19th Brazilian Congress of Thermal Sciences and Engineering
November 06th-10th, 2022, Bento Gonçalves - RS - Brazil

ENC-2022- 0637
A DATA-DRIVEN APPROACH FOR ATMOSPHERIC BOUNDARY LAYER FLOWS: LARGE EDDY SIMULATION AND DIMENSIONALITY REDUCTION

Pedro Roberto Barbosa Rocha
Marcos Sebastião de Paula Gomes
Department of Mechanical Engineering, Catholic University of Rio de Janeiro, PUC-Rio, 22451-900, Rio de Janeiro, RJ, Brazil
prbarbosr@gmail.com, mspgomes@puc-rio.br

Abstract. *Spatiotemporal analyses of atmospheric systems are extremely important for a variety of environmental studies, especially for those related to the wind energy sector, where there is a strong interest in finding optimal parameters for the operation of wind farms. For this optimization problem, many heavy numerical simulations are needed, which becomes unfeasible to be achieved in most situations since extremely high computer processing power and storage capacity are required. The present work tackled this challenge by applying dimensionality reduction tools to the outputs of a large eddy simulation (LES) of an atmospheric boundary layer (ABL) flow. This high-fidelity simulation was performed using ANSYS Fluent. To mimic the three-dimensional (3D) transient air flow over a mountainous terrain, a Gaussian-shaped bump was placed in the left-hand side of a rectangular domain and a velocity profile was prescribed at the inlet plane. The incremental principal components analysis (IPCA) algorithm built in Python was employed to reduce the dimensionality of the system and an excellent agreement was achieved between original and reconstructed fields. This efficient data reduction allows building reduced-order models (ROMs), which are much faster than their full-order counterparts and still portray the main features of the flow.*

Keywords: *large eddy simulation, dimensionality reduction, atmospheric boundary layer, reduced-order model*

1. INTRODUCTION

The comprehension of atmospheric processes is fundamental to many engineering applications, such as the design of structures exposed to wind loads, quite common in the wind energy sector. Also, Earth's air circulation is a major driving force for pollutant dispersion, a field in which solving an inverse problem is particularly desired so that the sources of these emissions may be located. In general, being able to predict accurately and efficiently weather and climate phenomena is quite challenging, since they exhibit nonlinear dynamical features with multiple scales in space and time, making their numerical simulations computationally expensive and usually requiring an unfeasible storage capacity. Even reproducing the dynamics of Earth's atmosphere with coarse mesh elements would imply in tens of millions of grid points and then in a massive amount of data to be stored for each simulation time instant. In view of this fact, reduced-order models (ROMs) built through scientific machine learning techniques arrive as a promising and much faster alternative to high-fidelity simulations, such as large eddy simulations (LESs), when dealing with multiple evaluations of the flow field for different sets of initial conditions and physical parameters.

Since its proposal in the early 1960s, LES has been widely adopted in the simulation of complex turbulent flows. Zhiyin (2015) made an in-depth review of this tool, showing its power in a plenty of applications, such as in realistic full annular gas turbine combustors. The author defends this method will be the mainstream of fluid engineering analyses in an average term, highlighting the challenges it must overcome in order to be as reliable and robust as other models. On the other hand, in the field of atmospheric sciences, it has already become quite popular. For instance, Porté-Agel et al. (2011) developed an LES framework to predict the interactions between the atmospheric boundary layer (ABL) flow and wind turbines and their effects. By parameterizing accordingly sub-grid scale turbulent fluxes and turbine-induced forces, they could obtain results in good agreement with experimental measurements. More recently, Han et al. (2016) performed large eddy simulations of ABL flows over complex terrains and obtained mean wind speeds and directions reasonably close to those measured by eight meteorological towers in a real wind farm site.

Detailed three-dimensional (3D) numerical experiments, fundamental for an accurate description of turbulent flow structures, usually demand a very high computing power. In other words, computational fluid dynamics (CFD) frameworks require a great amount of grid points and timesteps in which the flow variables, such as pressure and

velocities, are computed. Indeed, it is quite cumbersome to deal with these high-dimensional data in order to build surrogate models for a dynamical system. Then, it is indicated a data reduction so that the flow variables are represented in a lower-dimensional space. This reduction may be performed through a proper orthogonal decomposition (POD) of the dataset, or through its incremental version (iPOD). Huang et al. (2019) and McQuarrie et al. (2021) successfully applied POD on their combustion datasets before constructing a ROM. On the other hand, Robertson et al. (2018) and Matsumoto et al. (2019) employed iPOD techniques on large-scale flow systems. These authors followed a different approach: while the first computed the POD modes simultaneously to the Computational Fluid Dynamics (CFD) data, avoiding saving all the snapshots of the flow field to disk, the second preferred doing these computations in separate stages.

