International Journal of ENGINE RESEARCH

*Original Article*

*International J of Engine Research*
14(4) 307–319

DOI: 10.1177/1468087412455748
jer.sagepub.com
SAGE

# A practical guide for using proper orthogonal decomposition in engine research

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

**Abstract**

Proper orthogonal decomposition has been utilized for well over a decade to study turbulence and cyclic variation of flow and combustion properties in internal combustion engines. In addition, proper orthogonal decomposition is useful to quantitatively compare multi-cycle in-cylinder measurements with numerical simulations (large-eddy simulations). However, the application can be daunting, and physical interpretation of proper orthogonal decomposition can be ambiguous. In this paper, the mathematical procedure of proper orthogonal decomposition is described conceptually, and a compact MATLAB® code is provided. However, the major purpose is to empirically illustrate the properties of the proper orthogonal decomposition analysis and to propose practical procedures for application to internal combustion engine flows. Two measured velocity data sets from a motored internal combustion engine are employed, one a highly directed flow (each cycle resembles the ensemble average), and the other an undirected flow (no cycle resembles the average). These data are used to illustrate the degree to which proper orthogonal decomposition can quantitatively distinguish between internal combustion engine flows with these two extreme flow properties. In each flow, proper orthogonal decomposition mode 1 is an excellent estimate of ensemble average, and this study illustrates how it is thus possible to unambiguously quantify the cyclic variability of Reynolds-averaged Navier–Stokes ensemble average and turbulence. In addition, this study demonstrates the benefits of comparing two different samples of cycles using a common proper orthogonal decomposition mode set derived by combining the two samples, the effect of spatial resolution, and a method to evaluate the number of snapshots required to achieve convergence.

**Keywords**

Proper orthogonal decomposition, internal combustion engines, turbulence, cyclic variability

Date received: 05 April 2012; accepted: 30 May 2012

## Introduction

In-cylinder turbulence in reciprocating internal combustion (IC) engines remains a subject of great interest[1] and some controversy, since there is still no single consensus as to how to distinguish between the flow properties associated with turbulence kinetic energy dissipation from the properties associated with cyclic variability. Whereas turbulence is important for fuel–air mixing and burning rate enhancement to achieve high engine speeds, cycle-to-cycle variations have long been recognized as an important source of limited power output, increased fuel consumption, and high pollutant emission levels for IC engines. Thus, it is important to define metrics to quantify and separate turbulence from cyclic variability. This applies equally to both experimental and numerical studies. Further, it is important to quantitatively compare cycles sampled from different experiments or between experiments and simulations. Proper orthogonal decomposition (POD) is now being used in an attempt to fulfill these needs.

In recent years, POD has been considered as a powerful tool to objectively identify the coherent structures in turbulent flows. Lumley[2] introduced POD into turbulent flow research as a tool to separate 'large eddies'

[1]School of Mechanical Engineering, Shanghai Jiao Tong University, National Engineering Laboratory for Automotive Electronic Control Technology, Shanghai, China
[2]Department of Mechanical Engineering, The University of Michigan, USA
[3]University of Michigan-Shanghai Jiao Tong University Joint Institute, Shanghai Jiao Tong University, Shanghai, China

**Corresponding author:**
David L Reuss, 2026 W. E. Lay Automotive Laboratory, The University of Michigan, 1231 Beal Avenue, Ann Arbor, MI 48109, USA.
Email: dreuss@umich.edu

in shear flows. Graftieaux et al.[3] suggested using POD to separate pseudo-fluctuations attributed to the unsteady nature of the large-scale vortices from fluctuations due to small-scale turbulence. POD has been applied to in-cylinder velocity data of reciprocating engines by a number of groups.[4–7]. Bizon et al.[8] reported that the use of POD has allowed the analysis of cyclic variations of two-dimensional (2D) scalar combustion images. Stöhr et al.[9] demonstrated combined application of POD on results from particle image velocimetry (PIV), and planar laser-induced fluorescence of OH (OH-PLIF) measurements can provide detailed insights into flow–flame interaction in turbulent flames. Recently, Voisine et al.[10] employed phase-invariant POD to analyze motored flow 'breakdown' and cyclic variability during intake and early compression in a four-valve pent-roof engine. Seeking insight into the causes of misfires, POD was applied to velocity and equivalence ratio images, and was useful in identifying and analyzing the differences in flow and mixture conditions at spark timing between well-burning and misfiring cycles.[11]

Computations and high-speed laser imaging diagnostics provide the ability to resolve the rapid evolution of in-cylinder flows, and POD holds promise as one method for comparing stochastic samples from both sources of data. Reynolds-averaged Navier–Stokes (RANS) computations provide ensemble-averaged flow motion in IC engines, useful for engineering computations. In contrast, large-eddy simulation (LES) provides multiple cycles of stochastic, in-cylinder flow processes formerly unavailable, thereby capturing cycle-to-cycle variations.[12] However, for any numerical tool for turbulent flows, model validation is always important. High-speed laser imaging provides the ability to resolve the rapid evolution of in-cylinder flow, fuel–air mixing, and combustion within individual cycles and for hundreds of consecutive cycles.[13,14] These experimental data can be used to validate the LES model, since the engine cycles are stochastic, cycle samples from either simulations or experiments require statistical rather than deterministic means for comparison.[15] POD is proposed as a metric to make quantitative, objective comparisons of cycles sampled from large-eddy simulations and/or experiments.[16–18]

According to Cordier and Bergmann,[19] the mathematics of POD has been around since 1943, and POD has been applied to fluid mechanics since 1967.[2] The principles have been described and discussed in detail by Cordier and Bergmann,[19] Chaterjee,[20] and Holmes et al.[21] Nonetheless, many practical questions arise as to the actual procedures of applying POD and how to use and interpret the results; these questions motivate the empirical interpretations performed here. Some of those questions were addressed in an analysis of both synthetic and measured flow fields to empirically reveal the properties of POD.[22] The synthetic velocity distributions were created to illustrate fundamental properties of the POD analysis while avoiding the complexities of real IC engine flows. The measured data sets were chosen as examples representing two extremes of motored in-cylinder flow. One data set was the result of a highly directed intake flow, where every cycle appears similar to the ensemble average at top dead center (TDC); the other was a highly undirected intake flow, where no cycle appeared similar to the TDC ensemble average. These data were used to illustrate how to interpret POD analysis with and without first subtracting the ensemble average, the POD mode energy spectra (energy fraction and absolute energy), and the relationship between mode patterns using the relevance index, as introduced by Liu and Haworth.[17]

This paper builds on previous work and presents guidance for executing POD analysis and practical application. A MATLAB® code for performing the POD is included in appendix 2. The principles and techniques defined in a previous study[22] are applied here. The same two examples of two-component, 2D velocity distributions measured in a motored IC engine are used to illustrate (a) an unambiguous quantification of the cyclic variability of RANS ensemble average and RANS turbulence, (b) the benefits of comparing samples of different cycles using a common POD mode set derived by combining the two samples, (c) the effect of spatial resolution, and (d) a means to evaluate the number of snapshots required to achieve convergence. The two motored IC engine measured data sets employed in the previous study are used here again to illustrate the degree and metrics by which POD can quantitatively distinguish between the two extreme in-cylinder engine flows.

