







RESEARCH ARTICLE | FEBRUARY 14 2024

## Extracting vector magnitudes of dominant structures in a cyclic engine flow with dimensionality reduction

S. J. Baker ; X. H. Fang ; A. Barbato ; S. Breda ; M. Magnani; S. Fontanesi ; F. C. P. Leach ; M. H. Davy 



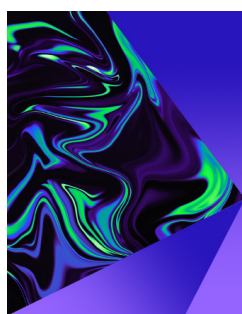
*Physics of Fluids* 36, 025131 (2024)

<https://doi.org/10.1063/5.0189368>



CrossMark

15 February 2024 09:28:17



## Physics of Fluids

Special Topic:  
Selected Papers from the 2023 Non-Newtonian  
Fluid Mechanics Symposium in China

**Submit Today**



# Extracting vector magnitudes of dominant structures in a cyclic engine flow with dimensionality reduction

Cite as: Phys. Fluids **36**, 025131 (2024); doi: [10.1063/5.0189368](https://doi.org/10.1063/5.0189368)

Submitted: 28 November 2023 · Accepted: 21 January 2024 ·

Published Online: 14 February 2024






View Online



Export Citation



CrossMark

S. J. Baker,<sup>1,a)</sup>  X. H. Fang,<sup>1,2</sup>  A. Barbato,<sup>3</sup>  S. Breda,<sup>4</sup>  M. Magnani,<sup>4</sup>  S. Fontanesi,<sup>4</sup>  F. C. P. Leach,<sup>1</sup>  and M. H. Davy<sup>1</sup> 

## AFFILIATIONS

<sup>1</sup>Department of Engineering Science, University of Oxford, Parks Road, Oxford OX1 3PJ, United Kingdom

<sup>2</sup>Department of Mechanical and Manufacturing Engineering, Schulich School of Engineering, University of Calgary, Calgary, Alberta T2L 1Y6, Canada

<sup>3</sup>Siemens Industry Software SRL, Via Scaglia Est. 15, Modena 41126, Italy

<sup>4</sup>Department of Engineering “Enzo Ferrari,” University of Modena and Reggio Emilia, Via P. Vivarelli 10, Modena 41125, Italy

<sup>a)</sup> Author to whom correspondence should be addressed: [samuel.baker@balliol.ox.ac.uk](mailto:samuel.baker@balliol.ox.ac.uk)

## ABSTRACT

In fluid mechanics research, data gathered from measurements and simulations may be challenging to interpret due to complexities such as transience, non-linearity, and high dimensionality. Velocity data from the airflow through an internal combustion engine often exhibit such properties; nevertheless, accurate characterizations of these airflows are required in order to correctly predict and control the subsequent combustion and emission processes in pursuit of net zero targets. The temporal mean is a common way of representing an ensemble of realizations of velocity fields, but the averaging process can artificially diminish the magnitudes of the resultant vectors. Accurate representation of these vector magnitudes is of particular importance, as the velocity magnitudes in the intake airflow are thought to be primary drivers of the subsequent variation in an engine flow, which influences emission formation and overall efficiency. As an alternative to the ensemble mean, this work proposes the application of a dimensionality reduction method known as the sparsity-promoting dynamic mode decomposition (SPDMD), which can extract core structures from an ensemble of velocity fields while retaining more realistic vector magnitudes. This is demonstrated for the first time with large-eddy simulation (LES) velocity data and compared to a corresponding set of experimental particle image velocimetry (PIV) data. The SPDMD 0 Hz modes are shown to be more representative of the velocity magnitudes present in both datasets. This facilitates more accurate quantification of the differences in vector magnitudes between simulations and experiments, and more reliable identification of which LES snapshots are closer to the PIV ensemble.

© 2024 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1063/5.0189368>

## I. INTRODUCTION

Due to the stochastic nature of turbulence, achieving a realization-by-realization match between velocity fields from high-fidelity simulations (i.e., large-eddy simulations and direct numerical simulations) and experiments is unlikely, so statistical comparisons are often utilized instead.<sup>1</sup> The ensemble mean (EM) velocity field is typically used to provide a representation of the general patterns in a collection of velocity fields and to compare one dataset to another. Naturally, the ensemble averaging calculation for velocity fields can lead to the cancellation of opposite vectors, which may diminish the overall vector magnitudes in the EM field. Therefore, care is needed when attempting

to use the EM to interpret the velocity magnitude information in a dataset, particularly when there are high levels of variation in the flow. This vector-cancelling feature of the EM is prevalent in the study of the highly turbulent airflow through internal combustion engines (ICEs) where cycle-to-cycle variations (CCVs) are often present. Parsimonious representations of these flows are needed in order to fairly compare results from computational fluid dynamics (CFD) simulations to measurements taken from particle image velocimetry (PIV). Accurate validation of cold flow simulations is of particular interest as the intake airflow is a major source of kinetic energy in the engine cylinder,<sup>2,3</sup> significantly influencing subsequent CCVs,<sup>4,5</sup>

combustion quality,<sup>6,7</sup> wall heat transfer,<sup>8</sup> and ultimately engine efficiency.<sup>9</sup> In particular, Van Dam and Rutland<sup>10</sup> showed that variability in engine flows is primarily due to variable velocity magnitudes rather than variable velocity directions. It is therefore of utmost importance to explore methodologies that can produce simple representations of vector fields with magnitudes that accurately reflect the original data.

Studies have shown the deficiencies of simple averaging processes when attempting to create reliable representations of highly variational engine flows.<sup>11–13</sup> This is particularly true when strong CCVs cause each of the individual snapshots to appear differently from the overall ensemble average. This has motivated the use of dimensionality reduction methods, which are a branch of unsupervised machine learning techniques that can extract dominant structures from complex datasets, of which the proper orthogonal decomposition (POD) and dynamic mode decomposition (DMD) are perhaps the most widespread in engine research.<sup>14,15</sup> POD is a technique that can decompose a high-dimensional dataset into a hierarchy of low-dimensional structures known as modes. It was first introduced to the field of fluid dynamics by Lumley in 1967.<sup>16</sup> In an early application of POD to engine flows, Fogleman *et al.*<sup>17</sup> proposed two implementations of POD: phase-dependent and phase-invariant. The development of these two approaches was motivated by the nature of engines as cyclic devices. Engine flows evolve periodically and are therefore cyclostationary processes.<sup>18</sup> Records of weather and climate patterns, such as daily temperature measurements, are other prominent examples of cyclostationary processes. Climate data could be analyzed by comparing the temperature at the same month across several years, or between consecutive months in a single year. Similarly, engine flows can be analyzed by considering snapshots taken at a fixed phase across consecutive engine cycles (phase-dependent analysis) or snapshots taken at consecutive phases within the same engine cycle (phase-invariant analysis).

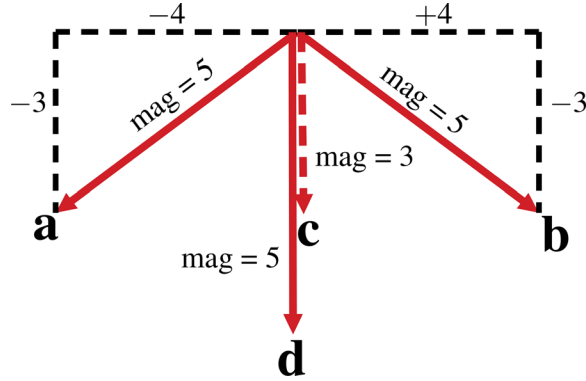
Phase-dependent POD provides a statistical description of the structures that appear at the same point in each cycle and has been used to quantify CCVs, characterize coherent structures, and investigate the turbulent energy cascade.<sup>11,19,20</sup> Phase-invariant POD can be used to track the progression of flow structures through consecutive time.<sup>21,22</sup> However, physically interpreting phase-invariant POD modes can be challenging. This is because engine flows are only cyclostationary rather than stationary, so they exhibit transient behavior within each cycle, and statistical descriptors such as the mean and variance lose a degree of physical significance.<sup>23,24</sup> Indeed, POD modes are only spatially coherent, so each spatial mode may consist of structures that are characterized by a mixture of different temporal dynamics.<sup>25</sup> A more appropriate tool for analyzing spatial and temporal dynamics together is the dynamic mode decomposition (DMD), developed by Schmid.<sup>26</sup> DMD produces spatial modes that each oscillate at a single frequency with an associated growth or decay rate, thus capturing the changes in spatial modes over time. DMD has been broadly applied to the study of flows,<sup>27–29</sup> but also in areas as diverse as epidemiology, medical imaging, and robotics.<sup>30</sup> One feature of DMD in its standard form is that it does not necessarily produce a reliable hierarchy of importance for the modes. The modes are scaled by an amplitude that is defined based on the first snapshot in the data; however, this relies heavily on the assumption that the snapshots evolve linearly in time from the initial condition, which may be only approximately valid for complex flows characterized by strong CCVs and measurement noise.<sup>31</sup> This motivated the development of an alternative form of DMD known as the

sparsity-promoting dynamic mode decomposition (SPDMD).<sup>32</sup> In SPDMD, a globally optimal solution for the mode amplitudes is calculated by solving an optimization problem that maximizes the sparsity of the solution while minimizing the error to the original data. SPDMD has recently been used to study a range of phenomena in fluid mechanics such as propeller wakes,<sup>33</sup> cavitation flows,<sup>34</sup> and vortex shedding.<sup>35</sup>

Studies of engine flow using any form of DMD are relatively scarce in the literature, possibly due to the difficulty in relating spatial structures to physical time in cyclostationary processes. For example, in order to fully interpret the results from a phase-dependent application of DMD, it is likely that one needs to assume that there is a link between flow fields at the same phase across consecutive cycles. However, a number of complex processes occur in the time it takes for an engine to return to the same phase, such as expansion, compression, and a replacement of the working fluid via the exhaust and fresh intake.<sup>9</sup> The degree to which the flow at a certain phase in one cycle is correlated with the flow at the same phase in the next cycle is therefore unclear. By assuming that the engine operates as a continuous process, Qin *et al.*<sup>36</sup> implemented phase-dependent DMD to investigate an engine flow, but did not explore the spatial structures of the DMD modes or relate them to physical features of the flow. An alternative phase-dependent implementation was proposed by Liu *et al.*<sup>37</sup> who treated each snapshot as an independent measurement. Therefore, they conducted a random permutation and averaging process to uncover the statistically global DMD features from cycle to cycle. Applications of DMD to engine flows are perhaps more naturally suited to phase-invariant approaches, where the physical links between consecutive snapshots are more apparent. Liu *et al.*<sup>38</sup> used phase-invariant DMD to investigate the dynamic properties of CCV structures. However, in order to bring physical meaning to the DMD modes, they required a combinatorial search to match the DMD modes to phase-invariant POD modes.

The core motivation of the present work is to produce a representation of a flow field that is analogous to using the vector **d** to represent the vectors **a** and **b** in Fig. 1. For this purpose, Shen *et al.*<sup>13</sup> compared phase-dependent POD to ensemble averaging and speed-based averaging. They concluded that among these options, POD-reconstructed flow fields provided the most suitable representations of the flow in the considered scenario. Fang *et al.*<sup>39</sup> highlighted the subjectivity in determining the threshold number for the number of POD modes retained in the reconstructions, and proposed the use of kernel principal component analysis (KPCA) as a more objective solution. Most recently, Baker *et al.*<sup>40</sup> analyzed a set of PIV data from an engine flow with SPDMD. The authors found that the 0 Hz SPDMD mode [also known as the direct current (DC) mode] was capable of extracting the average structure from the PIV data while retaining velocity magnitudes that were more representative of the original dataset. However, this work was limited to the use of a single realization of Reynolds-averaged Navier–Stokes (RANS) simulations for comparison with the PIV validation set, therefore restricting the potential utility of SPDMD.

The present study improves upon this work by exploring for the first time the ability of SPDMD to assess the similarity of two datasets that each contain multiple realizations, namely, PIV data from the well-known Darmstadt engine<sup>41,42</sup> and velocity data from large-eddy simulations (LES) of the same engine.<sup>43,44</sup> The previous work on this LES data developed validation methodologies that utilized ensemble averaging and POD; the use of the SPDMD in the present study extends these by facilitating a closer consideration of the velocity



**FIG. 1.** Diagram showing two vectors **a** and **b** with equal magnitudes, pointing in different directions. An arithmetic average of **a** and **b** gives the vector **c**, which has a diminished magnitude due to the cancellation of horizontal velocity components. The vector **d** represents a potentially more desirable form of the average, which can be achieved with dimensionality reduction methods. In the diagram, “mag” refers to the vector magnitude.

magnitudes in the data. Therefore, new insights are revealed about how to validate LES simulations against experimental data, quantify the similarity of the simulations to the PIV data, and locate sources of error in the datasets.

First, the representation quality of the ensemble mean (EM) flow fields is assessed for both the LES and PIV data. The SPDMD 0 Hz modes are also explored, and this method is shown to yield more representative velocity magnitudes in both cases. The impact of using the SPDMD 0 Hz modes to represent the flows rather than the EM is then explored. The effect of two important parameters in the DMD algorithm, namely, the mode threshold number and the total number of snapshots, is investigated in this application to aid and inform practitioners in the implementation of the SPDMD method in engine-relevant flows.

## II. NUMERICAL METHODS

### A. DMD and SPDMD

This section provides a brief overview of the aspects of DMD that are most relevant to this study. For a more detailed description, readers are guided toward Refs. 26, 30, and 31. In short, the DMD algorithm is capable of organizing high-dimensional datasets into discrete modes that are coherent in both space and time. This is achieved by taking two time-shifted data matrices as inputs ( $\mathbf{x}_k$  and  $\mathbf{x}_{k+1}$  for a time step  $k$ ), and finding a best-fit linear approximation of an operator (**A**) that maps one time-shifted matrix onto the other

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k. \quad (1)$$

A low-dimensional representation of the data can be constructed by taking a sum of three components: first, the leading spatial modes, given by the eigenvectors of **A** (denoted as **Φ**); second, corresponding frequencies of oscillation and growth/decay rates, governed by the eigenvalues of **A** (denoted as **Λ**); and finally, an amplitude vector that scales the modes and produces a hierarchy of importance among them (denoted as **α**)

$$\mathbf{x}_{k+1} \approx \sum_{m=1}^p \phi_m \lambda_m^k \alpha_m = \mathbf{\Phi} \mathbf{\Lambda}^k \mathbf{\alpha}. \quad (2)$$

The values of **Φ** and **Λ** are typically obtained by conducting a singular value decomposition (SVD) of the input data matrix; see Kutz and Brunton<sup>30</sup> for more details. The definition of the mode amplitudes can vary between different implementations of DMD. Most simply, the amplitudes can be calculated by setting  $k=0$  and taking an inverse of the **Φ** matrix. However, this approach relies heavily on the assumption that each snapshot of data evolves linearly from the previous snapshot,<sup>32</sup> which is likely too strong for phase-dependent implementations of DMD. In addition, this amplitude definition is not very robust to outlier measurements, which can contribute to large mode amplitudes and high decay rates due to their time-local nature.<sup>31</sup> In fluid velocity data, outliers are not usually representative of the full time series of measurements or reflective of the most important flow structures. Therefore, an alternative amplitude definition is sought.

