Feature Extraction from Turbulent Channel Flow of Moderate Reynolds Number via Composite DMD Analysis

Binghua Li 1,2, *, Jesús Garicano-Mena 2, *, Yao Zheng 1 and Eusebio Valero 2

1 Center for Engineering and Scientific Computation, School of Aeronautics and Astronautics, Zhejiang University, Zhejiang 310027, China.
2 E.T.S.I. Aeronáutica y del Espacio, Universidad Politécnica de Madrid, 28040 Madrid, Spain.

*Corresponding author e-mail: libh@zju.edu.cn, a jesus.garicano.mena@upm.es.

Abstract. In this contribution, we described a Dynamic Mode Decomposition (DMD) analysis of a turbulent channel flow database at a moderate friction Reynolds number $Re_t \approx 950$. More specifically, a composite-based DMD analysis was conducted, employing hybrid snapshots assembled by skin friction $C_f(t_k)$ and either instantaneous Reynolds stress $u'v'(x; t_k)$ or streamwise velocity fluctuation $u'(x; t_k)$ fields. The DMD modes thus obtained were sorted according to its relevance to the $C_f$: less than 2% of the modes suffice to reconstruct accurately either the streamwise velocity or the Reynolds stress profiles near the wall. Furthermore, we aim to extend our preliminary work on the analysis of the turbulent database, by considering snapshots encompassing a larger spatial subdomain and covering a longer temporal span. However, this study involved data matrices significantly larger than that one, which the memory footprint of this problem exceeds a typical workstation. Accordingly, we have resorted to the parallel, memory distributed DMD algorithm as a reinforcement. With this enhanced composite DMD algorithm, flow features of moderate and even large turbulent channel problems could be identified and characterized.

1. Introduction
Turbulent flows are a typical example of complex dynamic system involving a wide ranges of spatial and temporal scales. This fact poses stringent constrains on the tools which required to describe such problems. Indeed, a well-resolved, statistically independent description of a turbulent flow requires dense grids of computational/data-acquisition points and high temporal resolution. Particle image velocimetry (PIV) or direct numerical simulation (DNS) can provide such data.

Feature extraction algorithms can be useful in the classification of the information hidden in the turbulent data, be it obtained with experimental testing or through DNS computations. In recent years, Proper Orthogonal Decomposition (POD, [5] [6]) and Dynamic Mode Decomposition (DMD, [7] [8]) have arisen as probably the most widely spread techniques.

DMD techniques, stemming from the contributions of Schmid [8] and Rowley and collaborators [7], have arisen as prominent feature identification methods in the field of fluid dynamics. Many variants of the DMD method were developed to retrieve meaningful flow structures from either experimental or numerical flow data in a purely data-driven manner. Successful applications of variants of the method...
to a great variety of flow data can be found in [9][10][11][12]. In this contribution, we will resort to the distributed memory, MPI-parallelized DMD algorithm described in [4]. An application of this algorithm to composite snapshots -i.e., snapshots formed by considering two or more different magnitudes (velocity components, skin friction, $\lambda_2$ invariant, ...)-to compute laminar-to-turbulent transition is described in [13].

In this contribution we consider the composite DMD analysis of a turbulent channel flow database at a moderate friction Reynolds number $Re_f \approx 950$. Departing from the methodology proposed in [3], we extend the study described in [1] to a data sequence encompassing a larger spatial domain and covering a longer temporal span. The data matrices at play in this extended study are significantly larger than those in [1], and hence the parallel, memory distributed DMD algorithm proposed in [4] becomes crucial. With this enhanced composite DMD algorithm, flow features of the turbulent channel problem could be identified and characterized.

This contribution is organized as follows: next section describes both the DNS solver employed to generate the turbulent databases and the specific implementations of DMD strategies applied to analyze them. Section 3 discusses the results obtained. Finally, section 4 presents the conclusions of our work.

2. Numerical methodology

2.1. Databases description

The work conducted pertains to the $Re_f \approx 950$ database described in [2]. The characteristics of the database are summarized in Table 1. A detailed account on the DNS simulation can be found in Lozano-Durán & Jiménez [14][15]. The database includes more than $1.5 \times 10^4$ snapshots, totaling 23 terabytes (TBs) of storage even with high compressed data format of HDF5. As we shall discuss in section 3.1, several measures have been taken to reduce the memory footprint before conducting the DMD analysis.

