

Lossy Data Compression Effects on Wall-bounded Turbulence: Bounds on Data Reduction

Evelyn Otero^{1,2}  · Ricardo Vinuesa^{1,2} · Oana Marin³ · Erwin Laure⁴ · Philipp Schlatter^{1,2}

Received: 30 October 2017 / Accepted: 16 April 2018 / Published online: 25 May 2018
© The Author(s) 2018

Abstract Postprocessing and storage of large data sets represent one of the main computational bottlenecks in computational fluid dynamics. We assume that the accuracy necessary for computation is higher than needed for postprocessing. Therefore, in the current work we assess thresholds for data reduction as required by the most common data analysis tools used in the study of fluid flow phenomena, specifically wall-bounded turbulence. These thresholds are imposed a priori by the user in L_2 -norm, and we assess a set of parameters to identify the minimum accuracy requirements. The method considered in the present work is the discrete Legendre transform (DLT), which we evaluate in the computation of turbulence statistics, spectral analysis and resilience for cases highly-sensitive to the initial conditions. Maximum acceptable compression ratios of the original data have been found to be around 97%, depending on the application purpose. The new method outperforms downsampling, as well as the previously explored data truncation method based on discrete Chebyshev transform (DCT).

Keywords Lossy data compression · Data reduction · Turbulence statistics · Orthogonal polynomials · Resilience

✉ Evelyn Otero
otero@kth.se

¹ Linné FLOW Centre, KTH Mechanics, SE-100 44, Stockholm, Sweden

² Swedish e-Science Research Centre (SeRC), Stockholm, Sweden

³ MCS, Argonne National Laboratory, Lemont, IL 60439, USA

⁴ Center for High Performance Computing, PDC. KTH, SE-100 44, Stockholm, Sweden

1 Introduction

The field of computational fluid dynamics (CFD) is data intensive, particularly for high-fidelity simulations. Direct and large-eddy simulations (DNS and LES, respectively) of turbulent flows, which are framed in this high-fidelity regime, involve a wide range of flow scales leading to a high number of degrees of freedom. Besides the computational bottleneck brought by the large size of the problem, a less explored aspect in the context of flow simulations is the data management. Large amounts of disk space and also slow speed of I/O (input/output) impose limitations on large-scale simulations. Typically the computational requirements for well-resolved flow structures are far higher than those needed for post-processing. Despite the wealth of strategies for truncating the data, such as using wavelets [1–3], reduced-order models [4], or orthogonal transforms [5], only few studies [2, 3, 6–9] have considered an analysis of the impact of data truncation on the accuracy of the postprocessed data from a physical point of view. Here we focus on one particular lossy data compression strategy, which has been equipped with an a-priori error estimator in L_2 -norm, and explore different levels of accuracy of the compressed data. The purpose is to identify bounds within which computational data can be reduced without affecting the ulterior postprocessing, i.e., determination of relevant flow quantities such as turbulence statistics, power-spectral densities or coherent structures. Also we underline the superiority of a data compression strategy over a common approach in engineering applications, namely to downsample the data set when it becomes computationally cumbersome.

In the present work we use the spectral-element code Nek5000 [10] to produce the data under analysis. The method considered here is an orthogonal-transform-based method inspired by the discrete Chebyshev transform (DCT), widely used in the image compression community [11], as well as in CFD applications [12]. However, as we previously identified (see the work by Marin et al. [13]), it is preferable to employ a spectral transform tailored to the spectral discretization of the underlying code. Since the numerical code under consideration is based on a Gauss–Lobatto–Legendre (GLL) grid, we employ an orthogonal transform based on Legendre grids, to which we shall refer as the discrete Legendre transform (DLT). However similar ideas using orthogonal transforms can be used on grids that do not correspond to spectral discretizations, such as finite volumes or finite differences. The difficulty in these cases though would pertain to adapting to curvilinear surfaces and achieving efficiency, both issues that are easily addressed in conjunction with variational formulations and finite element type of discretization. Also on equidistant grids it would be more suitable to apply the discrete Cosine transform since the mapping to a spectral grid is not stable for very high polynomial orders. The current choice of compression algorithm adapted to the mesh and spectral discretization is nonetheless convenient in the context of turbulence simulations which are widely performed using spectral element discretizations.

Some relevant techniques used to analyze turbulent flows include the computation of turbulence statistics, which are based on averages of flow realizations in time and in any homogeneous direction; the assessment of power-spectral density maps, based on Fourier analysis performed in the homogeneous directions; and identification of coherent structures in the flow, which is based on the extraction of relevant flow features from the instantaneous flow fields. In the present study we consider these analysis techniques in a canonical flow case, namely the turbulent flow through a straight pipe [14], and also on a more complex geometry, i.e. the turbulent flow around a NACA4412 wing section [15]. Additionally, we consider the impact of data truncation on restarts from previous flow fields stored in disk, assessed in a highly sensitive case of transitional flow, namely the jet in cross flow [16].

The article is organized as follows: in Sections 2 and 3, a description of the lossy data compression under consideration, namely the DLT, is provided from a mathematical and the implementation point of view, respectively; the test cases used for investigation are described in Section 4; an assessment of the impact of such compression on visualizations, turbulence statistics, spectral analysis, and simulation resilience in several test cases is presented in Section 5; finally, a summary of the conclusions of the present study is given in Section 6.

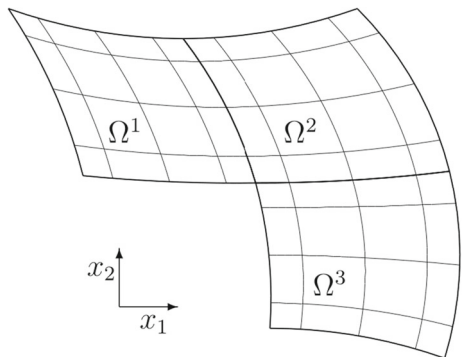
2 Mathematical Formulation of Data Compression Algorithm

Orthogonal transforms have been reported to yield high energy-compactness ratios [5]. This feature makes them very good candidates for data compression, especially on spectral grids which are widely used in direct numerical simulations (DNS) of turbulent flows [17]. Additional benefits, such as high computational efficiency due to tensor-product implementations on hexahedral grids [18], further increase their advantage due to the fact that data compression should be designed to be of negligible computational cost and low accuracy loss.

The data for a spectral-element implementation using a Galerkin formulation is typically mapped to reference elements where the computation is performed with spectral accuracy. For a domain Ω discretized in quadrilateral (hexahedral in 3 dimensions) elements Ω_e , the total mesh can be represented as in Fig. 1, such that $\Omega = \cup_e \Omega_e$. A local restriction of the global velocity field \mathbf{u} down to the element level is $\mathbf{u}_e = \mathbf{u}|_{\Omega_e}$.

The solution is defined via an expansion in orthogonal polynomials ϕ defined at the element level as $u = \sum_{i=0}^N a_i \phi_i$ which for higher dimensions is easily expanded via separability to $\mathbf{u} = \sum_{ijk=0}^N a_{ijk} \phi_{ijk}$. Therefore it is natural to make all theoretical considerations in one dimension and they naturally carry to any higher dimension. The partial differential equation $\mathcal{L}\mathbf{u} = \mathbf{f}$ is solved in weak form via a continuous Galerkin approach by minimizing the projection of the residual on a space of test functions \mathbf{v} , details can be found in e.g. [17]. The approach ensures C_0 continuity between elements, i.e. the function values are the same, but not derivatives. Here we concentrate on connecting the orthogonal transform to the polynomial expansion and use an orthogonal transform tailored to the numerical scheme employed by the code. The specific implementation in the code Nek5000 [10] uses Legendre polynomials for the expansion of the solution and hereby we use the Discrete Legendre Transform (DLT) as outlined in [13].

Fig. 1 Sample of spectral-element mesh. Figure extracted from [17]



The compression will be performed element-wise by applying an orthogonal transform to obtain the equivalent of the data field in spectral space. The transform here is denoted by T and can either be a discrete cosine or discrete Legendre transform, however we focus here on the DLT which is given by Legendre polynomials. A spectral space transformed field \mathbf{v} is related to the original field \mathbf{u} as $\mathbf{v} = T\mathbf{u}$, and the transform matrix T is given by:

$$T_{ij} = f_j \phi_i(x_j), \quad i, j = 0, \dots, N, \tag{1}$$

where in the case of the DLT the Legendre polynomials ϕ_i are obtained via the recurrence relation:

$$(i + 1)\phi_{i+1}(x) = (2i + 1)x\phi_i(x) - i\phi_{i-1}(x), \quad i = 1, \dots, N, \tag{2}$$

where $\phi_0(x) = 1, \phi_1(x) = x$, and:

$$f_j = \begin{cases} \sqrt{(2(j + 1) - 1)/2}, & 0 \leq j \leq N - 1, \\ \sqrt{(N - 1)/2}, & j = N. \end{cases} \tag{3}$$

The L_2 -norm over a field can be obtained via the discrete equivalent approximation given as $\int_{\Omega} \mathbf{u}^2 d\Omega \approx \mathbf{u}^T W \mathbf{u}$, where W is the integration weight matrix given by:

$$W_{ij} = \frac{2}{N(N + 1)} \frac{1}{\phi_N^2(x_i)} \delta_{ij}, \quad i, j = 0, \dots, N. \tag{4}$$

Furthermore, orthogonal transforms such as DLT (or DCT), since they stem from orthogonal polynomials, have the property that discretely $T^T W T = I$, where W is again the diagonal weight matrix.