For the present work, a three-dimensional (3D) atmospheric boundary layer (ABL) flow simulation was performed within Ansys Fluent software. To treat turbulence and reproduce reliably the transient flow structures, the LES method was employed. Then, an iPOD of the dataset was performed, reducing its dimensionality. The Gaussian-shaped bump geometry over which air flows unsteadily, mimicking a mountainous terrain, is shown in Fig. 1. The bump is 0.4 m in height and its base is 1.04 m in diameter. The size of the domain is 8 m x 3 m x 4 m, quite similar to the one adopted by Lignarolo et al. (2011) for investigating the impact of different wall boundary conditions on the accuracy of the LES approach. Here, our main goal was to evaluate the capability of the iPOD technique in adequately representing the 3D ABL flow system in a lower-dimensional space, which is fundamental in order to build a consistent data-driven reduced-order model.

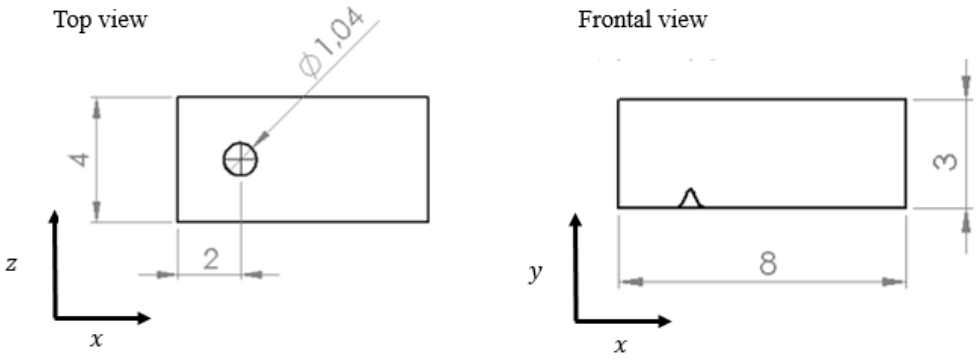

Figure 1. Three-dimensional flow domain

2. MATHEMATICAL AND NUMERICAL MODELING

2.1 Mathematical modeling

Large eddy simulations are largely employed for modeling ABL flows, which exhibit different spatial and temporal scales. In essence, LES neglects the smallest length scales in the numerical solution for the flow field due to the low-pass filtering of the Navier-Stokes equations. For an incompressible flow of a Newtonian fluid, the filtered mass and momentum conservation equations are, respectively:

$$\frac{\partial \overline{u_i}}{\partial x_i} = 0, \tag{1}$$

$$\frac{\partial \overline{u_i}}{\partial t} + \overline{u_j}\frac{\partial \overline{u_i}}{\partial x_j} = -\frac{1}{\rho}\frac{\partial \bar{p}}{\partial x_i} + \nu\frac{\partial^2 \overline{u_i}}{\partial x_j \partial x_j} - \frac{\partial \tau_{ij}}{\partial x_j}, \tag{2}$$

where $\overline{u_i}$ is the filtered velocity, $\bar{p}$ is the filtered pressure, ρ is the density and ν is the kinematic viscosity. For convenience, the upper horizontal bar that indicates the filtering operation will be omitted from now on. Also, the loss of information that comes from the this operation may be compensated by sub-grid models, which are included in the formulation through the sub-grid scale stress tensor τ_{ij}. To compute this tensor, the WALE model was employed.

2.2 Numerical modeling

The high-fidelity simulation of the flow was carried out in ANSYS Fluent software. The entire domain depicted in Fig. 1 was divided into 132,097 hexahedral mesh elements, with a refinement in the surroundings of the bump. A semi-empirical velocity profile u_p was prescribed at the inlet plane and it is described by

$$\bar{u}_p = \frac{u^*}{k}\ln\left(\frac{y}{y_0}\right), \tag{3}$$

where $u^* = 1\text{m/s}$ is the friction velocity, $k = 0.4$ is the von Karman constant and $y_0 = 10^{-4}\text{m}$ is a reference height.