### Table 1. Parameters of the database

| $Re_f$ | $L_x/h$ | $L_z/h$ | $\Delta x^+$ | $\Delta z^+$ | $\Delta y_{max}^+$ | $N_x N_z$ | $N_y$ | $\Delta t^+$ | $u_e$ |
|--------|---------|---------|--------------|--------------|-------------------|-----------|------|--------------|------|
| 932    | $2\pi$  | $\pi$   | 11           | 5.7          | 7.6               | 786       | 385  | 0.8          | 0.04539 |

2.2. Feature detection algorithms: composite DMD with parallelized realization

We present here a brief summary of the DMD technique, as proposed in [8]. Given a sequence of instantaneous flow fields numbered from 1 to $n_s$ (e.g., taking one or all recorded variables), the following data matrix can be constructed:

$$V_{n_s} = \{v(t_1), v(t_2), \ldots, v(t_{n_s})\}$$

(1)

Where the subindex and superindex identify, respectively, the first and last time instants of the sequence. The data is ordered in time, and separated by a constant sampling time interval $\Delta t^s$ such that: $t_{k+1} = t_k + \Delta t^s$ for all $k = 1, \ldots, n_s - 1$. In the case of linear stability analysis and within the exponential growth region, it is possible to define a linear operator $A$ (i.e., a numerical approximation of the linearized Navier-Stokes operator) such that $v(t_{k+1}) = A v(t_k)$. For non-linear systems, $A$ approximates the Koopman operator. Eq. 1 can then be rewritten as a Krylov sequence (see [16]): $V_{n_s} = \{v(t_1), Av(t_1), \ldots, A^{n_s-1}v(t_1)\}$, which can alternatively be written in matrix form as $AV_{n_s} = V_{2}^{n_s}$.

Next, the Singular Value Decomposition (SVD) of the matrix $V_{1}^{n_s-1} = USW^H$ is obtained; the superscript $H$ denotes conjugate transposition. As detailed in [4], the SVD step revolves around the memory distributed parallel TSQR (Tall-and-Skinny matrix QR decomposition) algorithm by Demmel et al. [17].

Using the SVD of the snapshot matrix, namely $AU^A = V_2^{n_s}$, leads to the projected matrix $\tilde{A} = U^A AU$. Note that the projected matrix is actually computed as $\tilde{A} = U^H V_2^{n_s} W S^{-1}$. The DMD operates under the assumption that the projected matrix $\tilde{A}$ conveys most of the information codified into operator $A$. 

The solution of the reduced eigenvalue problem $\mathbf{A}\mathbf{y}_i = \mu_i \mathbf{y}_i$ provides the reduced DMD modes $\mathbf{y}_i$ and the eigenvalues $\mu_i$. These eigenvalues define the growth rates ($\Re(\mu_i)$) and the frequencies ($\Im(\mu_i)$) mapped to the unit circle. The approximated eigenmodes of the matrix $A$ can then be recovered via a projection onto the original space, using relation $\phi_i = \mathbf{U}\mathbf{y}_i$. Eventually, the growth rates and frequencies in the complex half-plane can be recovered from the eigenvalues as $\lambda_i = \log(\mu_i) / \Delta t^s$.

The DMD decomposition allows to reconstruct the original data sequence as the linear superposition:

$$v(\vec{x}, t) = \sum_{i=1}^{n_s-1} \alpha_i \phi_i(\vec{x}) e^{\lambda_i t}.$$  \hspace{1cm} (2)

In this contribution, the amplitudes $\alpha_i$ are computed following the formulation in [9], which is derived from the minimization problem in the Fröbenius norm:

$$\min_{\alpha_i} \| \mathbf{UW}^T - \mathbf{YD}_\alpha \mathbf{V} \|^2_F.$$  \hspace{1cm} (3)

Where the columns in matrix $\mathbf{Y}$ are the eigenvectors $\mathbf{y}_i$, diagonal matrix $\mathbf{D}_\alpha$ contains the unknown amplitudes $\alpha_i$ and $\mathbf{V}$ is a Vandermonde matrix whose columns are generated by the successive powers of the column vector $[\mu_0^k, \ldots, \mu_{n_s-1}^k]^T$, with $k = 0, \ldots, n_s - 1$. Finally, note how using $n_r < n_s - 1$ in Eq. 2 leads to a reduced representation of the data sequence.

