Mixed Averaging Procedures

Mixed averaging procedures*

Michele Errante[1*], Markus Klein[2], Andrea Ferrero[1], Francesco Larocca[1], Guglielmo Scovazzi[3], Massimo Germano[3]

[1*]Department of Mechanical and Aerospace Engineering, Politecnico di Torino, Corso Duca degli Abruzzi 24, Torino, 10129, Italy.
[2]Department of Aerospace Engineering, University of the Bundeswehr Munich, Werner-Heisenberg-Weg 39, Neubiberg, 85577, Germany.
[3]Department of Civil and Environmental Engineering, Duke University, 121 Hudson Hall, Durham, 27708, NC, USA.

*Corresponding author(s). E-mail(s): michele.errante@polito.it; Contributing authors: markus.klein@unibw.de; andrea_ferrero@polito.it; francesco.larocca@polito.it; guglielmo.scovazzi@duke.edu; mg234@duke.edu;

Abstract

The *statistical* operators typically applied in postprocessing numerical databases for statistically steady turbulence are a mixture of physical averages in homogeneous spatial directions and in time. Alternative averaging operators may involve phase or ensemble averages over different simulations of the same flow. In this paper, we propose straightforward metrics to assess the relative importance of these averages, employing a mixed averaging analysis of the variance. We apply our novel indicators to two statistically steady turbulent flows that are homogeneous in the spanwise direction. In addition, this study highlights the local effectiveness of the averaging operator, which can vary significantly depending on the mean flow velocity and turbulent length scales. The work can be utilized to identify the most effective averaging procedure in flow configurations featuring at least two homogeneous directions. Thus, this will contribute to achieving better statistics for turbulent flow predictions or reducing computing time.

Keywords: Numerical simulation, Expected values, Statistical homogeneity, Mixed averaging procedures

---

*This version of the article has been accepted for publication, after peer review but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: https://doi.org/10.1007/s10494-023-00527-x.

Ethical statement:
- Funding: N/A
- Conflict of Interest: N/A
- Ethical approval: N/A
- Informed consent: N/A
- Author contribution: All authors contributed equally to this work.
- Data Availability Statement: Data sets generated during the current study are available from the corresponding author on reasonable request.

1 Introduction

The theory of turbulence primarily relies on the concept of ensemble averaging, and the statistical description of turbulent flows largely depends on that. Ideally, conducting the same experiment or the same simulation under identical conditions is the preferred theoretical scenario, but it is very difficult to realize practically. Time, space, or phase averages are the usual surrogates, acting as substitutes for ensemble averages when possible. Obviously, in the general case, for statistically unsteady turbulence, involving complex geometry, without homogeneous directions, ensemble averaging remains the sole option.

A simulation of turbulent flows based solely on ensemble averages is still far from realization, due to the unaffordable computational cost that would be required. Currently, mixed averaging procedures serve as substitutes for the ensemble average in practical simulations of steady turbulent flows with space symmetries. Shirian et al. [1] assess the quality of the statistics evaluated using averaging in statistically stationary chaotic phenomena, particularly in low Reynolds number, incompressible homogeneous and isotropic turbulence. Furthermore, various algorithms have been proposed to estimate the variance of the computed average field, in order to assess the statistical convergence of results. Noteworthy among these methodologies are the Non-Overlapping Batch Means (NOBM) method [2], the Overlapping Batch Means (OBM) method [3], and the Batch Means Batch Correlations (BMBC) method [4]. These approaches, while distinct, share a fundamental conceptual similarity, requiring the delineation of subsamples (batches) derived from the computed time history. However, the aforementioned approaches do not address the relative importance of different averaging procedures, which is the focus of the present work.

Mixed ensemble time or space or phase averages are currently applied in complicated geometries and complex problems, such as diesel sprays [5], where phase averages or ensemble averages of the same turbulent flows under different initial conditions are produced. Specifically, in that work, mixed ensemble and spatial averages, spatial and time averages, ensemble, spatial and time averages are compared in order to estimate their efficiency and reliability.

Testing the significance of the different averaging operators in such cases is not so straightforward. For instance, let's consider a category of turbulent flows crucial in engineering applications - those that exhibit homogeneity in both time t and in the spanwise direction y. In this case, the flow theoretically has infinite extensions, but in practice, it is limited by a finite extent of time T and a finite extent of space, denoted here as Y without loss of generality.