## Mathematical concept of POD

Conceptually, the POD technique decomposes the original velocity or scalar field into a sum of weighted, linear, basis functions or modes, $\varphi_m$. In this study, the modes are computed from two-component, 2D velocity distributions, but could be applied to three-component three-dimensional velocity distributions as in Liu and Haworth,[17] or one-component scalar distributions, which has been demonstrated by Sirovich and Kirby.[23] The modes are referred to as an empirical basis as they are generated from a sample of snapshots (here measured or computed velocity distributions, $V$) rather than prescribed a priori. The number of modes is equal to the number of snapshots. In addition, the modes are orthogonal to each other and normalized such that the magnitude of each mode is unity. The actual energy of each mode is contained in a matrix of coefficients, $c$, where there is one coefficient for each mode of each snapshot. The coefficients are generated by projection of each mode on to each snapshot. Thus, one can reconstruct each snapshot by summing all modes multiplied by the coefficients for that snapshot. Alternately, one can compute all of the energy contained in one mode for the entire sample by summing the square of all coefficients for that mode. Here, only the

mathematical descriptions necessary to define the nomenclature are described, since comprehensive descriptions are available in the literature.[19–21]

## Mathematical procedures of POD

The POD technique provides a linear approximation of a set of functions that enable an easier characterization of the complex original input data as a sum of weighted basis functions. For practical applications, the decomposition can be carried out using the classical method[2] or the snapshot method.[24] For experimental data such as laser Doppler anemometry, which provides a well-defined time description but with limited spatial resolution, the classical method should be employed. The results of the snapshot method are equivalent to those from the classical method, but the snapshot method is more efficient in cases where the number of spatial locations (grid nodes) in the velocity field is greater than the number of snapshots. The snapshot method was used here and a sample code written in MATLAB® 2009b is provided in appendix 2.

The fundamental idea of POD is to decompose a set of velocity distributions, $V^{(k)} = (u, v)^{(k)}_{i,j}$ or scalar quantities distributions $s^{(k)} = (s)^{(k)}_{(i,j)}$, (called snapshots, $i,j$ is the index of the grid points in the velocity or scalar quantities distributions, $k$ is the snapshot index) into a linear combination of $M$ spatial basis functions (POD modes,$\varphi_m$) and the corresponding coefficients $c_m^{(k)}$

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

with the constraint that the basis functions are orthonormal to each other.

The mathematic procedures of POD are shown as follows. The $u$ velocity of every single velocity distribution $V^{(k)}$ was reordered into a row and put into a matrix

$$U = \begin{bmatrix} U^{(1)} \\ U^{(2)} \\ \vdots \\ U^{(K)} \end{bmatrix} = \begin{bmatrix} u^{(1)}_{i=1,j=1} & u^{(1)}_{i=1,j=2} & \cdots & u^{(1)}_{i=1,j=J} & u^{(1)}_{i=2,j=1} & \cdots & u^{(1)}_{i=I,j=J} \\ u^{(2)}_{i=1,j=1} & u^{(2)}_{i=1,j=2} & \cdots & u^{(2)}_{i=1,j=J} & u^{(2)}_{i=2,j=1} & \cdots & u^{(2)}_{i=I,j=J} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ u^{(K)}_{i=1,j=1} & u^{(K)}_{i=1,j=2} & \cdots & u^{(K)}_{i=1,j=J} & u^{(K)}_{i=2,j=1} & \cdots & u^{(K)}_{i=I,j=J} \end{bmatrix} \tag{2}$$

where $I \times J$ is the number of total grid points in the velocity field and $K$ is the total number of snapshots. The $v$ velocity component was processed the same way as the $u$ velocity to form the matrix $V$. Then, the spatial correlation matrix for velocity distributions was defined as

$$C = \frac{1}{K}(UU^T + VV^T) \tag{3}$$

For scalar quantities distributions, $C = \frac{1}{K} SS^T$ (where $S$ was obtained the same way as equation (2)). The associated code for this procedure is shown in appendix 2, sections 1 and 2. The code is given for velocity distributions; the correlation matrix $C$ must be replaced by $C = \frac{1}{K} SS^T$ if the POD analysis is to be carried out for scalar quantity distributions.

The goal of using the POD technique is to find a sequence of orthonormal basis functions $\varphi_m$ (POD modes) representing the 'coherent structures' in such a way that the following function is minimized

$$\sum_{k=1}^{K} \left|\left| V^{(k)} - \sum_{m=1}^{M} c_m^{(k)} \varphi_m \right|\right|^2 \rightarrow \min \tag{4}$$

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

where $\|\cdot\|$ denotes the $L^2$ norm.

This minimization is realized by solving the eigenvalue problem of correlation matrix $C$

$$C\beta_m = \lambda_m \beta_m \tag{5}$$

Here, this step is performed with the MATLAB® code in appendix 2, section 3.

The basis functions are obtained by projecting $U$ and $V$ onto the eigenvector $\beta_m$ ($m = 1, 2, \ldots, M$), with subsequent normalization (section 4 in appendix 2). The basis functions represent the extracted flow pattern, which are often considered to be synonymous with 'coherent structures' in the literature. It is important to recognize that these structures do not necessarily reflect a real flow structure that can be observed in a flow field. Rather, each mode represents a component of a flow field that, by definition (equation (1)), is reconstructed by summing over all weighted modes. A real flow pattern only appears prominent in a POD mode when the flow structure is nearly the same and in the same location in each flow field (snapshot).[22]

The coefficients of each mode were computed by projecting the original velocity fields onto the computed basis functions (section 5 in appendix 2) that are then exported, along with the respective coefficients for display and further processing (section 6 in appendix 2). The $K \times M$ coefficient matrix $c_m^{(k)}$

$$c_m^{(k)} = \begin{bmatrix} c_1^{(1)} & c_2^{(1)} & \ldots & c_M^{(1)} \\ c_1^{(2)} & c_2^{(2)} & \ldots & c_M^{(2)} \\ \ldots & \ldots & \ldots & \ldots \\ c_1^{(K)} & c_2^{(K)} & \ldots & c_M^{(K)} \end{bmatrix} \tag{6}$$

contains the amplitude that the corresponding basis function contributes to a particular snapshot, $k$, in each row, and the amplitude corresponding to a particular mode, $m$, in each column. Because of the use of the $L^2$ norm for POD, it is natural to apply this to velocity distributions since $\frac{1}{2}\left(c_m^{(k)}\right)^2$ represents the energy contributed by the $m$th mode to the $k$th cycle. Based on equation (1), the velocity distribution from a given cycle $k$ can be reconstructed by summing all of the

modes multiplied by their corresponding coefficient for that cycle. Likewise, the kinetic energy per unit mass from all of the cycles captured by the $m$th mode is

$$KE_m = \frac{1}{2}\sum_{k=1}^{K}(c_m^{(k)})^2 \tag{7}$$

where $KE_m = KE/\rho = \frac{1}{2}V^2$. The energy fraction of the $m$th mode is given by

$$ke_m = KE_m / \sum_{m=1}^{M} KE_m \tag{8}$$

The modes are ordered with decreasing corresponding energy so that mode 1 carries the highest energy, and 'energy spectra' showing the energy as a function of ordered mode number are assembled. It is worth noting here that many POD discussions quantify the energy as $c^2$. Here, we present equation (7) as a reminder that kinetic energy, $\frac{1}{2}\rho V^2$ is required for energy balances when comparing different data sets. This is particularly important for phase-invariant analysis, remapping between grids, and when there are spatial density variations, where the mass associated with each vector can dominate differences in the kinetic energy.

Ensemble-averaged velocity distributions are related to the POD through the averaged coefficients and the POD modes using

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

where the operator $\langle c_m\rangle^K$ indicates the ensemble average over all coefficients for a given mode. The exact ensemble average is achieved when all modes are used and a low-order estimate is obtained with a truncated set.

## Metric for comparing velocity or scalar mode patterns

It is common in the literature to compare POD mode spectra (energy fraction or cumulative energy versus mode number) to assess the equivalency of spectra. A fair comparison of coefficients and energy spectra would assume the corresponding modes (patterns) are the same. Comparison of modes is often carried out by visual comparison, which is both qualitative and impractical. Thus, a metric to quantify the similarity of mode patterns is needed. Here, the 'relevance index'[17], $R_p$, which is obtained by projecting one velocity or scalar distribution $D^{(1)}$ on to another $D^{(2)}$ is employed

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

The numerator denotes the inner product of two distributions, 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 distributions are identical, $R_p = -1$ means two velocity distributions are exactly opposite, $R_p = 0$ if two POD modes are orthogonal. It can be used to compare two POD basis functions, or to compare a basis function with a snapshot, or to compare two snapshots. Since $R_p$ is normalized by the $L^2$ (energy for flow velocity), it is only a measure of the degree of similarity between two distribution patterns. This is demonstrated in Figure 1, which shows three vortices that are identical in structure, except for the velocity magnitudes. The vortices in Figure 1(b) and 1(c) are obtained by multiplying every vector in vortex of Figure 1(a) by 2 and 5, respectively. The $R_p$ between any two of these three velocity distributions is 1. This means that these three are identical structures, though they contain different amounts of kinetic energy. 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 structures. In addition, $R_p$ provides a single value metric-of-comparison for the entire distribution, which will be shown to be useful for comparing the structure of different modes.