Jovanović *et al.*<sup>32</sup> proposed a different amplitude definition in the sparsity-promoting DMD (SPDMD) algorithm. In this approach, the mode amplitudes are defined such that they reflect the degree of influence that the modes have over the full time series. This is done by attempting to reconstruct the original data with a reduced number of modes and calculating the reconstruction error. Mathematically, this is achieved by minimizing the sum of the reconstruction error and the sparsity of the solution

$$\underset{\alpha}{\text{minimize}} \quad J(\alpha) + \gamma \text{card}(\alpha), \quad (3)$$

where **α** is the unknown vector of amplitudes.  $J(\alpha)$  is an objective function representing the reconstruction error, for which the derivation can be found in Ref. 32. The operator “card” represents cardinality, measuring the number of non-zero elements in a set (also referred to as the sparsity). Therefore, Eq. (3) finds the fewest number of modes that retain the most amount of information in the dataset.  $\gamma$  is a user-defined regularization parameter that controls the trade-off between minimizing reconstruction error and sparsity.

However, calculating the cardinality requires a combinatorial search, which can quickly become intractable, so the  $L_1$ -norm is used as a convex relaxation

$$\underset{\alpha}{\text{minimize}} \quad J(\alpha) + \gamma \sum_{i=1}^m |\alpha_i| \quad (4)$$

for a total of  $m$  modes. This is a mixed-norm problem, as the reconstruction error  $J(\alpha)$  is calculated with the  $L_2$ -norm. Instead of attempting to solve this joint optimization problem simultaneously, a more efficient approach is employed, known as the alternating direction method of multipliers (ADMM).<sup>45</sup> In order to use the ADMM, there needs to be a separation of variables, which is achieved with the introduction of a dummy variable **β**,

$$\underset{\alpha}{\text{minimize}} \quad J(\alpha) + \gamma \sum_{i=1}^m |\beta_i|, \quad (5)$$

subject to  $\alpha - \beta = 0$ . Then, the augmented Lagrangian is introduced

$$\begin{aligned} \mathcal{L}_\rho(\alpha, \beta, \lambda) := & J(\alpha) + \gamma \sum_{i=1}^m |\beta_i| \\ & + \frac{1}{2} (\lambda^* (\alpha - \beta) + (\alpha - \beta)^* \lambda + \rho \|\alpha - \beta\|_2^2), \end{aligned} \quad (6)$$



where  $\lambda$  is the vector of Lagrange multipliers,  $*$  denotes the complex conjugate transpose, and  $\|\cdot\|_2$  is the  $L_2$ -norm.  $\rho$  is a user-defined parameter that applies a penalty to the difference between  $\alpha$  and  $\beta$ . With the Lagrangian defined, the ADMM can be applied with three steps: minimize  $\alpha$  with a fixed  $\beta$ ; update  $\alpha$  and then minimize  $\beta$ ; and update  $\lambda$ .

$$\begin{aligned}\alpha^{k+1} &:= \arg \min_{\alpha} \mathcal{L}_{\rho}(\alpha, \beta^k, \lambda^k), \\ \beta^{k+1} &:= \arg \min_{\beta} \mathcal{L}_{\rho}(\alpha^{k+1}, \beta, \lambda^k), \\ \lambda^{k+1} &:= \lambda^k + \rho(\alpha^{k+1} - \beta^{k+1}).\end{aligned}\quad (7)$$

In this way, ADMM provides an efficient solution to Eq. (5) by solving each optimization problem separately, switching between minimizing the reconstruction error (min  $\alpha$ ) and minimizing the sparsity (min  $\beta$ ). The algorithm begins with an initial point  $(\beta^0, \lambda^0)$  and iterates until two user-defined tolerances  $\varepsilon_1$  and  $\varepsilon_2$  are reached

$$\|\alpha^{k+1} - \beta^{k+1}\|_2 \leq \varepsilon_1 \quad \text{and} \quad \|\beta^{k+1} - \beta^k\|_2 \leq \varepsilon_2. \quad (8)$$

## B. Implementation

In this study, phase-dependent SPDMD is used to extract average (0 Hz) flow structures with magnitudes that are more representative of the overall datasets. Following standard DMD procedure,<sup>30,40</sup> a given set of PIV or LES data is arranged into a 2D matrix with each row representing a new measurement location in space ( $n$ ) and each column a new cycle/snapshot in time ( $m$ ). The full matrix is then split into two time-shifted matrices, one containing the first  $1 : M - 1$  columns and the other containing the last  $2 : M$  columns, where  $M$  is the total number of snapshots.  $M = 250$  for the PIV data and  $M = 50$  for the LES data.

The time-shifted matrices are fed into the SPDMD algorithm, yielding the modes, eigenvalues, and optimal amplitudes. The MATLAB code provided in the supplementary material of Ref. 32 was adapted for this purpose. Default values were kept for the user-defined parameters in the SPDMD algorithm:  $\rho = 1$ , max ADMM iterations = 10 000,  $\varepsilon_1 = 1 \times 10^{-6}$ ,  $\varepsilon_2 = 1 \times 10^{-4}$ . Jovanović *et al.*<sup>32</sup> recommend performing a sweep over  $\gamma$  across several orders of magnitude in order to determine a suitable value for the problem at hand. In the present study, a  $\gamma$  sweep is conducted and  $\gamma = 1$  is used, as discussed in Sec. VB. The steps involved in extracting the average structures using SPDMD are outlined in Procedure 1.

**PROCEDURE 1** Methodology for using SPDMD 0 Hz modes with velocity data.

- 1: Establish the input data matrices, with each column representing a new snapshot and each row representing a new location in space (see Refs. [30, 40]).
- 2: Decide on a threshold number for the number of modes.
- 3: Feed the data matrices into the SPDMD algorithm [32].
- 4: Identify the modes oscillating at 0 Hz ( $\phi_0$ , corresponding to  $\text{imag}(\log(\Lambda)) = 0$ ).
- 5: Reconstruct the 0 Hz flow field as the product  $\phi_0 \lambda_0 \alpha_0$  (from Equation 2).

## C. Histogram distance

This work considers the similarity of vector magnitudes between two different velocity fields. For a given vector field, the vector magnitude distribution can be visualized as a histogram. There are a number of different metrics available for measuring the similarity between two histograms; this study adopts the histogram distance (HD), also known as the histogram intersection, which is a method that has been widely used in image processing and computer vision applications.<sup>46</sup> This technique has the flexibility to account for the spatial distributions of a variety of non-normally distributed data, as well as being intuitive to understand and simple to represent visually. The HD quantifies the degree of overlap between the areas of two histograms and is calculated as follows:

$$\text{HD} = \frac{\sum_i (\min(h_1(i), h_2(i)))}{\min(\|h_1(i)\|, \|h_2(i)\|)} \quad (9)$$

for two histograms  $h_1$  and  $h_2$  evaluated at the  $i$ th bin. The value of the HD can vary between 0, indicating no overlap and therefore no similarity, and 1 for identical histograms. Note that the HD may be affected by the presence of outliers, and the degree of this sensitivity can be dependent on the bin size chosen for the histograms. In the present work, the bin sizes for the two crank angles studied were chosen as subjective compromises between producing enough bins to accurately represent the distributions of the data without becoming overly sensitive to outliers. For 470 CAD aTDCf, the bin size was chosen as 0.5 m/s between minimum and maximum values of 0 and 25 m/s, for a total of 50 bins. For 700 CAD aTDCf, the bin size was 0.2 m/s between minimum and maximum values of 0 and 8 m/s, for a total of 40 bins.

## III. EXPERIMENTAL SETUP

The particle image velocimetry (PIV) measurements used in this study were taken from the Darmstadt engine, which is an optically accessible single-cylinder engine from the Technische Universität Darmstadt (TUD).<sup>41</sup> Optical access to the combustion chamber is provided via a transparent piston and cylinder liner. The PIV data were gathered at the “OP. A” engine operating point, which is characterized by an engine speed of 800 rpm and an intake pressure of 0.95 bar. The piston crevice volume, 77.6 mm long and 0.5 mm thick, is known to impact the effective compression ratio. Table I contains further details of the setup.

TABLE I. Engine operating conditions.

Parameter	Description
Engine head	Spray-guided
Operating point	OP. A.
Displacement volume	499.6 cc
Bore $\times$ Stroke	86 $\times$ 86 mm
Compression ratio	8.7:1
Engine speed	800 rpm
Average intake pressure	0.95 bar
Average intake temperature	23.9 °C
Analysis crank angles	470; 700 CAD aTDCf

The PIV setup is detailed in Baum *et al.*<sup>41</sup> and summarized here. The PIV data were gathered by imaging the Mie scattering of the seeding particles, which were silicone oil droplets measuring  $\approx 1 \mu\text{m}$  in diameter. The seeding density was optimized to produce between 8 and 10 particle pairs in the final interrogation window size. The raw images were processed with the DaVis, LaVision software using a multi-pass algorithm with a decreasing window size from  $96 \times 96$  to  $32 \times 32$  with a 75% overlap. The cross-correlation peak values varied between 0.4 and 0.9. Vector post-processing was conducted with a peak ratio factor of 1.8 and a median filter operation to remove spurious vectors. The spatial resolution was 0.6023 mm based on the final interrogation window size. Overall, the accumulated uncertainty in the velocity calculation was 5%.

The PIV dataset consists of 250 snapshots of consecutive cycles measured on the so-called valve plane. This plane is offset from the cylinder center by 19 mm in order to align with the intake valves, as shown in Fig. 2. PIV measurements were taken every five crank angle degrees (CAD) from 360 to 720 CAD after firing the top dead-center (aTDCf). This study focuses on the results at two fixed phases, namely, 470 and 700 CAD aTDCf. At 470 CAD, the intake valves are at their maximum lift, and the intake jet can be observed, which is an important physical phenomenon in engine research. 700 CAD is close to a typical spark timing and therefore represents one of the final states of fuel-air mixing before combustion, where typically a higher CCV is present in both the experiments and large-eddy simulations. The time between images taken in consecutive cycles was 0.15 s based on an engine speed of 800 rpm.

#### IV. COMPUTATIONAL SETUP

3D turbulent flow simulations of the air through the engine cylinder were conducted with SimCenter STAR-CCM+ In-Cylinder Solution v2020.2, licensed by Siemens DISW. Turbulence was modeled within the large-eddy simulation (LES) framework, via the dynamic Smagorinsky subgrid-scale model<sup>47</sup> with an all- $y^+$  wall treatment, following indications from.<sup>48</sup> The all- $y^+$  wall treatment blends high- and low- $y^+$  approaches with an exponential weighing function; it is therefore designed for meshes where pointwise control of the near-wall cell centroid cannot be guaranteed and situations where the centroid falls in the buffer layer cannot be avoided (i.e.,  $1 < y^+ < 30$ ).

To improve computational efficiency, the overall domain was restricted to the space between the intake and exhaust pressure probes, which were treated as inlet and outlet boundaries. Time-dependent pressure and temperature boundary conditions were sourced from a 1D GT-Power model of the engine, which was previously validated in Refs. 43 and 49. Notably, no artificial perturbations were introduced into the boundary conditions for the LES simulations. This decision was based on findings from Fontanesi *et al.*,<sup>50</sup> which demonstrated that the turbulence generated by the airflow over the intake valves was in fact the primary cause of variability in the subsequent flow fields.

The PISO algorithm was used for pressure-velocity coupling. The pressure and temperature discretization schemes were a bounded central second-order scheme and a second-order monotone advection and reconstruction scheme (MARS), respectively. The ideal gas equation was used as the equation of state. Dynamic viscosity was calculated using Sutherland's law. Heat transfer across the domain wall boundaries was calculated with blended wall functions using the so-called two-layer all- $y^+$  wall treatment, for which the wall heat flux  $\dot{q}_w''$  is computed as follows:

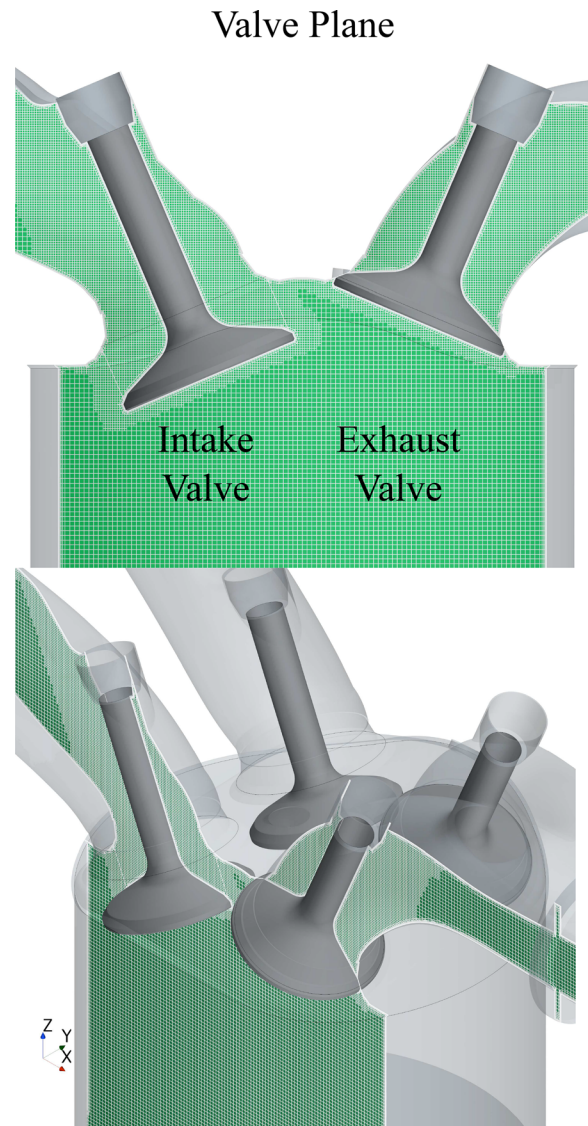


FIG. 2. Diagram showing the valve plane in green for both an in-plane view (top) and an isometric view (bottom). The plane is offset from the cylinder center, cutting through an intake valve and an exhaust valve on one side of the engine cylinder.

$$\dot{q}_w'' = \rho C_p u^* \frac{\hat{T}_C - T_w}{\hat{T}_C^+}, \quad (10)$$

where  $\rho$  is the density,  $C_p$  is the specific heat,  $u^*$  is the velocity scale (calculated iteratively),  $\hat{T}_C$  is the cell temperature,  $T_w$  is the wall temperature, and  $\hat{T}_C^+$  is the non-dimensional filtered cell temperature computed based on Kader's law. More details can be found in Ref. 51.

