Proceedings of the 15th ERCOFTAC Symposium on Engineering Turbulence Modelling and Measurements
Dubrovnic, Croatia, 22-24 September 2025 doi: 10.5281/zenodo.17280519

ENSEMBLE-BASED DATA ASSIMILATION OF PIV DATA FOR TURBULENT FLOW PAST A SURFACE-MOUNTED CUBE

N. P. Pallas[1], *V. Pappa*[1] *and D. Bouris*[1]

[1] *School of Mechanical Engineering, National Technical University of Athens, Greece*

npallas@mail.ntua.gr

Abstract

Computational and experimental methods have been used separately for years in the framework of fluid mechanics, even in situations where they are both available and quantify the same phenomenon. By combining experimental data and CFD simulations, data assimilation (DA) is capable of producing a new, improved solution, offering a new perspective in the field. In this scope, a localised, ensemble-based DA method, derived from the BLUE (Best Linear Unbiased Estimator) equations is presented here and applied to the case of 3D, turbulent and time-averaged flow past a surface mounted cube at a Reynolds number equal to 16500. Stereo PIV data are available and are assimilated into a CFD solution, which is obtained with a Reynolds-averaged Navier-Stokes (RANS) simulation in conjunction with the RNG $k-\epsilon$ turbulence model. The main novelty is that additional CFD runs for characterising uncertainty through generation of an ensemble are completely avoided, yielding significant reduction of the computational cost. The foundation for this approach is derived from theory, mathematical (i.e. sensitivity analyses) and physical (assumption of Reynolds number independence) insights as well as experimental data (hot-film anemometry data). Results pertaining to two measurement planes covering the near-wake and the roof regions of the cube are presented. The influence of the number of measurements is also examined for the former region. Significant correction was observed even for very few assimilated PIV measurements (about 150 measurement points) whenever CFD (model) uncertainty was clearly higher than the PIV (measurement) uncertainty.

1 Introduction

Computational and experimental sources of information have been used separately for years in the framework of fluid mechanics, even in cases where they are both available and pertain to the same phenomenon. Both of these sources of information are characterised by specific advantages and disadvantages, but data assimilation (DA) is capable of exploiting the advantages while mitigating the disadvantages. By combining experimental data and CFD results, DA produces a new, improved solution. In the general framework of DA, three main categories of methods are encountered: (a) state observer/nudging/measurement integration methods, (e.g. Pallas and Bouris, 2022), (b) sequential/Kalman Filter (KF)-based methods (e.g. Moldovan et al., 2021) and (c) variational methods. (e.g. Symon et al., 2017). Uncertainty quantification (UQ) is inherent in KF-based methods, while in variational and nudging endeavours the mathematical formulation must be properly adjusted to accommodate UQ. Additionally, although variational methods are characterised by very high accuracy, they are also linked to issues such as their somewhat difficult mathematical formulation (Moldovan et al., 2021) i.e. the use of adjoint operators. In the present attempt, a KF-based method is chosen.

A popular variant of the KF is the Ensemble KF (Evensen, 2009). This method usually relies on costly CFD simulations in order to quantify model uncertainty. A novelty of the proposed approach is that these additional CFD runs are completely avoided. The foundation for this approach is derived from theory (Cook, 1978), mathematical (i.e. sensitivity analyses) and physical (assumption of Reynolds number independence) insights as well as experimental data (hot-film anemometry data). The underlying assumption is that the model uncertainty mainly originates from inaccurate inlet boundary conditions (BCs). When the introduction of input model uncertainty is based on the available experimental data set, it is essentially information available a priori. This is implemented here to define a range of scaling factors that are applied to the initial solution in order to create the ensemble members. This approach is proven to be equivalent to perturbing the BCs and performing new CFD simulations.

The computational cost is further mitigated by implementing explicit localisation (Hunt et al., 2007) which also ensures a sufficient number of ensemble members, spanning the statistical space of the problem. The applied localisation significantly facilitates the parallelisation of the algorithmic procedure, which is done with OpenMP. As a result, the proposed method can be considered as a computationally efficient post-processing tool, capable of running on a personal computer and useful whenever there is a CFD solution and experimental data quantifying the same phenomenon. Although a time-averaged case is stud-

 doi: 10.5281/zenodo.17280519

ied here, it can be viewed as the initial correction step of KF-based DA approaches for unsteady flows. Mons et al. (2024) also presented the possibility of initially performing DA for a time-averaged flow field and then using the corrected solution for a subsequent resolvent analysis to estimate flow fluctuations. This approach is also useful in situations where the employed experimental equipment is not capable of providing time-resolved velocity fields e.g. with a low frequency Particle Image Velocimetry (PIV) system.