## POD analysis of in-cylinder flow

For the purpose of illustrating, two 200-cycle 2D PIV data sets of motored in-cylinder flows[25] were employed. These data sets were selected because they represent two extremes; one is a directed flow, which has every cycle similar to ensemble average flow, and the other one is undirected flow, which has no cycle appearing like the ensemble average. The turbulence properties of these flows were analyzed previously using Reynolds decomposition[26] and POD.[22] Both flows were measured in a single-cylinder, four-stroke optically accessible engine with a pancake-shaped combustion chamber. The PIV measurements were performed with a laser sheet parallel to the head at TDC compression, bisecting the 12 mm clearance height. High spatial resolution

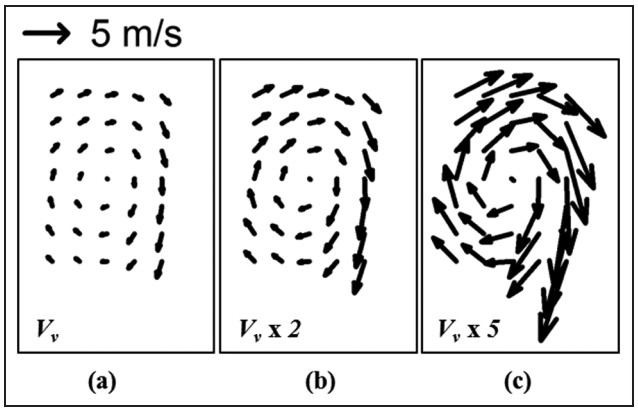


**Figure 1.** Three synthetic velocity distributions. The $R_p$ between any two of these three velocity distributions is 1.

film-based 2D PIV measurements captured the center 70 mm diameter of the 92 mm diameter cylinder, with 1 mm x 1 mm (128 x 128 pixels) interrogation windows on a 0.5 mm grid. Electro-optical image shifting with cross-correlation was used to resolve the directional ambiguity.[27] The samples were recorded every 12th cycle in approximately 70 snapshots for each test. A total of 200 velocity distributions were measured for both the directed flow that was generated with a shrouded intake valve and a relatively undirected flow using a standard valve. The engine was motored at 1200 r/min, 40 kPa MAP (Manifold absolute pressure).

It is appropriate to comment on the effect of PIV measurement velocity errors on the POD analysis. Modeling by Westerweel[28] predicts that 1% precision can be achieved with PIV (0.1 pixel displacement-peak location resolution with 8 pixel maximum displacement), However, Megerle et al.[29]demonstrated empirically that 2% is a more realistic value with real particle images. Though these analyses were performed for digital PIV, 2% error is considered a worst-case estimate of the photographic PIV for the measurements used here.[25] This PIV error would then influence the POD spectrum as the square or at an energy fraction of $4e^{-4}$, which would not significantly influence the lowest order modes. Of course, in IC engines, false peak detection can occur in regions of wall light scattering, due to inhomogeneous particle density, and in regions where local/instantaneous flow velocities (statistical extremes) are too large to be properly captured by the laser pulse separation; these measurement errors can dominate the flow and thereby appear in low-order POD modes.

### *POD with and without subtracting the ensemble average*

Chatterjee[20] noted that it is valid to perform POD with or without first subtracting the average, stating that only the interpretation of the results change. The relationship between POD analysis and Reynolds decomposition (ensemble average and turbulence) was quantified empirically in a previous study,[22] where POD analysis was performed on the same two data sets employed here, comparing the analysis with and without subtracting the ensemble average. The purpose was to assess when it is useful to subtract the ensemble average prior to performing the POD and when POD on the original data set is more meaningful.

For IC engine analysis, where the ensemble average can vary with the same time and spatial scales as the turbulence, POD analysis of velocity distributions without mean subtraction is advantageous. It was demonstrated that without subtracting the average, POD mode 1 provided an excellent estimate of ensemble average flow pattern, with $R_p > 0.999$ for both the directed and undirected flows.[22] Also, the kinetic energy content of mode 1 was within 0.1% and 8% of the ensemble average kinetic energy for the directed and undirected flows, respectively. Thus, based on equation (9), the kinetic energy of the ensemble average can be estimated by summing the POD mode 1 coefficients from each cycle (snapshot) as

$$KE_{ave} \approx \frac{1}{2K} \sum_{k=1}^{K} \left( c_1^{(k)} \right)^2 \tag{11}$$

Therefore, the coefficients of mode 1 can be used as a metric for the prevalence of the ensemble average flow in each cycle, which are both the most probable flow structure and the RANS decomposition mean.

Further, it was demonstrated that modes $2 \rightarrow 200$ contain the orthogonal basis functions and energy of the RANS turbulence,[22] which is estimated as

$$KE_{turb} \approx \sum_{m=2}^{200} \left[ \frac{1}{2K} \sum_{k=1}^{K} \left( c_m^{(k)} \right)^2 \right] \tag{12}$$

Chen et al.[22] also demonstrated that every flow structure from every snapshot is, to some degree, in every mode; the structures are distributed amongst the modes such that each snapshot can be uniquely reconstructed by summing over the product of modes and coefficients specific to that snapshot. Thus, physical interpretation of these modes (patterns) is not intuitive. Nonetheless, equation (12) provides a means to unambiguously compute the RANS turbulence energy of each cycle (snapshot); the coefficients can provide the spatially averaged RANS turbulence for each cycle and, since the modes retain the spatial information, can provide the spatial distribution of the RANS turbulence energy within each cycle.

## Quantifying cyclic variation using POD

In this section, we exploit the properties of equations (11) and (12) to quantify the cyclic variability of the RANS ensemble average and RANS turbulence based on POD, thus, using energy decomposition rather than temporal or spatial filtering with an arbitrary cut-off frequency.

As shown in the section 'Mathematical procedures of POD', $\frac{1}{2}\left(c_m^{(k)}\right)^2$ represents the energy that the $m$th mode contributes to the $k$th cycle. Since the mode 1 is an excellent estimate of the ensemble-averaged flow pattern, $\frac{1}{2}\left(c_1^{(k)}\right)^2$ for each of the 200 cycles can be used to quantify the kinetic energy of the ensemble average flow structure contained in each cycle. Similarly, it is possible to quantify the cyclic variability of the RANS turbulent kinetic energy. Since modes $2 \rightarrow 200$ contain the energy of RANS turbulence, equation (12) can be used to compute the energy contained in the RANS turbulence modes of each of the 200 cycles where

$$(KE_{turb})_k = \frac{1}{2} \sum_{m=2}^{200} (c_m^{(k)})^2 \tag{13}$$

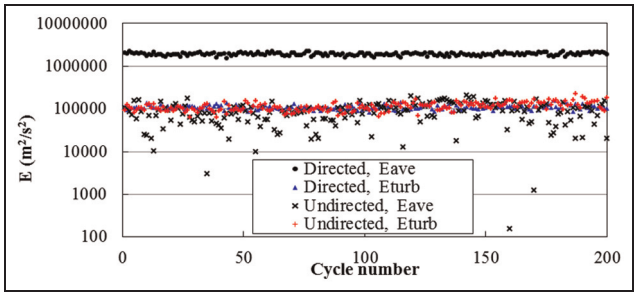


**Figure 2.** POD-derived ensemble average and turbulence energies for all 200 snapshots obtained from POD coefficients for directed and undirected flows.