The meshing tool utilized a cell cut-and-trim approach, along with a prismatic mesher for the near-wall grid. A morphing technique was utilized to account for the movement of engine components. The computational grid, shown in Fig. 2, was primarily hexahedral with a core grid size of 0.75 mm in the cylinder. Fixed and moving control

volumes were added near the valves to refine the grid size to 0.375 mm, while the mesh in the ports was coarsened to 1.5 mm. Eight prismatic cell layers were applied to all walls, where the first layer measured 10  $\mu\text{m}$  and the total thickness was set to 0.6 mm.

Overall, the maximum number of cells was approximately  $11 \times 10^6$  (of which  $5.1 \times 10^6$  cells were located inside the cylinder) at bottom dead-center (BDC), while the cells at top dead-center (TDC) numbered around  $5.5 \times 10^6$  (with  $2.7 \times 10^6$  inside the cylinder). The computational domain included a slight extension of the crevice volume, which reduced the compression ratio by just under 5%. This choice was made following extensive investigations reported in Refs. 49 and 52 to account for the unknown levels of blow-by losses in the physical engine due to the crevice volume.

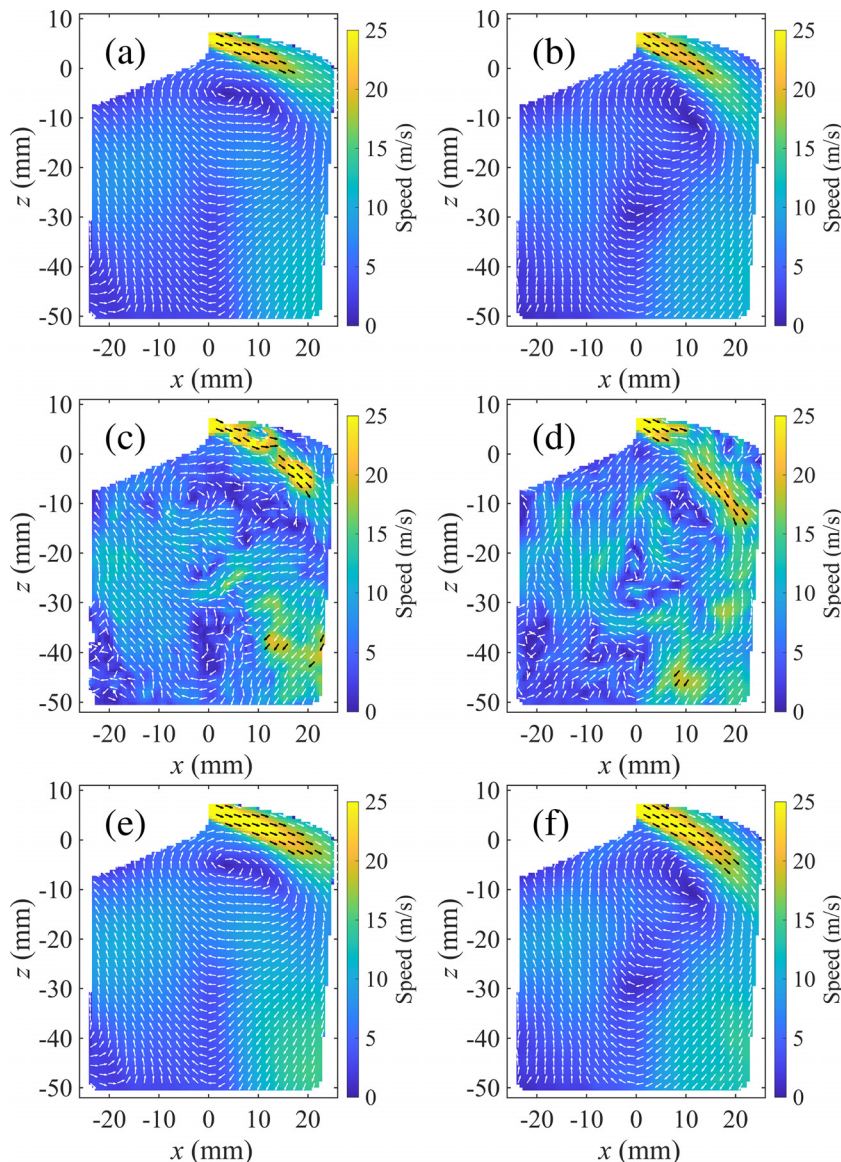
The simulations were initialized with a two-step approach. First, four consecutive Reynolds-averaged Navier–Stokes (RANS) simulation

cycles were performed to remove the effect of the initial conditions and to reach cyclic convergence. Second, a single LES cycle was run to obtain a consistent initial field for the subsequent simulations. Once initialized, 50 consecutive LES cycles were run, providing the LES dataset that was analyzed in this study. Instantaneous 2D velocity components were recorded at each of the grid points on both the valve plane and were then interpolated onto the coarser PIV grid for comparison against the experimental data.

## V. RESULTS

### A. Ensemble mean analysis

The ensemble mean (EM) flow fields from the PIV and LES at 470 CAD aTDCf are shown in Figs. 3(a) and 3(b), respectively. In both cases, the intake jet is shown as the high-speed yellow portion of the flow to the top-right of the flow fields. The two intake jets show a



**FIG. 3.** Flow fields at 470 CAD aTDCf. The plots show (a): the PIV ensemble mean (EM), (b): the LES EM, (c): an arbitrary PIV snapshot, (d): an arbitrary LES snapshot, (e): the PIV SPDMD 0 Hz mode truncated at 57 modes, (f): the LES SPDMD 0 Hz mode truncated at 14 modes. One in every three vectors is plotted for clarity. The colormap is scaled by the vector magnitudes, which we refer to as “speed” in this paper to emphasize the scalar nature of this quantity.



qualitative match in terms of length and magnitude; however, the resulting vortex center is positioned further to the right in the EM LES field. This points to the possible existence of differences in the intake jet that are not immediately made apparent by the EM fields.

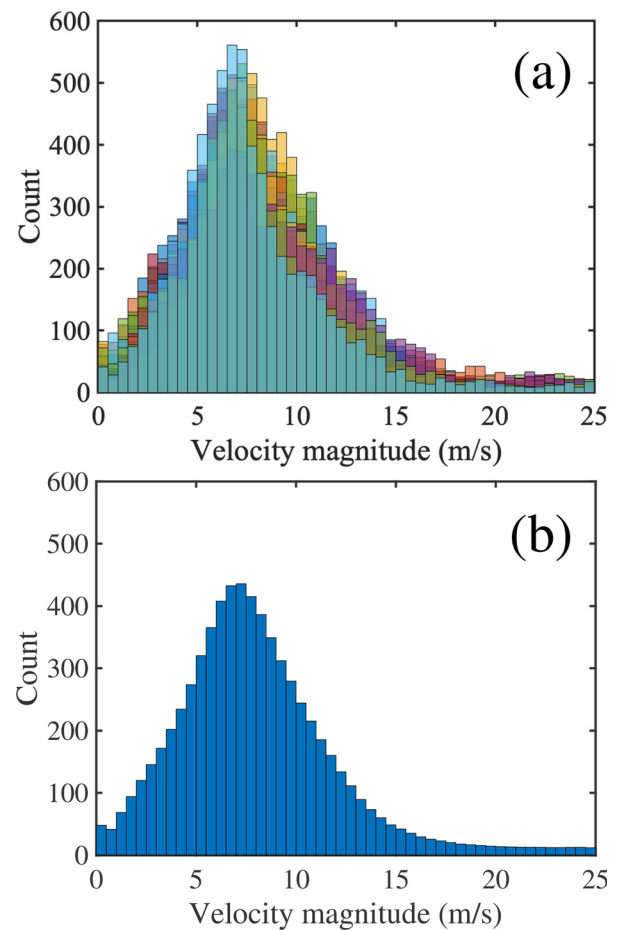
In order to further examine the intake jet, single snapshots from PIV and LES are shown in Figs. 3(c) and 3(d), respectively. In these images, the intake jets are shown to extend further into the flow fields for both cases. The fact that the intake jets are shorter in the EM fields suggests that the tips of the intake jet are more variable from snapshot to snapshot, causing them to get smoothed over in the EM fields. The single LES snapshot also shows the intake jet pointing further downward than the jet in the single PIV image, which may contribute to the different vortex center locations. Finally, a faster recirculation zone can be seen in the bottom-right of both of the single PIV and LES images that is not obvious in either EM field, suggesting that this section is another area of significant variability.

These differences between the EM fields and the single snapshots call into question the ability of the EM fields to fairly represent the instantaneous velocity magnitudes present in the original data. The quality of representation for the velocity magnitudes in the dataset can be visualized by considering histogram plots of the velocity magnitude (or flow speed) distributions. These magnitude histograms measure the number of occurrences of a particular magnitude bin, which is strictly positive, so they can be averaged without risking the cancellation of opposites that comes with averaging vectors directly. The result is a “cycle set histogram” that represents the velocity magnitudes across the entire dataset. Figure 4 shows the flow speed histograms for the first 20 individual PIV snapshots panel (a) and the average flow speed for the full 250 PIV ensemble panel (b), with the latter being used as the “ground truth” for the PIV data at this crank angle.

With this method, the EM velocity magnitude distributions can then be compared to their respective cycle set histograms, shown on the top row of Fig. 5 for both the PIV panel (a) and LES panel (b) flow fields. In both cases, the EM histograms in purple display a left skew (to lower velocity magnitudes) when compared to the cycle set histograms in blue. This is because the superposition of vectors in the EM calculation results in vector cancellation and a systematic under-prediction of the true velocity magnitudes. This effect can be quantified via the histogram distance (HD), which calculates the relative size of the overlapping area between two histograms.  $HD = 0.81$  between the PIV EM and the PIV cycle set, and  $HD = 0.82$  between the LES data and its ensemble average. The ability of the EM to represent velocity magnitudes depends on the variability in the dataset, so the HD values will decrease for more variable crank angles and test points. For example, the results for another important crank angle in this engine, 700 CAD aTDCf, are presented in Fig. 6. For this 700 CAD case, disparities can be seen between the PIV and LES EM fields [panels (a) and (b)] and single snapshots [panels (c) and (d)], particularly toward the bottom of the flow fields where there is a relatively strong recirculation zone off the top of the upward-moving piston. As a result, the HD between the EMs and the respective cycle set histograms is lower, at  $HD = 0.71$  for the PIV and  $HD = 0.75$  for the LES.

## B. SPDMD analysis

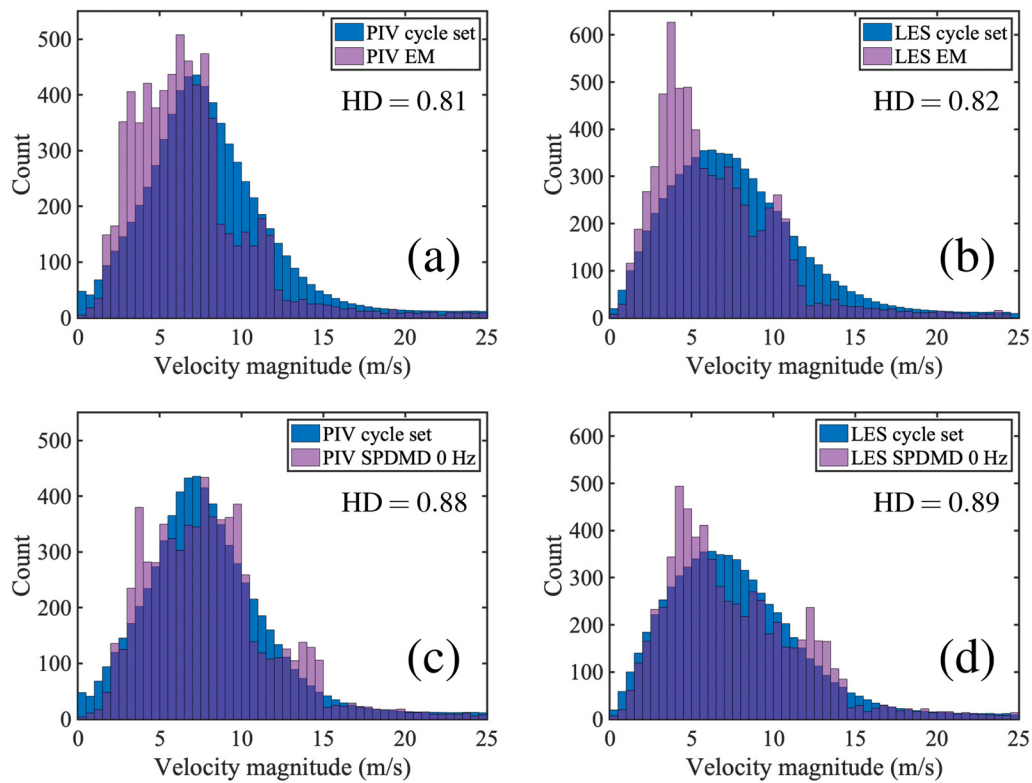
The 0 Hz (DC) modes from the SPDMD algorithm were then explored as alternatives to the EM. The choice of the  $\gamma$  parameter in the SPDMD algorithm had a negligible effect on the velocity