The proposed approach is applied to the case of time-averaged, 3D turbulent flow past a surface-mounted cube at a Reynolds number of 16500. A Reynolds-averaged Navier-Stokes (RANS) simulation is carried out, combined with the RNG $k-\epsilon$ turbulence model, using OpenFOAM software. Stereo PIV data are available for the same problem from an experiment performed in the closed-loop wind tunnel facility of the National Technical University of Athens (Pappa et al., 2023). Ensemble-based DA is performed here by utilising the Best Linear Unbiased Estimator (BLUE) equations. The influence of the number of measurements is also briefly discussed. It is shown that as few as ~ 150 measurements can lead to a correction that is comparable with that obtained by assimilating the whole PIV dataset (~ 10000 measurements).

2 Equations and Methodology

Equation (1) gives the correction $\mathbf{x}_\alpha$ (analysis) of the model solution $\mathbf{x_b}$, by assimilating the existing measurements $\mathbf{y}$, according to the BLUE method. $\mathbf{x_b}$ and $\mathbf{y}$ are arrays containing the CFD solution and the PIV measurements, respectively.

$$\mathbf{x}_\alpha = \mathbf{x_b} + K\left(\mathbf{y} - H_m\mathbf{x_b}\right) \tag{1}$$

H_m is a mapping matrix from the model to the measurement space. Note that for the interpolation from the model to the measurement space, a method based on the work of Shepard (1968) is applied, where the 4 model grid-points closest to a measurement point are utilised. The gain matrix, K, determines the degree of correction by quantifying the relative uncertainty between the two sources of information:

$$K = P_b H_m^{\mathsf{T}}\left(H_m P_b H_m^{\mathsf{T}} + R\right)^{-1} \tag{2}$$

where P_b and R are the model and measurement error covariance matrices, respectively. Ensemble-based methods make use of the following equation (Evensen, 2009), where the superscript j denotes a specific model ensemble member and N_{ens} is the size of the ensemble:

$$P_b = \frac{1}{N_{ens}}\sum_{j=1}^{N_{ens}}\left[\left(\mathbf{x_b} - \mathbf{x_b}^j\right)\left(\mathbf{x_b} - \mathbf{x_b}^j\right)^{\mathsf{T}}\right] \tag{3}$$

It is assumed that the matrix R, is diagonal (i.e. spatially uncorrelated measurements) with its diagonal entries equal to the measurement uncertainty.

The size of the ensemble confines the rank of the model error covariance matrix, P_b, since it can be at most equal to $N_{ens} - 1$ (Evensen, 2009; Hunt et al., 2007). Potential rank deficiency problems of P_b are tackled by performing explicit localisation (Hunt et al., 2007). This means that only the r closest measurements to a specific model variable are used for its correction. Thus, the initial correction (Equation 1), ceases to be an operation in the whole model space and becomes a point-wise one. Equations (1), (2) and (3) still hold though, after appropriate adjustments. They are applied separately for every different model variable (e.g. streamwise and vertical velocity components). As a result, potentially spurious spatial correlations between model state variables (e.g. velocities at different positions) beyond a certain distance are omitted. Cross-variable correlations (e.g. among different velocity components) are also omitted in the current paper. All the necessary searches implicated in the proposed approach are effectuated by using the k-nearest neighbours algorithm.