Now using Eqs. 1 and 4 we can construct the forward and backward transform. Once the data is transformed into spectral space, the decay is much faster and it is easy to perform a truncation such that it satisfies an a-priori error threshold imposed by the user.

Regarding the truncation procedure, note that since for general curvilinear meshes certain elements in the computational domain may be significantly distorted, the L_2 -norm of the global velocity field $\|\mathbf{u}\|_{L_2}$ needs to be weighted by $V = \int_{\Omega} d\Omega$, i.e. the volume of the domain. We denote this weighted norm as:

$$\|\mathbf{u}\|_v = \sqrt{\frac{\|\mathbf{u}\|_{L_2}^2}{\int_{\Omega} d\Omega}}. \tag{5}$$

Given the fact that the entire mesh is decomposed into hexahedral elements e , we have $\mathbf{u}|_{\Omega_e} = \mathbf{u}_e$. Then it is possible to write the global L_2 -norm as a sum of local norms as:

$$\|\mathbf{u}\|_v^2 = \sum_e \|\mathbf{u}_e\|_{v_e}^2, \tag{6}$$

where $v_e = \int_{\Omega_e} d\Omega_e$ is the volume per element. Straightforward algebra yields that a volume-weighted error threshold down to the element level is identical to the one over the entire mesh, i.e. $\|\mathbf{u}\|_v \leq \epsilon$ translates into $\|\mathbf{u}_e\|_{v_e} \leq \epsilon$.

Moreover, the fact that we operate on curvilinear meshes which are mapped to a reference element leads to a rescaling by the Jacobian of the orthogonality relation $T^{-1} W T = I$ as described in [13]. The truncation procedure relies on the fact that the L_2 -norm in spectral space over one element is equal to the L_2 -norm in real space given as $\|\mathbf{u}\|_{v_e}^2 = \|\mathbf{v}\|_{v_e}^2$. Thus, imposing a threshold over the entire velocity field \mathbf{u} , $\|\mathbf{u}\|^2 \leq \epsilon$, is equivalent to $\|\mathbf{v}\|_{v_e}^2 \leq \epsilon$.

To summarize, the algorithm for the DLT-based data compression proposed in [13] is illustrated in Fig. 2 distinguished from DCT with the subscript "DLT" and consists of the following steps:

- Convert data field to spectral space: $\mathbf{v} = T \mathbf{u}$.
- Truncate in spectral space: \mathbf{v} being truncated to $\tilde{\mathbf{v}}$ (remove low-magnitude spectral coefficients).
- Write to disk compressed data field: $\tilde{\mathbf{v}}$.
- Recover data field: $\tilde{\mathbf{u}} = T^{-1} \tilde{\mathbf{v}}$.

It is important to note that the truncation algorithm is identical irrespective of the applied transform, and all throughout the remainder of this manuscript we shall refer to the DLT or DCT algorithms, although the only difference is given by the orthogonal transform applied. This will be further discussed in the next section.

3 Implementation of the Method

3.1 Description of the compression procedure

Figure 2 gives a general overview of the compression procedure used for this investigation, where both the DLT and the DCT methods can be used to convert the data to spectral space (and vice versa). As it can be observed, the difference between the DLT and the DCT compression is the additional mapping into a Chebyshev grid required for the latter. The user has control over the level of compression by prescribing an error threshold as an input to the truncation. Once the truncation is performed, the resulting compression ratio is obtained. Therefore, the error threshold is an input, and the compression ratio is an output of the truncation process.

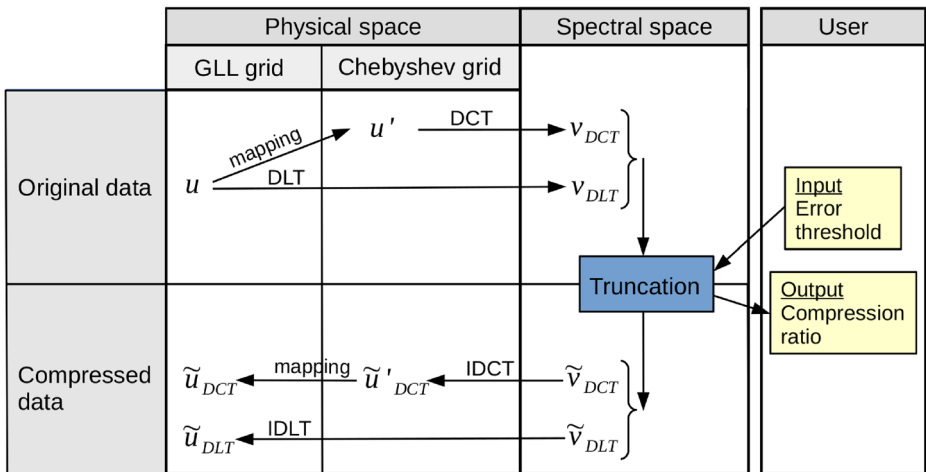


Fig. 2 Schematic representation of the compression procedure considering both DLT and DCT to convert the data to spectral space and vice-versa. Physical data expressed on the Chebyshev grid is denoted by $(\cdot)'$, and $(\tilde{\cdot})$ indicates compressed data

A simplified description of the truncation algorithm [13] is given in Algorithm 1, where $w = Tu_{el}$ (line 2) is the mode amplitude vector for element el , ϵ_{el} (line 4) is the error threshold adapted to the volume of the element el , w_{trunc} (line 6) is the resulting mode amplitude vector for element el after truncation, Cr_{el} (line 7) is the compression ratio for element el , and n_{tot} (line 7) is the total number of grid points per element.

Algorithm 1 Truncation algorithm on a spectral element el

```

1: procedure TRUNCATION
2:    $sort(w)$ 
3:    $w_{trunc} \leftarrow w$ 
4:    $n \leftarrow 0$ 
5:   while  $\sqrt{\sum_{j=1}^{n+1} w^2(j)} < \epsilon_{el}$  do
6:      $n \leftarrow n + 1$ 
7:      $w_{trunc}(n) \leftarrow 0$ 
8:    $Cr_{el} \leftarrow \frac{n}{n_{tot}}$ 

```

In fact the truncation is defined as opposed to downsampling exclusively in terms of mode amplitude and it is performed up to a user-controlled error level (line 4). Therefore, this compression method truncates data in an adaptive way by setting to zero the modes with lowest energy contribution to the flow (line 6). These modes are first sorted by increasing energy content through the function $sort$ (line 2). Finally, the error estimator allows to control the error incurred through the truncation by comparing the L_2 -norm of the error being applied, with the provided error threshold (line 4). The implementation of the whole algorithm was performed with high efficiency in mind by employing tensor products as described in [13].

In this work we define the overall compression ratio Cr as the arithmetic average of the compression ratios Cr_{el} over all the elements. The compression ratio can be then defined as $Cr = \frac{\text{lost data size}}{\text{original data size}}$ computed as the number of entries set to zero over the total number of grid points averaged over the elements. Thus, higher compression ratios imply smaller file sizes.

Figure 3 illustrates the behavior of the DLT truncation on a test case, namely the turbulent flow through a straight pipe, which is further described in Section 4 and in [14]. In this figure we also show the compression-ratio results obtained by using the DCT, as in our previous work [19]. Note that, as discussed in Fig. 2, the only difference between using DLT and DCT is the employed transform, and that the truncation part of the algorithm is identical in both cases. This figure shows that, for the same error threshold of 10^{-4} , the Cr curve exhibits qualitatively similar behavior when using both transforms, although DLT leads to higher compression ratios. As described above, the compression ratio Cr is defined for each local element, and therefore in Fig. 3 (left) it is possible to identify the spectral-element mesh distribution in the cross-sectional area of the pipe. Despite the wide variation in the values of Cr , which range from around 10.5% to 98.2% in a particular streamwise position of the pipe, it can be observed that the Cr field exhibits a certain structure, which is connected to that of the instantaneous flow field. The highest Cr values are located close to the pipe walls, then the compression ratio decreases farther from the wall before increasing again towards the core of the pipe. This is confirmed by the results in Fig. 3 (right), which show the Cr profile value averaged in time, streamwise and azimuthal directions as a function of the wall-normal direction y^+ . Note that in this study the superscript ‘+’ indicates scaling

in viscous units, i.e. in terms of the fluid kinematic viscosity ν and the friction velocity $u_\tau = \sqrt{\tau_w/\rho}$ (where τ_w is the wall-shear stress and ρ is the fluid density).

A very important conclusion inferred from Fig. 3 (right) is the fact that the highest compression ratios can be achieved when the turbulence level is low: Cr values above 80% are observed within the viscous sublayer ($y^+ < 7$ approximately). The increase in streamwise velocity fluctuations (see Section 5.2 for a detailed description of the turbulence statistics) significantly reduces the average compression ratio below 60%, before increasing again up to around $Cr = 70\%$. It is important to note that the particular trend observed in Fig. 3 (right) is also determined by the non-uniform grid distribution in the wall-normal direction, and the fact that the value of Cr is unique for each spectral element. In any case, this result is consistent with the fact that it is much more complicated to compress a highly-turbulent signal, spanning a wide range of scales, than a close-to-laminar one which contains little information in addition to its mean component. Also note how the lowest instantaneous Cr of 10.5% is much lower than the average one, a fact that is connected with the highly-varying instantaneous turbulence intensities in wall-bounded turbulent flows.