2.3. Composite DMD analysis of the turbulent databases

In this work two types of DMD analysis are considered: classical and composite DMD analysis. The classical DMD is performed on snapshots of instantaneous Reynolds stress distribution, $u'v'(\vec{x}, t_k)$, whereas the composite DMD analysis employs hybrid snapshots obtained by concatenating instantaneous skin friction at the wall $C_f(t_k)$ and the instantaneous Reynolds stress distribution $u'v'(\vec{x}, t_k)$.

Following the approach in [3][18], we are interested in simplified DMD expansions -i.e. using $n_r < n_s - 1$ in Eq. 2- that lead to an acceptable reduced representation of the data sequence. Two different criteria have been considered for the mode selection. The first criterion consists in retaining only those modes fulfilling $\frac{|\alpha_i|}{|\alpha_{\text{max}}|} \geq 10\%$. The second criterion relies on the weighted quantities $\beta_i \equiv (\phi_i \cdot e_{C_f}) \alpha_i$, which $e_{C_f}$ is the unit vector along the component of the skin friction: only those modes with $\frac{|\beta_i|}{|\beta_{\text{max}}|} \geq 10\%$ will be kept. As we shall see in section 3.2, this second criterion allows to recover accurately the Reynolds stresses distribution for $n_r \ll n_s$.

Finally, whatever the upper limit $n_r$ taken in expansion Eq. 2, the Reynolds stress reconstruction extracted from the DMD analysis is obtained from:

$$u'v'^{\text{DMD}}(\vec{x}, t) = \sum_{i=1}^{n_s-1} \alpha_i (\phi_i - (\phi_i \cdot e_{C_f}) e_{C_f}) e^{\lambda_i t},$$  \hspace{1cm} (4)

By averaging Eq. 4 along the homogeneous directions $x - z$ and time, we obtain the DMD reconstructed Reynolds stress profiles:

$$\langle u'v' \rangle_{x,z,t}^{\text{DMD}}(y) = \frac{1}{n_s\Delta t^s} \sum_{i=1}^{n_s-1} \alpha_i (\phi_i - (\phi_i \cdot e_{C_f}) e_{C_f}) \int_0^{n_s\Delta t^s} e^{\lambda_i t} dt,$$  \hspace{1cm} (5)

which $\langle \cdot \rangle_{x,z,t}$ denotes the temporal and spatial averaging operator. In the next section, the $\langle u'v' \rangle_{x,z,t}^{\text{DMD}}(y)$ profiles will be compared against those obtained from DNS.

3. Results and discussion

3.1. Data sequence definition and verification
Achieving an accurate and statistically converged representation of a turbulent process requires well resolved domains and averaging over a significant time period. As the $Re_{\tau}$ number increases, the range
of temporal and spatial scales present in the flow becomes wider and wider. In fact, for the moderate $Re_t \approx 950$ turbulent flow, performing a DMD analysis on the whole database in [2] is prohibitively expensive in terms of memory, see [1].

Accordingly, before conducting the DMD analysis, we attempt to obtain a reduced view of the channel flow database that is nevertheless representative of the underlying flow physics. Fig. 1 shows second-order moments obtained from shorter snapshot subsequences ($n_s = 500, 800$ and $1200$) applying different spatial decimation along the homogeneous directions $x$ and $z$ (i.e., retaining one out of every $s_x$ and/or $s_z$ grid points). As visible in Fig. 1(a), considering spatially decimated snapshots has barely an impact on second-order quantities. Fig. 1(b) shows in turn the effect of snapshot sequence length $n_s$: observe how the error incurred is relatively small, especially inside the $y^+ < 150$ region. Therefore, the data sequence considered in the rest of this work is the one consisting of the first $n_s = 1200$ snapshots spatially decimated along both $x$ and $z$ directions with $s_x = s_z = 2$. In this manner, the data to be processed by DMD is reduced from the original 15754 snapshots, each of them consisting of $n_x \times n_y \times n_z \approx 2.27 \times 10^8$ grid points to 1200 snapshots of size $n_x \times (\text{floor}(\frac{n_y}{s_y}) + 1) \times n_z \approx 1.26 \times 10^7$ points.