In order to create the model ensemble, one or more input parameters are usually perturbed. Common choices of input sources of uncertainty include unmodelled physics and uncertain BCs (Resseguier et al., 2021). Here, uncertain inlet BCs are assumed to be the dominant source and are used to generate the ensemble members. In problems with an Atmospheric Boundary Layer (ABL) as the approaching flow, the inlet BCs of Richards and Hoxey (1993) are frequently employed in conjunction with a $k-\epsilon$-type turbulence model. It is seen that two main parameters, i.e. the friction velocity u_* and the roughness length z_0, control the shape of these inlet BCs for the velocity, the turbulence kinetic energy and the turbulence dissipation rate. An obvious approach would be to perturb both of them. However, in the examined case, u_* is considered as the sole input uncertainty parameter, which alleviates additional complexity. This is supported by Cook (1978) where a significant variability in the extraction of u_* is stated, while z_0 is a much more stable geometric quantity, related to surface roughness. This choice of input uncertainty was also supported by means of sensitivity and error propagation analysis of the used inlet BCs to variations of u_* and z_0. Since the conclusions extracted by implementing both the aforementioned investigative tools are similar, results only pertaining to the former (i.e. sensitivity analysis) are shown here. The following expression is plotted in Figure 1, in order to examine the influence of either u_* or z_0 on the inlet velocity at the reference height i.e. the height of the cube $z_{ref} = H$:

$$\varepsilon_r = \frac{|U^i_{\text{ref}} - U_{\text{ref}}|}{U_{\text{ref}}} = g(r_p), \quad 0.3 \leq r_p \leq 1.2 \tag{4}$$

where $r_p = z_0^i/z_0$ or u_*^i/u_*. So, r_p is the ratio of the tested z_0^i or u_*^i values over the nominal values i.e. $z_0 = 0.00145$ m and $u_* = 0.22$ m/s, respectively,

 doi: 10.5281/zenodo.17280519

which were experimentally extracted and can be found in (Pappa et al., 2023). The range from 0.3 to 1.2 corresponds to the suburban regime (VDI, 2000) whenever z_0 is varied. It has been chosen as such since there is no particular reason for the value of z_0 to be uncertain to the point that the roughness characterisation of the terrain changes from one regime to another. This is further supported by Cook (1978) who stated that *the roughness length z_0 has a unique value in any given boundary layer*.

Figure 1 shows that the influence of u_* on the inlet BCs is clearly stronger than that of z_0. For example, a $\pm 20\%$ variation of u_* yields a $\pm 20\%$ variation of the inlet velocity at the height of the cube (reference velocity) while $\pm 20\%$ variation of z_0 produces a markedly lower variation of the reference velocity i.e. $\pm 5\%$. Similar conclusions are valid also for the inlet turbulence dissipation rate (ϵ) profile, as calculated by using the equations of Richards and Hoxey (1993). The former discussion, is irrelevant for the inlet profile of the turbulence kinetic energy k since it is only a function of u_* and not z_0 and therefore neither a comparative sensitivity study nor an error propagation analysis is necessary. Although the conclusions extracted here pertain only to the velocity at the height of the cube, an analysis (not shown here) has been carried out, which spanned the full height of the domain, showing that the influence of u_* on the inlet velocity is noticeably more pronounced than that of z_0, except very close to the ground i.e. $z/z_{ref} < 0.1$. Finally, it is noted that by perturbing u_*, all inlet quantities change at the same time and in a physically meaningful way. This also applies to the imposed top BC i.e. constant shear stress $\tau = \rho u_*$. Based on all previous arguments, any uncertainty linked to the calculation of z_0 is ignored.

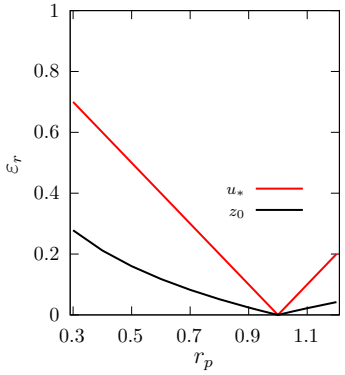

Figure 1: Relative error of the (streamwise) reference velocity, i.e., $\varepsilon_r = |U^i_{\text{ref}} - U_{\text{ref}}|/U_{\text{ref}}$ (%) with respect to the ratio $r_p = z^i_0/z_0$ (black) and u^i_*/u_* (red) within the range $[0.3, 1.2]$ corresponding to the moderately rough regime.