More specifically, without relying on probabilistic assumptions, we define, in this particular case, for the components of the velocity field $u_i(t,x,y,z)$, the quantities

$$\begin{aligned}
\langle u_i\rangle_t(x,y,z) &= \frac{1}{T}\int_0^T u_i(x,y,z,t')dt' \\
\langle u_i\rangle_y(x,z,t) &= \frac{1}{Y}\int_0^Y u_i(x,y',z,t)dy' \qquad (1)\\
\langle u_i\rangle_{ty}(x,z) &= \frac{1}{TY}\int_0^T\int_0^Y u_i(x,y',z,t')dt'dy' = \langle u_i\rangle_{yt}(x,z) \qquad (2)
\end{aligned}$$

and we introduce the associated Generalized Central Moments [6] given by

$$\begin{aligned}
\tau_t(u_i,u_j) &\equiv \langle u_i u_j\rangle_t - \langle u_i\rangle_t\langle u_j\rangle_t \\
\tau_y(u_i,u_j) &\equiv \langle u_i u_j\rangle_y - \langle u_i\rangle_y\langle u_j\rangle_y \\
\tau_t(\langle u_i\rangle_y,\langle u_j\rangle_y) &\equiv \langle\langle u_i\rangle_y\langle u_j\rangle_y\rangle_t - \langle\langle u_i\rangle_y\rangle_t\langle\langle u_j\rangle_y\rangle_t \\
\tau_y(\langle u_i\rangle_t,\langle u_j\rangle_t) &\equiv \langle\langle u_i\rangle_t\langle u_j\rangle_t\rangle_y - \langle\langle u_i\rangle_t\rangle_y\langle\langle u_j\rangle_t\rangle_y \\
\tau_{ty}(u_i,u_j) &\equiv \langle\langle u_i u_j\rangle_t\rangle_y - \langle\langle u_i\rangle_y\langle u_j\rangle_t\rangle_y = \tau_{yt}(u_i,u_j) \qquad (3)
\end{aligned}$$

It is easy to see that in terms of these quantities, we can express the following identities:

$$\begin{aligned}
\tau_{ty}(u_i,u_j) &= \tau_t(\langle u_i\rangle_y,\langle u_j\rangle_y) + \langle\tau_y(u_i,u_j)\rangle_t \\
&= \tau_y(\langle u_i\rangle_t,\langle u_j\rangle_t) + \langle\tau_t(u_i,u_j)\rangle_y = \tau_{yt}(u_i,u_j); \qquad (4)
\end{aligned}$$

the aforementioned relations will be instrumental to the definition of the new indices in Eq. (5).

The meaning of these decompositions of the total turbulent stress $\tau_{ty}(u_i,u_j)$ related to the mixed average in space and time can be found in the framework of the statistical Law of the Total Variance [7]. Specifically, $\langle\tau_y(u_i,u_j)\rangle_t$ and $\tau_t(\langle u_i\rangle_y,\langle u_j\rangle_y)$ represent the fractions of the Reynolds stress due to the time averaging *within* and *between* the space average, and $\langle\tau_t(u_i,u_j)\rangle_y$ and $\tau_y(\langle u_i\rangle_t,\langle u_j\rangle_t)$ represent the fractions of the Reynolds stress due to the space averaging *within* and *between* the time average. Utilizing these insights, we can define two measures of statistical homogeneity: the first is related to time average, and the second to spanwise average, expressed as the ratios of the traces:

$$M_t(x,y) = \langle\tau_t(u_i,u_i)\rangle_y/R_{ii} \quad ; \quad M_y(x,y) = \langle\tau_y(u_i,u_i)\rangle_t/R_{ii} \qquad (5)$$

where $R_{ij} \equiv \tau_e(u_i,u_j)$ are the Reynolds stresses.

The introduced indices provide a computationally straightforward and physically intuitive method for assessing the efficiency of mixed averaging procedures among the different averaging components. These homogeneity indices contribute to improved statistics in turbulent flow predictions and have the potential to mitigate computational costs. Employing the newly introduced indices, in the following, we analyze a database comprising two statistically steady turbulent flows homogeneous in the spanwise direction: the turbulent plane jet flow and the turbulent flow past a cascade of blades, used as benchmark cases in the application of the homogeneity indices.