The error threshold considered for the turbulent pipe in Fig. 3, 10^{-4} , was also employed to compress the fields of the turbulent flow around the NACA4412 wing section (Fig. 4) described in Section 4 and [20]. As opposed to the pipe, which is an internal flow where the only location with zero turbulent fluctuations is essentially the wall, in the external flow around a wing section there is a significant region of the domain, beyond the boundary layers developing around the wing, where the fluctuations are zero as well. This significantly affects the possibilities of data compression, as shown for the spanwise domain centerplane in Fig. 4. It can be observed how the obtained Cr is above 99.9% throughout most of the computational domain, essentially outside the turbulent boundary layers. On the other hand, the average Cr is 82%, only slightly higher than that obtained for the pipe. This is due to the fact that the Cr is much lower around the boundary layers, which require a significantly larger number of elements to be accurately simulated than the farfield. Since the compression ratio does not depend on the element size but on the number of grid points per element (which is the same for all the elements), the resulting difference in Cr is not as high as

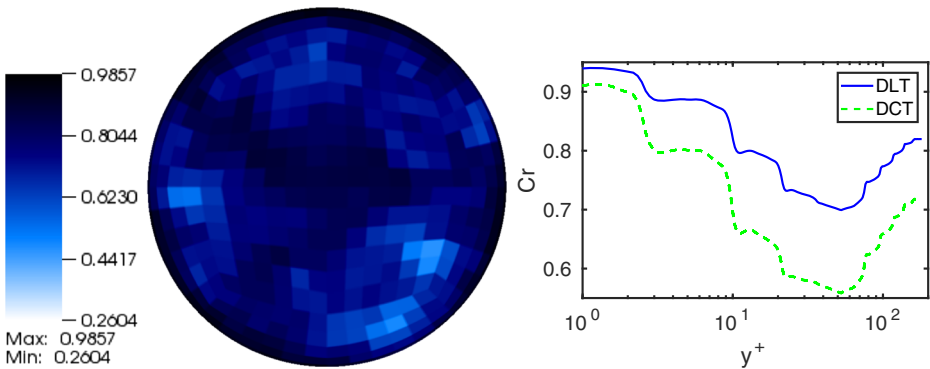


Fig. 3 DLT truncation with an input error threshold of 10^{-4} , resulting in an averaged compression ratio of 79%. (Left) Instantaneous compression ratio distribution per element at a certain streamwise location in the pipe-flow case described in Section 4. (Right) Compression ratio profile averaged in time, streamwise and azimuthal directions as a function of the inner-scaled wall-normal distance. Results from both DLT and DCT are shown for comparison purposes

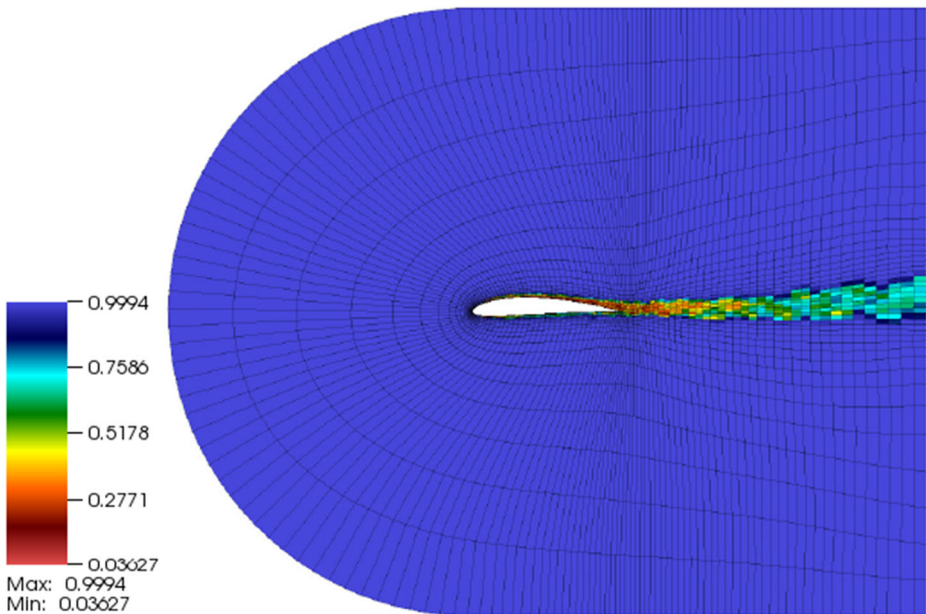


Fig. 4 DLT truncation with an input error threshold of 10^{-4} , resulting in an averaged compression ratio of 82%. Instantaneous compression ratio distribution per element at the spanwise centerplane in the NACA4412 wing section case described in Section 4. The spectral-element mesh (without GLL points) is also shown

expected. Moreover, Cr increases in the wake as the outflow is approached, due to the significant decrease of turbulent fluctuations farther downstream from the wing trailing edge. It is important to highlight that a number of different factors determine the achievable Cr for a certain error threshold: the flow around the wing section was computed with LES, a fact that explains the relatively lower Cr values observed on the boundary layers, compared with those in the pipe. Moreover, the friction Reynolds number on the wing suction side ranges from $Re_\tau \simeq 76$ up to around 139 (below the Reynolds number in the pipe).

3.2 Comparison of DLT and downsampling

In DLT, the truncation algorithm is based on the amplitude of the modes and not on the mode number as it is the case in the downsampling method. The DLT truncation eliminates data in an adaptive way by removing the weak modes with less energetic contribution to the whole flow field. Since turbulent flows contain a wide range of scales, both in space and in time, it is possible that for a particular instantaneous flow realization certain higher-order modes contain more energy than some of the lower-order modes.

This is illustrated schematically in Fig. 5, which shows the difference between both methods. By performing downsampling, the contribution of high-order modes with significant energy contribution may be eliminated, while preserving less energetic lower-order modes. On the other hand, the DLT-based truncation preserves all the modes above a certain amplitude level, regardless of their position in the spectrum. Therefore, it should in principle be possible to obtain higher Cr values using DLT than using downsampling, for the same level of accuracy.

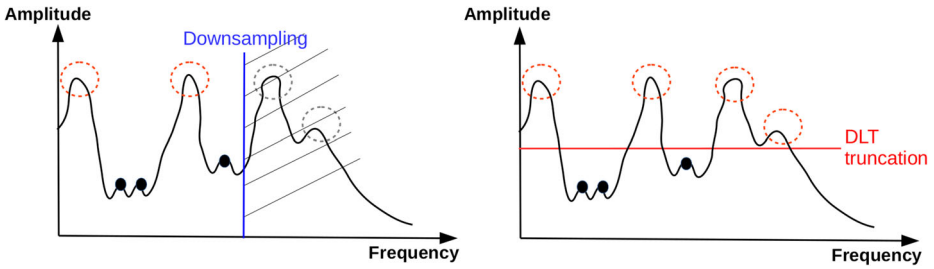


Fig. 5 Schematic representation of spectral truncation process for (left) downsampling and (right) DLT

3.2.1 Description of the downsampling procedure

The compression ratio applied with our downsampling implementation is controlled by a diagonal matrix D , and applied to the solution in the spectral space as:

$$\tilde{u}_{DS} = V^{-1} \underbrace{DVuV^T D^T}_{\text{Spectral space}} (V^{-1})^T, \tag{7}$$

Physical space

where V is the nodal-to-modal transform matrix. Note that, in the notation of Fig. 2, $v_{DS} = VuV^T$ and $\tilde{v}_{DS} = Dv_{DS}D^T$. The compression ratio specific to the downsampling is then defined as:

$$Cr_{DS} = \frac{n_{tot}^3 - d^3}{n_{tot}^3}, \tag{8}$$

with n_{tot}^3 being the number of grid points per element in three dimensions and d the number of diagonal elements set to 1 in D . Note that the filtering is also done in spectral space, as well as the truncation (see Eq. 7 and Fig. 2). Moreover, contrary to the DCT implementation, the spectral space used for the downsampling and the DLT truncation is the same.

The results presented in Section 5 will focus on the DLT performance, which will be compared with the one obtained by means of downsampling.

4 Summary of Test Cases

The first case under consideration is the turbulent flow through a smooth straight pipe, obtained through a fully-resolved DNS by El Khoury et al. [14]. Although this is one of the three main canonical flow cases in wall-bounded turbulence, it has not been investigated numerically as thoroughly as channels [21, 22] or zero-pressure-gradient boundary layers [23, 24], due to the numerical problems associated with the singularity at the pipe center. It is however a very important flow case since it allows direct comparisons with experimental studies due to the lack of problems such as the development of side-wall boundary layers or secondary flows [25]. El Khoury et al. [14] considered a bulk Reynolds number $Re_b = 5,300$, based on bulk velocity and pipe diameter, which corresponds to a friction Reynolds number of $Re_\tau = 181$ (which is formed in terms of the pipe radius and the friction velocity

u_τ , defined in Section 3.1). The spectral-element code Nek5000 [10] was used to perform the simulations, and a total of 36,480 spectral elements were considered to discretize the domain. Since the polynomial order was $N = 7$, the computational mesh contained a total of 18.67 million grid points. An instantaneous streamwise velocity field and the spectral-element mesh are shown in Fig. 6 (left). The flow is periodic in the streamwise direction, in which the domain has a length of 25 pipe radii, and that a total of 205 instantaneous flow fields were stored in the simulation. Additional details about the simulation procedure are provided in [14].