To generate the model ensemble members, u_* is assumed to follow a Gaussian distribution given by: $u_* \sim N(\overline{u_*}, \sigma^2_{u_*})$, where $\overline{u_*}$ and σ_{u_*} are the nominal value and the uncertainty of the friction velocity, respectively. The nominal values of u_* and z_0 can be calculated by a fitting of ABL velocity data (here hot-film data) to the logarithmic law (see Richards and Hoxey, 1993). σ_{u_*} is approximated by the following equation, derived by error propagation:

$$\sigma_{u_*} = \sqrt{\left(\frac{\partial u_*}{\partial U}\right)^2 \sigma_U^2} = \left[\frac{\kappa}{\ln(z/z_0)}\right] \cdot \sigma_U \qquad (5)$$

Hence, if the uncertainty σ_U of the inlet velocity is a priori known, then the corresponding uncertainty σ_{u_*} is calculated. The quantity σ_U can be approximated as $\sigma_U(z) = I_U(z) \cdot U(z)$ where $I_U(z)$ is the inlet turbulence intensity along the height z. This approximation implies that the turbulent nature of the flow is the dominant contribution to uncertainty. This is an exaggeration, but the aim here is to calculate σ_{u_*} based on concrete a priori information. Then, by applying Equation (5), a profile of σ_{u_*} along the height z is extracted, which is shown in Figure 2. Since u_* is a unique parameter of the inlet (i.e. no variation of u_* with respect to the height z), a unique value of σ_{u_*} should also be chosen. Therefore it is assumed that $\sigma_{u_*} = min[\sigma_{u_*}(z)] \approx 11\%$ which is a conservative choice since it will lead to a smaller correction from the DA methodology. In this way, it is ensured that no overestimation of uncertainty takes place throughout the height of the domain.

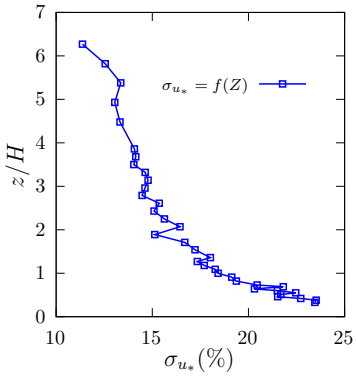

Figure 2: Friction velocity uncertainty, σ_{u_*}, non-dimensionalised by the nominal value of the friction velocity $u_* = 0.22$ m/s. All results are plotted against the non-dimensional height z/H where H is the height of the cube.

Introducing variations in u_* with constant z_0 leads to a scaling of mean velocities, as implied by the necessity for u_* and z_0 to satisfy the logarithmic law, and thus to a Reynolds number (Re) variation. Therefore, every sampled value of $u_* \sim N(\overline{u_*}, \sigma^2_{u_*})$, corresponds to a "scaled" "log law" velocity profile and thus a different Re number. Hence, if Re-independence is established for the examined range, then the common approach of creating ensemble members via independent CFD simulations could be avoided and, equivalently, substituted by the above mentioned Re-scaled model solutions, leading to a significant reduction of the computational cost.

 doi: 10.5281/zenodo.17280519

3 Numerical and Experimental Setup

The computational setup widely follows the guidelines suggested by Franke et al. (2007). The dimensions of the computational domain along with the coordinate system are given in Figure 3. This domain size results in a blockage ratio less than 0.5%.

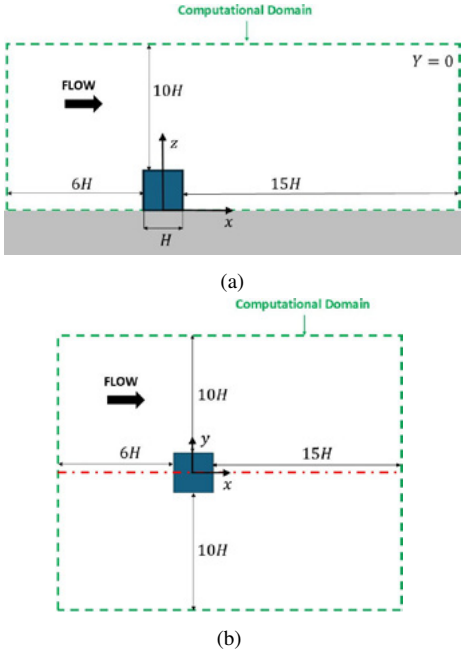

Figure 3: Schematic illustration of the computational domain: (a) plane of symmetry, (b) top view. The flow is from left to right.