2 The turbulent plane jet flow

The turbulent plane free jet represents an important statistically steady turbulent flow benchmark provided with a spatial symmetry. The details regarding the computational domain, the numerical technique, and boundary conditions for the plane jet configuration can be found in [8].

As previously mentioned, the primary focus of this exploration is the analysis of the homogeneity indices. Figure 1 displays the trace of the Reynolds stresses obtained from space-time averaging along with the homogeneity indices M_t and M_y. In this specific simulation, the M_t index exceeds M_y, suggesting that statistics are mainly captured by sampling in time. Nevertheless, it is evident that sampling in y direction plays still an important role in some regions of the computational domain.

Additionally, another focal point of this study is examining the relationship between the defined indices and the number of independent samples N^{SIS}, where the superscript SIS stands for statistically independent samples. This quantity is simply assumed as $N^{SIS} = N_t^{SIS} N_y^{SIS}$, with N_t^{SIS} and N_y^{SIS} representing the independent samples in time and in the spanwise direction, respectively, estimated as follows

$$N_t^{SIS} = \min\{N_t, T/\tau_{conv}\} \quad ; \quad N_y^{SIS} = \min\{N_y, \mathcal{Y}/\mathcal{L}_{\in\in}\} \, ; \quad \tau_{conv} = L_{11}/\langle U\rangle \tag{6}$$

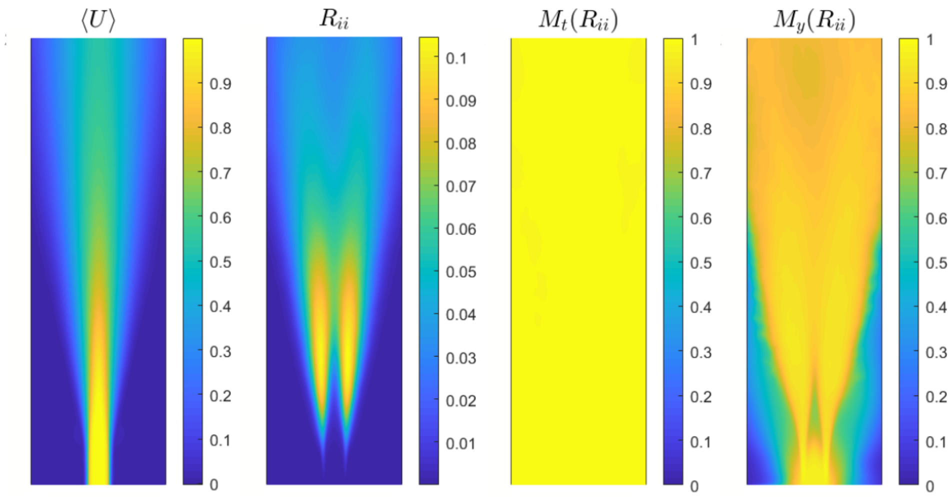

Fig. 1 Left to right: Mean axial velocity, trace of Reynolds stresses R_{ii}, homogeneity index M_t and M_y. Only a section of the computational domain is shown in y direction ($[-3, 3] \times [0, 20]$).

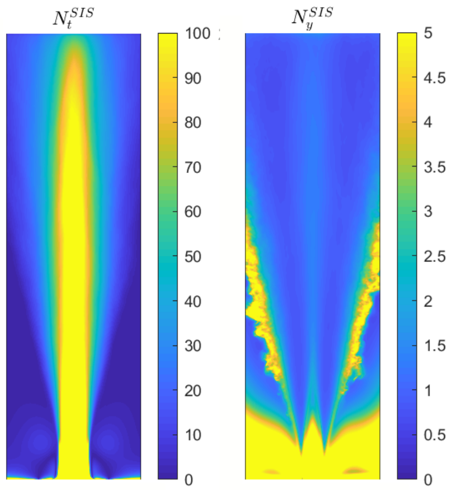

Fig. 2 Left to right: Estimated number of independent samples in time t and y direction. Only a section of the computational domain from the turbulent plane jet is shown in lateral y direction ($[-3, 3] \times [0, 20]$).

where T is the total simulation time, $\mathcal{Y}$ is the size of the computational domain in the spanwise (y) direction, and N_t and N_y are the number of samples in time and in the spanwise direction, respectively. L_{11} and L_{22} represent directional length scales in the axial and spanwise directions: $L_{11} = 0.43R_{11}^{3/2}/\varepsilon_{11}$, $L_{22} = 0.43R_{22}^{3/2}/\varepsilon_{22}$, using as a first approximation the relation $L_{11} \approx 0.43k^{3/2}/\varepsilon$ given in [9].