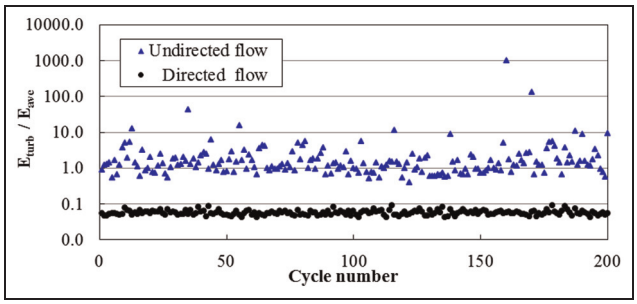


**Figure 3.** The ratio between turbulence energy and ensemble average energy for all 200 snapshots, for directed and undirected flows.

Figure 2 shows the cyclic variability of the energy associated with mode 1 ($KE_{ave}$) and the absolute turbulence flow energy in modes 2–200 ($KE_{turb}$) for all 200 cycles. These values of course represent the spatially averaged over the field-of-view for each cycle (snapshot). The prominent features are that the $KE_{ave}$ of the undirected flow is much lower at TDC than for the directed flow due to higher dissipation, which was revealed by the measured and computed results in this engine by Reuss et al.[30] in a previous RANS-based analysis. Also, $KE_{ave}$ of the undirected flow is at the same absolute magnitude as $KE_{turb}$, which in turn is about the same for both the directed and undirected flows, as was found by Funk et al.[26] Figure 3 shows the spatially averaged turbulence intensity within each cycle. Clearly, the turbulence intensity of the undirected flow is much higher due to the low $KE_{ave}$. These observations are all consistent with the traditional RANS analysis of Funk et al.[26]

Figures 2 and 3 demonstrate that POD analysis can provide four new metrics not provided by the usual RANS decomposition. *First*, POD has provided a mechanism to unambiguously separate the RANS average energy from the RANS turbulence energy on a cycle-by-cycle basis, thus providing the cyclic variability of each. *Second*, POD shows the RANS average (most expected structure) for the directed flow dominates the turbulence in every cycle, as could previously only be concluded through cycle-by-cycle visual observation.[23] For the undirected flow, the cyclic variability of the RANS average far exceeds the cyclic variability of the RANS turbulence; this discrimination could not be achieved using the RANS decomposition since all root mean square (r.m.s.) fluctuations about the mean would be included in the turbulence. *Third*, the coefficients provide cyclic variability of the spatially averaged value of RANS turbulence for each cycle. In addition, since the modes still contain the spatial information, the local increase (decrease) of either can also be derived within each cycle, provided that high-speed imaging data are available. Thus, POD modes 2 and above provide new metrics for correlating and modeling the flow properties with fuel mixing, heat transfer, turbulent flame propagation, etc., since they retain simultaneous spatial and temporal (cycle-to-cycle) information not available from the RANS decomposition. In particular, POD can provide a mechanism to conditionally sample and correlate cycles with the rare/extreme flow events. It can provide a quantifiable metric to test if errant combustion cycles are dominated by the ensemble average cyclic variation or the turbulence cyclic variation, as well as how far from the cycle mean they must be to induce 'bad' combustion cycles. Having identified the features of the errant cycles, the POD modes provide the possibility that the errant flow feature can be identified. *Finally*, Table 1 demonstrates that POD can provide cycle-averaged statistics, quantifying cyclic variability parameters for comparing different test conditions. Thus, probability distribution functions (PDFs) of the coefficients can be used to quantify and discriminate the cyclic repeatability of engine designs and operating conditions.

A cautionary note is in order. The proposed metric for cyclic variability requires that equations (11) and (12) hold true, as for the two data sets used here. It is reasonable to imagine an extreme data set where the mean is near zero, in which case, the most energetic mode (mode 1) could not be the mean. Thus, it is always necessary to test for the equivalence of mode 1 and the ensemble average prior to proceeding with the analysis.

## Comparison of data sets using POD

One of the main purposes of using POD is the analysis of differences and commonalities between data sets. This could be a comparison of computed LES cycles with measured PIV cycles, or sets of all measured or all computed data. It is usual in the literature to assess equivalence using the energy spectra (energy fraction and cumulative energy versus mode number). However, since the modes are an empirically formed basis, the modes from two different samples are not necessarily the same. Thus, it is worth demonstrating the importance of using the relevance index to quantify the equivalency of the mode patterns in addition to the energy spectra.

**Table 1.** Average, standard deviation, and coefficient of variation (COV) of ensemble average energy ($E_{ave}$), turbulence energy ($E_{turb}$), and the ratio between turbulence energy and ensemble average energy ($E_{turb}/E_{ave}$) for all 200 snapshots, for the directed and undirected flows plotted in Figure 2.

| | Directed | | | Undirected | | |
|---|---|---|---|---|---|---|
| | $E_{ave}$ | $E_{turb}$ | $E_{turb}/E_{ave}$ | $E_{ave}$ | $E_{turb}$ | $E_{turb}/E_{ave}$ |
| Average | 1,898,337 | 107,602 | 0.057 | 87,776 | 113,838 | 7.9 |
| Standard deviation | 144,286 | 12,660 | 0.009 | 43,734 | 26,342 | 72.1 |
| COV | 8% | 12% | 16% | 50% | 23% | 915% |

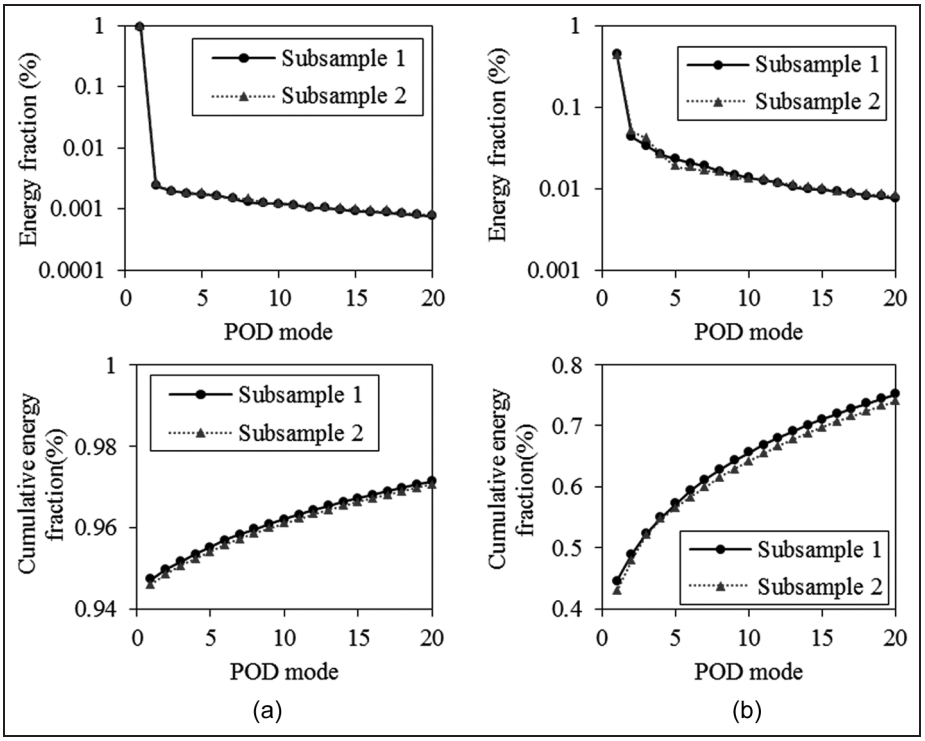


**Figure 4.** Energy fraction and cumulative energy fraction comparison between subsamples for (a) directed flow and (b) undirected flow.

## Separate or combined-sample for application of POD

An important question is whether to perform the POD on the two data sets separately or to combine them into a single sample then perform the POD.[11,16] As with any statistical method, the two data sets must have the same number of samples to create an unbiased comparison and, in the case of POD, must be on (or mapped on to) identical grids to create comparable modes. Two truly identical data sets should separately generate two identical POD basis sets. $R_p$ could be used to show the basis functions are identical, and the energy spectra can be used to demonstrate that they have the same mode energy. Two data sets could be different either because no cycles are the same or because some, but not all cycles, are the same. In this case, having a single POD mode set allows the use of the coefficients to analyze the equivalency of these two data sets.

