_{i,j}, v_{i,j})^{(k)}$. Here, i, j are the indices of the grid points in the PIV measurement plane and k is the index of the velocity field. In engine flows the samples can be phase invariant where the samples are collected continuously in time for many cycles, or phase dependent where samples are collected at the same crank angle for many cycles [9]. For this study, only the phase-dependent sampling is considered; thus, $V^{(k)}$ is the velocity field at a given crank angle from the kth cycle. The properties described in the results here are also valid for scalar fields, three-dimensional and phase-invariant samples of engine flows.

The POD produces a linear basis set consisting of M basis functions φ_m and the corresponding coefficients $c_m^{(k)}$, that can reconstruct all K velocity distributions,

$$V^{(k)} = \sum_{m=1}^{M} c_m^{(k)} \varphi_m, \tag{1}$$

where m is the mode index, with the total number of modes equal to the total number of snapshots, $M = K$. The procedures of determining the orthonormal POD basis functions φ_m are detailed in [11, 12, 16]. In this study, the basis functions were created with a Matlab code; comparison with a commercial code (LaVision Davis 7.2) and the code of Liu and Haworth [9] demonstrated computational speed within 2% and accuracy to 12 decimal places. The code minimizes the following function:

$$\sum_{k=1}^{K} \left\| V^{(k)} - \sum_{m=1}^{M} c_m^{(k)} \phi_m \right\|^2 \to \min \tag{2}$$

subject to

$$(\varphi_i, \varphi_j) = \delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases},$$

where $\|\|$ denotes the L^2 norm. The basis functions contain the 'flow patterns', normalized in L^2 space so that the sum of the squares of all the vectors in an individual basis function φ_m is unity [11]:

$$\sum_{i=1}^{I} \sum_{j=1}^{\mathrm{J}} \left(\mu_{i,j}^2 + v_{i,j}^2\right) = 1, \tag{3}$$

where μ and v are the x and y components of φ, respectively. Furthermore, each basis function is orthogonal to all others, i.e.

$$\int \varphi_m(x) \varphi_p(x)\, \mathrm{d}x = \delta_{m,p}. \tag{4}$$

These two properties define the basis functions as orthonormal.

The $K \times M$ coefficient matrix $c_m^{(k)}$,

$$c_m^k = \begin{bmatrix} c_1^1 & c_2^1 & \dots & c_M^1 \\ c_1^2 & c_2^2 & \dots & c_M^2 \\ \dots & \dots & \dots & \dots \\ c_1^K & c_2^K & \dots & c_M^K \end{bmatrix}, \tag{5}$$

contains the amplitude that the corresponding basis function contributes to a particular snapshot. The coefficients are computed by projecting the original K velocity fields onto the M computed basis functions. Thus, in this study the velocity field from the kth engine cycle can be reconstructed[4] by summing all M modes multiplied by their respective coefficient for that cycle using equation (1). Since φ_m is normalized and $c_m^{(k)}$ is the amplitude, $\frac{1}{2}\left(c_m^{(k)}\right)^2$ quantifies the kinetic energy per unit mass the mth mode contributes in the kth velocity field.

The POD mode *spectrum* is often used to describe the energy fraction each mode captures of the total ensemble of snapshots. The mass specific kinetic energy from all of the cycles captured by the mth mode is

$$\mathrm{KE}_m = \frac{1}{2} \sum_{k=1}^{K} \left(c_m^{(k)}\right)^2 = \frac{1}{2} K \cdot \lambda_m. \tag{6}$$

Thus, equation (6) relates the coefficients to the eigenvalue [11]. The *energy fraction* of the mth mode is

$$ke_m = \mathrm{KE}_m / \mathrm{KE}_{\mathrm{total}}, \tag{7}$$

where

$$\mathrm{KE}_{\mathrm{total}} = \sum_{m=1}^{M} \mathrm{KE}_m. \tag{8}$$

Thus, $\mathrm{KE}_{\mathrm{total}}$ is the total kinetic energy summed over all K velocity distributions. The mode order of the *POD spectrum* is defined as $m = 1 \to M$ according to their energy fraction, mode 1 having the largest energy fraction and $\mathrm{KE}_m > \mathrm{KE}_{m+1}$ for all m. This spectrum defines the concept of 'low-order reconstruction' and 'low-order modeling' described in the literature. In particular, most of the energy in the flow often can be captured from the first few modes. Holmes *et al* suggest that a lower order estimate contains at least 90% of the total energy and neglect no modes with more than 1% of the total energy [11]. Low-order reconstruction is consistent with Lumley's original aspiration quoted in the introduction, where it is desired to decompose the flow into large energetic coherent structures and smaller turbulence motion associated with smaller scales.

As noted in the introduction, here POD is applied prior to or after subtracting the average velocity. If the POD is applied without subtracting the average, the reconstructed ensemble-averaged velocity field is related to the POD through the cycle-averaged coefficients and the POD modes using

$$\langle V_{M'} \rangle = \sum_{m=1}^{M'} \left[\frac{1}{K} \sum_{k=1}^{K} c_m^{(k)} \right] \varphi_m = \sum_{m=1}^{M'} \langle c_m \rangle^{(K)} \varphi_m, M' = 1 \to M, \tag{9}$$

where the operator $\langle c_m \rangle^{(K)}$ indicates the ensemble average of c over all cycles (snapshots) for a given mode. Equation (9)

[4] Reconstruction will be used throughout this paper to indicate the computation of the velocity distribution as per equation (1), for one or more modes, m, and for one or all snapshots, k, as indicated for each instance.

demonstrates that a low-order estimate is achieved with a truncated set of M' modes and the exact ensemble average is achieved when all modes are used and $M' = M$. The quantitative implications of this will be demonstrated in the results section.

In the analysis that follows, it is useful to quantify the degree to which the flow patterns of either two velocity fields or two basis functions are similar or dissimilar. This is important because there is less value comparing energy of two POD modes if the POD modes are not the same. To this end the relevance index R_p will be used [9]. R_p is obtained by projecting one velocity field $V^{(1)}$ onto another velocity field $V^{(2)}$,

$$R_p = \frac{(V^{(1)}, V^{(2)})}{\|V^{(1)}\|\|V^{(2)}\|}. \tag{10}$$

The numerator denotes the inner product of two velocity fields, and $\|\cdot\|$ denotes the L^2 norm (for POD modes, equal to 1). The value of the relevance index varies from -1 to 1. $R_p = 1$ if the two velocity fields are identical, $R_p = -1$ if the two velocity fields are identical but exactly opposite in sign, and $R_p = 0$ if two POD modes are orthogonal. Using POD modes with or without multiplication by the POD coefficients (amplitude) gives the same R_p. Thus, R_p is an objective metric of the similarity of the patterns without regard to the energy content (magnitude), which is not always visually obvious in displayed vector structure.

2. Interpretation of POD of synthetic flow fields

Simple synthetic velocity fields are created here by superposition of individual 2D structures on a 2D grid. They are employed here to isolate five fundamental properties of the POD. The discussion is not unique to engine flows and also can be extended to POD of scalar/intensity fields. The properties of these synthetic velocity fields were conjured in response to questions raised while investigating POD of the real turbulence. The fundamental synthetic element was a vortex V_v. It is asymmetric in its spatial dimensions and contains shear (not solid body rotation). The maximum velocity within the vortex is 2 m s^{-1} and the total energy is 22 m^2 s^{-2}. Synthetic flows were then constructed by superimposing the number and magnitude of the vortices onto a background as described below. Of course, numerous synthetic flows could be conjured and for good reasons; the ones used here include features observed in the measured flows that are also discussed in this paper.