![Figure 1. Reynolds shear stress profiles $\langle u'v' \rangle(y)$: data sequence sensitivity to length $n_s$ and spatial decimation $s_x$, $s_z$. In (a), sensitivity to spatial decimation $s_x = s_z$; in (b), sensitivity to length $n_s$.](image)

Note that the instantaneous flow-fields considered in this analysis are from $t_1 = 806.044$ to $t_{n_s} = 825.307$. The temporal span is $t_{n_s} - t_1 \approx 19.26$. However, the sampling times of the snapshots provided in this database are visibly non-uniform. Prior to conducting the DMD analysis, the data sequence was linearly interpolated so as to have a uniform sampling time given by $t_s \approx 0.0161$.

Finally, recall that in [3] the impact that the range of $y^+ \in [0, y^+_\text{max}]$ had on the performance of the composite DMD analysis for a $Re_t = 200$ database was considered. It was observed how DMD analyses with $y^+_\text{max} = 50$ led to good representation of near wall features. The same conclusions were achieved in the preliminary analysis of the $Re_t \approx 950$ database described in [1]. However, since our goal is extending the work in [1], we will consider $y^+_\text{max} = 150$, in an attempt to capture not only the near-wall region but also features into logarithm layer.
**Figure 2.** Composite DMD analysis on $C_f - u'_v'$ data sequences with $y_{max}^+ = 50$: results for $n_s = 500$, $s_x = s_z = 4$ using serial DMD (from [1]) vs results for $n_s = 1200$, $s_x = s_z = 2$ using parallel DMD (present work). In (a), $\mu$-plane. In (b), weighted amplitudes $|\beta_i|$ vs angular pulsation $\Im(\lambda_i)$. The black, solid line indicates threshold $|\beta_i|/|\beta_{max}| \geq 10\%$.

Fig. 2(a) compares the spectra and weighted average distribution for two composite DMD analyses. The first analysis, reproduced from [1], is conducted on a relatively short snapshot sequence ($n_s = 500$) and considers a spatial decimation $s_x = s_z = 4$ (e.g. one out every four points along the homogeneous directions is retained) using the serial algorithm. The second analysis, produced for this contribution, is conducted on a longer snapshot sequence ($n_s = 1200$) with spatial decimation $s_x = s_z = 2$. Completion of this analysis is not possible without the parallel DMD algorithm from [4].

In both cases, the eigenvalues lie mostly near the unit circle. This is consistent with the statistically stationary nature of the turbulent channel flow. The amplitudes factor $\beta_i$ distribution in Fig. 2(b) are comparable to each other.

**Figure 3.** DMD spectra obtained from analysis based on $u'_v'$ snapshots (classical) and hybrid $C_f - u'_v'$ snapshots (composite): (a), $\mu$-plane. In (b), corresponding amplitude $|\alpha_i|$ and $|\beta_i|$ vs angular pulsation $\Im(\lambda_i)$. The black, solid line indicates threshold $|\alpha_i|/|\alpha_{max}| \geq 10\%$ and $|\beta_i|/|\beta_{max}| \geq 10\%$. 
3.2. DMD analysis on $y_{\text{max}}^+ = 150$

Following the previous discussion, both classical and composite DMD analyses were finally applied to a ns = 1200 data sequence with $n_x = n_z = 2$ and $y_{\text{max}}^+ = 150$. The classical snapshots are simply the $u'v'(\bar{x}, t_k)$ field, whereas the composite analysis considers hybrid $C_f(t_k) - u'v'(\bar{x}, t_k)$ snapshots. The computations were conducted on the Universidad Politécnica de Madrid CeSViMa supercomputer, using up to 128 CPUs (Intel Xeon Gold 6230, 20 cores @ 2.1 GHz) and 640 Gb memory.

From the eigenvalues distribution, Fig. 3, we find, once again, that most of the modes lie near the locus $|\mu| = 1$. Still some modes located inside the unit circle. From the eigenvalues distributions shown in Fig. 3(a), we also observe that classical and composite DMD eigenvalues practically coincide, see also [3].