To illustrate the benefits of analyzing a combined set of data rather than separate data sets, the 200 PIV samples were divided into two samples: 100 cycles were randomly selected from the 200 cycle sample to form subsample 1, and the remaining 100 cycles used to form subsample 2, for each flow configurations. The POD results from the two separate subsamples are shown in Figure 4, which shows energy fraction and cumulative energy fraction captured by the first 20 POD modes. Figure 5 shows the modes from subsample 1 projecting onto the modes from subsample 2, for (a) directed and (b) undirected flows, respectively. As shown in Figure 4, the energy fractions captured by the modes are very similar for the two subsamples. The differences are less than 1.6%. However, the relevance index analysis shown in Figure 5 presents a different picture in that all modes except mode 1 are different in their spatial structure. This suggests that comparing energy fractions alone can be misleading: the POD modes can be really

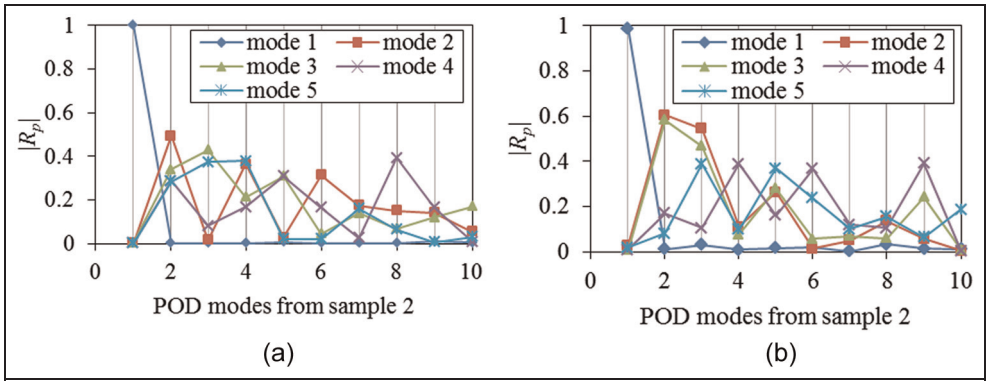


**Figure 5.** Modes from subsample 1 projecting onto the modes from subsample 2 for (a) directed and (b) undirected flows.

different even if the energy fraction is similar. Further, this demonstrates the added complexity in quantifying and interpreting the comparison of two data sets with two different mode sets. Since the flow patterns have been distributed between different modes (cf. Figure 5), it is difficult to unambiguously quantify equivalency.

Conducting the POD after combining the two data sets into a single sample will create a single POD mode set, thereby allowing the use of the coefficients or energies to analyze the equivalency of these two data sets. To illustrate this, the two data sets used in Figures 4 and 5 are analyzed using the coefficients based on the POD analysis from the *combined* data sets. The two PDFs, conditionally sampled on the first and second subsamples, are shown in Figure 6. Because the coefficients of the two subsamples are within the same range of values and the coefficients were from same mode set, a level of equivalency is demonstrated.

The use of combined data sets, but applied to scalar measurements, was illustrated by Chen et al.[11] In that study, POD analysis of equivalence ratio and velocity were used to investigate the differences between normally burning and misfired cycles using the coefficients from a single mode set generated from a combined data sample.

## Impact of data resolution

The spatial resolution from experimental data (PIV) and simulation data (LES) is not necessarily the same. Therefore, it is useful to investigate the effect of spatial resolution on POD analysis. Here, the POD results from a fine and coarse grid are compared. The original velocity field spatial resolution is 1 mm on a 0.5 mm grid.[25] The resolution of the data sets was reduced by first applying a 1 mm Gaussian filter to the 0.5 mm grid, then mapping the data on to a 4mm grid by retaining only every 64th ($8 \times 8$) grid node.

Figure 7 shows the cumulative energy fraction obtained from POD for these two spatial resolutions. Data for the 4 mm resolution set converges with mode number slightly faster than that of the 1 mm resolution, i.e. the lower order modes contain more of the total energy. Low-pass spatial filtering removes small-scale structures, presumed to be at lower energy (higher modes numbers), and brings the filtered field closer to a spatially averaged result. Thus, the first mode is expected to have a higher energy fraction as it is a better estimate of the ensemble average.

The total energy (based on $V^2$) listed in Figure 7 is a reminder of the necessity to keep track of the mass of kinetic energy, as was discussed with equations (7) and (8). Decreasing the grid spacing means each vector represents 64 times more mass on the 4 mm $\times$ 4 mm grid and, therefore, should have 0.5 mm $\times$ 0.5 mm $\times$ 64 times the kinetic energy based on mass. Figure 7 shows that the low-resolution data (4 mm grid) has 1/64.8 and 1/69.9 less energy for the directed and undirected flows, respectively, rather than the 1/64 expected from the grid spacing (mass) alone. The difference to the expected value is due to the dissipation from the spatial-averaging filter. Some numerical dissipation can be expected for any grid remapping and thus affect the low-energy higher modes, as demonstrated by the plot in Figure 7.

The effect of the resolution on the mode pattern is shown in Figure 8. The relevance indexes of the first 40 modes of the unfiltered, 0.5 mm grid are projected onto the low-resolution subgrid (4 mm), only where the subgrid nodes exist. The results from the directed flow show that the corresponding modes are almost identical ($R_p$ close to 1) for the first 10 modes and $R_p > 0.9$ for the first 15 modes. Thus, the lower order (high-energy) POD modes are not significantly affected by the spatial resolution of the smaller spatial scales. A similar conclusion can be made for the undirected flow. The exceptions are modes 6 and 7, which had a significant decrease of $R_p$, though still near 0.9. Drawing on the principles that were based on POD analysis of the synthetic velocity distributions,[22] we speculate that the lower values of modes 6 and 7 show the lower resolution due to the elimination of either a few very high-energy small-scale velocity structures, or many small structures that occurred with the same pattern and location in many cycles.

POD is energy based, and the higher the mode number, the lower the fraction of energy contribution to the