The second test case is the well-resolved LES of turbulent flow around a NACA4412 wing section at a Reynolds number based on inflow velocity and chord length $Re_c = 100,000$, computed by Vinuesa and Schlatter [20]. This case constitutes a moderately complex scenario for the study of wall-bounded turbulence due to the streamwise pressure gradients imposed on the boundary layers developing on the suction and pressure sides of the wing, the incipient separation around the trailing edge and the development of a turbulent wake. The assessment of pressure-gradient and flow-history effects on turbulent boundary layers is a challenging problem which has received considerable attention in the wall-bounded turbulence community in recent years [26]. In the case under consideration, an angle of attack of 5° is introduced, and the spanwise periodic direction has a length of 10% of the wing chord. The computational mesh, which is shown in Fig. 6 (right), consists of 28,000 spectral elements with $N = 11$, which amount to a total of 48.4 million grid points. Additional simulation details, including the implementation in Nek5000, can be found in [15, 27].

Finally, we also consider results from a simulation of jet in cross-flow (JCF), reported in [16, 28]. This case was chosen due to its sensitivity to the initial conditions, and we select a stable configuration [16, 28] with a polynomial order of $N = 7$, and a velocity ratio of $R = 0.63$ defined as $R = V/U_\infty$. Note that V and U_∞ are the peak jet velocity and the free-stream velocity, respectively. A detailed view of the computational mesh, as well as an instantaneous visualization of the vortical structures present in the flow, are shown in Fig. 7.

After defining the various test cases, the impact of data compression on the various analysis techniques is assessed below.

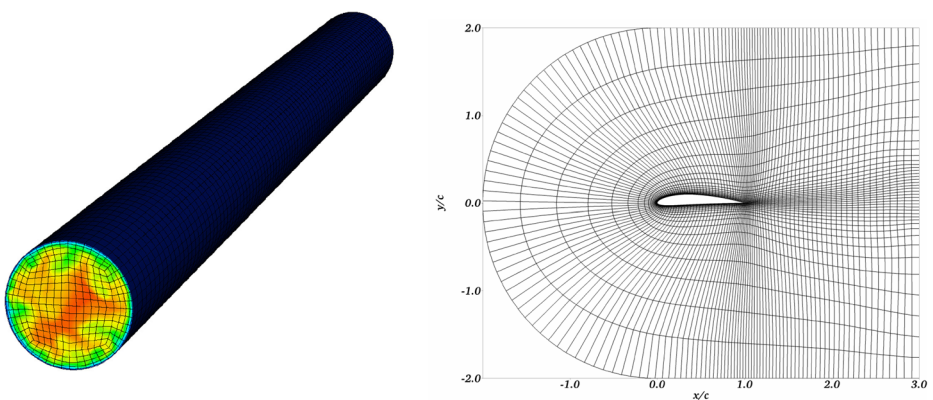


Fig. 6 (Left) Instantaneous field showing streamwise velocity and the spectral-element mesh of the pipe flow case. (Right) Two-dimensional visualization of the computational domain used to simulate the NACA4412 wing section case

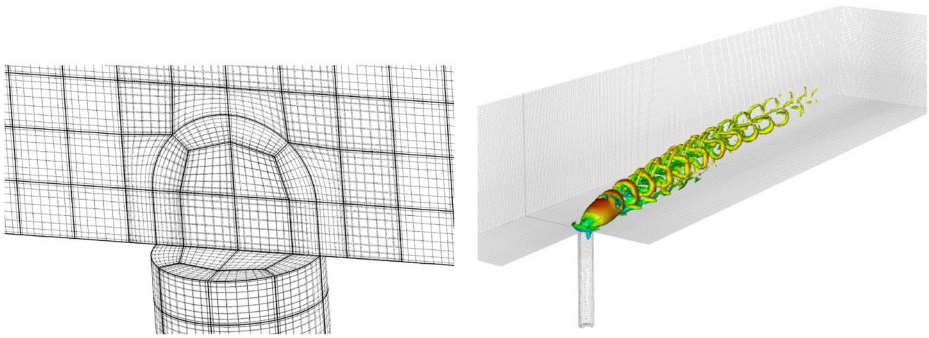


Fig. 7 (Left) Computational mesh of the jet in cross-flow mesh case and (right) instantaneous vortical structures identified in the flow using the λ_2 method [29]

5 Data Analysis on Truncated Data Sets

In this section we investigate the impact of data compression on various analysis techniques commonly used to characterize wall-bounded turbulent flows. This analysis is performed on the test cases described in the previous section (Section 4).

5.1 Flow visualization and vortex identification

In Fig. 8 we show the instantaneous in-plane horizontal velocity in a Cartesian frame of reference at a specific streamwise position of the pipe, for a number of compression ratios ranging from 0% to 97.5%. In this figure it can be observed that, for simple visualization purposes, a compression ratio of around 95.7% is acceptable. Higher compression ratios show discontinuities across element boundaries, as observed in the case with $Cr = 97.5\%$.

The impact of compression on the vortical structures present in the turbulent flow through the pipe, and around the wing section, is studied in Fig. 9. The vortical structures [30] are identified using the λ_2 criterion [29], fixing the value of the represented isosurface in viscous units. Doing so, we account for the changes in friction velocity derived from the compression. It is interesting to observe that in the pipe case the vortical structures obtained with 88.5% compression are in very good agreement with those obtained in the original field. The additional Cr cases, 95.7% and 97.5%, exhibit progressively larger distortion in the structure topology and as well more noisy fields. Regarding the turbulent boundary

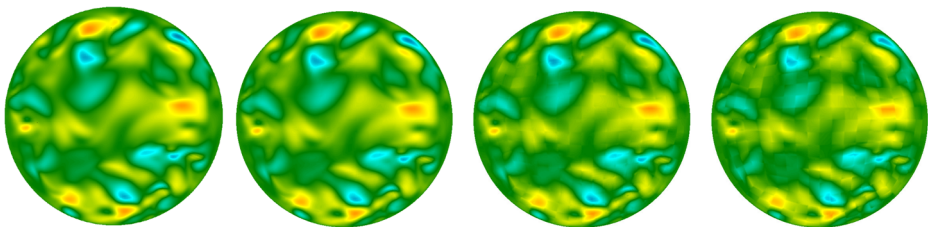


Fig. 8 Instantaneous in-plane horizontal velocity at a streamwise position of the pipe, for compression ratios of (from left to right) 0%, 88.5%, 95.7% and 97.5% using DLT

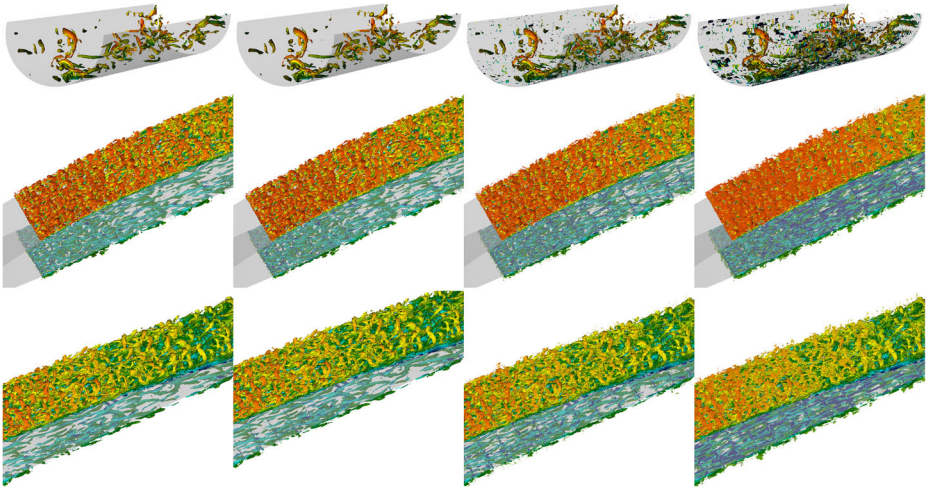


Fig. 9 Vortical structures identified with the λ_2 criterion [29] in the turbulent flow (top row) through the pipe and (middle and bottom rows) around the wing section. The compression ratios under consideration are (from left to right) 0%, 88.5%, 95.7%, and 97.5% for the pipe, and 0%, 84.9%, 88.1%, and 89.2% for the wing. The isosurfaces used to visualize the coherent structures were kept fixed in viscous scaling for increasing Cr values

layers around the wing, the visualization with $Cr = 84.9\%$ exhibits very good agreement with the uncompressed data for the two regions of the wing shown in Fig. 9: the first one at a streamwise distance of around 30% of the chord from the leading edge, and the second one at around 80%. Note that whereas the first region is subjected to a small streamwise pressure gradient, the second one exhibits a strong adverse pressure gradient, a fact that significantly affects the topology and energy content of the vortical structures. As in the case of the pipe, the larger Cr values of 88.1% and 89.2% lead to progressively more distorted structures, as well as larger noise levels. Regarding the compression levels yielding a certain error in the instantaneous visualizations, it is important to note that whereas the pipe flow was obtained from a DNS, the wing fields were obtained from a LES. This implies that the lower resolution employed in the latter already establishes a more restrictive bound regarding attainable data compression for a certain level of accuracy.