2.1. Example 1: POD mode capture of coherent structures

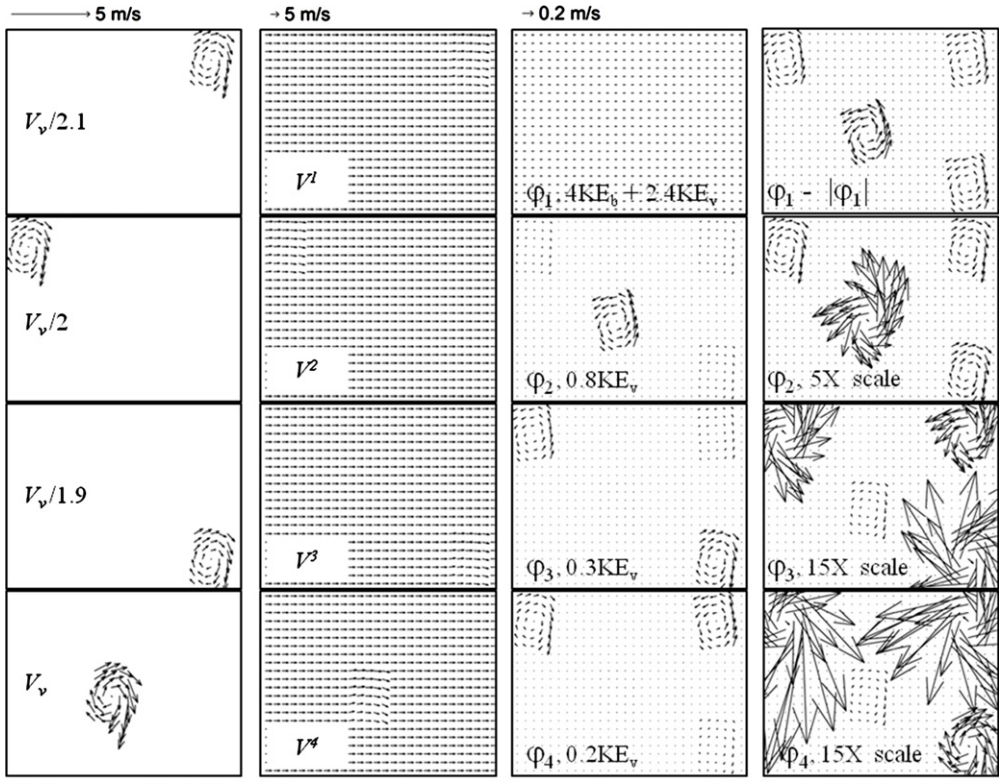
Figure 1 demonstrates how coherent structures are captured by and distributed throughout the modes. The four velocity fields were synthesized with five flow structures, four randomly placed vortices and a uniform background flow. Each of the vortices has identical structure V_v, but the magnitude of the vortex is different for each of the four fields as noted in figure 1(*a*). For this example, they are superimposed upon (added to) identical uniform velocity distributions of 4 m s^{-1}, $(u, v) = (4, 0)$ m s^{-1}, which have a total energy of 4928 m^2 s^{-2}, thus, creating the input velocity fields as shown in figure 1(*b*). The vortex magnitudes were chosen with a maximum velocity less than the mean flow. The magnitudes of the vortex velocities were scaled (as indicated in figure 1(*a*)) so that each has a unique energy content. The superimposed vortices in the four input velocity fields are difficult to detect, appearing only as blurry modulations. These properties were chosen to create idealized grid turbulence, where the temporal and spatial scales of the average are clearly separated from those of the coherent structures. The four basis functions $\varphi_{1\text{–}4}$ resulting from the POD analysis are shown in figures 1(*c*) and (*d*); note that the ensemble average has not been subtracted. In figure 1(*d*), the display perspectives have been changed to show that some elements of all four vortex structures are present in all four modes. The spatial average, $|\varphi_1|$, has been subtracted from φ_1 to bring out the spatial structures, and the display scale was changed for the other three modes. The coefficients are listed in figure 1(*e*), which shows the sign and magnitude by which each φ_n is scaled. This example demonstrates two important properties of the POD.

Property 1. The flow pattern of every structure from every velocity field (background and vortex) is contained in every mode. The modes are not themselves individual coherent structures as some times implied by verbal descriptions in the literature. This is demonstrated in figure 1(*d*).

Property 2. Given that every structure is present in every basis function, the velocity magnitudes in the basis functions (flow patterns) and the sign and magnitude of the coefficients are distributed such that the velocity structures in a given snapshot are reconstructed, while structures not present in a given snapshot are eliminated (have zero magnitude).

First, consider the uniform background that is in the positive x direction in the velocity field but in the negative x direction in φ_1. Thus, the coefficients $c_1^{(k=1-4)}$ are negative. Furthermore, the coefficients are large for all k, since the background flow has far more energy than the vortices and is identical for all k. Now consider reconstruction of $V^{(4)}$ using all four modes, $\varphi_{m,m=1\text{–}4}$ and coefficients $c^{(4)}_{m=1-4}$ according to equation (1) (figure 1(*e*), bottom row). Since $V^{(4)}$ contains the strongest vortex, most of its energy is captured in the amplitude of φ_1 and φ_2 and the coefficients $c_1^{(4)}$ and $c_2^{(4)}$; higher order modes contribute much less modulation of the vortex $V_v^{(4)}$. However, the sum of the contributions from all four modes times the respective coefficients must sum to zero for the other three vortices, $V_v^{(1\text{–}3)}$, since they do not exist in $V^{(4)}$. Furthermore, the magnitude of the uniform flow must be modulated between the modes, such that in each fully reconstructed velocity field $V^{(k)}$ it has the same uniform value. Similar insight can be gained by comparing the coefficients used for reconstructing the other input velocity fields, $V^{(1\text{–}3)}$.

This example also demonstrates an important principle about displaying the structure of velocity distributions. In particular, it is necessary to plot the distributions from a perspective that reveals the low-energy structures. The absence of the appearance of a coherent structure in a mode does not necessarily mean that the mode contains only random noise, but only that it may have a small magnitude (low energy) compared to other structures in that mode.



(e)

	m1	m2	m3	m4
k1	-99.36	1.64	-0.78	-2.53
k2	-99.37	1.71	-1.94	1.88
k3	-99.37	1.80	2.68	0.62
k4	-99.54	-5.15	0.04	0.03

Figure 1. Four synthetic velocity distributions. (*a*) Input vortices, (*b*) input velocity fields (vortex + uniform flow), (*c*) POD basis functions displayed at the same scale, (*d*) basis functions displayed with different perspectives and (*e*) the coefficient matrix corresponding to the four modes and four velocity fields. V_v has a maximum of 2 m s^{-1} and a KE = 22 m^2 s^{-2}. The uniform background velocity is 4 m s^{-1}, KE$_b$ = 4928 m^2 s^{-2}.

2.2. Example 2: POD of repeated identical structures

Figure 2(*a*) demonstrates how the energy of identically repeated structures is distributed among the modes. Using the same fundamental vortex element as example 1, five unique vortex elements were created from a single flow pattern with amplitudes scaled to create the energy variations noted in figure 2(*a*); the second vortex was repeated in 15 snapshots and the third vortex repeated in 5. These vortices were superimposed on 23 low-energy random background velocity fields used here only to provide quantitative energy differences between identically repeated vortices. $V^{(1)}$ contains the single most energetic structure, which occurs in only one snapshot. $V^{(2)}$–$V^{(16)}$ contain the weakest structures but these occur often (15 times) at the same location and with the same orientation. The magnitude and repeated occurrence of the other three unique distributions are given in figure 2(*a*); $V^{(17)}$–$V^{(21)}$ are stronger than $V^{(2)}$–$V^{(16)}$ but occur only five times and, finally, the vortices in $V^{(22)}$ and $V^{(23)}$ have identical structure but different signs and magnitudes. The first five of the 23 POD modes for example 2 are shown in figure 2(*b*). As in example 1, all five structures are contained in all 23 modes; for example, the vortex pattern $V^{(2–16)}$ is weakly discernable in φ_2 and φ_5. Since the coefficient matrix is large (23 × 23), the coefficients of the first four modes are provided graphically in figure 3(*a*). Comparison of the coefficient values in figure 3(*a*) with the basis functions and snapshots in figure 2 support assertions in property 2 that was discussed in section 2.1.

Property 3. Coherent structures with identically repeated position and orientation are placed in modes according to their ensemble total energy.

Example 2 reveals the following additional properties of the POD.

One strong structure can appear in a higher mode than weaker structures that occur often. The inverse is also true. This is demonstrated by the vortices in snapshots $V^{(1)}$, $V^{(2)}$–$V^{(16)}$ and $V^{(17)}$–$V^{(21)}$ appearing in the order of φ_2, φ_3 and φ_1, respectively, according the energy totaled across all snapshots. It is of course important to remember from example 1 that each of these vortex flow patterns are present in every basis function, but with different amplitudes.