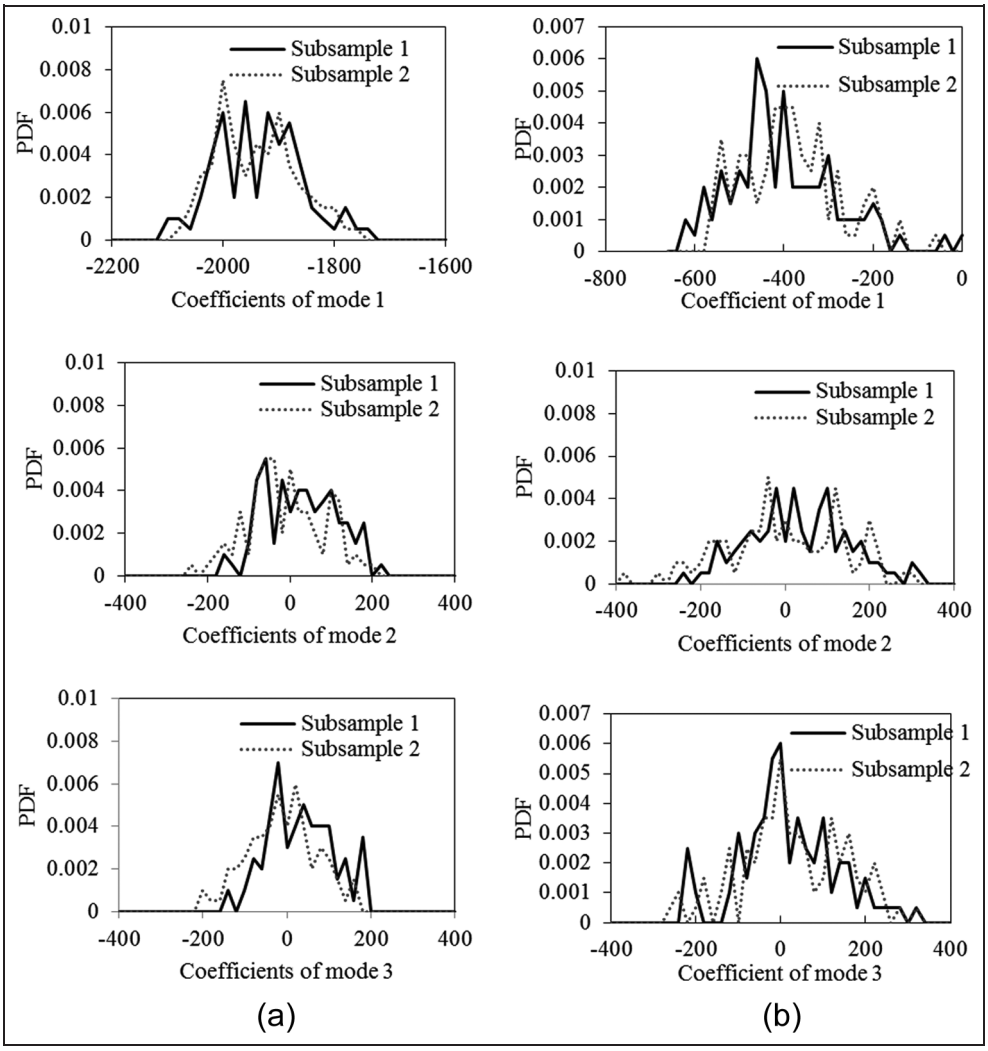


**Figure 6.** The two PDFs conditionally sampled on the first and second subsamples, but using the coefficients from the combined POD analysis for (a) directed and (b)undirected flows.

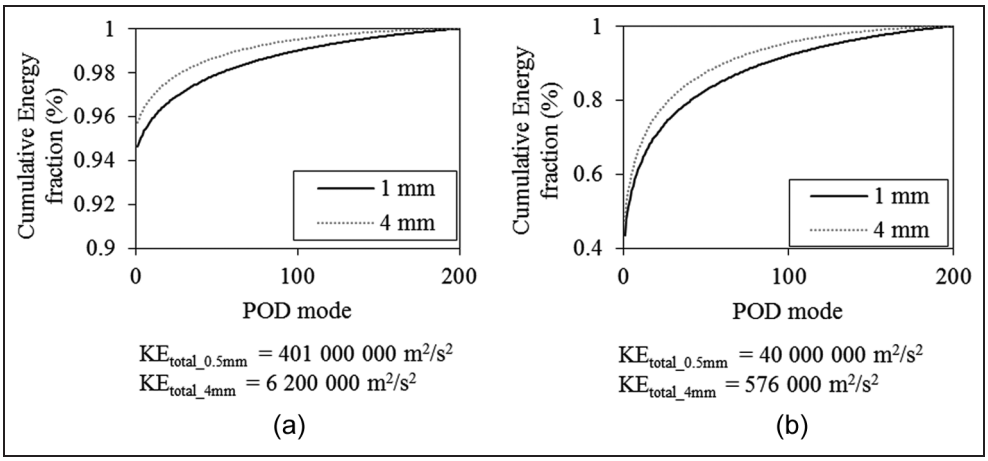


**Figure 7.** Cumulative energy fraction for two spatial resolutions.

total flow. These findings are in agreement with the application of spatial filtering in Funk et al.[26] in that less energy is contained in small-scale structures and therefore the spatial filtering has little effect on the energy fraction contained in the low-order (most energetic) modes. This should not be considered universal;

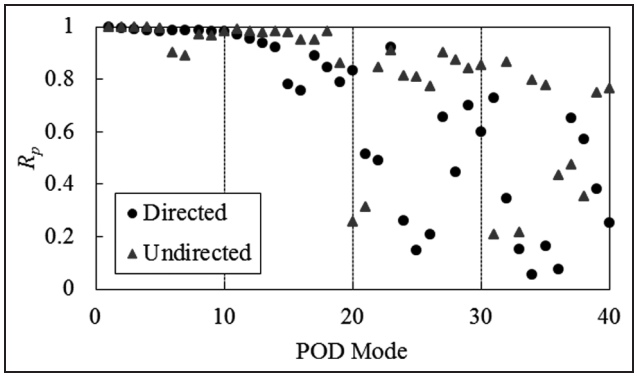


**Figure 8.** First 40 modes of spatial resolution 4 mm case projected onto the subgrid (4 mm) of modes from spatial resolution 0.5 mm case.

a different result might be found in flows with highly energetic small-scale structures.

## Number of required input files

A method is required to assess the sample size required to create POD modes that have converged. The discussion of convergence is not simple since the number of created modes is equal to the number of input snapshots. Thus, adding additional snapshots to an input data set will redistribute the energies amongst the new larger number of modes, according to equation (1). It is usual in the literature to compare POD energy spectra to observe mode convergence, thus energy fraction is a logical choice to use here as well. However, as with the use of sequential data sets, comparing energy fraction spectra of dissimilar mode patterns is a poor test of equivalency. The utilization of the relevance index is again useful to quantify the similarity of mode patterns, to assess when they have converged.

To demonstrate a method for evaluating mode convergence with the number of snapshots, modes from different sample sizes are compared with the same numbered modes from the full 200 snapshots in Figure 9. Flow pattern (relevance index), energy fraction spectra, and cumulative energy fraction are shown for both the directed and undirected flow. Taken together, the relevance index and energy fraction show that at least 120 snapshots are required in order to get a converged solution. However, the energy spectrum converges faster than the flow pattern (relevance index). This is clearly shown by the 80 sample set, where the energy has converged, and $R_p$ has not. Thus, $R_p$ is again shown to be a more sensitive metric for discrimination of the equivalency of two mode sets as they approach convergence.

Traditional statistical methods can also be used to estimate the error of a sample mean and variance as a function of sample size. This would then of course estimate the statistical error in mode 1 and the sum of modes $2 \rightarrow M$, respectively. Assuming large samples and Gaussian distributions, Baker et al.[31] estimate the mean error as StdDev/ and r.m.s. error as StdDev/ $\left(\sqrt{N} * \text{mean}\right)$ and $\sqrt{2/(N-1)}$, respectively. Using the values in Table 1, estimates of the statistical error in the ensemble average of the directed and undirected flows, respectively, are determined to be 2% and 12% for 20 samples. This then explains the lack of variation in mode 1 with sample size. The statistical error in the turbulence of both flows is estimated to require 120 samples just to achieve 13%, which is consistent with the improving but not yet converged $R_p$. Figure 9 demonstrates that the energy spectra are insensitive metrics as they are converging before reasonable values of statistical convergence are achieved. However, $R_p$ spectra do approach convergence at the same rate as the statistical error approaches small errors.

This demonstrates that the number of available snapshots could significantly affect the conclusions that are drawn from the POD analysis, and the convergence of the modes should be verified before using both the POD modes (relevance index) and energy fraction. This is true for both the directed and undirected flows presented here. It should be noted that the number of snapshots required in these two data sets should not be considered universal; the data here is only an illustration of a procedure to quantitatively evaluate the numbers of modes and snapshots required to achieve convergence.

## Conclusions

This paper has provided procedures and interpretation for the application of POD to issues unique to reciprocating IC engine flows. The mathematical procedure of POD has been described conceptually, and a compact MATLAB® code provided. Two in-cylinder IC engine motored flows were used to illustrate practical procedures for its application and to illustrate the properties of the POD analysis, empirically. The two-component, 2D data sets were measured in a previous study using PIV.[25] These two examples represent extremes of in-cylinder flow. One data set was the result of a highly directed intake flow, where every cycle appears similar to the ensemble average at TDC; the other was a highly undirected intake flow, where no cycle appeared similar to the TDC ensemble average.

Expanding on the results of a previous study,[22] the present work demonstrates the following.