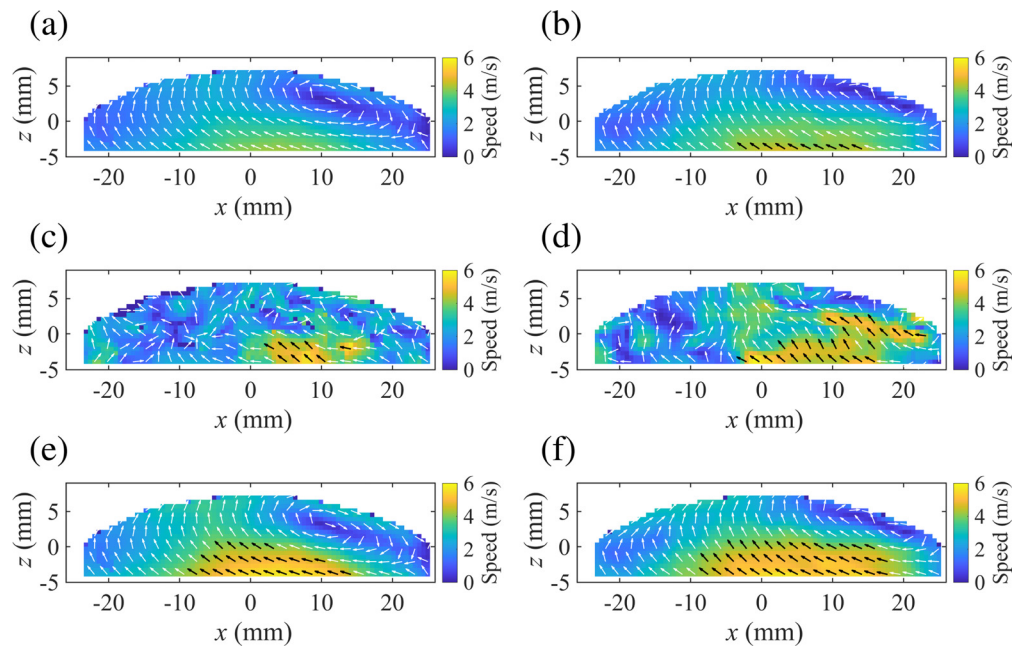
**FIG. 4.** Velocity distributions plotted as histograms for (a): the first 20 PIV snapshots, where the histogram for each individual snapshot is plotted in a different color and superimposed atop one another, and (b): the average of all 250 PIV histograms, dubbed the “cycle set histogram.” The data are from 470 CAD aTDCf.

magnitudes given by the 0 Hz modes for the 250-cycle PIV dataset. This is because  $\gamma$  controls the number of modes that are retained through the sparsity promotion, but this study is only concerned with a single mode (the 0 Hz mode). Indeed, for a fixed mode threshold number of 50 modes, increasing  $\gamma$  from 0.01 to 10000 only changed the HD between the resultant 0 Hz modes by 0.005. Therefore,  $\gamma$  was simply set to one for the remainder of the study.

Use of the singular value decomposition (SVD) in the DMD algorithm means that the DMD analysis can also be sensitive to the truncation or hard threshold for the number of DMD modes retained in the analysis.<sup>53</sup> In physical situations, choosing a threshold that retains only the most dominant modes can be akin to denoising the data, as dominant modes are more likely to contain a coherent signal, while the lesser modes often contain measurement noise and outlier data.<sup>54</sup> When using SVD-based approaches for PIV data, a threshold should be chosen that includes enough modes to fully capture the behavior of the system, without losing generality due to the influence of outliers. Objectively choosing a threshold for physical systems is a challenging task, as the most appropriate value can be case-dependent.<sup>39</sup>

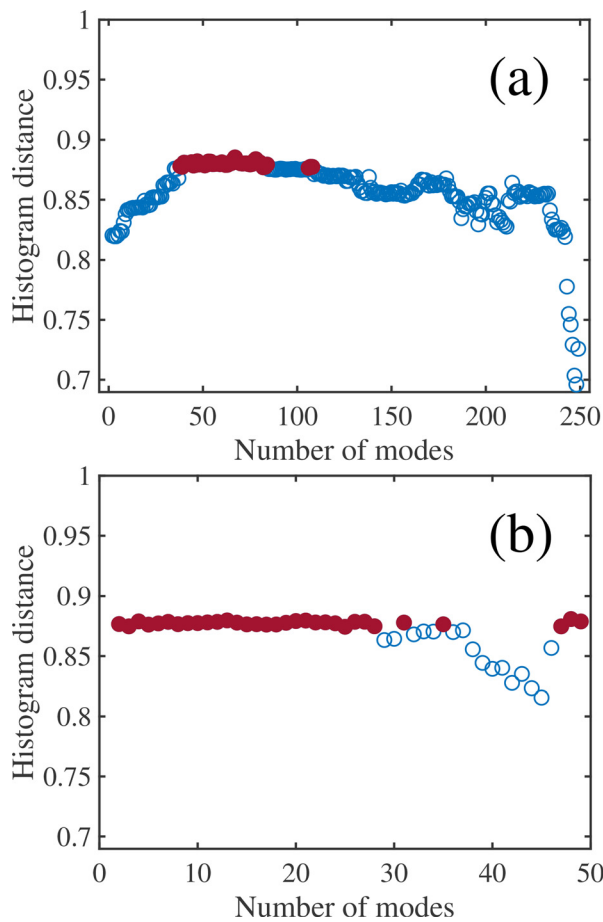


**FIG. 5.** Comparisons between the average velocity histograms at 470 CAD aTDCf and the velocity distributions of (a): the PIV EM, (b): the LES EM, (c): the PIV SPDMD 0 Hz mode, and (d): the LES SPDMD 0 Hz mode. Histogram distances (HD) for each case are also reported.



**FIG. 6.** Flow fields at 700 CAD aTDCf. The plots show (a): the PIV ensemble mean (EM), (b): the LES EM, (c): an arbitrary PIV snapshot, (d): an arbitrary LES snapshot, (e): the PIV SPDMD 0 Hz mode truncated at 189 modes, (f): the LES SPDMD 0 Hz mode truncated at 10 modes. One in every three vectors is plotted for clarity.





**FIG. 7.** Sweeps calculating the HD between the SPDMD 0 Hz mode and the cycle average histograms for varying mode threshold numbers for (a): PIV data and (b): LES data at 470 CAD aTDCf. The filled red circles represent HD values that are within 1% of the maximum HD value.

Therefore, in this work, sweeps over the choice of mode threshold number were conducted for both the PIV and the LES data, to find the threshold that resulted in the 0 Hz modes with the largest HD against the average histograms. The results for 470 CAD are shown in Fig. 7, where the filled red circles represent thresholds with high HDs that were within 1% of the maximum HD found in each sweep. For the PIV data, a continuous streak of high HD values can be found between mode thresholds of 38 and 84. For the LES data, there was a longer continuous streak of high HDs found between 2 and 28 modes. In both cases, therefore, there were many good options for the truncation number. As an attempt to provide consistency to the choice of threshold for the two datasets, the distribution of the singular values for each case was investigated, as plotted in Fig. 8. The mode thresholds of 38 and 84 correspond to 41% and 61% of the cumulative variance in the PIV dataset, while for the LES data, the thresholds of 2 and 28 modes correspond to a wider range of 22% and 74% of the variance. As a middling value for both cases, a criterion of 50% of the variance was used to select the threshold for each. The thresholds used for the SPDMD analysis at this crank angle were therefore 57 for the PIV

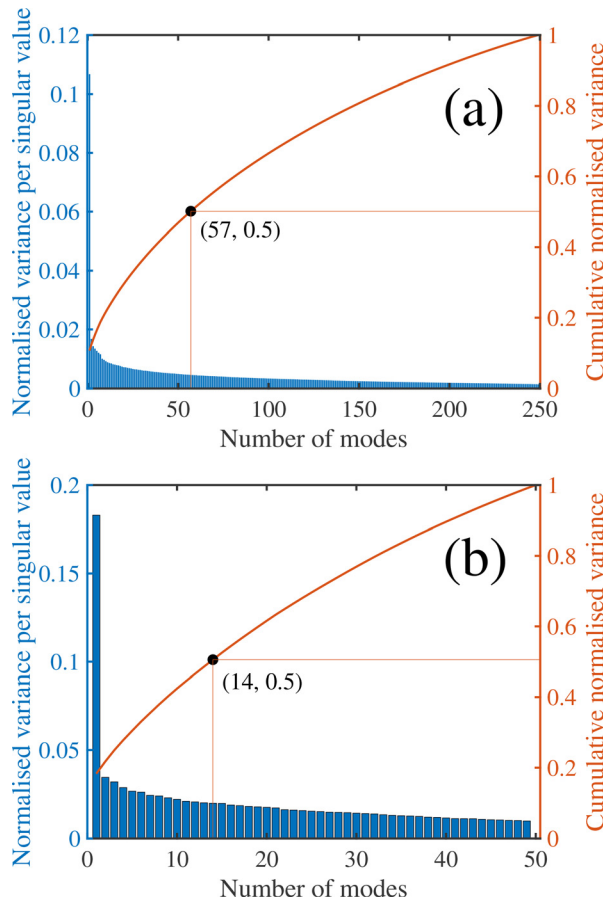
data, and 14 for the LES data, as marked in Fig. 8. Note that this criterion is not intended to be valid for other datasets or test points, and only serves as a method of selecting a single threshold from the variety of good options in this case.

The 0 Hz modes at the threshold numbers of 57 and 14 are plotted as flow fields in Figs. 3(e) and 3(f) for the 470 CAD case. In these fields, the ADMM algorithm has rescaled the average flow by optimizing the velocity magnitudes across each of the PIV and LES snapshots. As a result, the intake jets for the SPDMD 0 Hz fields have longer lengths and faster speeds. In addition, a difference in the angle of the intake jets is made more apparent, which is a potential cause of the different vortex center locations between the PIV and LES data. Finally, faster recirculation zones have been retained in the bottom-right corners of both SPDMD fields, which is also more representative of the dynamics in the individual PIV and LES snapshots. For 700 CAD, the HD sweeps for the PIV and LES data are shown in Fig. 9. Here, the results look vastly more disparate than for 470 CAD, which is a sign that the PIV and LES data have different levels of variability at 700 CAD. The optimal thresholds for this crank angle were 189 for the PIV data, corresponding to 93% of the cumulative variance, and 10 for the LES data, corresponding to 50% of the cumulative variance. The threshold for the PIV data at this crank angle is a fairly extreme result, and is potentially symptomatic of significant CCVs, as the low-order DMD modes are not very representative of the data and many modes are needed for an accurate reconstruction. The level of variation in the velocity magnitudes of these datasets is further explored later in Fig. 10. The SPDMD 0 Hz modes at thresholds of 189 and 10 are plotted in Figs. 6(e) and 6(f), respectively. The effect of using the modes is similar to the 470 CAD case, with higher velocity magnitudes being retained in the recirculation zones in particular.

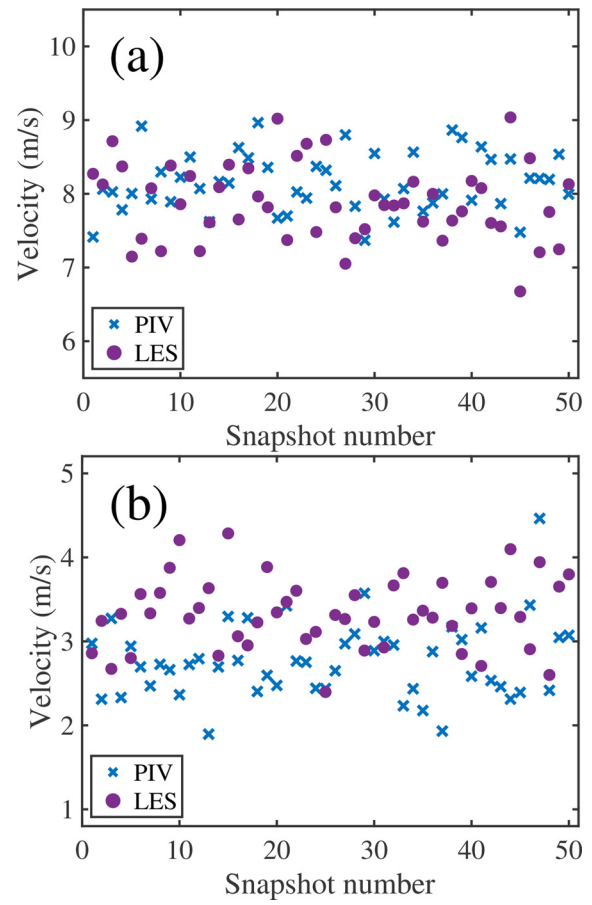
The corresponding histograms between the 0 Hz SPDMD fields and the full cycle sets are plotted along the bottom row of Fig. 5 for 470 CAD. Unlike the EMs, there is less overall skew for the SPDMD histograms in purple, indicating a more accurate representation of the velocity magnitudes in the dataset overall. A slight left skew for LES SPDMD histograms is observed here which could be an indication that more realizations might be needed for the particular condition. However, for the existing data, SPDMD shows a significant improvement in preserving the flow field magnitude information compared to the ensemble mean. For the PIV data,  $HD = 0.88$ , and for the LES data,  $HD = 0.89$ . The difference is more pronounced at 700 CAD, where the HDs are increased to  $HD = 0.93$  for the PIV data and  $HD = 0.85$  for the LES, as shown in Fig. 11. The increases in HD due to the SPDMD 0 Hz modes are consistent with the previous findings of Baker *et al.*<sup>40</sup> who investigated PIV data from a different optical engine and found that the use of the SPDMD 0 Hz modes was able to increase the HD from 0.57 with the EM to 0.89 with the SPDMD 0 Hz modes.