At the outlet plane, the gauge pressure was set to zero, while at the lower surface, including the bump, a no-slip boundary condition was adopted. A zero flux of all quantities across other boundaries was assumed, meaning they have a null normal velocity component. The flow was initialized with only the x-velocity being higher than zero ($\sim 23\text{m/s}$). The filtered momentum equation was solved via the bounded central differencing scheme for spatial discretization and the transient formulation was based on a second order accurate implicit scheme. The pressure-velocity coupling was based on the coupled scheme.

The dataset is comprised by a total of 10,000 snapshots (or timesteps), with a maximum of 50 iterations being done for each one. The interval between two consecutive time instants is 10^{-4}s. The convergence was reached when the residuals for the velocity components and the continuity equation admitted values lower than 10^{-6}. Finally, the dataset generated by the high-fidelity simulation was projected onto a lower dimensional iPOD basis, a subject addressed in the next section.

3. DIMENSIONALITY REDUCTION

To reduce the dimensionality of the 3D ABL flow data, the incremental principal component analysis (IPCA) from Scikit-learn (Ross et al., 2008; Pedregosa et al., 2011) was the chosen iPOD technique. It is based on the singular value decomposition (SVD) of the data, whose dominant singular vectors comprise the basis onto which the high-dimensional data is projected. Let $D^{k\times n}$ be a dataset matrix representing the spatiotemporal evolution of a discretized physical system, where k indicates the timestep and n refers to the grid point. By applying SVD, this matrix is written as

$$D = U\Sigma V^T, \tag{4}$$

where $U^{k\times k}$ is a unitary matrix where the temporal information of the system is embedded, $\Sigma^{k\times n}$ is a diagonal matrix containing the singular values of D, σ_i, and $V^{n\times n}$ portrays the right singular vectors of D. The energetic content of each singular vector is measured by its associated singular value. That implies the r dominant singular vectors have $\sum_{i=1}^{r}\sigma_i^2 \,/\, \sum_{i=1}^{n}\sigma_i^2$ of the system`s energy. Then, to reduce the dimensionality of D, this matrix is multiplied by the r first columns of V, V_r, the POD basis. To reconstruct the original field, the reduced one must be multiplied by V_r^T.

Original flow variables (pressure and x, y and z velocities) were normalized from -1 to 1 and centered at mean values, composing the dataset matrix D. To apply IPCA, the dataset was divided in 30 batches, which means the SVD was carried out incrementally 30 times. Mean, maximum and minimum values for the variables were computed in 100 batches so that global ones could be found and applied in the normalization procedure. It was not feasible to employ PCA instead of its incremental version due to computational limitations. The Python notebook responsible for executing the incremental proper orthogonal decomposition is publicly available at *https://github.com/prbarbosr/3D-Atmospheric-boundary-layer*.

Let $s_{i,j}$ and $s_{i,j}^{rec}$ be the original and reconstructed state variables of interest, where $s \in \{p, u_x, u_y, u_z\}$, at the i-th timestep and j-th grid point. Then, the normalized root-mean-square reconstruction error for variable s, $\varepsilon_{rec,s}$, is given by Eq. (5). The global reconstruction error, ε_{rec}, was obtained through a simple arithmetic mean among the values for $\varepsilon_{rec,s}$. Here, n_x and n_t refer to the number of grid points and training (or testing) timesteps, respectively.

$$\varepsilon_{rec,s} = \frac{1}{\max(s_{i,j}) - \min(s_{i,j})}\sqrt{\frac{\sum_{j=1}^{n_x}\sum_{i=1}^{n_t}\left(s_{i,j}^{rec} - s_{i,j}\right)^2}{n_x n_t}} \tag{5}$$

4. RESULTS AND DISCUSSION

Initially, the LES approach may be evaluated qualitatively through Fig. 2, which shows it was able to capture multiscale flow features (e.g., the recirculation just after the bump). Domain size was selected with care to avoid top and lateral boundary conditions from having a high influence on the results. The initial transient data composed by 4,000 timesteps were discarded. Figure 2(a) presents the z-velocity at the plane $z = 2\text{m}$ for the first snapshot. Note that the flow becomes much more turbulent after the obstacle and that, near its top, the velocity increases to ensure mass conservation. In Fig. 2(b), which shows streamlines near the bump at the same plane and time instant, it is observed a transient flow recirculation generated by a positive pressure gradient in that region.