In line with the observation that N_t^{SIS} greatly exceeds N_y^{SIS}, the index M_t surpasses M_y, emphasizing that time sampling dominates statistics (see Figure 2). While Eq. 6 offers a quick estimation of independent samples (assuming isotropic flow in equilibrium), accurately determining decorrelation time or distance can be complex due to challenges such as undershoots, oscillations, and poor autocorrelation convergence at large separations. The indices proposed in Eq. 5 provide a more straightforward evaluation.

3 Gas turbine cascade

The second turbulent flow benchmark is the transitional and separated flow field around the T106C subsonic low pressure turbine (LPT) cascade.

The simulation is carried out with the isentropic exit Mach number $M_{2,is}$ and the isentropic exit Reynolds number $Re_{2,is}$ set to $M_{2,is} = 0.65$ and $Re_{2,is} = 80,000$, respectively. The T106C cascade geometry is characterized by a pitch t to chord c ratio of $t/c = 0.95$ and an inlet flow angle β_1 of 32.7° with respect to the axial direction.

The three-dimensional, compressible Navier-Stokes-equations are solved in non-dimensional form using the Discontinuous Galerkin finite elements framework, without any subgrid-scale model. Hence, the following computations can be classified as Implicit Large Eddy Simulation (ILES). For spatial discretization, a second order

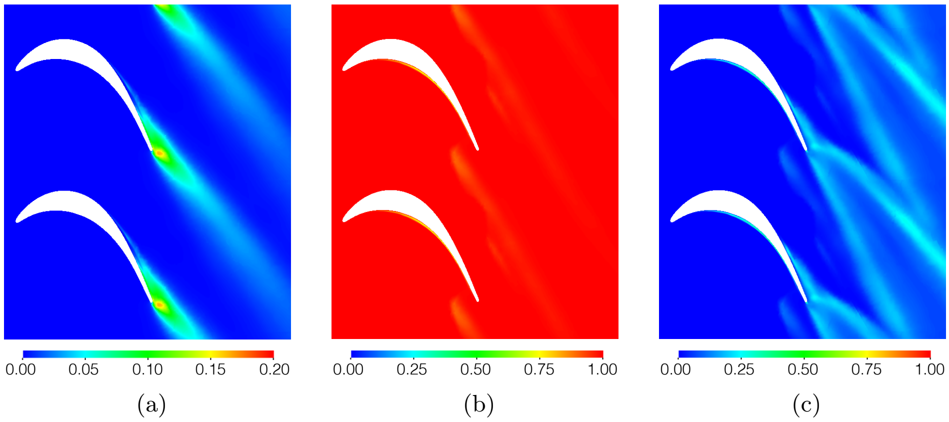

Fig. 3 R_{ii} (a), homogeneity index M_t (b) and M_y (c) for the T106C cascade.

accurate DG scheme is used, while time integration is performed with an explicit second order Runge-Kutta-method, considering a CFL number of 0.45. Further details about the flow, the computational domain, and the numerical framework used for the simulations are provided in [10] and omitted here for brevity.

The Reynolds stresses R_{ii}, obtained through space-time averaging, along with the homogeneity indices M_t and M_y, are presented in Figure 3. Since the indices M_t and M_y are defined assuming the presence of turbulence, it is important to emphasize that these plots were generated by setting a threshold on the local turbulence intensity I. This implies that M_t and M_y were computed for points in space where the local turbulent intensity exceeded the value of $I = 0.015$. It is apparent that the index M_t is larger than M_y, consistently with the observations in the preceding section, reinforcing that statistics are primarily captured by sampling in time.

4 Discussion

The analysis suggests that time averaging is generally more efficient than spatial averaging in the considered testcases. Nevertheless, averaging in time is not always desirable because it can require very long wall clock times (WCT). To reduce WCT, options such as ensemble averaging or extending the domain in a spatially homogeneous direction can be explored.

In our study, we assume a flow with temporal and spatial homogeneity in the y-direction (with domain dimensions $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$) and a linear scaling of floating-point operations with mesh size $N_x \cdot N_y \cdot N_z$. The total computational time (TCT) depends on the time to reach a statistically steady state, measured in flow through units for initialization (FTI) and the number of flow through times for collecting statistics (FTS).