### C. Comparison of EM and SPDMD representations

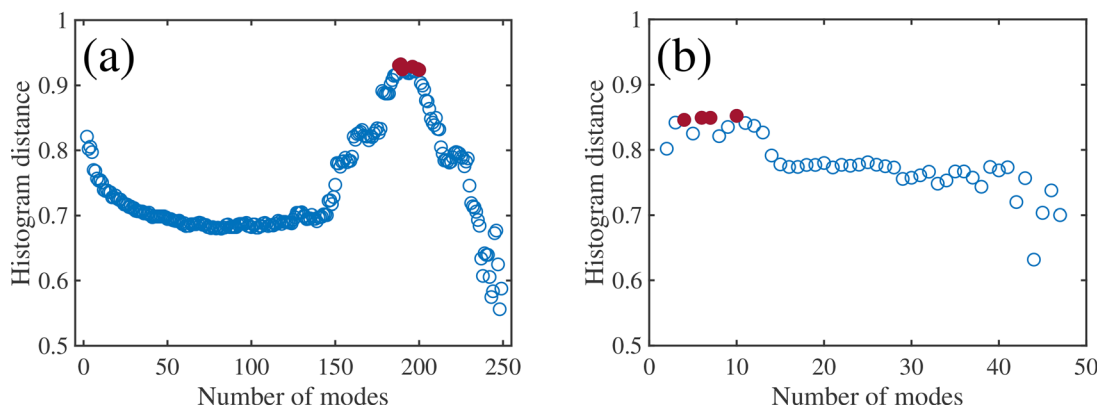
The impact of using either the EM or the SPDMD 0 Hz fields for comparing the PIV and LES data is shown in Fig. 12. In the case of 470 CAD, the overall shapes are the same regardless of whether the EM or SPDMD 0 Hz modes are chosen; indeed,  $HD = 0.83$  for both cases. This is because the SPDMD increased the HD of both the LES and PIV data by 0.07, so the histograms have been stretched along the  $x$  axis by similar amounts in the SPDMD comparison. However, for the 700 CAD case, the SPDMD 0 Hz modes have rescaled the PIV and LES data by different amounts, leading to the HD being 0.82 when EM



**FIG. 8.** Singular value decomposition plots at 470 CAD aTDCf for (a): 250 PIV snapshots, and (b): 50 LES snapshots. The variance captured in each singular value is given by the bar chart on the left axis, and the cumulative variance is given by the line graph on the right axis. The number of modes corresponding to 50% of the cumulative variance is marked for each dataset.



**FIG. 10.** Planar-averaged velocity magnitudes for the first 50 snapshots of PIV data (blue crosses), and all 50 of the LES snapshots (purple circles) for (a): 470 CAD aTDCf and (b): 700 CAD aTDCf. Only 50 PIV snapshots are shown for clarity.



**FIG. 9.** Sweeps calculating the HD between the SPDMD 0 Hz mode and the cycle average histograms for varying mode threshold numbers for (a): PIV data and (b): LES data at 700 CAD aTDCf. The filled red circles represent HD values that are within 1% of the maximum HD value.

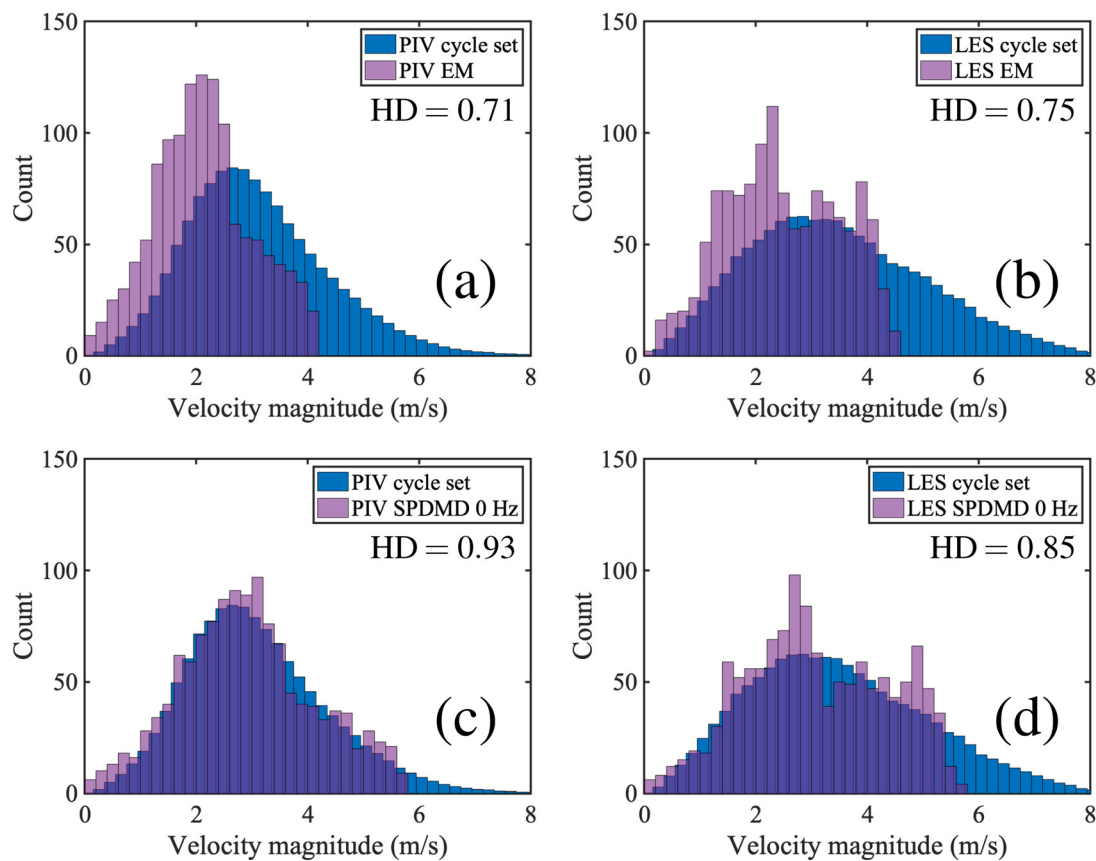
fields are compared, but 0.86 when SPDMD 0 Hz fields are considered. This is further evidence that the variabilities of the PIV and LES data are more similar at 470 CAD than at 700 CAD.

To test this hypothesis more directly, planar averages were taken of the velocity magnitudes for each of the single LES and PIV snapshots. The average velocities for each LES snapshot are plotted alongside the first 50 PIV snapshots in Fig. 10 for both crank angles. At first glance, the planar average velocity magnitudes appear to have a similar spread for both PIV and LES datasets at 470 CAD. Indeed, a quantitative comparison reveals that the relative standard deviations were similarly low at 4.6% for the PIV data and 6.6% for the LES. This supports the findings of Barbato *et al.*<sup>44</sup> who found that these experimental and LES datasets had similar levels of variability in the peak cylinder pressures. For 700 CAD, the relative standard deviations were more disparate, at 26.4% for the PIV data and 12.6% for the LES data. As a result, the SPDMD 0 Hz modes stretched the PIV histogram by a larger amount than the LES histogram, as shown in Fig. 12. Therefore, the size of the improvement in HD due to the SPDMD does appear to be affected by the level of variability in a dataset; furthermore, if two datasets have velocity magnitudes with different levels of variability, then it could be expected that the SPDMD would increase the magnitudes of one dataset more than the other, resulting in a change in the HD between the two SPDMD 0 Hz modes relative to the HD between the

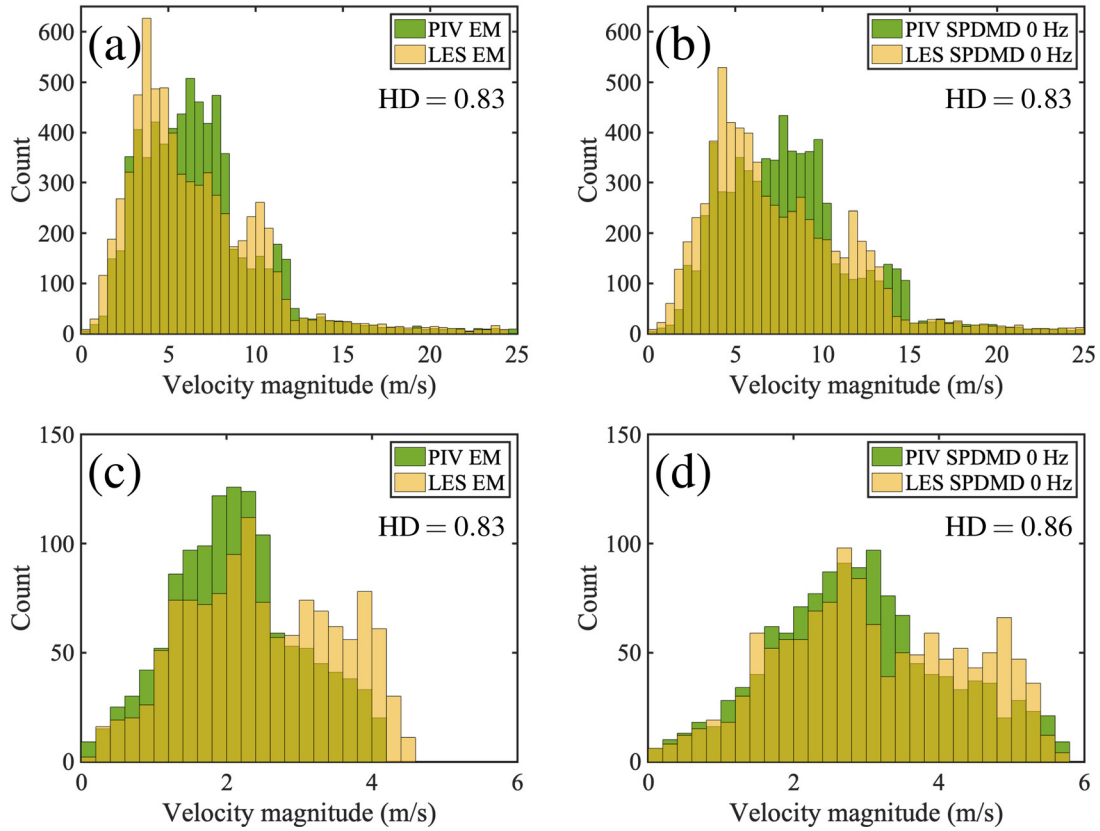
two EM fields. More evidence needs to be gathered to test this hypothesis, which will be further investigated in future work.

The effect of these re-scalings is demonstrated more clearly in Figs. 13 and 14, which show contour plots of the speed differences between the LES and PIV data at each point in the grid. Figure 13(a) compares the LES EM and PIV EM fields at 470 CAD, while Fig. 13(b) compares the respective SPDMD 0 Hz modes. Positive values indicate that the LES results under-predicted the PIV data, while negative values indicate under-predictions by the LES.

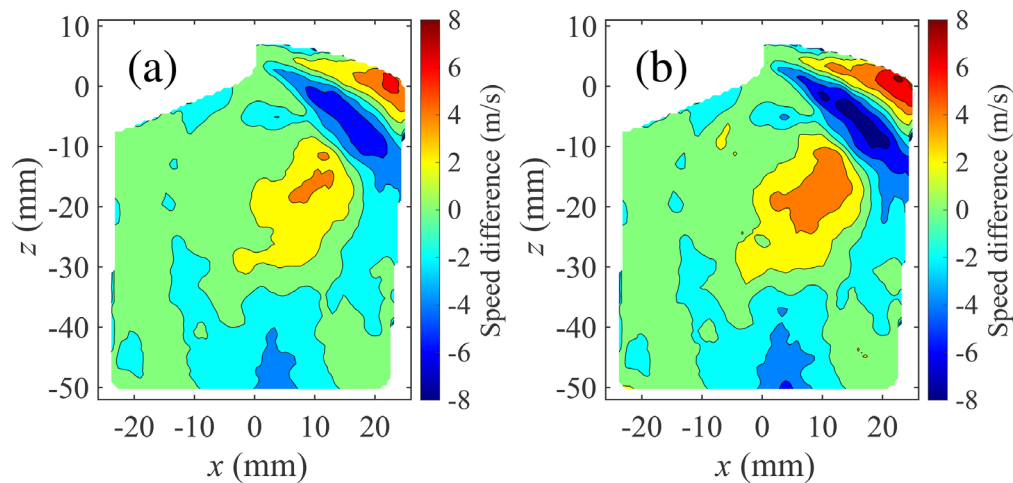
Similar regions of discrepancy are highlighted in both figures, with the largest speed differences being around the intake jet region, toward the top-right of the flow fields. However, the magnitudes of these discrepancies are different, even for a case where the SPDMD 0 Hz modes have rescaled the velocity magnitudes by similar amounts. For example, at 470 CAD, a comparison between EM fields would result in the conclusion that the LES results under-predicted the PIV data by at most 6 m/s in the region (20, 0), whereas the under-predictions reach as high as 8 m/s in the same region when the SPDMD 0 Hz modes are considered. In a similar vein, the EM field comparison shows that the LES over-predicted the PIV by 5 m/s in the region (15, -10), while the SPDMD 0 Hz modes indicate that this over-prediction was actually 7 m/s. This has consequences for the development of LES models, where it is important that sources of variability are correctly diagnosed, and



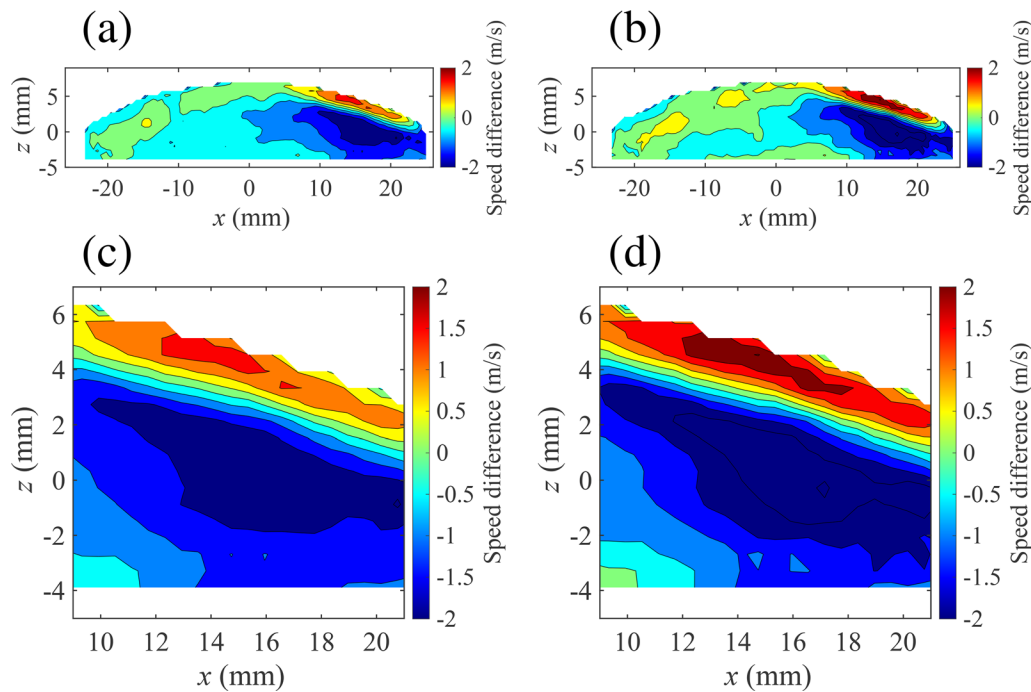
**FIG. 11.** Comparisons between the average velocity histograms and the velocity distributions at 700 CAD aTDCf of (a): the PIV EM, (b): the LES EM, (c): the PIV SPDMD 0 Hz mode, and (d): the LES SPDMD 0 Hz mode.