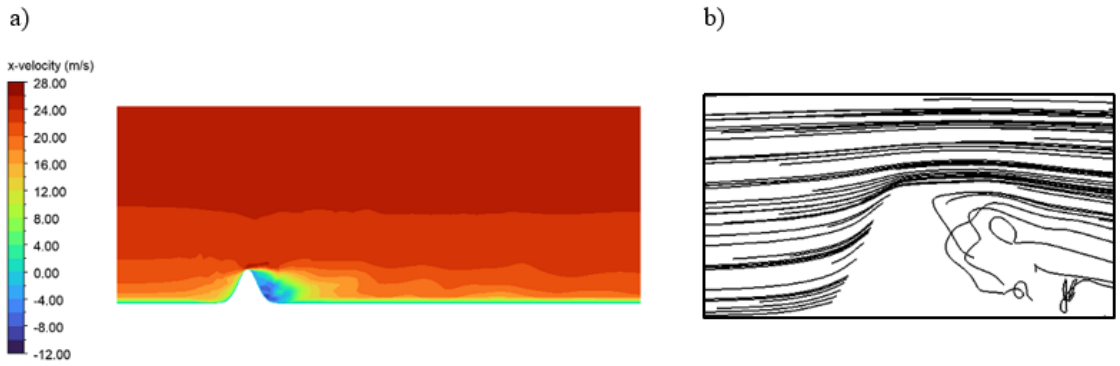

Figure 2. Z-velocity (a) profile and (b) streamlines at the plane $z = 2$m for the first snapshot. The right-hand side figure represents the region near the bump.

To build the iPOD basis, 5,700 training snapshots were employed. The energetic content of the r dominant singular vectors (or modes) that constitute this basis was assessed through the ratio $\sum_{i=1}^{r} \sigma_i^2 \,/\, \sum_{i=1}^{n} \sigma_i^2$. As this ratio approaches 1, more features of the dynamical system are captured by the iPOD basis. Figure 3(a) shows its dependence on the selected value for r. A single mode was capable of representing almost 70% of the system`s energy, while 17 modes captured more than 90%. The global reconstruction error, calculated through Eq. (5), is depicted in Fig. 3(b) for different numbers of modes. As expected, the error decreases when r increases.

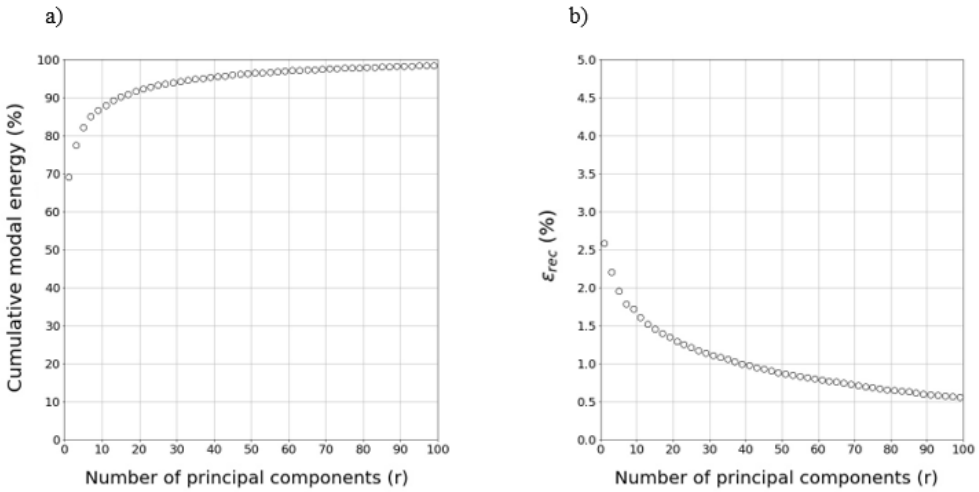

Figure 3. Influence of the number of principal components on the (a) cumulative modal energy and (b) global reconstruction error

The normalized root-mean-square errors between original and reconstructed fields for training and testing periods were computed through Eq. (5) for each state variable (p, u_x, u_y and u_z) and are exhibited in Tab. 1. Pressure and z-velocity are the easiest and the most difficult variables to be reconstructed in general. Although the errors for the testing interval are higher than those for the training one, they are quite acceptable due to the complexity of the dynamical system.