1. POD can be used to unambiguously quantify the cyclic variability of the RANS average and RANS turbulence energies using coefficients of the POD modes. The RANS average and RANS turbulence energies can be computed, either for the ensemble of all cycles at a given hardware/test condition, or for individual cycles. Since the POD modes retain the spatial information of the flows, it is possible to quantify the local spatial values as well. The undirected flow configuration demonstrated that

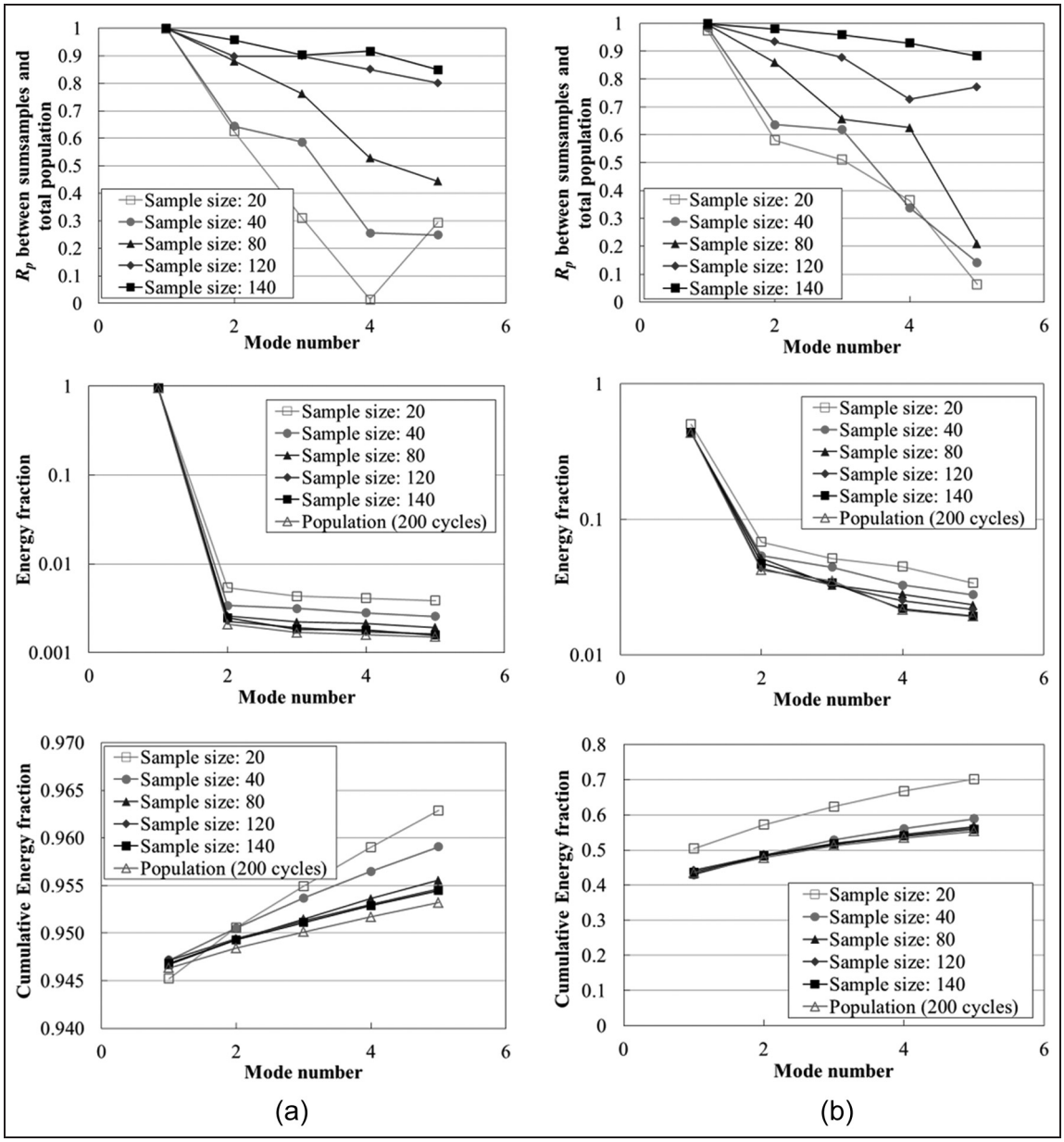


**Figure 9.** First five modes of subsamples projected onto same numbered modes of full 200 snapshots, energy fraction, and cumulative energy fraction for (a) directed and (b) undirected flow.

this procedure unambiguously separates the cyclic variability of the RANS average as long as mode 1 is a good estimate of the ensemble average, which otherwise would have been included in the r.m.s. fluctuations of the RANS turbulence.

2. POD provides a mechanism for comparing two data samples, e.g. experiments to experiments or experiments to simulations. It was demonstrated that POD energy spectra (energy fraction and cumulative energy) are poor metrics of equivalence, since the spectra of unequal mode sets (patterns) can have similar spectra. Combining samples prior to performing the POD provides a common POD mode set so that the comparison can be quantified with the coefficients alone.
3. Data sets were compared before and after a combination of spatial filtering and subsampling to a coarser grid. Results demonstrated that this largely did not affect the lowest 15 of 200 mode patterns or energy for both of the data sets used here.
4. A procedure was demonstrated to evaluate the number of snapshots required to achieve mode convergence. It was demonstrated that POD energy spectra (energy fraction and cumulative energy) do not sufficiently discriminate between large and small samples when compared with traditional statistical errors. Convergence of the relevance index of Liu and Haworth[17] is required to ensure the number of snapshots is sufficiently large for the POD modes to converge.

## Acknowledgements

The authors would also like to acknowledge discussions with Professor Min Xu from Shanghai Jiao Tong University.

## Funding

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. Hao Chen is grateful for financial support from Shanghai Jiao Tong University to enable a 10 month visit to the University of Michigan.