5.2 Turbulence statistics

In Fig. 10 we analyze the effect of various compression ratios on several turbulence statistics calculated based on the 205 instantaneous fields from the pipe case. We computed the inner-scaled mean velocity profile U^+ , as well as the streamwise and azimuthal velocity fluctuations (denoted by $\overline{u^{2+}}$ and $\overline{w^{2+}}$, respectively) and the Reynolds-shear stress $\overline{uv^+}$. To this end, we first computed the mean profile and Reynolds stresses based on time and streamwise averages of the three-dimensional fields. The resulting two-dimensional field was then interpolated on a number of radial profiles distributed along the azimuthal direction, and the in-plane velocity components were rotated before performing the averaging in the homogeneous azimuthal direction. Additional details regarding the computation of turbulence statistics and interpolations in Nek5000 are provided in [31]. The turbulence

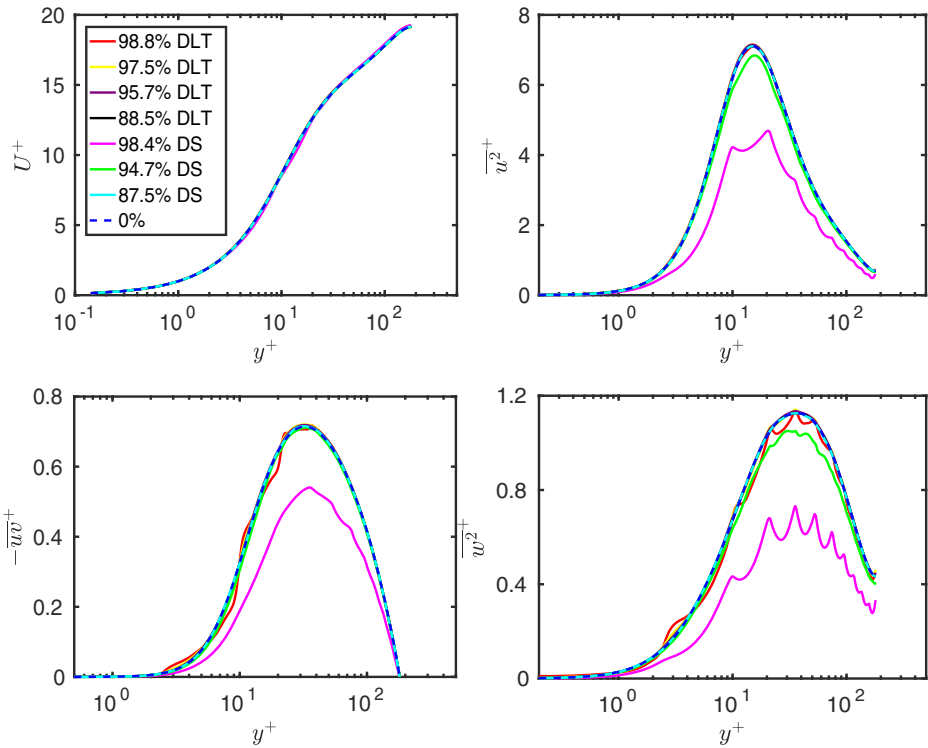


Fig. 10 Turbulence statistics for the turbulent pipe case at $Re_\tau = 180$, obtained from fields with different levels of compression. The results obtained with DLT and downsampling are labelled as DLT and DS, respectively. All the cases include the corresponding Cr , where the uncompressed data corresponds to 0%

statistics were computed based on the original, uncompressed fields, and also based on the same fields after applying several levels of compression by means of both DLT and downsampling. Figure 10 shows that using DLT it is possible to apply compression ratios up to 97.5% and obtain excellent agreement with the original data in the four profiles under consideration. On the other hand, using downsampling already leads to noticeable deviations with $Cr = 94.7\%$, especially in the azimuthal velocity fluctuation profile. This figure also indicates that the mean flow is less sensitive to the effect of compression than the fluctuations, since all the cases under consideration are in reasonably good agreement with the uncompressed data, which is not the case in the Reynolds-stress profiles. This observation is related to the findings by Örlü and Schlatter [32], who documented the attenuation in the fluctuating wall-shear stress after applying downsampling to a turbulent boundary layer data set obtained through DNS. Also note that when considering a larger $Cr \simeq 98\%$ the DLT method clearly outperforms the downsampling, since the latter shows significant deviations with respect to the reference profiles in all the fluctuating quantities.

In the following we perform a more detailed analysis of the effect of the compression on the statistics, namely regarding the increased errors with respect to the uncompressed case (which is considered as the baseline result). In particular, we consider the indicator e defined below, which is the relative difference between compressed and original data evaluated at

each wall-normal location, and integrated over y^+ . This indicator is defined as follows for a given quantity of interest ϕ :

$$e = \frac{1}{Re_\tau - 1} \int_1^{Re_\tau} \frac{\sqrt{(\phi_c^+ - \phi_o^+)^2}}{\phi_o^+} dy^+, \tag{9}$$

where ϕ_o^+ and ϕ_c^+ are the original and compressed quantities of interest, respectively. The following quantities of interest are considered in the present study: U^+ , $\overline{u^{2+}}$, $\overline{v^{2+}}$, $\overline{w^{2+}}$ and $\overline{uv^+}$. This equation can be used as well with any other turbulence statistics to be analyzed. Note that the integration in Eq. 9 is started at $y^+ = 1$ to avoid the contribution of the very small velocities below this wall-normal location, which lead to very large local values of e that spuriously affect the overall calculation of the error.

In Fig. 11 we show the value of e for each of the five quantities of interest mentioned above as a function of Cr , for compressed fields obtained based on DLT and downsampling. Note that we also include statistics results based on the DCT algorithm, which were previously discussed in [19]. As expected, the values of e increase for all the statistical quantities with Cr , and interestingly the increase of the error associated with the mean velocity profile is less steep than that from the other quantities. This is consistent with the observations

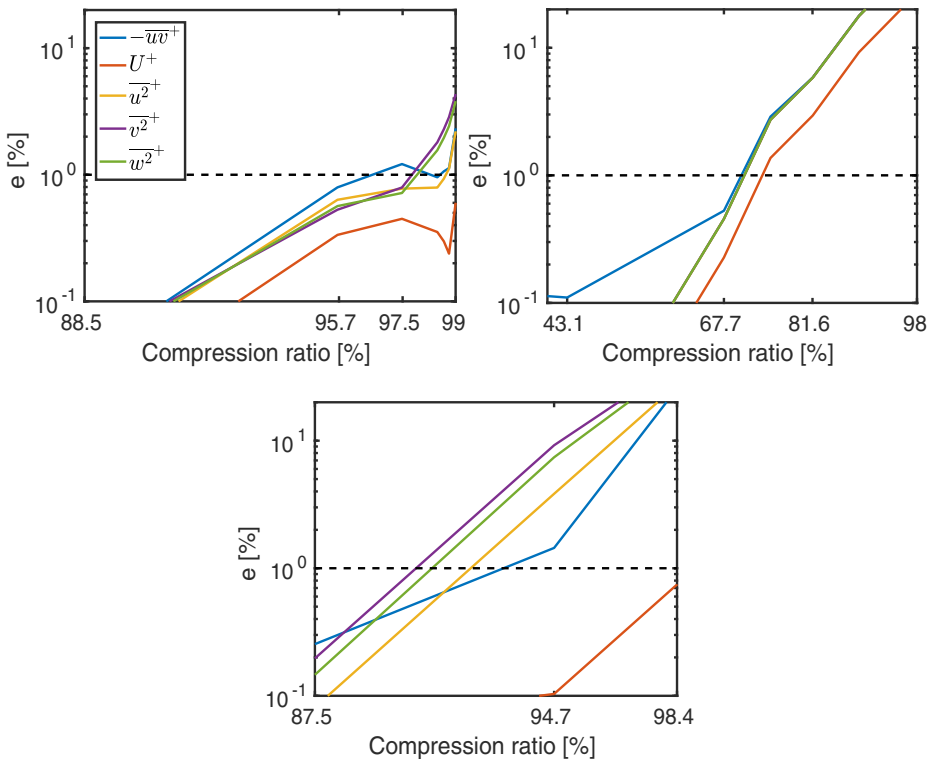


Fig. 11 Averaged relative error of the turbulence statistics as a function of compression ratio, obtained with (top-left) DLT, (top-right) DCT and (bottom) downsampling. The dashed line indicates the maximum acceptable value of $e = 1\%$

made from simple inspection of the profiles, and can be explained by the fact that the compression operation is linear; we can exchange averaging and filtering (i.e. they commute), and thus filtering of the smooth mean flow is likely to induce less errors. Also note that the mean flow converges much faster in time than the fluctuating quantities when computing flow statistics.

Furthermore, in the present work we consider that the maximum admissible level of e is 1%, a threshold which can be used to define a bound on the compression level before the resulting statistics are considered to be inaccurate. Note that typical statistical uncertainty in the evaluation of turbulence quantities from DNS data is on the order of 1%, a fact that motivates this choice as an acceptable error level induced by the compression. Based on this threshold, the compression obtained by means of DLT would yield acceptable values for all the statistics up to $Cr = 97.5\%$, which is consistent to what was observed in Fig. 10, and interestingly if one restricts the analysis to the streamwise mean flow and velocity fluctuations it would be possible to increase Cr up to almost 99% keeping the value of e below 1%. Regarding the compression based on downsampling, a maximum compression ratio of about 90% can be applied with all the statistics within acceptable levels of error, whereas the value of Cr could be increased to 94.7% if only the mean flow and the Reynolds-shear stress are considered. Note that these results indicate that it is possible to apply an additional 7.5% compression ratio (97.5% compared to 90%), while keeping the error levels acceptable, when using DLT with respect to downsampling. Interestingly, these values are significantly larger than the ones obtained when using DCT, which allows a maximum level of compression of 67.7%.

Although the reasons for this discrepancy between DLT and DCT are not fully understood at the moment, we believe that it is related to the fact that the data reduction is performed on the Chebyshev grid in the case of DCT, rather than on the GLL grid where the data is actually represented (as it is the case of DLT and downsampling). In fact these results clearly show, as well as [13], that having an orthogonal transform that corresponds to the original grid, i.e. here GLL grid (Fig. 2), is better than mapping to the Chebyshev grid (Figs. 2 and 11) even though the DCT is a good approximation of the optimal Karhunen-Loève transform [5, 33, 34], allowing a higher compression ratio [35]. The additional error incurred in practical applications on complex curvilinear grids by mapping to a Chebyshev grid (or for the cosine transform on the original equidistant grid) proves to be detrimental for postprocessing. The reason for this lies in the fact that for the same compression ratio the error is higher when a mapping is required, see Fig. 12 and [13].