**FIG. 12.** Histogram comparisons among (a): the PIV and LES EMs at 470 CAD aTDCf, (b): the PIV and LES SPDMD 0 Hz modes at 470 CAD aTDCf, (c): the PIV and LES EMs at 700 CAD aTDCf, and (d): the PIV and LES SPDMD 0 Hz modes at 700 CAD aTDCf.



**FIG. 13.** Plots showing the differences in velocity magnitude between the PIV and the LES data for (a): the PIV and LES EM fields, and (b): the SPDMD 0 Hz modes at 470 CAD aTDCf. The speed difference is calculated as PIV speeds–LES speeds, so that positive values represent regions where the LES under-predicted the PIV data, and negative values indicate over-predictions by the LES. Each contour line represents 2 m/s of speed difference.



**FIG. 14.** Plots showing the differences in velocity magnitude between the PIV and the LES data for (a) and (c): the two EM fields, and (b) and (d): the two SPDMD 0 Hz modes at 700 CAD aTDCf. The speed difference is calculated as PIV speeds–LES speeds, so that positive values represent regions where the LES under-predicted the PIV data, and negative values indicate over-predictions by the LES. Each contour line represents 0.5 m/s of speed difference.

that any discrepancies are accurately quantified in order to ensure that LES models are faithfully capturing the dynamics of the physical experiments. Similar results can be seen for 700 CAD, with the EM and SPDMD 0 Hz comparisons yielding different results for the level of discrepancy in velocity magnitude predictions, particularly in the vicinity of the tumble vortex to the right of the flow field.

#### D. Inspecting single LES snapshots

The effect of using SPDMD 0 Hz modes to represent data rather than the EM is also apparent when attempting to compare single LES snapshots against an ensemble of PIV data. The velocity histogram for the single LES snapshot shown in Fig. 3(d) is plotted alongside the PIV EM and PIV SPDMD 0 Hz histograms in Fig. 15. Here, the results look quite different; the comparison between the single LES snapshot and the PIV EM is more disparate with  $HD = 0.71$ , while the comparison against the SPDMD 0 Hz mode gives a closer match with  $HD = 0.84$ . Therefore, this LES snapshot is more representative of the PIV ensemble than it would have seemed if the EM had been used as the sole validation target. This has implications for how particularly unrepresentative LES snapshots should be identified.

For example, consider a comparison between Fig. 3(a): the PIV EM and Fig. 3(d): the LES single snapshot. The low HD of 0.71 between these two flow fields would identify this LES snapshot as being especially unrepresentative of the PIV data. In particular, the length of the intake jet and the higher-speed recirculation zones may be highlighted as specific problem areas. However, if the SPDMD 0 Hz mode in Fig. 3(e) is chosen as the benchmark representing the PIV data, the opposite conclusions about the LES snapshot may be drawn.

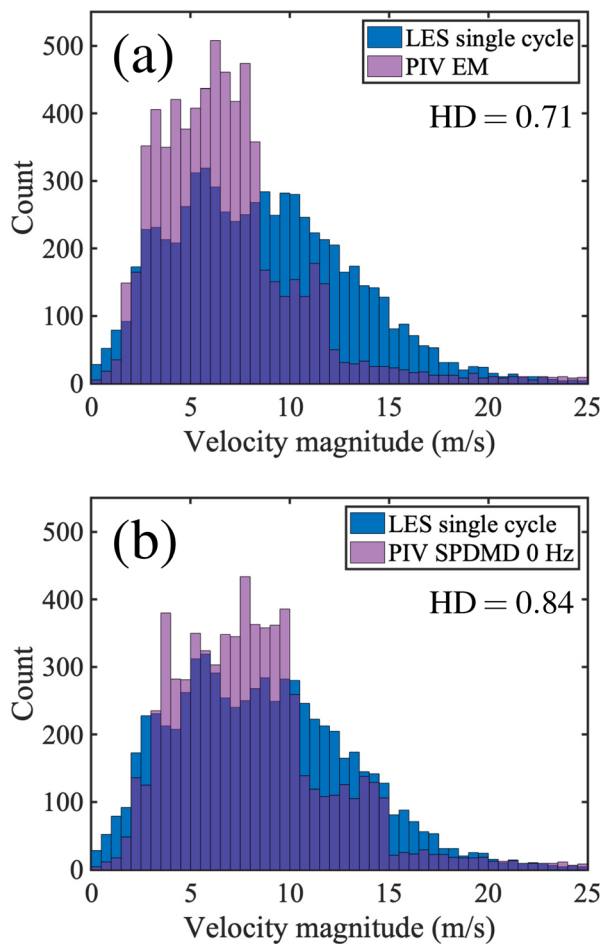
In fact, this snapshot turns out to be one of the more representative ones within the LES dataset in terms of velocity magnitudes. This highlights the fact that care needs to be taken when using the EM as the validation target for LES results, and when drawing conclusions about the accuracy of velocity magnitude predictions.

#### E. Reliability of SPDMD 0 Hz modes with dataset sizes

The effect of the number of available snapshots on the SPDMD 0 Hz modes was also explored, as this was a key difference between the LES and PIV data. The convergence of the real parts of the DMD eigenvalues and the calculation of the mode residuals are two widely used metrics for analyzing the convergence of DMD spectra.<sup>26,55</sup> For a statistically stationary process, the real parts of the eigenvalues will converge to unity, so this is an appropriate test for assessing the validity of the 0 Hz SPDMD modes. On the other hand, the residual calculation gives a more complete picture of DMD convergence, as it quantifies the difference between the reconstruction of the last snapshot and the actual data.<sup>55</sup> However, this study focuses purely on the 0 Hz SPDMD mode, and a reconstruction of the data outside of the 0 Hz mode is not required. Therefore, the real parts of the DMD eigenvalues are used as the convergence metric instead.

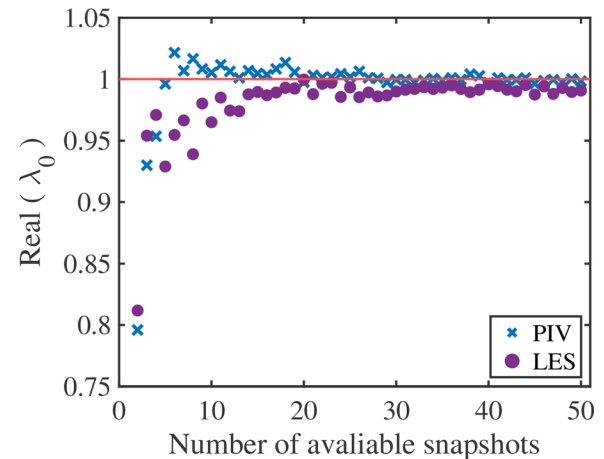
Figure 16 plots the real parts of the eigenvalues associated with the 0 Hz SPDMD modes ( $\lambda_0$ ) for an increasing number of available snapshots. Specifically, separate  $\lambda_0$  values were calculated for the first  $m$  and  $m + 1$  snapshots and analyzed accordingly. Because the number of snapshots varied for each calculation, it was impractical to fix a constant SPDMD mode threshold number for each case. Instead, a criterion was chosen such that the threshold numbers were set to the





**FIG. 15.** Histogram comparisons between (a): a single snapshot from LES and the PIV EM, and (b): a single snapshot from LES and the PIV SPDMD 0 Hz mode, all at 470 CAD aTDCf.

number of singular values that contained at least 50% of the cumulative variance in the data. For example, when the first 20 LES snapshots were considered, the 6 largest singular values accounted for 52.8% of the cumulative variance in the data. Therefore, 6 was chosen as the SPDMD mode threshold number for  $m = 20$ . 50% was chosen as the variance threshold as it is known to produce reliable SPDMD 0 Hz modes for these datasets over a number of different crank angles; however, different variance threshold criteria ranging from 10% to 95% were also tested and found to have a negligible effect on the overall convergence patterns. In Fig. 16, it can be seen that convergence to unity is achieved with relatively small dataset sizes, indicating that SPDMD can produce reliable 0 Hz modes for datasets containing as few as 20 snapshots for both the LES and PIV data. Convergence to the unit circle indicates that there are enough snapshots to generate a steady state signal and that the total time spanned by the dataset at approximately 20 cycles is longer than the slowest characteristic time-scale in the data.<sup>55</sup> This result is consistent with the findings of Grenga *et al.*<sup>55</sup> who investigated direct numerical simulation (DNS) results of a turbulent planar premixed hydrogen/air jet flame with DMD. Their



**FIG. 16.** Convergence of the real part of the SPDMD 0 Hz eigenvalues to unity for an increasing number of snapshots in the dataset at 470 CAD aTDCf. The convergence behavior for the PIV data is given by the blue crosses, the LES data by the purple circles, and the line  $\text{Real}(\lambda_0) = 1$  is plotted in red.

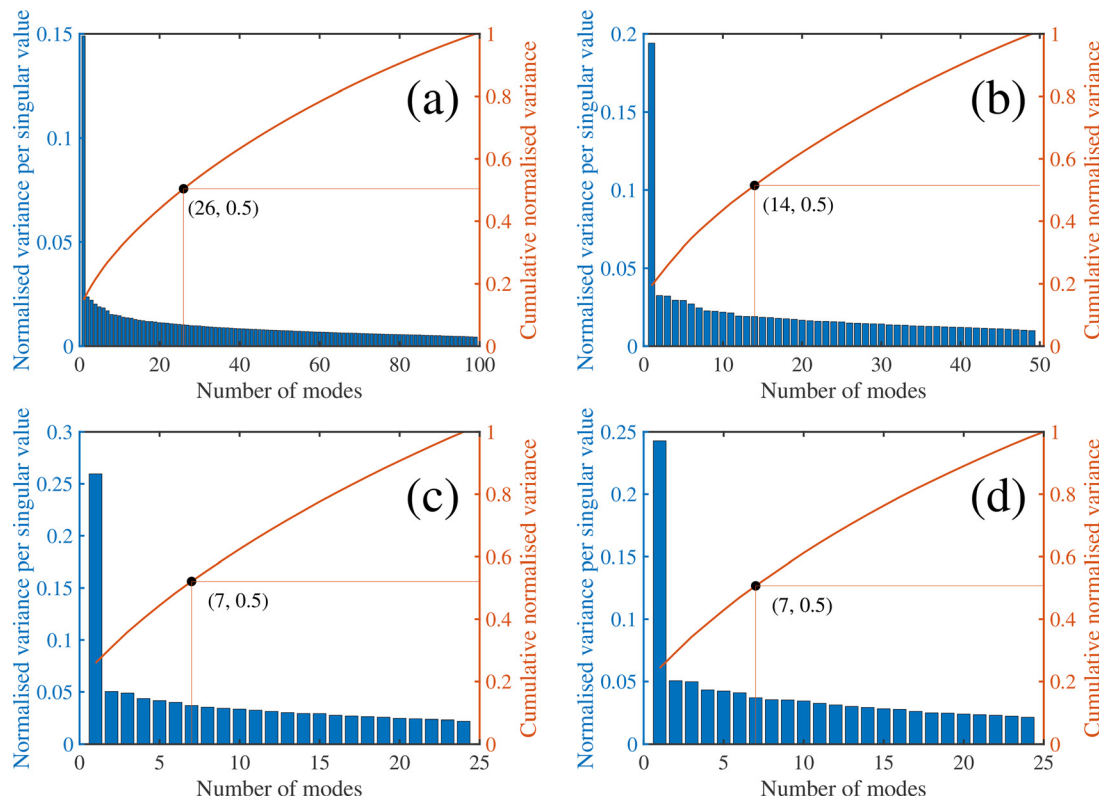
results showed that real ( $\lambda_0$ ) changed by negligible amounts for dataset sizes varying from 21 snapshots to 401.

#### F. Further effects of dataset size

Further discussion is warranted regarding the decision to compare 250 PIV snapshots to 50 LES realization. The full 250 PIV snapshots were utilized for the majority of this study for practical reasons, as the dataset sizes are relatively small, and therefore, caution should be taken when discarding information by only considering a smaller subsample of the data. There is also the question of how to select a subsample of PIV snapshots; options were considered such as simply choosing the first 50 PIV snapshots, the 50 least variable PIV snapshots,<sup>44</sup> or conducting a larger statistical analysis using a combination of subsamples. Due to this added complexity, the authors decided to place more emphasis on how well the SPDMD 0 Hz modes can represent their respective datasets separately, and what these modes can be used for. However, some interesting results can still be found if a simple subsampling option is chosen and the first 25, 50, and 100 PIV snapshots are analyzed. The singular value decompositions and HD sweeps are shown for each of the PIV subsamples as well as for the first 25 LES cycles in Fig. 17.

The PIV data behave more similarly to the LES data when 50 snapshots are considered for each. For example, the amount of variance contained in the first singular value and the number of singular values needed to reach 50% of the cumulative variance (denoted here as SV50) are presented in Table II. In all cases, the proportion of the variance that is contained in the first singular value is relatively low; in an LES study of the transparent combustion chamber (TCC) engine, Liu *et al.*<sup>23</sup> reported variances in the first singular value that ranged between 0.28 and 0.98, with the lowest values being reported at crank angles near to TDC. Crank angles with the intake valve open ranged around 0.7. This may suggest that the intake flow in the Darmstadt engine is more variable than in the TCC, as larger numbers of modes are needed in order to capture the flow behavior.

For the present engine, as more snapshots are considered, the proportion of the variance that is contained in the first singular value



**FIG. 17.** Singular value decomposition plots for (a): the first 100 PIV snapshots, (b): the first 50 PIV snapshots, (c): the first 25 PIV snapshots, and (d): the first 25 LES snapshots, all at 470 CAD aTDCf. The variance captured in each singular value is given by the bar chart on the left axis, and the cumulative variance is given by the line graph on the right axis. The number of modes corresponding to 50% of the cumulative variance is marked for each case.