Property 4. Two coherent structures with identical flow patterns that differ only by magnitude (energy) and sign are

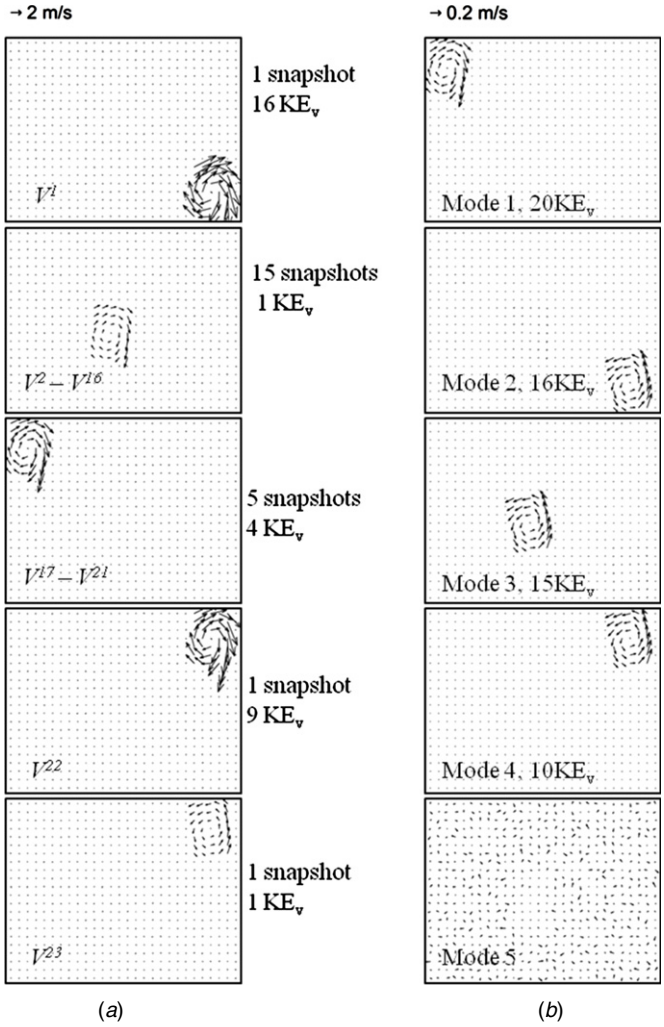


Figure 2. Twenty three synthetic velocity distributions with identical vortices superimposed on a random background. (*a*) The input velocity fields for POD analysis and (*b*) first five POD modes and with mode energy. V_v has a maximum of 2 m s^{-1} and a KE = 22 m^2 s^{-2}. The background velocity, u and v, is randomly distributed on the interval $[-0.1, 0.1]$ m s^{-1} with KE $\simeq$ 2 m^2 s^{-2}.

dominant in a single basis function; the snapshot-to-snapshot difference is captured by the corresponding coefficients.

This is demonstrated by $V^{(22)}$ and $V^{(23)}$, which have identical position and orientation (the appearance of the tilted orientation is a result of the display, where the tails of the vectors are at the grid nodes and the vectors differ in sign and magnitude). Both are dominant in φ_4 and, based on the coefficients $c_4^{(n)}$ shown in figure 3(*a*), φ_4 only contributes significantly in snapshots $k = 22$ and 23. This exemplifies the amplitude modulation of the basis function's flow pattern by the coefficient. Inspection of the POD coefficients from example 2 in figure 3(*a*) reinforces properties 1 and 2 as well. Every mode contributes to the reconstruction of every snapshot, k, since c is never identically zero. Though the contributed energy may be negligible, it is necessary to negate the presence of the coherent structures in the other modes.

The PDFs of the coefficients for modes 2 and 3 are shown in figure 3(*b*) to reinforce this premise. The vortex of $V^{(1)}$ occurred only once, and is dominant in φ_2. Thus, $c_2^{(k)}$ had a large negative value once, but is present many times near zero to negate its presence in other cycles (snapshots). Mode 3 captures the 15 occurrences of the weak vortex ($V^{(2)}$–$V^{(16)}$) and thus the PDF has many occurrences at a large negative value, but still many with small values near zero to negate its presence in snapshots where it is not present. Figure 3(*b*) also demonstrates the usefulness of the coefficient PDF's to quantify the cyclic variability of each mode's contribution to a particular flow field realization. This will be revisited with the discussion of measured engine flows.

2.3. Examples 3 and 4: POD of translated and rotated overlapping structures

In examples 1 and 2, the vortices in different velocity fields were clearly separated in space. In real flows, the location of coherent structures are not necessarily separated from snapshot to snapshot, but would be expected to overlap, be slightly translated or have a change in orientation. In examples 3 and 4, shown in figure 4, POD is performed on two velocity fields; the velocity fields contain the same vortex structure and energy as example 1 but with no background flow. In example 3, the vortex in $V^{(2)}$ is translated to the right of $V^{(1)}$, with three of the five columns overlapping. In example 4 the vortex in $V^{(2)}$ is not translated but velocity distribution is rotated (on identical grids) by $\pi/6$ compared to $V^{(1)}$; recall that the structures are non-symmetric and have shear, and thus are not identical under rotation. In both examples, the second (lower energy mode) is necessary to properly reconstruct the original velocity fields $V^{(1)}$ and $V^{(2)}$, even though the second modes appear to be non-physical turbulent flow patterns. Thus, small translation and/or rotation of a structure between snapshots create non-physical patterns in the modes. From that the following statement is derived.

Property 5. POD does not necessarily create modes that show physical flow structures. The modes must be summed to reconstruct the physical velocity distributions.

Based on examples 2–4, it is noted that only if the structures are identical between snapshots (or a very strong structure in one snapshot) will a mode create a flow pattern that is equivalent to the physical coherent structures. This is shown most clearly by Chen *et al* [21], where an entrainment vortex appeared at the trailing edge of a fuel injection event; the vortex was very apparent in the first mode, and in the second, which would modulate its strength. This also demonstrates the converse; a basis function may not 'appear' to be a coherent structure, but this does not mean it is noise or unimportant to creation of a coherent structure in a specific snapshot (cycle). The point here is that if a coherent structure occurs randomly in space (spatial phase) it may not be apparent in any mode if it is modulated by other structures.

3. POD analysis of measured in-cylinder engine flows

Compared to the synthetic velocity distributions, in-cylinder velocity fields contain the superpositions of a much larger number of multiple randomly distributed structures on many

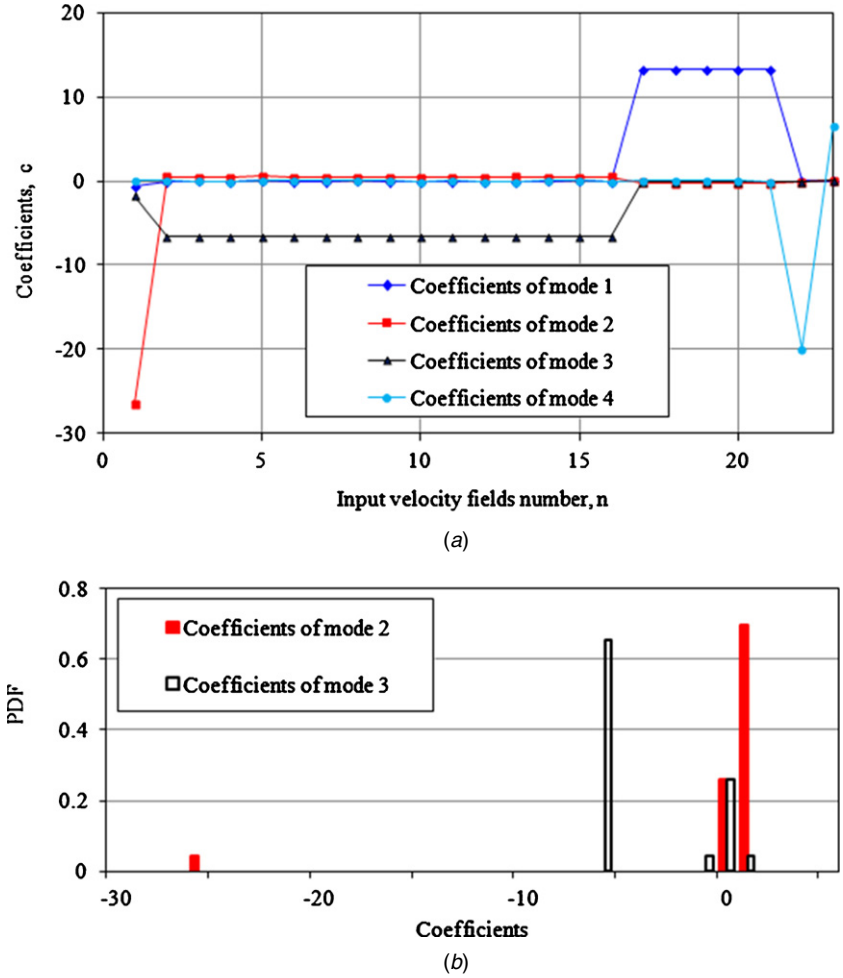


Figure 3. (*a*) Coefficients of first four POD modes for all 23 input velocity fields (the coefficients of higher modes are around 0, not shown here.) (*b*) PDFs of coefficients for mode 2 and mode 3 in (*a*).