However, note how sorting according to $\beta_i$ factor a smaller number of modes of a certain relevance is retrieved: i.e., the $\beta_i - 3(\alpha_i)$ distribution is narrower. This means that the number of modes associated either with large values of $|\beta_i|$ is always smaller than the number of modes associated with large $|\alpha_i|$, see Figs. 3(b). This phenomenon was also observed in a composite DMD analysis conducted on $Re_t = 200$ channel flow database [3].

The advantage of using parameter $\beta_i$ over $\alpha_i$ to classify the different DMD modes retrieved becomes apparent from Figure 4: whereas the superposition of a fair amount of classical DMD modes associated with large values of $|\alpha_i|$ is incapable of providing an accurate reconstruction of the DNS Reynolds stress profile, observe how the combination of few ($n_r = 19$, less than 2%) composite DMD modes associated with large values of $|\beta_i|$ reconstructs quite accurately the Reynolds stress profile. This supports the conjecture that modes associated to larger values of $|\beta_i|$ are more relevant to the physics investigated than those associated to large values of $|\alpha_i|$.

The averaged reconstructed Reynolds stress transversal distribution (consider Eq. 5 with $\langle \circ \rangle_{x,t}$) can also be visualized, see Fig. 5. Fig. 5(a) and Fig. 5(b) show such reconstructions for composite ($C_f - u'v'$) DMD analyses of $Re_t = 200$ ([3]) and $Re_t \approx 950$ turbulent channel flows using modes with $|\beta_i|/|\beta_{\text{max}}| \geq 10\%$. Fig. 5(c) is the same, but obtained from the parallel composite DMD analysis technique described above. Comparison Figs. 5(b) and (c) shows how features extracted in [1] can still be found at corresponding spanwise locations. However, features that are relevant at the near-wall region are seen, thanks to the parallel composite DMD algorithm, to extend beyond $y^+ = 100$.

Comparison with the composite DMD analysis of the near-wall region for a $Re_t = 200$ channel flow, see Fig. 5(a), shows that the spanwise spacing between features is still in the range $\Delta z^+ \approx$
100–150. This is consistent with the findings in e.g. [10][19]. However, the centers of these features seem to lie slightly beyond \( y^+ = 50 \). This may indicate that, at this moderate Reynolds number \( (Re_t \approx 950) \), motions associated to skin friction comparatively larger scale than those at lower Reynolds.

In conclusion, by allowing to face larger subdomains for longer temporal spans, the parallel, composite DMD technique presented in this manuscript allows to better characterize the \( Re_t \approx 950 \) database.

---

**Figure 5.** Composite DMD analysis of two turbulent channel flow databases, using hybrid \( C_f(t_k) - u'v'(\tilde{x}, t_k) \) snapshots. Reconstructed DMD flow field using Eq. 5: (a), \( n_r = 9 \); (b), \( n_r = 1 \); (c), \( n_r = 19 \).

4. Conclusion

In this contribution, we have reported a parallel, composite DMD investigation on a turbulent channel flow at moderate Reynolds number \( (Re_t \approx 950) \). This work is a follow up of that presented in [1]: the parallel DMD algorithm described in [4] has been introduced to enable the analysis of the turbulent database considering a larger subdomain and for a longer time span.

Pursuing an efficient use of the computational resources available, the data size was carefully reduced to still ensure the relevant turbulent flow physics is retained. Both reductions of the spatial resolution and temporal extent have been undertaken. A snapshot sequence of \( n_s = 1200 \) snapshots, spatial decimation \( s_x = s_z = 2 \) and considering half channel was finally deemed an acceptable input for our DMD investigation.

With the composite DMD analysis, and the classification of the modes according to factor \( \beta_i \), we identified a small number of dynamic modes \( (n_r = 19, \text{about } 1.6\% \text{ of the original sequence length}) \) and still reconstruct the shear stress properly. Flow features present in the near-wall region but extending well beyond \( y^+ = 100 \) and connected to skin friction were identified. These features share some properties with similar ones identified in \( Re_t = 200 \) database, as in [3].

In sum, this work has described the application of the parallelized composite DMD technique to a higher Reynolds number turbulent channel flow, showing the possibility of investigating large database using DMD techniques. The results pave the way to apply further composite DMD analysis, e.g. using \( C_f(t_k) - u'v'(\tilde{x}, t_k) \) or \( C_f(t_k) - \lambda_2(\tilde{x}, t_k) \) on large Reynolds number database sequences, say \( Re_t > \)
1000~2000. This task will surely benefit as well from the spatial agglomeration strategies recently presented in [18].