**TABLE II.** The impact of sample size on the proportion of the total variance contained in the first singular value, and the number of singular values needed in order to capture 50% of the cumulative variance in the data (denoted as SV50) for 470 CAD aTDCf.

	Variance in first singular value	SV50
PIV 250	0.11	57 (23%)
PIV 100	0.15	26 (26%)
PIV 50	0.19	14 (28%)
PIV 25	0.26	7 (28%)
LES 50	0.18	14 (28%)
LES 25	0.24	7 (28%)

decreases, and the value of SV50 is also reduced as a percentage of the total number of singular values. The reduction in the relative SV50 suggests that some redundancy is entering the dataset, as the raw SV50 values increase at a slower rate than the total number of snapshots. The fact that fewer modes are needed to capture 50% of the variance in the data (relative to the total number of snapshots) may indicate that the data are becoming more converged. The decrease in the variance contained in the first singular value may be due to the fact that larger datasets are more complex overall and therefore more challenging to capture in a single mode. However, a more thorough

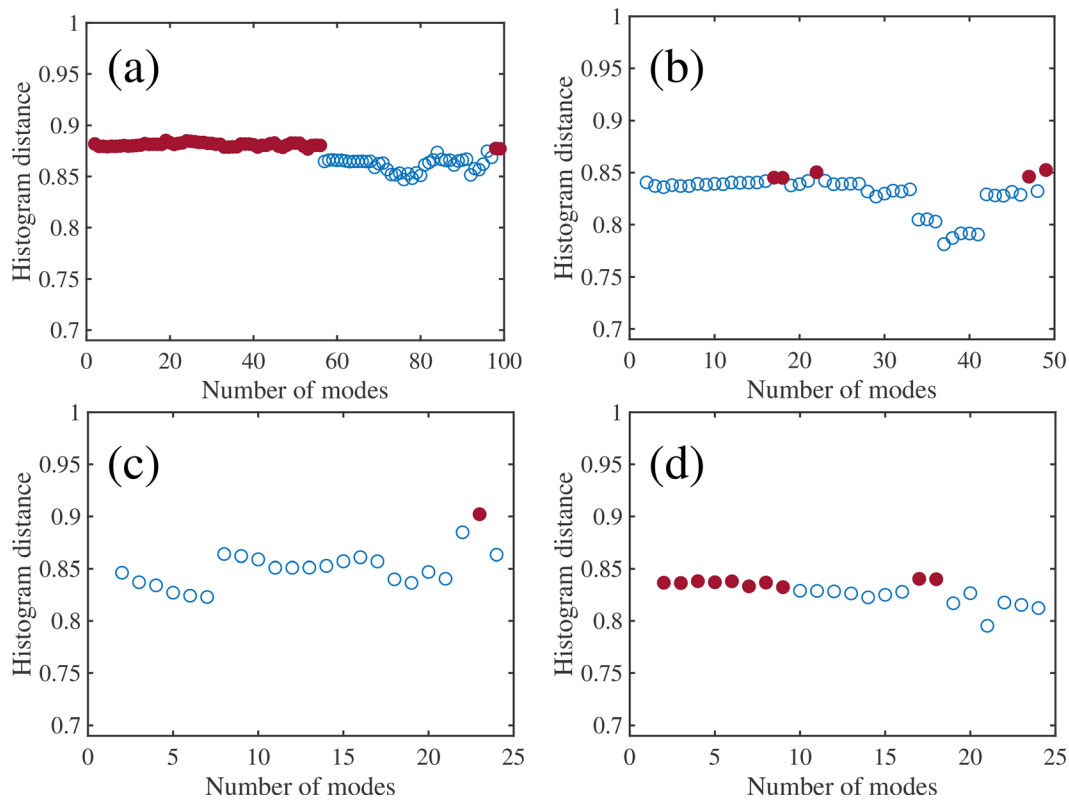
investigation into the various possible sampling strategies would be required in order to verify these hypotheses, which is beyond the scope of this work.

The corresponding HD sweeps for varying mode threshold numbers considering subsamples of the PIV and LES snapshots are shown in Fig. 18. There does not seem to be a pattern in the distribution of the HDs with varying sample size when all of the HD plots are considered together, so it is likely that hard thresholding the SPDMD reconstructions will remain a case-dependent exercise. However, as previously stated, a larger statistical analysis would need to be conducted before a firm conclusion can be drawn.

## VI. DISCUSSION

The results in this work show that the SPDMD 0 Hz modes can provide more realistic representations of vector magnitudes than the EM for a dataset that is characterized by some degree of variability. For the application of validating LES data against PIV data, this is shown to be useful in a number of ways, including:

1. gaining insight into the dynamics of the physical system, as shown in Figs. 3 and 6,
2. quantifying differences in vector magnitudes between simulations and experiments, as shown in Figs. 13 and 14,
3. identifying LES snapshots that are particularly representative of the PIV ensemble, as shown in Fig. 15.



**FIG. 18.** Sweeps calculating the HD between the SPDMD 0 Hz mode and the cycle average histograms for varying mode threshold numbers for (a): the first 100 PIV snapshots, (b): the first 50 PIV snapshots, (c): the first 25 PIV snapshots, and (d): the first 25 LES snapshots, all at 470 CAD aTDCf. The filled red circles represent HD values that are within 1% of the maximum HD value.

Point 1 refers to the more representative vector magnitudes that are displayed in the SPDMD 0 Hz modes, for example, as shown in Figs. 3 and 6. The figures indicate that an isolated consideration of the EM fields could result in a misleading conclusion about what the speed of the intake jet should be. One could argue that this “error” cancels out if the EM is used to compare both datasets, as the EM diminishes the intake jets in the PIV and LES data by similar amounts. However, this is not guaranteed to be the case, especially if the variability in one dataset is significantly different from another. For example, for the 700 CAD case, the relative standard deviation in the planar-averaged velocity magnitudes was 12.6% for the LES data, but 26.4% for the PIV data. This difference in variability leads to a more significant difference in the HD between the EM fields and SPDMD 0 Hz modes, as shown in Fig. 12. Therefore, the SPDMD 0 Hz modes can provide increased confidence in the velocity magnitudes given by the flow fields that are used to represent the ensembles of data. A note is needed here regarding the use of SPDMD 0 Hz modes and the quantification of variability in the data, or CCVs. Although the size of the increase in the HD to the set of individual snapshots due to the SPDMD is affected by the level of variability in the dataset, other more direct methods for calculating CCV would likely be more suitable, such as quantifying the pressure variation<sup>56</sup> or through a POD analysis.<sup>57</sup> Rather, the primary purpose of the SPDMD 0 Hz modes here is to provide a more

reliable representation of a vector field ensemble for use as a validation target.

The contributions of this study also provide evidence for the robustness of the proposed application of SPDMD across different datasets. With this work, along with a previous study by Baker *et al.*,<sup>40</sup> the approach has been shown to yield increased HDs to the sets of individual snapshots for both LES and PIV data, representing a variety of physical behaviors across different engines and phase angles. In another aspect of this previous study, Baker *et al.*<sup>40</sup> directly compared the SPDMD approach to the use of POD-based reconstructions of the flow in terms of each method’s ability to produce reliable validation targets from a PIV dataset. Although the POD-based reconstructions were found to be capable of avoiding the issue of diminished vector magnitudes in some cases, POD has the property of producing a distinct POD-based reconstruction for each snapshot in the dataset. This one-to-one relationship between the number of snapshots and the number of validation targets results in an excessive number of validation targets being produced. Therefore, within the realm of CFD validation, POD is perhaps better suited to the quantification of in-cylinder flow CCVs or the comparison of isolated POD modes (i.e., comparing like POD modes across datasets) as conducted by Barbato *et al.*,<sup>44</sup> rather than using the POD-based reconstructions as validation targets. Even so, the isolated POD modes do not necessarily resemble

the individual snapshots of the physical flow. On the other hand, the SPDMD 0 Hz modes offer an objective many-to-one solution for the formation of a validation target from an ensemble of data that resembles the individual snapshots of the physical flow and without the diminishing of vector magnitudes that is associated with ensemble averaging.

An ongoing challenge with this method lies in how to define objectively the threshold for the number of modes to retain in the SPDMD analysis. Currently, it appears that the best practice is to simply test a number of different thresholds and select one with a high HD to the individual snapshots, as shown in Fig. 7. The thresholding challenge is common to many methods with roots in the singular value decomposition (SVD), including POD. In some cases, a hard threshold can be defined where there is a clear separation between the high- and low-variance modes in the form of an “elbow” in the plot of singular values,<sup>58</sup> but this is rarely the case for turbulent engine flows.<sup>39</sup> Gavish and Donoho<sup>53</sup> provide a theoretical basis for finding the optimal hard threshold for a low-rank matrix with Gaussian noise, but it is unclear how well this translates to a dataset with an unknown level of noise. Future work will investigate thresholding options for such physical datasets.

Finally, although this method has been applied to the study of a cyclostationary airflow, the SPDMD 0 Hz modes likely also have advantages over the traditional average in a range of other physical scenarios. For example, the application of this technique to the study of spray formation in continuous time will also constitute future work.

## VII. CONCLUSION

In this work, flow field velocity data from PIV measurements were compared to LES results using two methods: the ensemble mean (EM) and the SPDMD 0 Hz modes. The ability of each method to accurately represent the velocity magnitudes present in each dataset was quantified with the histogram distance (HD). While the EM fields under-represented the magnitudes seen in both the PIV and LES datasets, the SPDMD 0 Hz modes were able to remove the under-predictions and yield higher HD values to the original data, indicating that the SPDMD 0 Hz modes were more representative of both datasets.

At 700 CAD, the use of the SPDMD 0 Hz modes had an impact on the comparison of HDs between the PIV and LES datasets. However, at 470 CAD, the SPDMD 0 Hz modes increased the HD for both the PIV and LES datasets by similar amounts, indicating that these two datasets had similar levels of variability in the velocity magnitudes. As a result, a comparison between EMs gave similar HDs to a comparison between SPDMD 0 Hz modes. Even so, the two methods provided different estimates for the sizes of the speed differences across the LES and PIV data. In this case, the SPDMD 0 Hz modes showed that the LES data tended to mis-predict the velocity magnitudes around the intake jet region by larger amounts than it would appear from the comparison of EM fields. Accurate quantification of these mis-predictions is necessary in order to correctly diagnose inaccuracies in velocity data and aid in the development of LES models.

The importance of having a validation target that fairly represents the PIV data was also investigated by exploring the different conclusions that could be drawn about the accuracy of a single LES snapshot. For the LES snapshot chosen, a comparison to the PIV EM indicated that the velocity magnitudes in the LES snapshot were fairly unrepresentative of the PIV data, with a relatively low HD = 0.71. However,

this same LES snapshot was shown to have more representative magnitudes if compared with the PIV SPDMD 0 Hz mode, where the HD increased to 0.84. The choice of validation target is therefore an important factor in attempting to assess the accuracy of individual LES snapshots and locate sources of error.

A convergence test was run to investigate how the number of snapshots in the dataset affected the SPDMD 0 Hz modes. For both the PIV and LES data, the SPDMD 0 Hz modes converged after approximately 20 snapshots. This suggests that SPDMD is able to produce a reliable 0 Hz mode for relatively small dataset sizes. Finally, the singular value distributions and HD sweeps are explored for subsamples of the datasets, but further study is required in order to draw firm conclusions about the validity of comparing a subsample of PIV data against LES data.

## ACKNOWLEDGMENTS

The authors would like to thank the Darmstadt Engine Workshop Research Group for providing the experimental data used in this study. This research was funded in whole or in part by the Engineering and Physical Sciences Research Council Prosperity Partnership, Grant No. EP/T005327/1. For the purpose of Open Access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript (AAM) version arising from this submission. The Prosperity Partnership is a collaboration among Jaguar Land Rover, Siemens Digital Industries Software, the University of Bath, and the University of Oxford. The authors would also like to thank the Department of Engineering Science technicians and maintenance teams for facilities support. Xiao Hang Fang gratefully acknowledges the financial support from the University of Calgary Transdisciplinary Scholarship Connector Grants and the John Fell Oxford University Press Research Fund. Due to confidentiality agreements with research collaborators, data supporting this paper can only be made available to bona fide researchers subject to a non-disclosure agreement. Details of the data and how to request access are available from the “Oxford Research Archive” repository at <https://ora.ox.ac.uk/>.

## AUTHOR DECLARATIONS

### Conflict of Interest

The authors have no conflicts to disclose.

### Author Contributions

**Samuel Baker:** Conceptualization (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Xiao Hang Fang:** Conceptualization (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Project administration (equal); Supervision (equal); Writing – original draft (equal); Writing – review & editing (equal). **Alessio Barbato:** Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Software (equal); Writing – original draft (equal); Writing – review & editing (equal). **Sebastiano Breda:** Data curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Software (equal); Writing – original draft (equal); Writing – review & editing (equal). **Mauro Magnani:** Data



curation (equal); Formal analysis (equal); Investigation (equal); Methodology (equal); Software (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Stefano Fontanesi**: Data curation (equal); Funding acquisition (equal); Project administration (equal); Resources (equal); Software (equal); Supervision (equal); Writing – original draft (equal); Writing – review & editing (equal). **Felix Leach**: Funding acquisition (equal); Project administration (equal); Resources (equal); Software (equal); Supervision (equal); Writing – original draft (equal); Writing – review & editing (equal). **Martin Howard Davy**: Funding acquisition (equal); Project administration (equal); Resources (equal); Software (equal); Supervision (equal); Writing – original draft (equal); Writing – review & editing (equal).

## DATA AVAILABILITY

Due to confidentiality agreements with research collaborators, data supporting this paper can only be made available to bona fide researchers subject to a non-disclosure agreement. Details of the data and how to request access are available from the “Oxford Research Archive” repository at <https://ora.ox.ac.uk/>.