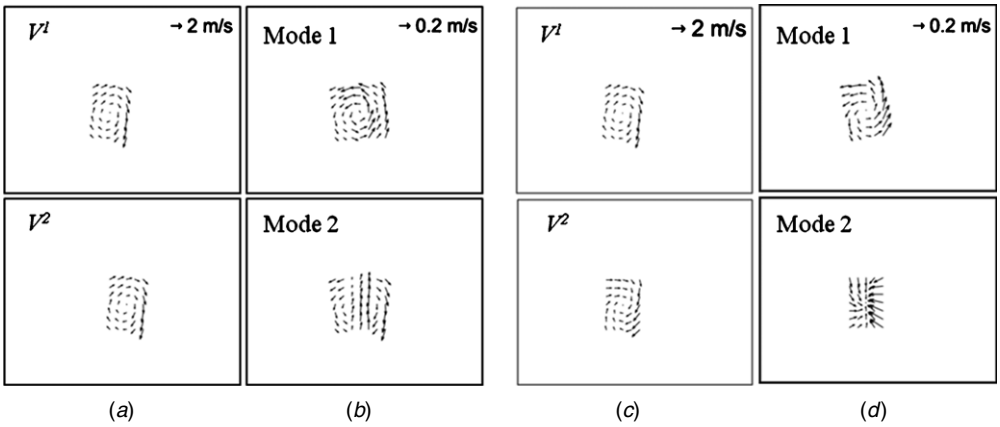


Figure 4. Input velocity distributions of two identical vortices after translation (*a*) and POD modes (*b*) (example 3). Input velocity distributions of two identical vortices after rotation about its center (*c*) and POD modes (*d*) (example 4).

spatial scales. In the following, measured velocity distributions are used to compare the energy content and flow patterns of the POD versus Reynolds decompositions to address the question whether POD should be performed with or without prior subtraction of the ensemble mean. For this purpose, two 200-cycle 2D PIV data sets for directed and undirected in-cylinder flows [2] are employed. These data were measured to reveal the cyclic variability of flow structures, and were analyzed previously employing the traditional Reynolds decomposition [17]. These data were chosen as they represent two extremes: the directed flow having every cycle appearing like the ensemble average, and the undirected having no cycles appearing like the ensemble average [2]. Both 200-cycle data sets were taken in an optical single-cylinder four-stroke

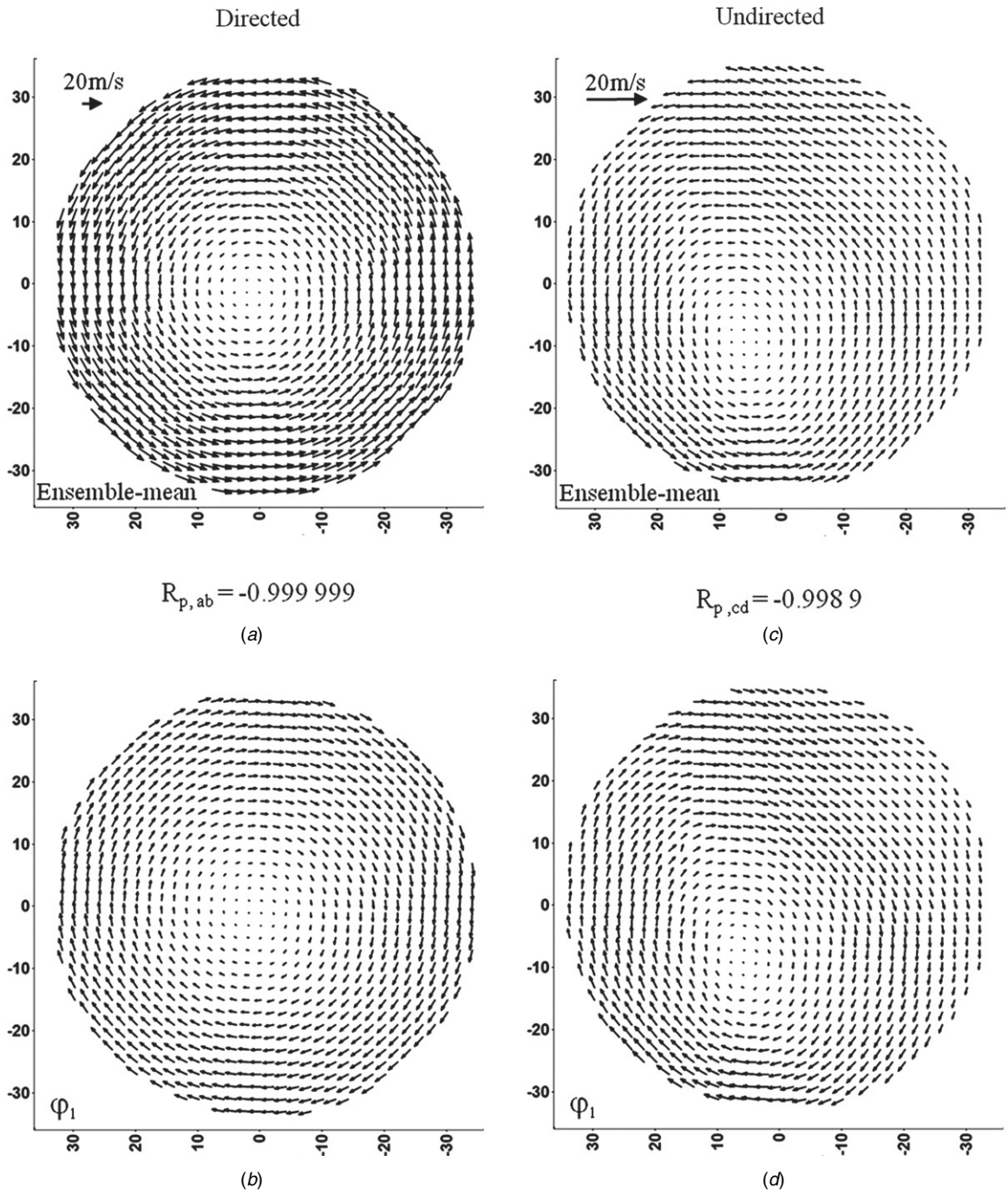


Figure 5. $\langle V \rangle$ and φ_1 for the directed engine flow (*a*) and (*b*), and for the undirected engine flow (*c*) and (*d*). One out of 16 vectors is shown. R_p is for the projection of φ_1 onto $\langle V \rangle$.

two-valve engine with a pancake shaped combustion chamber [18]. The PIV measurements were made with laser sheets parallel to the head at TDC (top dead center) compression, bisecting the 12 mm clearance height. Film-based, two-dimensional PIV measurements captured the 70 mm diameter center of the 92 mm diameter cylinder, with 1 mm × 1 mm (digitized to 128 and 128 pixel) interrogation spots on a 0.5 mm grid. Electro-optical image shifting with cross-correlation [19] was used to resolve the directional ambiguity. The samples were recorded approximately every 12th cycle from three separate tests (in approximately 70 snapshots per test) resulting in 200 snapshots for each data set. The engine was motored at 1200 rpm and 40 kPa MAP for both cases.

This following section quantifies the relationship between the traditional RANS decomposition (ensemble average and turbulence), the POD modes of the 200 original velocity snapshots, V, and the POD modes of the 200 velocity snapshots after subtracting the average, $V–\langle V \rangle$.