Table 1. Reconstruction error for each state variable of interest (all the values are expressed in % and $r = 61$)

Interval	$\varepsilon_{rec,p}$	ε_{rec,u_x}	ε_{rec,u_y}	ε_{rec,u_z}	ε_{rec}
Training	0.65	0.81	0.84	0.83	0.78
Testing	1.12	1.76	1.77	1.85	1.62

Original and reconstructed temperature and velocities fields along the training period for two probe locations are presented in Fig. 4. Note the iPOD basis onto which the LES data was projected is comprised by the $r = 61$ dominant singular vectors. Locations 1 and 2 were selected randomly. Similarly, Fig. 5 presents original and reconstructed fields at the same locations, but it refers to the testing snapshots (i.e., to the unseen data). The small oscillations in the reconstructed

time series at location 2 are justified by the data compression. By increasing the number of modes, these oscillations tend to disappear. A good agreement was achieved for both training and testing regions, indicating the iPOD basis was capable of representing faithfully seen and unseen 3D ABL flow data in a lower-dimensional space.

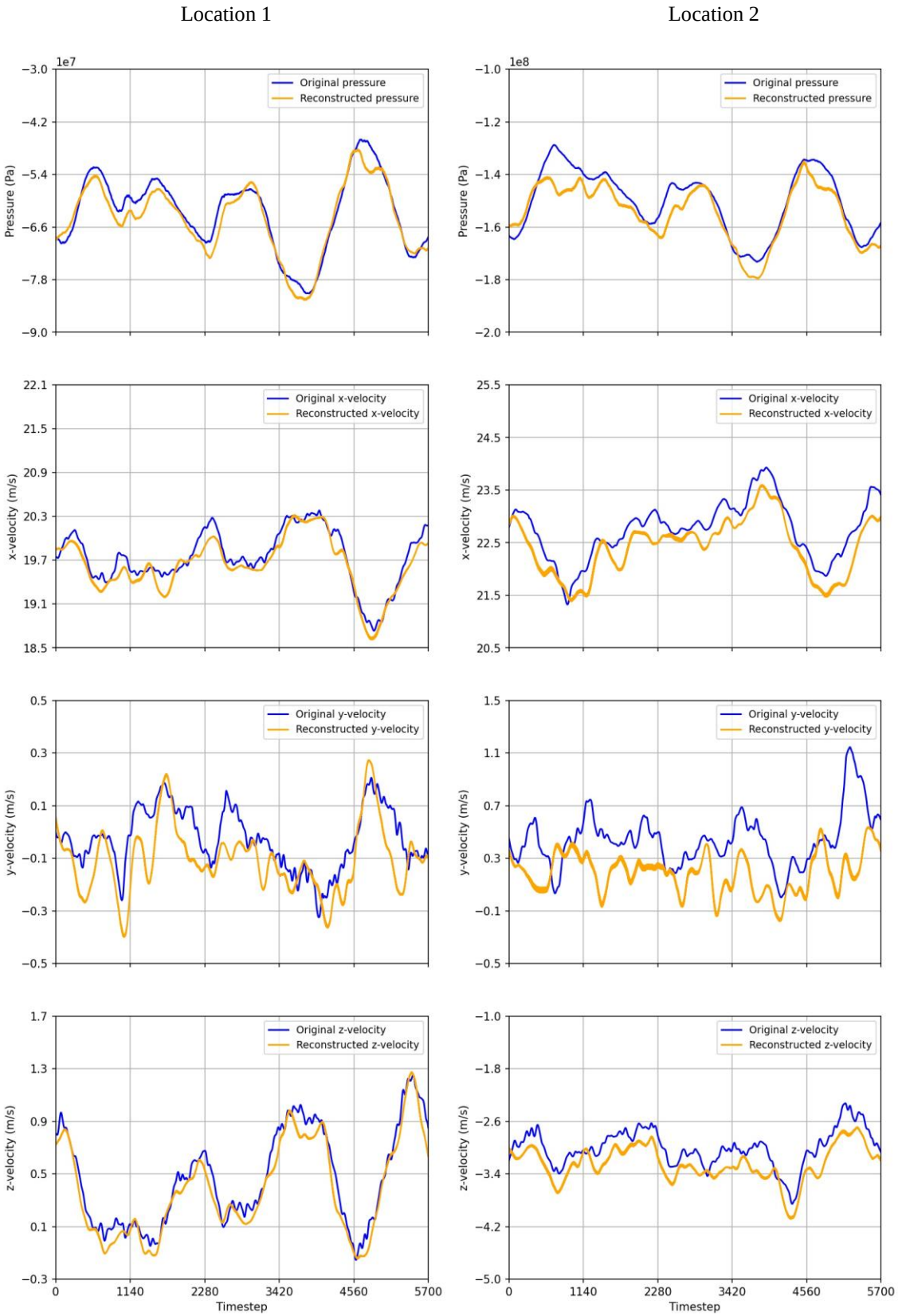

Figure 4. Original and reconstructed pressure and velocities fields along the training period for two probe locations