## REFERENCES

- <sup>1</sup>S. B. Pope, *Turbulent Flows* (Cambridge University Press, 2000).
- <sup>2</sup>F. Hartmann, S. Buhl, F. Gleiss, P. Barth, M. Schild, S. A. Kaiser, and C. Hasse, “Spatially resolved experimental and numerical investigation of the flow through the intake port of an internal combustion engine,” *Oil Gas Sci. Technol. – Rev. IFP Energies nouvelles* **71**, 2 (2016).
- <sup>3</sup>C. Hasse, V. Sohm, and B. Durst, “Detached eddy simulation of cyclic large scale fluctuations in a simplified engine setup,” *Int. J. Heat Fluid Flow* **30**, 32–43 (2009).
- <sup>4</sup>H. Chu, C. Welch, H. Elmestikawy, S. Cao, M. Davidovic, B. Böhm, A. Dreizler, and H. Pitsch, “A combined numerical and experimental investigation of cycle-to-cycle variations in an optically accessible spark-ignition engine,” *Flow, Turbul. Combust.* **110**, 3–29 (2023).
- <sup>5</sup>C. Welch, M. Schmidt, L. Illmann, A. Dreizler, and B. Böhm, “The influence of flow on cycle-to-cycle variations in a spark-ignition engine: A parametric investigation of increasing exhaust gas recirculation levels,” *Flow, Turbul. Combust.* **110**, 185 (2022).
- <sup>6</sup>W. Zeng, S. Keum, T.-W. Kuo, and V. Sick, “Role of large scale flow features on cycle-to-cycle variations of spark-ignited flame-initiation and its transition to turbulent combustion,” *Proc. Combust. Inst.* **37**, 4945–4953 (2019).
- <sup>7</sup>L. Engelmann, J. Laichter, P. Wollny, M. Klein, S. A. Kaiser, and A. M. Kempf, “Cyclic variations in the flame propagation in an spark-ignited engine: Multi cycle large eddy simulation supported by imaging diagnostics,” *Flow, Turbul. Combust.* **110**, 91–104 (2023).
- <sup>8</sup>F. Berni, G. Cicalese, and S. Fontanesi, “A modified thermal wall function for the estimation of gas-to-wall heat fluxes in cfd in-cylinder simulations of high performance spark-ignition engines,” *Appl. Therm. Eng.* **115**, 1045–1062 (2017).
- <sup>9</sup>R. Stone, *Introduction to Internal Combustion Engines* (Springer, 1999), Vol. 3.
- <sup>10</sup>N. Van Dam and C. Rutland, “Understanding in-cylinder flow variability using large-eddy simulations,” *J. Eng. Gas Turbines Power* **138**, 102809 (2016).
- <sup>11</sup>H. Chen, D. L. Reuss, D. L. Hung, and V. Sick, “A practical guide for using proper orthogonal decomposition in engine research,” *Int. J. Engine Res.* **14**, 307–319 (2013).
- <sup>12</sup>N. J. Beavis, S. S. Ibrahim, and W. Malalasekera, “A numerical study of intake valve jet flapping in a gasoline direct injection engine,” *Int. J. Powertrains* **7**, 38–52 (2018).
- <sup>13</sup>L. Shen, C. Willman, R. Stone, T. Lockyer, R. Magnanon, and G. Virelli, “On the use of particle image velocimetry (PIV) data for the validation of Reynolds averaged Navier–Stokes (RANS) simulations during the intake process of a spark ignition direct injection (SIDi) engine,” *Int. J. Engine Res.* **23**, 1061–1081 (2022).
- <sup>14</sup>F. Rulli, S. Fontanesi, A. d’Adamo, and F. Berni, “A critical review of flow field analysis methods involving proper orthogonal decomposition and quadruple proper orthogonal decomposition for internal combustion engines,” *Int. J. Engine Res.* **22**, 222–242 (2021).
- <sup>15</sup>F. Zhao and D. L. Hung, “Applications of machine learning to the analysis of engine in-cylinder flow and thermal process: A review and outlook,” *Appl. Therm. Eng.* **220**, 119633 (2022).
- <sup>16</sup>J. L. Lumley, “The structure of inhomogeneous turbulent flows,” in *Atmospheric Turbulence and Radio Wave Propagation* (Nauka, 1967), pp. 166–178.
- <sup>17</sup>M. Fogleman, J. Lumley, D. Rempfer, and D. Haworth, “Application of the proper orthogonal decomposition to datasets of internal combustion engine flows,” *J. Turbul.* **5**, 023 (2004).
- <sup>18</sup>W. A. Gardner, A. Napolitano, and L. Paura, “Cyclostationarity: Half a century of research,” *Signal Processing* **86**, 639–697 (2006).
- <sup>19</sup>S. Roudnitsky, P. Druault, and P. Guibert, “Proper orthogonal decomposition of in-cylinder engine flow into mean component, coherent structures and random gaussian fluctuations,” *J. Turbul.* **7**, N70 (2006).
- <sup>20</sup>S. Wu, S. Patel, and M. Ameen, “Investigation of cycle-to-cycle variations in internal combustion engine using proper orthogonal decomposition,” *Flow, Turbul. Combust.* **110**, 125–147 (2023).
- <sup>21</sup>K. Liu and D. C. Haworth, “Development and assessment of pod for analysis of turbulent flow in piston engines,” Technical Report No. 2011-01-0830 (SAE Technical Paper, 2011).
- <sup>22</sup>P. Abraham, K. Liu, D. Haworth, D. Reuss, and V. Sick, “Evaluating large-eddy simulation (LES) and high-speed particle image velocimetry (PIV) with phase-invariant proper orthogonal decomposition (POD),” *Oil Gas Sci. Technol. – Rev. IFP Energies nouvelles* **69**, 41–59 (2014).
- <sup>23</sup>K. Liu, D. C. Haworth, X. Yang, and V. Gopalakrishnan, “Large-eddy simulation of motored flow in a two-valve piston engine: Pod analysis and cycle-to-cycle variations,” *Flow, Turbul. Combust.* **91**, 373–403 (2013).
- <sup>24</sup>L. Shen, K.-Y. Teh, P. Ge, F. Zhao, and D. L. Hung, “Temporal evolution analysis of in-cylinder flow by means of proper orthogonal decomposition,” *Int. J. Engine Res.* **22**, 1714–1730 (2021).
- <sup>25</sup>K. Taira, S. L. Brunton, S. T. Dawson, C. W. Rowley, T. Colonius, B. J. McKeon, O. T. Schmidt, S. Gordeyev, V. Theofilis, and L. S. Ukeiley, “Modal analysis of fluid flows: An overview,” *AIAA J.* **55**, 4013–4041 (2017).
- <sup>26</sup>P. J. Schmid, “Dynamic mode decomposition of numerical and experimental data,” *J. Fluid Mech.* **656**, 5–28 (2010).
- <sup>27</sup>Y. Fu, X. Lin, L. Li, Q. Chu, H. Liu, X. Zheng, C.-H. Liu, Z. Chen, C. Lin, T. K. Tse *et al.*, “A POD-DMD augmented procedure to isolating dominant flow field features in a street canyon,” *Phys. Fluids* **35**, 025112 (2023).
- <sup>28</sup>Y. Liu, J. Long, Q. Wu, B. Huang, and G. Wang, “Data-driven modal decomposition of transient cavitating flow,” *Phys. Fluids* **33**, 113316 (2021).
- <sup>29</sup>Y. Yuan, K. Zhou, W. Zhou, X. Wen, and Y. Liu, “Flow prediction using dynamic mode decomposition with time-delay embedding based on local measurement,” *Phys. Fluids* **33**, 095109 (2021).
- <sup>30</sup>J. N. Kutz, S. L. Brunton, B. W. Brunton, and J. L. Proctor, *Dynamic Mode Decomposition: Data-Driven Modeling of Complex Systems* (SIAM, 2016).
- <sup>31</sup>P. J. Schmid, “Dynamic mode decomposition and its variants,” *Annu. Rev. Fluid Mech.* **54**, 225–254 (2022).
- <sup>32</sup>M. R. Jovanović, P. J. Schmid, and J. W. Nichols, “Sparsity-promoting dynamic mode decomposition,” *Phys. Fluids* **26**, 024103 (2014).
- <sup>33</sup>L. Wang, X. Liu, N. Wang, and M. Li, “Modal analysis of propeller wakes under different loading conditions,” *Phys. Fluids* **34**, 065136 (2022).
- <sup>34</sup>Y. Wu, R. Tao, Z. Yao, R. Xiao, and F. Wang, “Application and comparison of dynamic mode decomposition methods in the tip leakage cavitation of a hydrofoil case,” *Phys. Fluids* **35**, 023326 (2023).
- <sup>35</sup>H. Ping, H. Zhu, K. Zhang, D. Zhou, Y. Bao, Y. Xu, and Z. Han, “Dynamic mode decomposition based analysis of flow past a transversely oscillating cylinder,” *Phys. Fluids* **33**, 033604 (2021).
- <sup>36</sup>W. Qin, L. Zhou, D. Liu, M. Jia, and M. Xie, “Investigation of in-cylinder engine flow quadruple decomposition dynamical behavior using proper orthogonal decomposition and dynamic mode decomposition methods,” *J. Eng. Gas Turbines Power* **141**, 081004 (2019).
- <sup>37</sup>M. Liu, F. Zhao, X. Li, M. Xu, and D. L. Hung, “Dynamic mode decomposition for extracting cycle-to-cycle variation of SIDi engine in-cylinder flow under motoring condition,” in *Internal Combustion Engine Division Fall Technical*



- Conference (American Society of Mechanical Engineers, 2020), Vol. 84034, p. V001T06A003.
- <sup>38</sup>M. Liu, F. Zhao, and D. L. Hung, "A coupled phase-invariant POD and DMD analysis for the characterization of in-cylinder cycle-to-cycle flow variations under different swirl conditions," *Flow, Turbul. Combust.* **110**, 31–57 (2023).
  - <sup>39</sup>X. Fang, L. Shen, C. Willman, R. Magnanon, G. Virelli, M. H. Davy, and R. Stone, "Manifold reduction techniques for the comparison of crank angle-resolved particle image velocimetry (PIV) data and Reynolds-averaged Navier–Stokes (RANS) simulations in a spark ignition direct injection (SID) engine," *Int. J. Engine Res.* **23**, 1275–1294 (2022).
  - <sup>40</sup>S. Baker, X. Fang, L. Shen, C. Willman, J. Fernandes, F. Leach, and M. Davy, "Dynamic mode decomposition for the comparison of engine in-cylinder flow fields from particle image velocimetry (PIV) and Reynolds-averaged Navier–Stokes (RANS) simulations," *Flow, Turbul. Combust.* **111**, 115 (2023).
  - <sup>41</sup>E. Baum, B. Peterson, B. Böhm, and A. Dreizler, "On the validation of les applied to internal combustion engine flows: Part 1: Comprehensive experimental database," *Flow, Turbul. Combust.* **92**, 269–297 (2014).
  - <sup>42</sup>M. Schmidt, C.-P. Ding, B. Peterson, A. Dreizler, and B. Böhm, "Near-wall flame and flow measurements in an optically accessible SI engine," *Flow, Turbul. Combust.* **106**, 597–611 (2021).
  - <sup>43</sup>A. Barbato, S. Fontanesi, and A. D'Adamo, "Impact of grid density and turbulence model on the simulation of in-cylinder turbulent flow structures: application to the Darmstadt engine," Technical Report No. 2021-01-0415 (SAE Technical Paper, 2021).
  - <sup>44</sup>A. Barbato, C. Iacovano, and S. Fontanesi, "Cold-flow investigation of the Darmstadt engine with focus on statistical convergence: Experimental and large eddy simulation analysis," *Flow, Turbul. Combust.* **110**, 59–89 (2023).
  - <sup>45</sup>S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein *et al.*, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *FNT Mach. Learn.* **3**, 1–122 (2010).
  - <sup>46</sup>M. J. Swain and D. H. Ballard, "Color indexing," *Int. J. Comput. Vision* **7**, 11–32 (1991).
  - <sup>47</sup>M. Germano, U. Piomelli, P. Moin, and W. H. Cabot, "A dynamic subgrid-scale eddy viscosity model," *Phys. Fluids A* **3**, 1760–1765 (1991).
  - <sup>48</sup>I. Ko, F. Rulli, S. Fontanesi, A. d'Adamo, and K. Min, "Methodology for the large-eddy simulation and particle image velocimetry analysis of large-scale flow structures on TCC-III engine under motored condition," *Int. J. Engine Res.* **22**, 2709–2731 (2021).
  - <sup>49</sup>C. Iacovano, F. Berni, A. Barbato, S. Fontanesi *et al.*, "A preliminary 1D–3D analysis of the Darmstadt research engine under motored condition," in *E3S Web of Conferences* (EDP Sciences, 2020), Vol. 197, pp. 1–12.
  - <sup>50</sup>S. Fontanesi, S. Paltrinieri, A. D'Adamo, and S. Duranti, "Investigation of boundary condition and field distribution effects on the cycle-to-cycle variability of a turbocharged GDI engine using les," *Oil Gas Sci. Technol. – Rev. IFP Energies nouvelles* **69**, 107–128 (2014).
  - <sup>51</sup>Siemens, *Simcenter STAR-CCM+ Documentation* (Siemens, 2020).
  - <sup>52</sup>M. Dias Ribeiro, A. Mendonça Bimbato, M. Araújo Zanardi, J. A. Perrella Balestieri, and D. P. Schmidt, "Large-eddy simulation of the flow in a direct injection spark ignition engine using an open-source framework," *Int. J. Engine Res.* **22**, 1064–1085 (2021).
  - <sup>53</sup>M. Gavish and D. L. Donoho, "The optimal hard threshold for singular values is  $4/\sqrt{3}$ ," *IEEE Trans. Inf. Theory* **60**, 5040–5053 (2014).
  - <sup>54</sup>B. P. Epps and E. M. Krivitzky, "Singular value decomposition of noisy data: Noise filtering," *Exp. Fluids* **60**, 1–23 (2019).
  - <sup>55</sup>T. Grenga, J. F. MacArt, and M. E. Mueller, "Dynamic mode decomposition of a direct numerical simulation of a turbulent premixed planar jet flame: Convergence of the modes," *Combust. Theory Modell.* **22**, 795–811 (2018).
  - <sup>56</sup>A. d'Adamo, S. Breda, F. Berni, and S. Fontanesi, "Understanding the origin of cycle-to-cycle variation using large-eddy simulation," *SAE Int. J. Engines* **12**, 45–100 (2018).
  - <sup>57</sup>K. Bizon, G. Continillo, K. Leistner, E. Mancaruso, and B. Vaglieco, "Pod-based analysis of cycle-to-cycle variations in an optically accessible diesel engine," *Proc. Combust. Inst.* **32**, 2809–2816 (2009).
  - <sup>58</sup>S. L. Brunton and J. N. Kutz, *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control* (Cambridge University Press, 2019).