3.1. POD and the ensemble average

The flow patterns of the first POD modes created from the original velocity distributions, V, are compared to

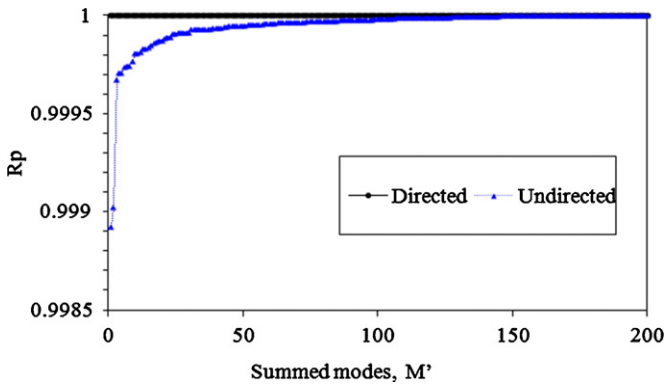


Figure 6. Relevance index (equation (10)) showing the rate that the flow pattern of the reconstructed ensemble average flow pattern, $\langle V_{M'}\rangle$, (equation (9)) approaches the true ensemble-average, $\langle V\rangle$, as more modes are used for the reconstruction (M' increases from $1 \rightarrow M$).

ensemble-averaged $\langle V\rangle$ flow patterns in figure 5. The relevance indices R_p of φ_1^V projected onto $\langle V\rangle$ are included as well. (Note that the superscript to the basis function here indicates whether the full velocity fields, V, or the Reynolds decomposed fields, V–$\langle V\rangle$, are used). The appearance of the patterns (centered-swirl for the directed flow and offset-swirl center for undirected flow) and the near unity relevance indices R_p demonstrates that the mode 1 flow pattern is an excellent (low order) estimate of the ensemble-averaged flow pattern for both flows. The negative value of R_p indicates that φ_1^V has an opposite direction, with the expectation that the coefficients will be negative (as will be shown later). However, figure 6 demonstrates that they are not fully identical since $R_p \rightarrow 1$ asymptotically as the higher modes are added for both the directed and undirected flows. This is intuitively reasonable, since the true average is the ensemble average of all modes from all cycles as per equation (9).

The kinetic energy content of POD mode 1 and the spatially averaged kinetic energy of the ensemble average $\langle$KE$\rangle$ can be compared in table 1 for both the directed and undirected flows. Mode 1 kinetic energy, KE_1, was computed from the coefficients as in equation (6) and divided by $K = 200$ to provide the ensemble-averaged value. Table 1 shows that the energy content of mode 1 is approximately equal to, but slightly larger than, that of the ensemble-averaged value for both the directed and undirected flows. The combination of (*a*) the flow-pattern equivalency of φ_1^V and $\langle V\rangle$ (nearly unity R_p in figure 5) and (*b*) the approximate equivalency of the energy leads to the conclusion that mode 1 is an excellent estimate of, but not identical to, the ensemble average, $\langle V\rangle$.

3.2. POD and the RANS turbulence

Based on the result of section 3.1, it is reasonable to deduce that modes $2 \rightarrow K$ (=200) contain an estimate of the Reynolds decomposed turbulence structures and energy, minus the fraction of the turbulence energy contained in mode 1 (recall that mode 1 has slightly more energy than the ensemble average). Using the partitioning of energy content as a first measure, the near-equivalency of the turbulence energy and the sum of the energy in modes $2 \rightarrow K$ is demonstrated in table 1, which shows that the kinetic energy of the RANS turbulence, $\mathrm{KE}_{\mathrm{rms}}$, is approximately equal to that of modes $2 \rightarrow$ M, $\mathrm{KE}_{2\text{–}200}/K$.

Performing the *POD after subtracting the ensemble average* from the instantaneous velocity distributions (snapshots) will, by definition, contain only the RANS turbulence. The POD modes show flow pattern components and the turbulence energy, but the ensemble-averaged velocity, as computed with equation (9), must be identically zero. Subtle differences in the turbulence extracted from POD results with and without first subtracting the ensemble average are noted in a comparison of the 199 POD modes ($2 \rightarrow 200$) from POD of the full velocity fields with the 200 POD modes of V–$\langle V\rangle$.

If the ensemble average was identical to mode 1, one would expect the $\varphi_m^{V-\langle V\rangle}$ and φ_{m+1}^V to have identical flow patterns. Figures 7 and 8, which plot the lowest order basis functions with and without subtracting the average for both the directed and undirected flows, show this is not true. Though close in appearance, they are not identically equivalent for either the directed or undirected flows. The differences are expected for three reasons: (i) $\langle V\rangle$ is similar but not identical in structure to mode 1, (ii) the energy and patterns of the coherent structures are redistributed according to energy and by the requirements to reconstruct the original snapshots (cf section 2.2), and (iii) coherent structures of the turbulence that are not identical from cycle-to-cycle are separated into 200 (instead of 199), which will be shown later to produce a different energy spectrum.

A visual comparison of the patterns in all modes is qualitative and exhausting, and the use of the relevance index

Table 1. Energy comparison of Reynolds decomposition and POD with and without subtracting the ensemble average.

		Directed KE	% $\langle V\rangle$	% rms	Undirected KE	% $\langle V\rangle$	% rms
Reynolds decomp.	$\langle$KE$\rangle$	1896 000	100		81 300	100	
	$\mathrm{KE}_{\mathrm{rms}}$	110 353		100	120 319		100
POD of original velocity	$\mathrm{KE}_{\mathrm{total}}/K$	2006 000			201 600		
	KE_1/K	1898 000	100.1		87 800	108.0	
	KE_2/K	4 200			8500		
	$\mathrm{KE}_{2\text{–}200}/K$	108 000		98	113 800		95
	ke_1	0.95			0.44		
POD after subtracting the ensemble average	$\mathrm{KE}_{\mathrm{total}}/K$	110 354		100	120 320		100
	KE_1/K	6 000			10 400		
	ke_1	0.054			0.086		

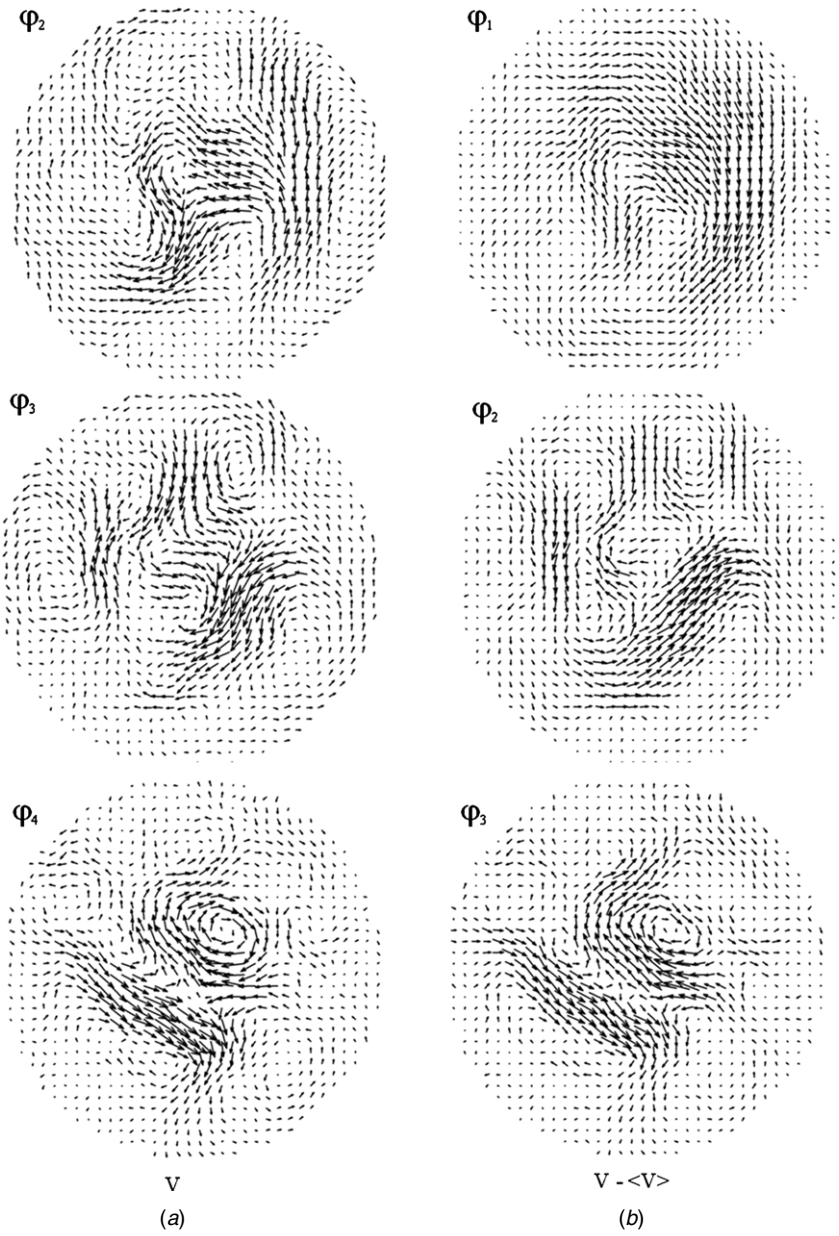


Figure 7. Basis functions for POD of the directed flow (*a*) without subtracting the ensemble average, V, and (*b*) after subtracting the ensemble average, $V-\langle V\rangle$.