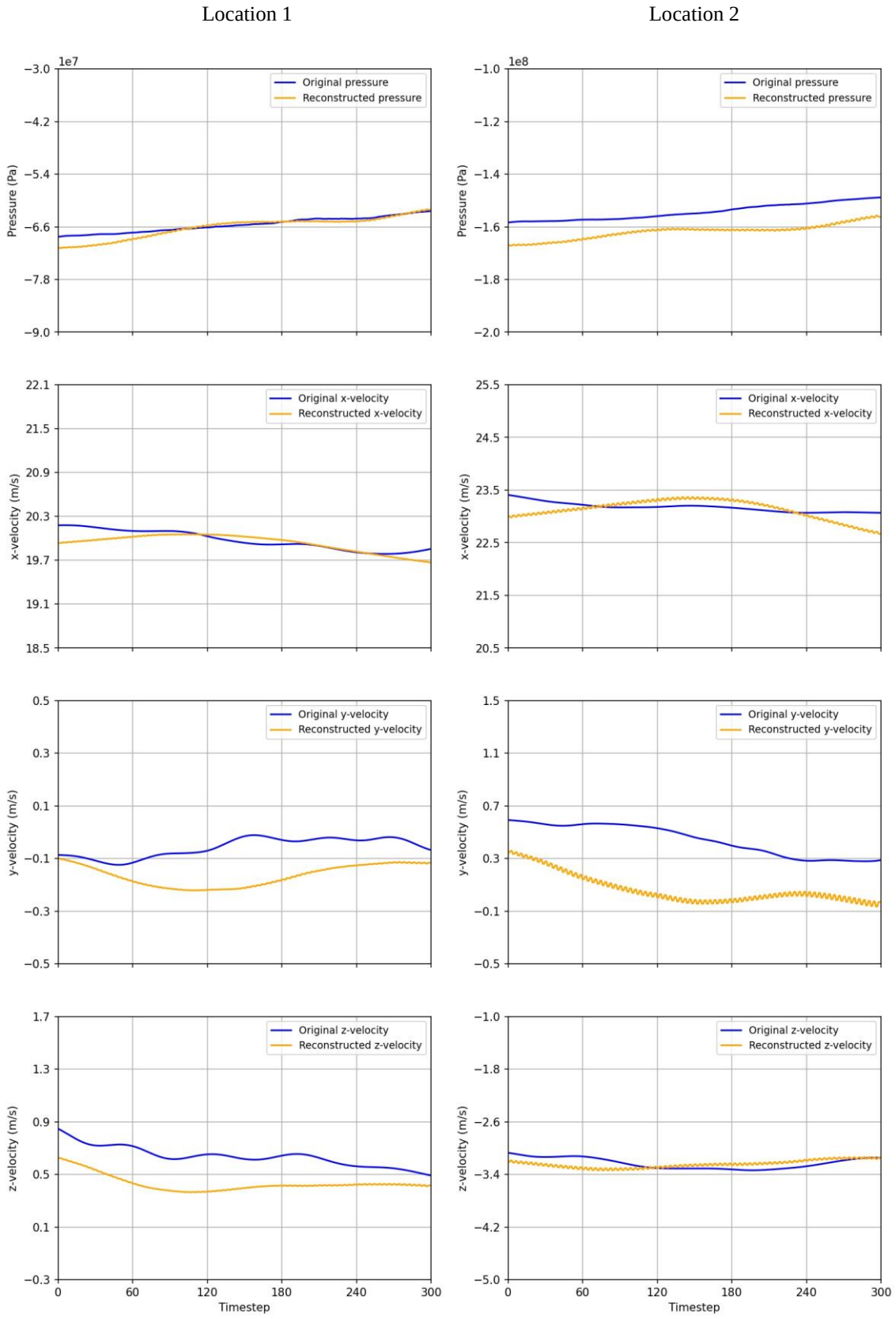

Figure 5. Original and reconstructed pressure and velocities fields along the testing period for two probe locations

5. CONCLUSION

A three-dimensional large-eddy simulation of an air flow over a bump was carried out in ANSYS Fluent, generating an extensive dataset containing the spatiotemporal evolution of pressure and velocities fields. Then, an incremental proper orthogonal decomposition of this high-dimensional data was performed so that it could be represented in a low-dimensional space. A good agreement between original and reconstructed fields was obtained for both training and testing periods, with normalized root-mean-square errors not larger than 2%, considering 61 dominant singular vectors. Pressure was the easiest state variable to be reconstructed, while z-velocity was the most difficult one. The results suggest the aforementioned data reduction technique is adequate for fluid flow systems and may be applied on model order reduction frameworks.

6. ACKNOWLEDGEMENTS

The authors thank the Brazilian Government agencies CAPES and CNPq for the continuous support during the development of this research.

7. REFERENCES

Han, Y., Stoellinger, M., Naughton, J., 2016. “Large eddy simulation for atmospheric boundary layer flow over flat and complex terrains”. *Journal of Physics: Conference Series*, Vol. 753, No. 3.

Huang, C., Duraisamy, K., Merkle, C.L., 2019. “Investigations and improvement of robustness of reduced-order models of reacting flow”. *AIAA Journal*, Vol. 57, No. 12, pp. 5377–5389.

Lignarolo, L., Gorlé, C., Parente, A., Benocci, C., 2011. “Large eddy simulation of the atmospheric boundary layer using OpenFOAM”. In *13th International Conference on Wind Engineering*. Amsterdam, Netherlands.