Estimating FTS as $N^{SIS} L^2/(\mathcal{X}\mathcal{Y})$, where L represents a characteristic turbulent length scale and N^{SIS} denotes the number of required samples to reach a desired confidence for the statistical error, noting that one flow through time takes (for a constant grid spacing) N_x/CFL time steps, the total computational time TCT scales as:

$$TCT \propto \frac{N_x^2 \cdot N_y \cdot N_z}{CFL} \left(FTI + \frac{N^{SIS} L^2}{\mathcal{X}\mathcal{Y}} \right). \tag{7}$$

Increasing the domain length ($\mathcal{Y}$) leads to a proportional increase in the number of grid points (N_y) while keeping the second contribution in the equation constant ($N_y \sim \mathcal{Y}$). Similarly, employing an ensemble averaging strategy with M simultaneous simulations from varied initial fields using M times more computational cores shows that the second contribution in Eq. 7 remains constant. However, the time for reaching a steady state has to be achieved for each flow realization, contributing to a total increase. Consequently, averaging in time is the most efficient approach in terms of saving computational resources, energy and to reduce associated carbon dioxide emissions, which are non-negligible for HPC computing.

WCT is proportional to the required computational demand divided by the number of cores. Hence, increasing $\mathcal{Y}$ and proportionally the number of cores reduces the wall clock time for collecting statistics, keeping the time for initialization constant. The similar argument and asymptotic limit holds for ensemble averaging of M flow realizations with a factor of M more computational cores, i.e., the minimum WCT in both cases scales with $N_x^2 \cdot N_y \cdot N_z \cdot FTI/(CFL \cdot N_{cores})$. This discussion highlights the importance of reducing the time to reach a statistically steady state, for which methods like multi-grid strategies or synthetic initial data are relevant.

5 Conclusions

We propose a basic mixed averaging analysis of Reynolds stress in statistically steady turbulent flows with spanwise homogeneity, using straightforward indices to assess the importance of spanwise and time averages. While time averaging is more effective, the inability to parallelize in time and long run times presents challenges. Combining ensemble and time averaging with varied initial data may offer a solution, though further research is needed to relate these indices to conventional statistical convergence methods.

References

[1] Shirian, Y., Horwitz, J.A.K., Mani, A.: On the convergence of statistics in simulations of stationary incompressible turbulent flows. Computers & Fluids **266**, 106046 (2023) https://doi.org/10.1016/j.compfluid.2023.106046

[2] Conway, R.W., Johnson, B.M., Maxwell, W.L.: Some problems of digital systems simulation. Management Science **6**(1), 92–110 (1959) https://doi.org/10.1287/mnsc.6.1.92

[3] Meketon, M.S., Schmeiser, B.: Overlapping batch means: Something for nothing? Proceedings of the 16th conference on Winter simulation (IEEE Press), 226–230 (1984)

[4] Russo, S., Luchini, P.: A fast algorithm for the estimation of statistical error in dns (or experimental) time averages. Journal of Computational Physics **347**, 328–340 (2017) https://doi.org/10.1016/j.jcp.2017.07.005

[5] Farrace, D., Panier, R., Schmitt, M., Boulouchos, K., Wright, Y.M.: Analysis of averaging methods for large eddy simulations of diesel sprays. SAE Int. J. Fuels Lubr. **8**(3), 568–580 (2015) https://doi.org/10.4271/2015-24-2464

[6] Germano, M.: Turbulence: the filtering approach. Journal of Fluid Mechanics **238**, 325–336 (1992) https://doi.org/10.1017/S0022112092001733

[7] Blitzstein, J.K., Hwang, J.: Introduction to probability. CRC Press LLC (2014)

[8] Klein, M., Sadiki, A., Janicka, J.: Investigation of the influence of the Reynolds number on a plane jet using direct numerical simulation. International Journal of Heat and Fluid Flow **24**(6), 785–794 (2003) https://doi.org/10.1016/S0142-727X(03)00089-4

[9] Pope, S.B.: Turbulent Flows. Cambridge University Press, Cambridge (2000)

[10] Errante, M., Ferrero, A., Larocca, F.: High-order DG method for an implicit LES of a gas turbine cascade. AIP Conference Proceedings (2023)