is helpful for a quantitative comparison. Figure 9 quantifies the differences between the patterns of modes $\varphi_{1\to4}^{V-\langle V\rangle}$ and $\varphi_{1\to15}^{V}$. Note that the absolute value of R_p is plotted because the variations between ±1 otherwise obscure the fractional values; the importance here is the equivalence of the mode patterns, and not the allocation of the signs of the coefficients. There are distinctive differences between the directed and undirected flows in $\varphi_m^{V-\langle V\rangle} = \varphi_m^V$ with little correlation with adjacent modes, i.e. the mode patterns have simply been shifted to higher modes. However, the modes obtained after subtracting the average for the undirected flow show some correlation with adjacent modes; thus, the coherent structures present in the original velocity snapshots have been distributed differently between the POD modes. It is not obvious that there is physical significance to be attached to the results in figure 9. The only definitive point is that the coherent structures associated with the turbulence, with and without subtracting the average, are decomposed into different basis functions. The fact that the patterns are not identical ($|R_p|\neq1$) is reasonable since the averages and mode 1 are not identical either. This demonstrates that even with real flow data, individual basis functions should not be interpreted as velocity structures; rather flow structures require reconstruction from potentially many modes.

As one would expect from energy conservation, table 1 shows that the $\mathrm{KE}_{\mathrm{total}}/K$ of $V-\langle V\rangle$ is nearly equal to the energy of the Reynolds turbulence $\mathrm{KE}_{\mathrm{rms}}$. Also, KE_1 of $V-\langle V\rangle$ is less than 10% of $\mathrm{KE}_{\mathrm{total}}$ suggesting that many modes are required to reconstruct the most energetic structures of the Reynolds turbulence. Often, relative energy spectra, computed using equations (6) and (7), and shown here in figures 10(*a*) and (*b*), are used in discussions of POD results. The absolute

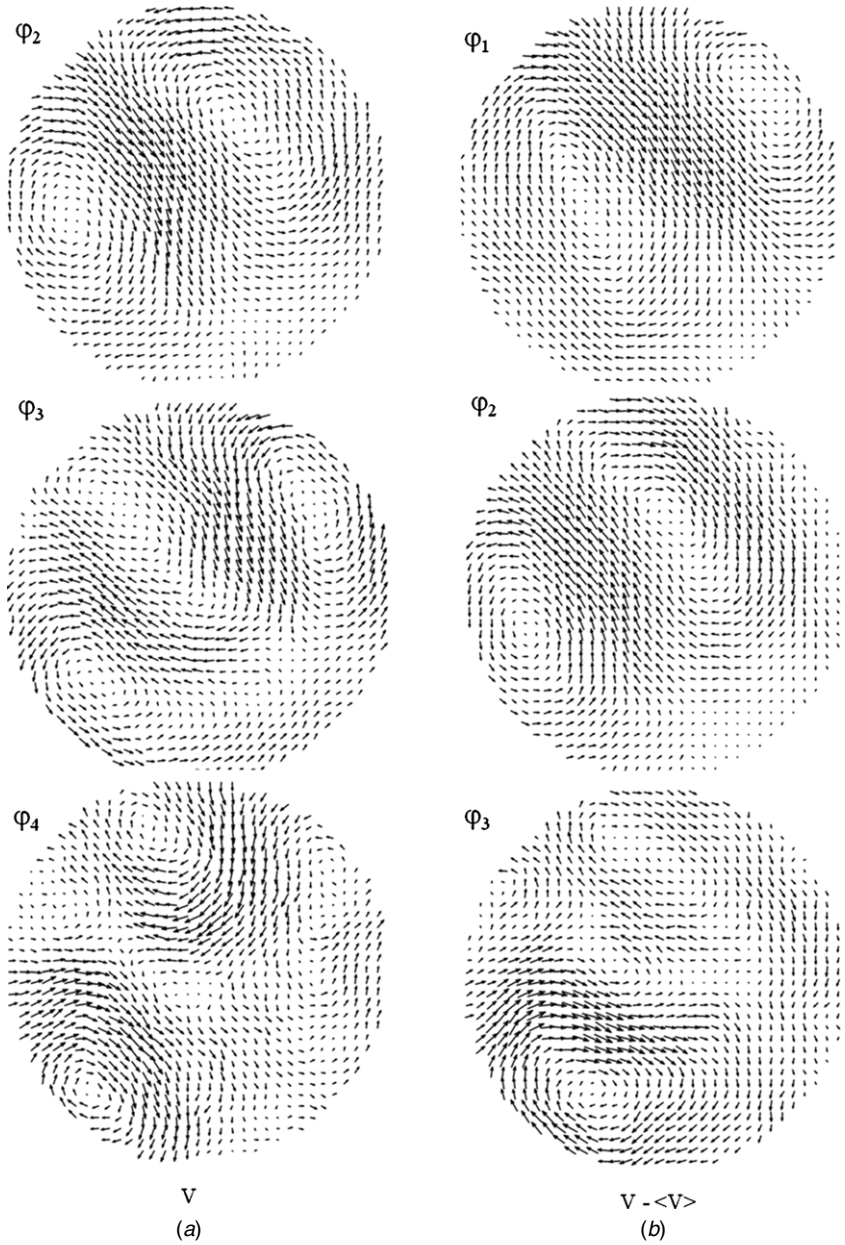


Figure 8. Basis functions for POD of the undirected flow (*a*) without subtracting the ensemble average and (*b*) after subtracting the ensemble average.

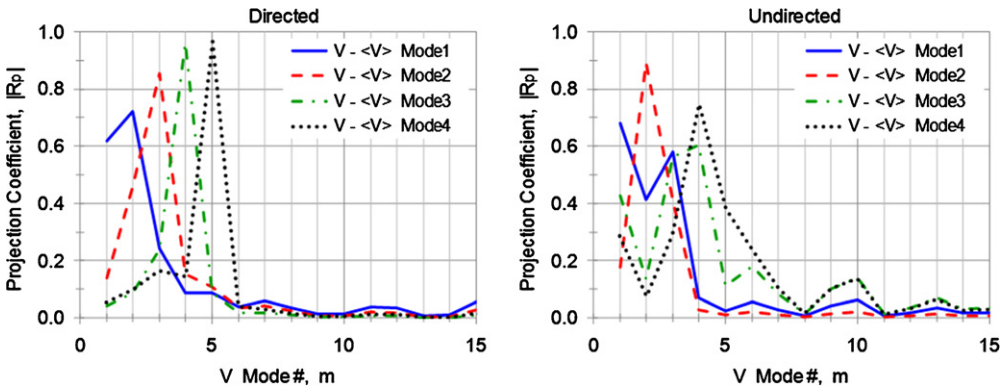


Figure 9. The relevance indices R_p of φ_m $(V–\langle V\rangle)$ projected onto $\varphi_1(V)$–$\varphi_{15}(V)$ quantify the similarities in the mode patterns of POD with and without subtracting the ensemble average. Shown are the projections of mode $m = 1 \rightarrow 4$ of $V–\langle V\rangle$ onto modes $m = 1 \rightarrow 15$ of V.