Matsumoto, D., Kiewat, M., Niedermeier, C.A., Indinger, T., 2019. “Application of Incremental Proper Orthogonal Decomposition for the Reduction of Very Large Transient Flow Field Data”. *International Journal of Automotive Engineering*, Vol. 10, No. 1, pp. 117–124.

McQuarrie, S.A., Huang, C., Willcox, K.E., 2021. “Data-driven reduced-order models via regularized operator inference for a single-injector combustion process”. *Journal of the Royal Society of New Zealand*, Vol. 51, pp. 194–211.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. “Scikit-learn: Machine Learning in Python”. *Journal of Machine Learning Research*, Vol. 12, pp. 2825–2830.

Porté-Agel, F., Wu, Y.T., Lu, H., Conzemius, R.J., 2011. “Large-eddy simulation of atmospheric boundary layer flow through wind turbines and wind farms”. *Journal of Wind Engineering and Industrial Aerodynamics*, Vol. 99, No. 4, pp. 154–168.

Robertson, E.D., Wang, Y., Pant, K., Grismer, M.J., Camberos, J.A., 2018. “A flow feature detection framework for large-scale computational data based on incremental proper orthogonal decomposition and data mining”. *International Journal of Computational Fluid Dynamics*, Vol. 32, No. 6, pp. 261–277.

Ross, D., Lim, J., Lin, R., Yang, M., 2008. “Incremental Learning for Robust Visual Tracking”. *Journal of Computer Vision*, Vol. 77, No. 1, pp. 125–141.

Zhiyin, Y., 2015. “Large-eddy simulation: Past, present and the future”. *Chinese Journal of Aeronautics*, Vol. 28, No. 1, pp. 11–24.

8. RESPONSIBILITY NOTICE

The authors are the only responsible for the printed material included in this paper.