In addition to the analysis above, where the statistics obtained without compression are used as a baseline, it is possible to define another reference result to assess the impact of the various compression techniques. As discussed for instance in [36], the inner-scaled total shear in fully-developed turbulent channel flows (as well as in straight pipe flows) is a linear function of the outer-scaled wall-normal coordinate y/h (where h is the channel half-height or the pipe radius). This is a direct result of integrating the streamwise momentum balance equation, and is therefore a feature of a *canonical* turbulent channel or pipe flow. This result was used by Vinuesa et al. [37] to define a convergence criterion for wall-bounded turbulence simulations. Essentially, the turbulence statistics can be considered to be converged when the deviation between the total shear and a linear profile, denoted by ε , is below a certain value. The parameter ε , which is defined as follows:

$$\varepsilon = \left[\frac{1}{h} \int_0^h \left(\frac{dU^+}{dy^+} - \overline{uv}^+ - \left(1 - \frac{y}{h}\right)^2 \right)^2 dy \right]^{1/2}, \tag{10}$$

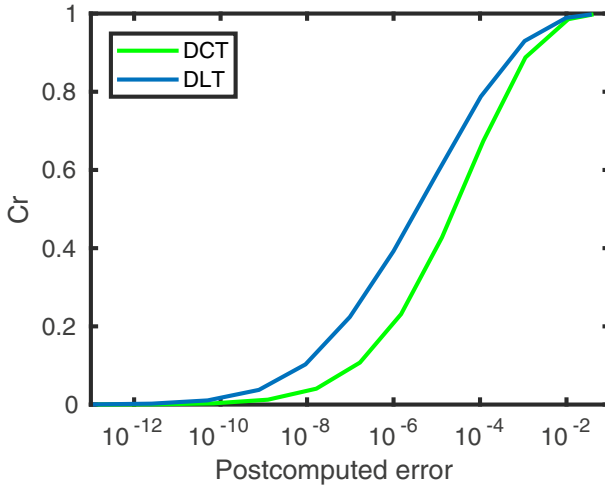


Fig. 12 Performance comparison between DLT and DCT truncation measured through the compression ratio obtained per error level. Compression applied on fields from the turbulent pipe case at $Re_\tau = 180$

can also be considered to evaluate the impact of the compression on the computed turbulence statistics. In Fig. 13 (left) we show the various terms in Eq. 10, namely the viscous and turbulent stresses, their sum and the linear profile, for the case without compression. At the wall, the total stress corresponds to the viscous term, the contribution of which quickly decays together with the increase of the Reynolds-shear stress profile. As expected, the sum of both is in very good agreement with the linear total stress profile, which implies that the statistics from the uncompressed data are representative of the canonical fully-developed pipe flow. The total stress is also computed for the fields compressed using the DLT technique, and it is interesting to note that although the case with 98.5% compression

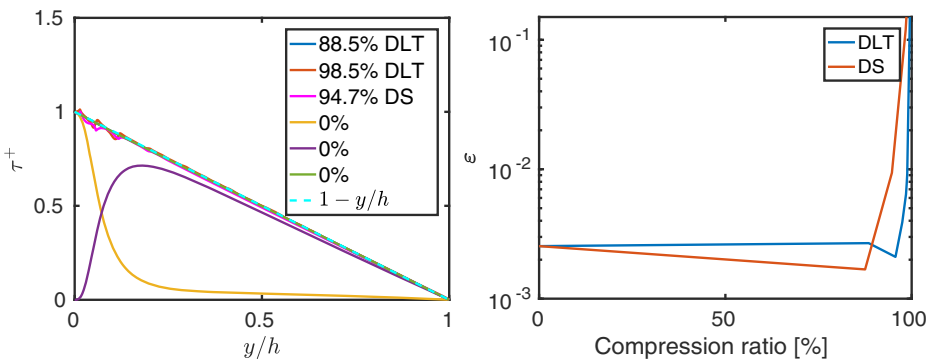


Fig. 13 (Left) Inner-scaled stress, where the 0% lines correspond to the uncompressed profiles of : (yellow) dU^+/dy^+ , (purple) $-\overline{uv}^+$ and (green) the total stress $dU^+/dy^+ - \overline{uv}^+$. The rest of solid lines correspond to the total stress with various compression levels and techniques. (Right) Evolution of the indicator ϵ , defined in Eq. 10, with Cr and compression technique

exhibits deviations from the linear profile, especially up to $y/h \simeq 0.15$, the case with 88.5% compression is in excellent agreement throughout the whole pipe. Note that significant deviations with respect to the linear profile are also observed when a value of $Cr = 94.7\%$ is considered for downsampling. These deviations are explored in further detail by analyzing the evolution of ε with Cr , both from the DLT and the downsampling compression methods. In the work by Vinuesa et al. [37], the authors established that a value of $\varepsilon \simeq 10^{-3}$ was necessary for the turbulence statistics to be completely converged. When using the original flow fields without any compression to calculate turbulence statistics, we obtained a slightly larger value of $\varepsilon = 2.5 \times 10^{-3}$, comparable to that of the simulation by Kim et al. [38], and slightly lower than the ones in the work by Moser et al. [39]. On the other hand, note that the statistics reported by El Khoury et al. [14] were not computed based on the 205 instantaneous flow fields, but rather on more frequent sampling performed during the simulation, as described in [31]. The value of ε obtained from the statistics reported by El Khoury et al. [14] is $\varepsilon = 1.3 \times 10^{-3}$, which indicates that the statistics are completely converged. In any case, the purpose of the present work is to assess the effect of compression with respect to the original data, and the statistics based on the instantaneous flow fields can be considered to be sufficiently converged for this purpose.

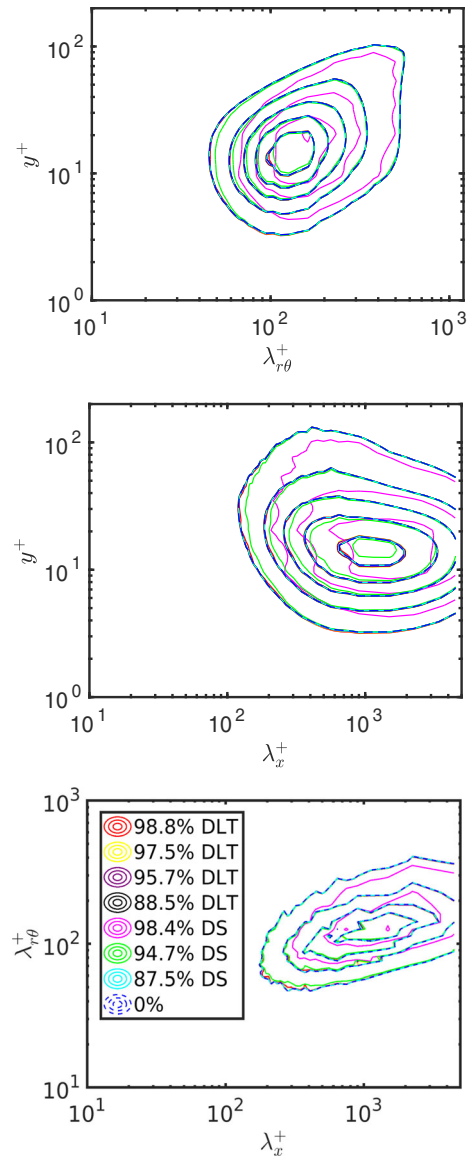
In Fig. 13 (right) it can be observed that ε remains approximately constant and equal to the uncompressed value 2.5×10^{-3} in DLT, and before a slight reduction to $\varepsilon \simeq 2 \times 10^{-3}$ the value of ε increases strongly beyond $Cr = 97.5\%$. In the case of downsampling, a mild reduction with Cr is observed, from $\varepsilon = 2.5 \times 10^{-3}$ to around 2×10^{-3} , before the increase at $Cr = 87.5\%$. Despite the mild reduction in both DLT and downsampling techniques, which might be associated to the fact that the initial uncompressed fields were not completely converged from the statistical point of view, these results indicate that the maximum acceptable Cr values are 97.5% and 87.5% for DLT and downsampling, respectively. Note that these results are in excellent agreement with the ones discussed above in terms of the error indicator e .

5.3 Spectral analysis

Although the turbulence statistics discussed above provide a good indication of the impact of the compression on the flow field, an analysis of the power-spectral density maps in the pipe yields even more detailed information. This is due to the fact that the power-spectral density distributions allow to assess, for each wall-normal position, the scales on which the energy resides, thus providing a more complete picture of the various contributions to the flow field. It is also important to note that, although in the context of the present work we computed turbulence statistics based on instantaneous fields in order to compare various compression levels, in practice those statistics would be computed directly during the simulation as described in [37]. On the other hand, spectral analysis of the flow is typically performed on the instantaneous fields, and therefore it is important to determine the maximum level of possible compression without negatively impacting the spectra, since this would be an analysis method that would greatly benefit from the storage savings associated with compression.