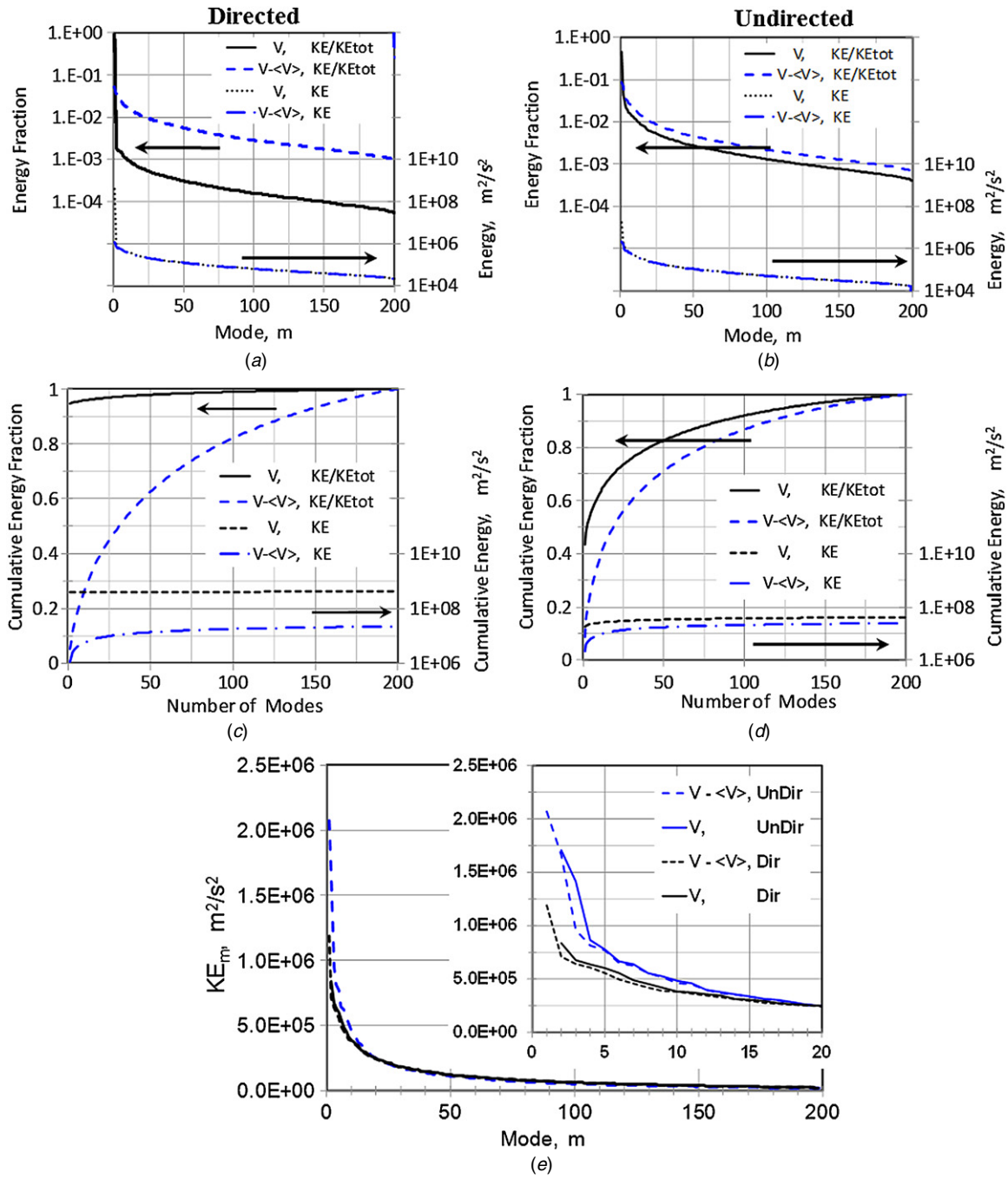


Figure 10. Energy-fraction spectra (*a*) and (*b*), and cumulative energy (*c*) and (*d*) for the directed and undirected engine flows. (*e*) Comparison of absolute energy.

energy spectra are included to emphasize what is lost in the normalization; specifically, there is a great difference in the total energy between the directed and undirected flows, but the absolute energies of the modes greater than $M = 15$ are nearly the same. The shape of the energy spectrum is quite insensitive to subtracting the average compared to the differences in the basis functions (cf figures 7–9). The biggest change is that for the highly directed flow (where the average is apparent in every realization), the KE_1 is dominant and contains most of the total energy. As a result, there is just an offset in the energy fraction, and the absolute energy, KE_m, in the higher modes $m > 1$ is nearly identical as shown in figure 10(*e*). In fact, when the average is subtracted before the POD analysis, there is less than a factor of 2 difference in the spectra of the absolute energy between the directed and undirected flows. This agrees with the analysis of the same data set by Funk *et al* [17], who have used traditional Reynolds decomposition and spatial filtering approaches to quantify differences between the low- and high-swirl flows. The cumulative energy in figures 10(*c*) and (*d*) shows that prior to subtracting the average, the directed-flow mode 1 contains well over 90% of the energy, where as the undirected flow requires over 75 modes to capture 90% of KE_{tot}. In summary, figure 10 provides two important results. Firstly, the spectra are relatively insensitive to changes in the flow. Secondly, in contrast to the directed flow the mode energy of the undirected flow is less than 1% after only 20 modes,

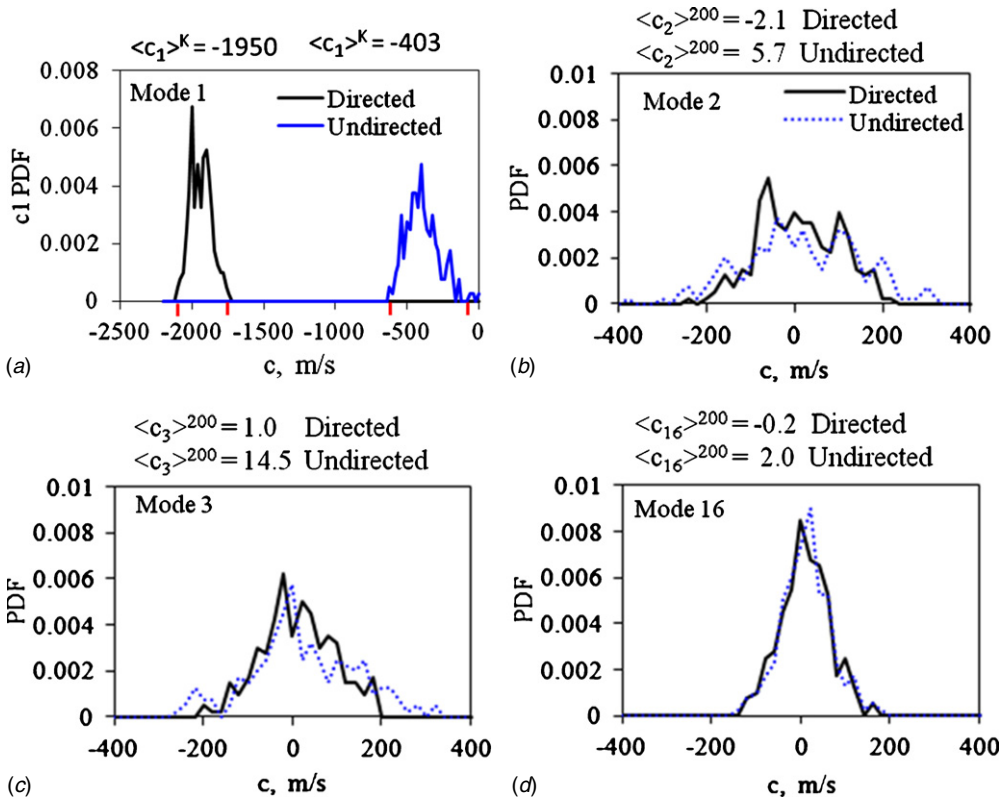


Figure 11. Coefficient PDFs for modes based on POD of V, i.e. without subtraction of the ensemble average.

but requires over 100 modes to achieve 90% of the Reynolds turbulence energy (cumulative energy fraction = 1 for V–$\langle V\rangle$).

3.3. Arguments for not subtracting the mean before performing POD

The previous two sections addressed the differences in POD results when either subtracting the ensemble mean or not before the decomposition is performed. In this section, the analysis of the two engine flows will be used to illustrate why it is useful, if not essential, to not subtract the average in some cases before performing POD.