## References

1. Granet V, Vermorel O, Lacour C, et al. Large-eddy simulation and experimental study of cycle-to-cycle variations of stable and unstable operating points in a spark ignition engine. *Combust Flame* 2012; 159: 1562–1575.
2. Lumley JL. The structure of inhomogeneous turbulence. In: Tararski Ya (ed) *Atmospheric turbulence and wave propagation*. Moscow: Nauka, 1967.
3. Graftieaux L, Michard M and Grosjean N. Combining PIV, POD and vortex identification algorithms for the study of unsteady turbulent swirling flows. *Measure Sci Tech* 2001; 12: 1422–1429.
4. Fogleman M, Rempfer D and Lumley JJ. Application of the proper orthogonal decomposition to datasets of internal combustion engine flows. *J Turbul* 2004; 5: 1–18.
5. Kapitza L, Imberdis O, Bensler HP, et al. An experimental analysis of the turbulent structures generated by the intake port of a DISI-engine. *Exp Fluids* 2010; 48(2): 265–280.
6. Marc D, Boree J, Bazile R, et al. Tumbling vortex flow in a model square piston compression machine: PIV and LDV measurements. SAE paper 972834, 1997.
7. Baby X and Floch A. Investigation of the in-cylinder tumble motion in a multi-valve engine: effect of the piston shape. SAE paper 971643, 1997.
8. Bizon K, Continillo G, Mancaruso E, et al. POD-based analysis of combustion images in optically accessible engines. *Combust Flame* 2010; 157: 632–640.
9. Stöhr M, Sadanandan R and Meier W. Phase-resolved characterization of vortex–flame interaction in a turbulent swirl flame. *Exp Fluids* 2011; 51(4): 1153–1167.
10. Voisine M, Thomas L, Boree J, et al. Spatio-temporal structure and cycle to cycle variations of an in-cylinder tumbling flow. *Exp Fluids* 2011; 50: 1393–1407.
11. Chen H, Reuss DL and Sick V. Analysis of misfires in a direct injection engine using proper orthogonal decomposition. *Exp Fluids* 2011; 51(4): 1139–1151.
12. Haworth DC. Large-eddy simulation of in-cylinder flows. *Oil Gas Sci Technol* 1999; 54(2): 175–185.
13. Sick V, Drake MC and Fansler TD. High-speed imaging for direct-injection gasoline engine research and development. *Exp Fluids* 2010; 49(4): 937–947.
14. Peterson B, Reuss DL and Sick V. High-speed imaging analysis of misfires in a spray-guided direct injection engine. *Proc Combust Inst* 2011; 33(2): 3089–3096.
15. Pope SB. Ten questions concerning the large-eddy simulation of turbulent flows. *New J Phys* 2004; 6: 35.
16. Sick V, Chen H, Abraham PS, et al. Proper-orthogonal decomposition analysis for engine research. In: Spicher U (ed) *Direkteinspritzung im Ottomotor VIII Forschungsergebnisse und aktueller Entwicklungsstand bei der Benzin-Direkteinspritzung*. Renningen: Expertverlag, 2012, pp. 1–12.
17. Liu K and Haworth D. Development and assessment of POD for analysis of turbulent flow in piston engines. SAE paper 2011-01-0830, 2011.
18. Enaux B, Granet V, Vermorel O, et al. Large eddy simulation of a motored single-cylinder piston engine: numerical strategies and validation. *Flow Turb Combust* 2011; 86: 153–177.
19. Cordier L and Bergmann M. *Proper orthogonal decompsition: an overview*. Lecture Series 2003–04 in VKI, Post-processing of experimental and numerical data, 2003.
20. Chatterjee A. An introduction to the proper orthogonal decomposition. *Current Sci* 2000; 78: 808–817.
21. Holmes P, Lumley JL and Berkooz G. Turbulence coherent structures dynamical systems and symmetry. Cambridge, UK: Cambridge University Press, 1996.
22. Chen H, Reuss DL and Sick V. On the use and interpretation of proper orthogonal decomposition of in-cylinder engine flows. *Meas Sci Tech* 2012; 23(8): 085302
23. Sirovich L and Kirby M. Low-dimensional procedure for the characterization of human faces. *JOSAA* 1987; 4: 519–524.
24. Sirovich L. Turbulence and the dynamics of coherent structures 1. Coherent structures. *Quart Appl Math* 1987; 45: 561–571.
25. Reuss DL. Cyclic variability of large-scale turbulent structures in directed and undirected IC engine flows. SAE paper 2000-01-0246, 2000.
26. Funk CO, Sick V, Reuss DL, et al. Turbulence properties of high and low swirl in-cylinder flows. SAE paper 2002-01-2841, 2002.
27. Reuss DL, Two-dimensional particle-image velocimetry with electrooptical image shifting in an internal combustion engine. In: *Proc SPIE optical diagnostics in fluids and flows*. SPIE, 1993, pp. 413–1150.
28. Westerweel J. Fundamentals of digital particle image velocimetry. *Measure Sci Tech* 1997; 8: 1379–1392.
29. Megerle M, Sick V and Reuss DL. Measurement of digital PIV precision using electrooptically-created particle-image displacements. *Measure Sci Tech* 2002; 13: 997–1005.
30. Reuss DL, Kuo T, Khalighi B, et al. Particle image velocimetry measurements in a high-swirl engine used for evaluation of computational fluid dynamics calculations. SAE paper 952381, 1995.
31. Baker RJ, Hutchinson P and Whitelaw JH. Velocity measurements in the recirculation region of an industrial burner flame by laser anemometry with light frequency shifting. *Combust Flame* 1974; 23: 57–71.

# Appendix 1

## Nomenclature

| | |
|---|---|
| $C$ | spatial correlation matrix, equation (3) |
| $c_m^{(k)}$ | POD coefficient for snapshot number $k$ and the mode number $m$ |
| $< c_m >^K$ | ensemble average over all coefficients for mode number $m$ |

| | |
|---|---|
| $D^{(1)}$, $D^{(2)}$ | scalar or velocity distributions |
| $E_{ave}$ | ensemble average energy |
| $E_{turb}$ | turbulence energy |
| $i, j$ | index of the grid points in the snapshot |
| $I, J$ | number of total grid points in the snapshot |
| $k$ | snapshot number |
| $ke_m$ | energy fraction captured by POD mode number $m$ |
| $K$ | total number of snapshots |
| $KE$ | kinetic energy, $\frac{1}{2}\rho V^2$ |
| $KE_{ave}$ | energy estimated with POD mode 1 |
| $KE_m$ | kinetic energy captured by POD mode number $m$ |
| $(KE_{turb})_k$ | absolute turbulence flow energy estimated in POD modes 2–200 |
| $M$ | total number of POD modes |
| $m$ | POD mode number |
| $N$ | sample size |
| $R_p$ | relevance index, equation (10) |
| $S$ | Snapshot of a scalar distribution |
| | StdDev standard deviation |
| $s^{(k)}$ | scalar quantities distribution number $k$ |
| $U$ | matrix for $u$ velocity of snapshots |
| $u_{i,j}^{(k)}$ | velocity component at grid point $(i, j)$ in snapshot number $k$ along $x$ axis |
| $V$ | snapshot of the total-velocity distribution |
| $v_{i,j}^{(k)}$ | velocity component at grid point $(i, j)$ in snapshot number $k$ along $y$ axis |
| $V^{(k)}$ | velocity distribution number $k$ |
| $< V >$ | ensemble-averaged velocity distribution |
| $\beta_m$ | eigenvector number $m$ |
| $\delta$ | Dirac delta |
| $\lambda_m$ | eigenvalue number $m$ |
| $\rho$ | gas density |
| $\varphi_m$ | POD mode number $m$ |
| $\|\cdot\|$ | $L^2$ norm |

# Appendix 2

## POD code for velocity distributions

The POD codes for the work described in this paper were developed using the commercial software MATLAB®, Ver. 2009b. The code for velocity distributions is provided below and is available at http://hdl.handle.net/2027.42/92348. Comments to explain the code are shown in *italics*.

```
functionVelocityDistributionPOD(Snapshots Address)
```

## Method of snapshots

*Section 1 – input snapshots. Each snapshot (txt file) contains four columns. The first two columns are the velocity distribution grid point coordinates for the x and y directions, respectively. The last two columns are u and v velocities, respectively.*

```
files = dir([SnapshotsAddress,'*.txt']);
n_snapshots = size(files,1);
for j=1:n_snapshots
   fid = fopen([SnapshotsAddress,files(j).name], 'r');
   data = fscanf(fid,'%f %f %f %f',[4,inf]);
   x = data(1,:); % x coordinate
   y = data(2,:); % y coordinate
   U(j,:) = data(3,:); % u velocity
   V(j,:) = data(4,:);% v velocity
   fclose(fid);
end
```

*Section 2 – compute spatial correlation matrix C*

```
c1 = U*U';
c2 = V*V';
C = (c1 + c2)/n_snapshots;
```

*Section 3 – solve the eigenvalue problem: C * EigenVector = EigenValue * EigenVector*

```
[beta, lmd] = svd(C);
```

*Section 4 – calculate basis functions*

```
phix = U'*beta;
phiy = V'*beta;
% Normalize basis functions
GridNum = size(x,2);
for j=1:n_snapshots
PhiNor = 0;
   for i=1:GridNum
   PhiNor = PhiNor + phix(i,j)^2 + phiy(i,j)^2;
   end
PhiNor = sqrt(PhiNor);
phix(:,j)= phix(:,j)/PhiNor;
phiy(:,j)= phiy(:,j)/PhiNor;
end
```

*Section 5–calculate coefficients*

```
TimCoeU = U*phix;
TimCoeV = V*phiy;
TimCoe = TimCoeU + TimCoeV;
```

*Section 6 – export basis functions*

```
for a=1:n_snapshots
   FilNamPhi = 1000 + a;
   PhiOut = fopen([SnapshotsAddress,num2str(FilNam
   Phi),'.txt']', 'wt');
   fprintf(PhiOut, '#DaVis 7.2.2 2D-vector 16 145 145
   "position" "mm" "position" "mm" "velocity" "m/
   s"\n');
   phia =[x;y;phix(:,a)';phiy(:,a)'];
   fprintf(PhiOut, '%20.9f %20.9f %20.9f %20.9f\n',
   phia);
   fclose(PhiOut);
end
% Write coefficients into excel file
xlswrite([SnapshotsAddress,' TimCoe.xlsx'],TimCoe);
```