Power-spectral density maps are shown in Fig. 14 for the original data set, as well as for several compression levels with DLT and downsampling. In Fig. 14 (top) we show the premultiplied azimuthal power-spectral density of the streamwise velocity $k_{r\theta} \Phi_{uu}/u_\tau^2$ as a function of the wall-normal direction. The uncompressed data shows the characteristic near-wall spectral peak, located at $y^+ \simeq 15$ and $\lambda_{r\theta}^+ \simeq 120$ (where $\lambda_{r\theta}^+$ is the inner-scaled wavelength in the azimuthal direction), indicative of the spacing between near-wall streaks

Fig. 14 (Top) Inner-scaled premultiplied azimuthal power-spectral density $k_{r\theta} \Phi_{uu}/u_r^2$. Five contours equally spaced from 0.4 to 4 are shown. (Middle) Inner-scaled premultiplied streamwise power-spectral density $k_x \Phi_{uu}/u_r^2$. Five contours equally spaced from 0.2 to 2 are shown. (Bottom) Inner-scaled premultiplied two-dimensional power-spectral density at $y^+ = 15$, $k_{r\theta} k_x \Phi_{uu}/u_r^2$. Four contours equally spaced from 0.2 to 0.8 are shown



reported for instance by Kline et al. [40] or Gupta et al. [41]. Note that the integration of this power-spectral density map over all the azimuthal wavelengths would result in the streamwise velocity fluctuation profile reported in Fig. 10.

By increasing the Cr with DLT it can be noted that the resulting power-spectral density distribution is in excellent agreement with the uncompressed data up to $Cr = 97.5\%$, and only the larger compression ratio of 98.8% reveals small differences in the spectrum. Regarding the downsampling, the maximum compression ratio in good agreement with the uncompressed data is 87.5%, with significant deviations emerging at higher Cr . A very

interesting observation can be made by comparing DLT at $Cr = 98.8\%$ and downsampling at $Cr = 94.7\%$: whereas in the former the contours start to smoothly deviate from the reference data, in the latter a significant impact in all the small scales (below wavelengths of around $\lambda_{r\theta}^+ \simeq 40$) is observed. This is even more noticeable when considering $Cr = 98.8\%$ with downsampling, where all the contours below $\lambda_{r\theta}^+ \simeq 60$ become significantly affected. In this case, the near-wall spectral peak becomes completely distorted as well. This reflects the essence of the DLT and downsampling methods, as also illustrated in Fig. 5: whereas in the DLT method we keep the modes with the largest energy content, in downsampling all the modes beyond a certain threshold are made zero, regardless of their energy contributions. This is the reason why, at progressively larger Cr values, a broader range of the small-scale spectral range becomes significantly distorted by the compression based on downsampling.

The premultiplied streamwise power-spectral density $k_x \Phi_{uu}/u_\tau^2$ shown in Fig. 14 (middle) confirms the observations made for the azimuthal spectrum, namely the maximum acceptable Cr values of 97.5% and 87.5% for DLT and downsampling, respectively. This spectrum also reveals the different effect of DLT and downsampling on the smaller scales, where a progressively smaller distortion on the shortest streamwise wavelengths close to the wall is noticeable for higher Cr when employing downsampling. The near-wall spectral peak is located in this case at $y^+ \simeq 15$ and $\lambda_x^+ \simeq 1,000$, which is in agreement with the pipe-flow DNS results by El Khoury et al. [42]. The premultiplied two-dimensional power-spectral density $k_{r\theta} k_x \Phi_{uu}/u_\tau^2$ is shown in Fig. 14 (bottom) at $y^+ = 15$. This spectral map combines the information from the two previous one-dimensional spectra, exhibiting a clear near-wall spectral peak at $\lambda_{r\theta}^+ \simeq 120$ and $\lambda_x^+ \simeq 1,000$, also in excellent agreement with the values reported by El Khoury et al. [42] for pipe flow up to $Re_\tau = 1,000$. This spectral map also shows the effect of the downsampling-based compression method on the smaller scales, which is consistent with the two previous one-dimensional spectra. These results indicate that DLT significantly outperforms downsampling, since at very high compression ratios up to 97.5% the small-scale energy content is well preserved.

5.4 Restart and resilience

Finally, we analyze the impact of data compression on the simulation restart using the jet in cross-flow case. As already indicated in Section 4, this case has been chosen for its strong sensitivity to the initial conditions, which will allow us to clearly detect any impact coming from the additional noise added by the compression of the restart files.

Figure 15 (top) shows the evolution in time of the streamwise velocity U for different restart points, from fields without and with compression for comparison. As indicated in this figure, we perform a restart at around 5,000 convective time units from the beginning of the simulation, using files with different levels of compression, namely with compression ratios of 90%, 95.6%, and 97.4%. In Fig. 15 (bottom) we compare the solution from a restart with uncompressed files and from compressed files. It can be observed that the impact of restarting from compressed files with 90% compression ratio is negligible. Higher compression ratios such as 97.4%, as shown in Fig. 15 (top), clearly reflect the sensitivity of the case to the initial conditions, producing right from the restart point a first phase of strong instability. A lower compression ratio of 95.6%, as shown in Fig. 15 (bottom), certainly provides a much more stable restart, but still not close enough to the original values requiring a very long transient phase.

This implies that in less sensitive cases, such as turbulent simulations, highly-compressed files could potentially be used for restart purposes, which would lead to significant benefits in terms of checkpointing and resilience in large-scale simulations.

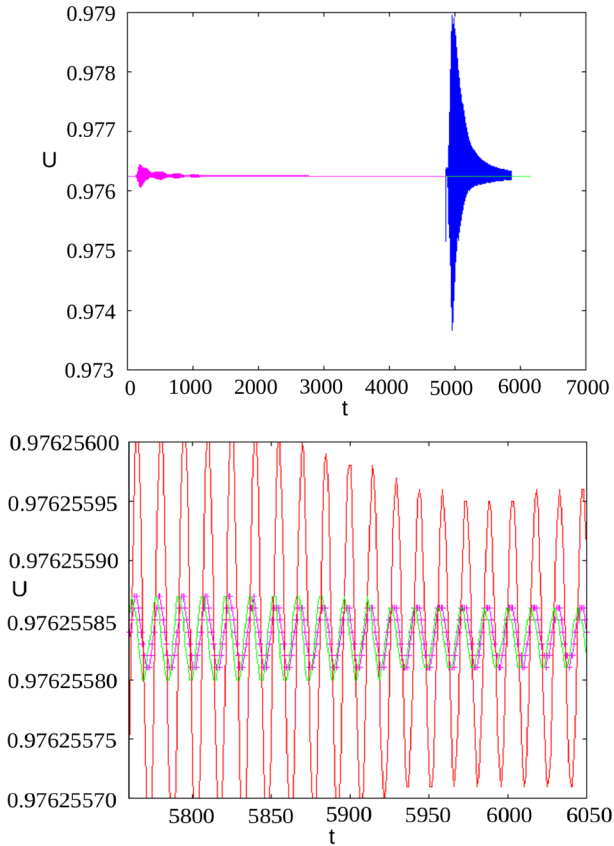


Fig. 15 Restart procedure on the JCF case. (Top) Time evolution of the streamwise velocity U with restart points from fields without (magenta) and with compression ratios: 90% (green) and 97.4% (blue). (Bottom) Zoom applied towards the end of the simulation. Comparison of the restart from a field without compression (magenta) and the restart using compressed files with 90% (green), 95.6% (red) compression ratio

6 Conclusions

In the present work we propose a lossy data compression algorithm, with which we have obtained high levels of acceptable compression ratios for the turbulent flow through a straight pipe and the turbulent boundary layers developing around a wing section. In the case of a turbulent flow through a straight pipe, it was possible to truncate up to about 98% of the data for the computation turbulence statistics, whereas with downsampling the maximum acceptable compression ratios were about 90%. The DLT based truncation method also outperforms DCT, which could compress up to a maximum of 70%. These results were also consistent with the spectral analysis, which in addition also reflected that downsampling significantly affects the smallest scales in the power-spectral density distributions.

For visualization and vortex identification purposes, the results were slightly more sensitive to the compression. It was possible to compress up to around 90% of the original data with DLT, while preserving the topology of the vortex clusters in the turbulent pipe flow case. On the other hand, the maximum compression ratio is case-dependent, and lower Cr

values were obtained in the case of the turbulent flow around a NACA4412 wing section. It is also important to note that whereas the former data set was based on a DNS, the latter was obtained from a well-resolved LES, a fact that is related to the lower attainable compression ratios. Moreover, we have observed that a strong data compression could be applied for a restart procedure in simulations of transition, which are typically highly-sensitive to the initial conditions. Hence, the DLT method highly compresses the data, while preserving the most relevant flow physics with control on the error incurred. The new DLT based method [13], which removes data in an adaptive way depending on the flow properties in the domain, strongly outperformed downsampling as well as the previous DCT based truncation method [19].

Further work could concern a quantitative analysis of the impact of the compression on the coherent structures to, among other aspects, derive a correlation between the amount of structures and the compression ratio. Similarly, the shape and size of identified structures, as a function of compression, could be assessed to give further quantitative bounds on the maximum compression. Relating to more implementational details of the method, the actual reading and writing of the compressed data should be extended with an Huffman encoding [13] for a final user application in the simulation software (in our case Nek5000 [10]). It should be noted that in principle our approach is suitable for any element-wise discretization in which a spectral representation within an element can be formulated. In the case of other basis functions than the Legendre polynomials, further investigation could be necessary to deeply understand the effect of the additional mapping to e.g. a Chebyshev grid on the compression performance, as observed with DCT truncation. Finally an interesting extension of the present work would be to assess, for canonical flow case, the attainable Cr for a certain error threshold as a function of Reynolds number. By knowing the necessary data size in boundary layers for various Re , potentially a model for the necessary information in more general cases could be derived.