Acknowledgments
The support provided by China Scholarship Council (CSC, No.201806320222) during a visit of Mr. Binghua Li to UPM-ETSIAE is acknowledged. This work is funded by Projects: Drag Reduction via Turbulent Boundary Layer Flow Control (DRAGY, GA-690623), 2016-2019, China-EU Aeronautical Cooperation project, co-funded by Ministry of Industry and Information Technology (MIIT), China, and Directorate-General for Research and Innovation (DG RTD), European Commission. The authors also acknowledge the computer resources, technical expertise and assistance provided by the Supercomputing and Visualization Center of Madrid (CeSViMa).

References
[1] B. Li, J. Garicano-Mena, and E. Valero. Feature extraction from turbulent channel flow databases via composite dmd analysis. J. Phys.: Conf. Ser., 1522:012008, 2020.
[2] UPM turbulent database. https://torroja.dmt.upm.es/turbdata. Accessed: 2020-04-10.
[3] J. Garicano-Mena, B. Li, E. Ferrer, and E. Valero. A composite dynamic mode decomposition analysis of turbulent channel flows. Phys. Fluids, 31(11):115102, 2019.
[4] T. Sayadi and P.J. Schmid. Parallel data-driven decomposition algorithm for large-scale datasets: with application to transitional boundary layers. Theor. Comput. Fluid Dyn., 30(5):415-428, 2016.
[5] G. Berkooz, P. Holmes, and J.L. Lumley. The proper orthogonal decomposition in the analysis of turbulent flows. Annu. Rev. Fluid Mech., 25(1):539-575, 1993.
[6] S. Volkwein. Proper orthogonal decomposition: Theory and reduced-order modelling. Lecture Notes, University of Konstanz, 4(4):1-29, 2013.
[7] C.W. Rowley, I. Mezić, S Bagheri, P Schlatter, and D.S. Henningson. Spectral analysis of nonlinear flows. J. Fluid Mech., 641:115-127, 2009.
[8] P.J. Schmid. Dynamic mode decomposition of numerical and experimental data. J. Fluid Mech., 656:5-28, 2010.
[9] M.R. Jovanović, P.J. Schmid, and J.W. Nichols. Sparsity-promoting dynamic mode decomposition. Phys. Fluids, 26(2):024103, 2014.
[10] A. Cassinelli, M. de Giovanetti, and Y. Hwang. Streak instability in near-wall turbulence revisited. J. Turbul., 18(5):443-464, 2017.
[11] S. Le Clainche, J.M. Vega, and J. Soria. Higher order dynamic mode decomposition of noisy experimental data: The flow structure of a zero-net-mass-ux jet. Exp. Therm. Fluid Sci., 88:336-353, 2017.
[12] J. Kou and W. Zhang. An improved criterion to select dominant modes from dynamic mode decomposition. Eur. J. Mech. B Fluids, 62:109-129, 2017.
[13] T. Sayadi, P.J. Schmid, J.W. Nichols, and P.Moin. Reduced-order representation of near-wall structures in the late transitional boundary layer. J. Fluid Mech., 748:278-301, 2014.
[14] A. Lozano-Durán and J. Jiménez. Effect of the computational domain on direct simulations of turbulent channels up to $Re_t = 4200$. Phys. Fluids, 26(1):011702, 2014.
[15] A. Lozano-Durán and J. Jiménez. Time-resolved evolution of coherent structures in turbulent channels: characterization of eddies and cascades. J. Fluid Mech., 759:432-471, 2014.
[16] Y. Saad. Numerical methods for large eigenvalue problems. Manchester University Press, 1992.
[17] J. Demmel, L. Grigori, M. Hoemmen, and J. Langou. Communication-optimal parallel and sequential qr and lu factorizations. SIAM J. Sci. Comput., 34(1):A206-A239, 2012.
[18] B. Li, J. Garicano-Mena, Y. Zheng, and E. Valero. Dynamic mode decomposition analysis of spatially agglomerated flow databases. Energies, 13(9):2134, 2020.
[19] M. Lagha, J. Kim, J.D. Eldredge, and X. Zhong. A numerical study of compressible turbulent boundary layers. Phys. Fluids, 23(1):015106, 2011.