Figure 11 shows the PDFs of the coefficients $c_m^{(k)}$ for four modes of the POD of the original velocity fields from both engine operating conditions. In figure 11(*a*), the coefficients of mode 1 for directed and undirected flows are negative because in both cases φ_1 rotates opposite to the average flow direction (cf figure 5). The directed-flow PDF shows that the coefficients for mode 1 are always considerably larger than those of higher modes and that mode 1 therefore dominates in all cycles. In contrast, the values of the undirected-flow mode 1 coefficients overlap with those of higher modes, and in fact the undirected flow has some cycles with near-zero coefficients for mode 1. The PDFs of c_1 quantify what was determined by visual inspection in the previous work of Reuss [2] in that (1) the ensemble average is always present in the directed flow and (2) though the undirected flow showed large-scale swirl structures in many cycles, they never looked like the ensemble average, and in some cases were not present at all. This demonstrates a benefit of performing the POD without subtracting the ensemble average: the coefficients of mode 1 reveal the extent to which the mean flow is present and it's cycle-to-cycle variability.

When applied to V–$\langle V\rangle$, the POD creates modes with flow patterns (basis functions) of the RANS turbulence, which must average to zero at every point in the distribution according to equation (9). As expected [20], the average coefficients, $c_m^{V-\langle V\rangle^K}$, are all effectively zero ($10^{-10} \rightarrow 10^{-15}$) for all modes of both flows; residual values are attributed to numerical inaccuracies and rounding errors in the computations. For POD of V, the average coefficients, $\langle c_m^V\rangle^K$, for $m > 1$ are near, but not identically, zero as indicated by the PDFs in figure 11. The non-zero values of $\langle c_m^V\rangle^K$ reflect the difference between mode 1 and the ensemble average of the original flow fields. This is also found in the overall energy budget as listed in table 1. The non-zero mean of the coefficients and their PDFs contain information that can assist in conditional sampling of data sets to identify cycles with unusual conditions, e.g. leading to engine misfires [21].

4. Conclusions

A set of synthetically created flow fields was used to analyze the basic properties of POD mode structures and conclusions that can be obtained from those. The synthetic velocity distributions demonstrated that every structure from every cycle is superimposed in every basis function; the modes do not separate individual structures. A structure is apparent in a single mode only if it has very high energy in a single snapshot, or the structure is identical (shape, location, and

orientation) in many cycles so that the ensemble sum of their energy is high. Any translation or change in orientation (of asymmetric or shearing) structures will distribute the flow energy to other modes. Flow patterns observed in a mode do not necessarily correspond to physical flow structures; those emerge in reconstructed flow fields using a sufficient number of modes.

Experimental data from two optical engine experiments with highly directed and undirected engine flows were used to demonstrate that POD without first subtracting the ensemble average produces a mode 1 energy and flow pattern that are nearly identical to those of the ensemble mean. This was true for both the highly directed flow where the ensemble-averaged flow is apparent in every cycle and in the undirected flow where the ensemble-averaged flow pattern is never observed in a single cycle. This conclusion was based on the fact that both the flow pattern (relevance index, R_p) and energy of mode 1 were nearly identical to the ensemble average. The modes of the POD computed from the original flow fields V (without subtracting the ensemble average) were compared to POD modes of V–$\langle V \rangle$, which is by definition POD of the RANS turbulence; modes $2 \rightarrow M$ of POD of V contain the RANS turbulence based on energy content and R_p, the mode basis functions (flow patterns); the basis functions were nearly identical for the directed flow (merely shifted one mode number), whereas the patterns were distributed to adjacent modes for the undirected flow. It is concluded that POD of V is more useful than V–$\langle V \rangle$ since it is possible to quantify the extent to which the individual cycles are similar to the ensemble average, yet contains the mode structure of the RANS turbulence.

Based on comparisons of the directed versus undirected flows here, it is observed that the POD energy spectra of the RANS turbulence are insensitive metrics for comparison of even two radically different flows and that POD of the full velocity fields is more useful in this case.

Acknowledgment

This work was supported by General Motors R&D within the GM-UM Collaborative Research Laboratory on Engine Systems Research at The University of Michigan. HC is grateful for financial support from Shanghai Jiao Tong University to enable a ten-month visit to the University of Michigan and the support of Min Xu and Yuyin Zhang from Shanghai Jiao Tong University. The authors would also like to express their gratitude to Jacques Boree (Université de Poitiers) for numerous discussions on POD.

References

[1] Heywood J B 1988 *Internal Combustion Engine Fundamentals* (New York: McGraw-Hill)

[2] Reuss D L 2000 Cyclic variability of large-scale turbulent structures in directed and undirected IC engine flows *SAE Technical Paper* 2000-01-0246

[3] Baby X, Dupont A, Ahmed A, Deslandes W, Charnay G and Michard M 2002 A new methodology to analyze cycle-to-cycle aerodynamic variations *SAE Technical Paper* 2002-01-2837

[4] Druault P, Guibert P and Alizon F 2005 Use of proper orthogonal decomposition for time interpolation from PIV data: application to the cycle-to-cycle variation analysis of in-cylinder engine flows *Exp. Fluids* **39** 15

[5] Roudnitzky S, Druault P and Guibert P 2006 Proper orthogonal decomposition of in-cylinder engine flow into mean component, coherent structures and random Gaussian fluctuation *J. Turbul.* **7** 1–19

[6] Fogleman M, Lumley J L, Rempfer D and Haworth D 2004 Application of the proper orthogonal decomposition to datasets of internal combustion engine flows *J. Turbul.* **5** 023

[7] Kapitza L, Imberdis O, Bensler H P, Willand J and Thevenin D 2010 An experimental analysis of the turbulent structures generated by the intake port of a DISI-engine *Exp. Fluids* **48** 265–80

[8] Voisine M, Thomas L, Boree J and Rey P 2011 Spatio-temporal structure and cycle to cycle variations of an in-cylinder tumbling flow *Exp. Fluids* **50** 1393–407

[9] Liu K and Haworth D 2011 Development and assessment of POD for analysis of turbulent flow in piston engines *SAE Technical Paper* 2011-01-0830

[10] Sick V, Reuss D, Rutland C, Haworth D, Oefelein J, Janicka J, Kuo T-W and Freitag X Y 2010 A common engine platform for engine LES development and validation *LES4ICE: Int. Conf. on LES for Internal Combustion Engine Flows* ed C Angelberger (Rueil-Malmaison: IFP Energies Nouvelles)

[11] Holmes P, Lumley J L and Berkooz G 1996 *Turbulence, Coherent Structures, Dynamical Systems and Symmetry* (Cambridge, UK: Cambridge University Press)

[12] Chatterjee A 2000 An introduction to the proper orthogonal decomposition *Curr. Sci.* **78** 808–17

[13] Sirovich L and Kirby M Low-dimensional procedure for the characterization of human faces *J. Opt. Soc. Am.* A **4** 519–24

[14] Lumley J L 1967 The structure of inhomogeneous turbulence *Atmospheric Turbulence and Wave Propagation* ed A M Yaglom and V I Tararsky (Moscow: Nauka) pp 166–78

[15] Sirovich L 1987 Turbulence and the dynamics of coherent structures: 1. Coherent structures *Q. Appl. Math.* **45** 561

[16] Cordier L and Bergmann M 2003 Proper orthogonal decomposition: an overview *Post-processing of Experimental and Numerical Data (Von Karman Institute Lecture Series)*

[17] Funk C O, Sick V, Reuss D L and Dahm W J A 2002 Turbulence Properties of high and low swirl in-cylinder flows *SAE Technical Paper* 2002-01-2841

[18] Kuo T-W and Reuss D L 1995 Multidimensional port and cylinder flow calculations for the transparent-combustion-chamber engine *Proc. ASME IVE Engine Modeling* vol 23 pp 19–29

[19] Reuss D L 1993 Two-dimensional particle-image velocimetry with electro-optical image shifting in an internal combustion engine *Proc. SPIE* **2005** 413

[20] Kriegseis J, Dehler T, Gnirß M and Tropea C 2010 Common-base proper orthogonal decomposition as a means of quantitative data comparison *Meas. Sci. Technol.* **21** 085403

[21] Chen H, Reuss D L and Sick V 2011 Analysis of misfires in a direct injection engine using proper-orthogonal decomposition *Exp. Fluids* **51** 1139