Acknowledgments Financial support from the Stiftelsen för Strategisk Forskning (SSF) and the Swedish e-Science Research Centre (SeRC) via the SESSI project is acknowledged. RV and PS acknowledge the funding provided by the Knut and Alice Wallenberg Foundation and the Swedish Research Council (VR). The present work was also partially supported by the ExaFLOW H2020 Project (Project Number 671571). All the data analysis in the present work was performed on resources provided by the Swedish National Infrastructure for Computing (SNIC). Also parts of this work were supported by the U.S. Department of Energy, Office of Science, Advanced Scientific Computing Research under Contract DE-AC02-06CH11357 and the Exascale Computing Project (Contract No. 17-SC-20-SC).

Funding Information Stiftelsen för Strategisk Forskning (SSF), Swedish e-Science Research Centre (SeRC), Knut and Alice Wallenberg Foundation, and Swedish Research Council (VR).

Compliance with Ethical Standards

Conflict of interests The authors declare that they have no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Trott, A., Moorehead, R., McGinley, J.: Wavelets applied to lossless compression and progressive transmission of floating point data in 3-D curvilinear grids Visualization'96 (1996)
2. Sakai, R., Sasaki, D., Nakahashi, K.: Parallel implementation of large-scale CFD data compression toward aeroacoustic analysis. *Comput. Fluids* **80**, 116–127 (2013)
3. Lee, D., Beam, R.M., Warming, R.F.: Supercompact multiwavelets for flow field simulation. *Comput. Fluids* **30**, 783–805 (2001)
4. Blanc, T.J.: Analysis and compression of large CFD data sets using proper orthogonal decomposition, Brigham Young University (2014)
5. Rao, K.R., Yip, P.: Discrete cosine transform: algorithms, advantages, applications. Academic Press Professional, San Diego (1990)
6. Lehmann, H., Werzner, E., Mendes, M.A.A., Trimis, D., Jung, B., Ray, S.: In situ data compression algorithm for detailed numerical simulation of liquid metal filtration through regularly structured porous media. *Adv Eng Mater* **15**, 1260–1269 (2013)
7. Laney, D., Langer, S., Weber, C., Lindstrom, P., Wegener, A.: Assessing the effects of data compression in simulations using physically motivated metrics. *Sci. Program.* **22**, 141–155 (2014)
8. Baker, A.H., et al.: Evaluating lossy data compression on climate simulation data within a large ensemble. *Geosci. Model Dev.* **9**, 4381–4403 (2016)
9. Li, S., Sane, S., Orf, L., Mininni, P., Clyne, J., Childs, H.: Spatiotemporal wavelet compression for visualization of scientific simulation data. In: Proc. 2017 IEEE international conference on cluster computing (CLUSTER), Honolulu (2017)
10. Fischer, P.F., Lottes, J.W., Kerkemeier, S.G.: NEK5000: open Source spectral element CFD solver Available at: <http://nek5000.mcs.anl.gov> (2008)
11. Ahmed, N., Natarajan, T., Rao, K.R.: Discrete cosine transform. *IEEE Trans. Comput.* **C-23**, 90–93 (1974)
12. Schmalzl, J.: Using standard image compression algorithms to store data from computational fluid dynamics. *Comput. Geosci.* **29**, 1021–1031 (2003)
13. Marin, O., Schanen, M., Fischer, P.F.: Large-scale lossy data compression based on an a priori error estimator in a spectral element code ANL/MCS-p6024-0616 (2016)
14. El Khoury, G.K., Schlatter, P., Noorani, A., Fischer, P.F., Brethouwer, G., Johansson, A.V.: Direct numerical simulation of turbulent pipe flow at moderately high Reynolds numbers. *Flow Turbul. Combust.* **91**, 475–495 (2013)
15. Hosseini, S.M., Vinuesa, R., Schlatter, P., Hanifi, A., Henningson, D.S.: Direct numerical simulation of the flow around a wing section at moderate Reynolds number. *Int. J. Heat Fluid Flow* **61**, 117–128 (2016)
16. Peplinski, A., Schlatter, P., Henningson, D.S.: Global stability and optimal perturbation for a jet in cross-flow. *Eur. J. Mech. B/Fluids* **49**, 438–447 (2015)
17. Deville, M.O., Fischer, P.F., Mund, E.H.: High-Order methods for incompressible fluid flow. Cambridge University press, Cambridge (2002)
18. Orszag, S.A.: Spectral methods for problems in complex geometries. *J. Comput. Phys.* **37**, 70–92 (1980). [https://doi.org/10.1016/0021-9991\(80\)90005-4](https://doi.org/10.1016/0021-9991(80)90005-4)
19. Otero, E., Marin, O., Vinuesa, R., Schlatter, P., Siegel, A., Laure, E.: The effect of lossy data compression in computational fluid dynamics applications: resilience and data postprocessing. In: Proc. Direct and Large-Eddy Simulation 11 (DLES-11), Pisa (2017)
20. Vinuesa, R., Schlatter, P.: Skin-friction control of the flow around a wing section through uniform blowing. In: Proc. European drag reduction and flow control meeting (EDRFCM), Rome (2017)
21. Lozano-Durán, A., Jiménez, J.: Effect of the computational domain on direct simulations of turbulent channels up to $re_\tau = 4200$. *Phys. Fluids* **26**, 011702 (2014)
22. Lee, M., Moser, R.D.: Direct numerical simulation of turbulent channel flow up to $re_\tau 5200$. *J. Fluid Mech.* **774**, 395–415 (2015)
23. Schlatter, P., Örlü, R.: Assessment of direct numerical simulation data of turbulent boundary layers. *J. Fluid Mech.* **659**, 116–126 (2010)
24. Sillero, J.A., Jiménez, J., Moser, R.D.: One-point statistics for turbulent wall-bounded flows at Reynolds numbers up to $\delta^+ 2000$. *Phys. Fluids* **25**, 105102 (2013)
25. Vinuesa, R., Duncan, R.D., Nagib, H.M.: Alternative interpretation of the Superpipe data and motivation for CICLoPE: The effect of a decreasing viscous length scale. *Eur. J. Mech. B/Fluids* **58**, 109–116 (2016)
26. Bobke, A., Vinuesa, R., Örlü, R., Schlatter, P.: History effects and near equilibrium in adverse-pressure-gradient turbulent boundary layers. *J. Fluid Mech.* **820**, 667–692 (2017)

27. Vinuesa, R., Hosseini, S.M., Hanifi, A., Henningson, D.S., Schlatter, P.: Pressure-gradient turbulent boundary layers developing around a wing section. *Flow Turbul. Combust.* **99**, 613–641 (2017)
28. Peplinski, A., Vinuesa, R., Offermans, N., Schlatter, P.: ExaFLOW use cases for Nek5000: incompressible jet in cross-flow and flow around a NACA4412 wing section. In: H2020 FETHPC-1-2014 D3.1 (2016)
29. Jeong, J., Hussain, F.: On the identification of a vortex. *J. Fluid Mech.* **285**, 69–94 (1995)
30. del Álamo, J.C., Jiménez, J., Zandonade, P., Moser, R.D.: Self-similar vortex clusters in the turbulent logarithmic region. *J. Fluid Mech.* **561**, 329–358 (2006)
31. Vinuesa, R., Fick, L., Negi, P.S., Marin, O., Merzari, E., Schlatter, P.: Turbulence statistics in a spectral-element code: a toolbox for high-fidelity simulations, Technical report ANL/MCS-TM-367 (2017)
32. Örlü, R., Schlatter, P.: On the fluctuating wall-shear stress in zero pressure-gradient turbulent boundary layer flows. *Phys. Fluids* **23**, 021704 (2011)
33. Dony, R.D.: Karhunen-Loève transform, the transform and data compression handbook. CRC Press LLC, Boca Raton (2001)
34. Unser, M.: On the approximation of the discrete karhunen-loève transform for stationary processes. *Signal Process.* **7**, 231–249 (1984)
35. Lazaridis, P., Bizopoulos, A., Tzekis, P., Zaharis, Z., Debarge, G., Gallion, P.: Comparative study of DCT and discrete legendre transform for image compression. In: Proc. X international conference society for electronics, telecommunications, automatics and informatics (ETAI), Ohrid (2011)
36. Pope, S.: Turbulent flows. Cambridge University Press, Cambridge (2000)
37. Vinuesa, R., Prus, C., Schlatter, P., Nagib, H.M.: Convergence of numerical simulations of turbulent wall-bounded flows and mean cross-flow structure of rectangular ducts. *Meccanica* **51**, 3025–3042 (2016)
38. Kim, J., Moin, P., Moser, R.D.: Turbulence statistics in fully developed channel flow at low Reynolds number. *J. Fluid Mech.* **177**, 133–166 (1987)
39. Moser, R.D., Kim, J., Mansour, N.N.: Direct numerical simulation of turbulent channel flow up to $Re_\tau = 590$. *Phys. Fluids* **11**, 943–945 (1999)
40. Kline, S.J., Reynolds, W.C., Schraub, F.A., Runstadler, P.W.: The structure of turbulent boundary layers. *J. Fluid Mech.* **30**, 741–773 (1967)
41. Gupta, A.K., Laufer, J., Kaplan, R.E.: Spatial structure in the viscous sublayer. *J. Fluid Mech.* **50**, 493–512 (1971)
42. El Khoury, G.K., Schlatter, P., Brethouwer, G., Johansson, A.V.: Turbulent pipe flow: statistics, Re -dependence, structures and similarities with channel and boundary layer flows. *J. Phys.: Conf. Ser.* **506**, 012010 (2014)