The lateral boundaries are treated as symmetry planes while at the outlet, Neumann BCs are applied for all variables except pressure, which is considered to be zero (reference pressure). Wall functions (Launder and Spalding, 1983), as implemented in OpenFoam, are imposed on every wall boundary of the domain for the turbulence dissipation rate (ϵ), while the velocity at the first cell away from the wall is calculated by using a value of turbulent viscosity that originates from the employed wall functions. Neumann BCs are imposed for pressure and turbulence kinetic energy (k). The inlet (Dirichlet) and top (constant shear stress) BCs have been already discussed.

A cartesian and collocated grid is used in the present study with $\sim$ 3.3 M hexahedral 3D cells. The grid is not uniform since appropriate grading is applied, with grid refinement near the walls of the cube and the ground. A maximum cell aspect ratio equal to 64 is ensured. Grid independence studies (not shown here) proved that the mesh is adequately dense. The SIMPLE-based solver *simpleFoam* of OpenFOAM is used for pressure-velocity coupling.

Regarding the experiment, three different planes of PIV data are available along with inlet velocity profiles measured with hot-film anemometry at $Re = 16500$. Here, DA results pertaining to only two of the three planes are given. The hot-film measurements provide the following (nominal) values of $u_* = \overline{u_*} = 0.22$ m/s and $z_0 = 0.00145$ m. These measurements enable the implementation of Equation (5) which yields $\sigma_{u_*}/\overline{u_*} = min[\sigma_{u_*}(z)]/\overline{u_*} \approx 11\%$, as already mentioned. The minimum resolved velocity, playing the role of the measurement uncertainty (i.e. diagonal entries of matrix R), is equal to 0.11 m/s. More information regarding the experiment can be found in (Pappa et al., 2023).

4 Results

To test for Re-independence, non-dimensional results corresponding to $Re = 10000, 16500$ and 22000 are compared to each other. Re numbers are defined with respect to H (cube's height) and $U_{ref} = U(H)$. The latter Re numbers correspond to $\sigma_{u_*} = \overline{u_*} - 3\sigma_{u_*} = 0.14$, $\overline{u_*} = 0.22$ and $\overline{u_*} + 3\sigma_{u_*} = 0.29$, respectively. No significant velocity differences were observed within this Re range e.g. less than 5% of U_{ref} at $X/H = 1$ ($0.5H$ from the leeward wall of the cube) and $Y/H = 0$ (plane of symmetry). The respective results are given in Figure 4 for both the streamwise (U/U_{ref}) and the vertical (W/U_{ref}) velocities. Although results are shown here only at one position, Re number independence was also verified for other positions. DA is applied to the case of $Re = 16500$.

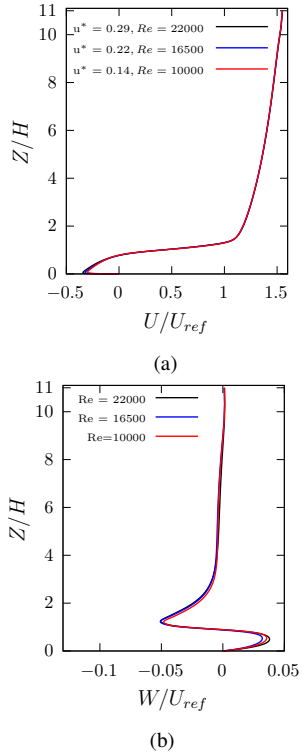

Figure 4: Vertical profiles for (a) U/U_{ref} and (b) W/U_{ref} at $X/H = 1$ (near-wake region) and $Y/H = 0$ (centre-plane which is also plane of symmetry) for $\sigma_{u_*} = \overline{u_*} - 3\sigma_{u_*} = 0.14$, $\overline{u_*} = 0.22$ and $\overline{u_*} + 3\sigma_{u_*} = 0.29$ corresponding to $Re = 10000$ (red), 16500 (blue) and 22000 (black), respectively.

DA results are given in Figure 5 for a plane cov-

 doi: 10.5281/zenodo.17280519

ering the near-wake region, which lies on the plane of symmetry of the cube (2D flow). Note that a constant number of $r = 10$ closest measurements are assimilated, following the localised approach mentioned in Section 2. A significant improvement is observed after implementing the suggested DA method in terms of streamwise velocity contours. However, the streamlines after DA do not strictly agree with the PIV ones. This might be attributed to the weak correction observed for the vertical velocity component due to the fact that its model uncertainty becomes comparable to 0.11 m/s (measurement uncertainty). Contours of the vertical velocity are given in Figure 6.

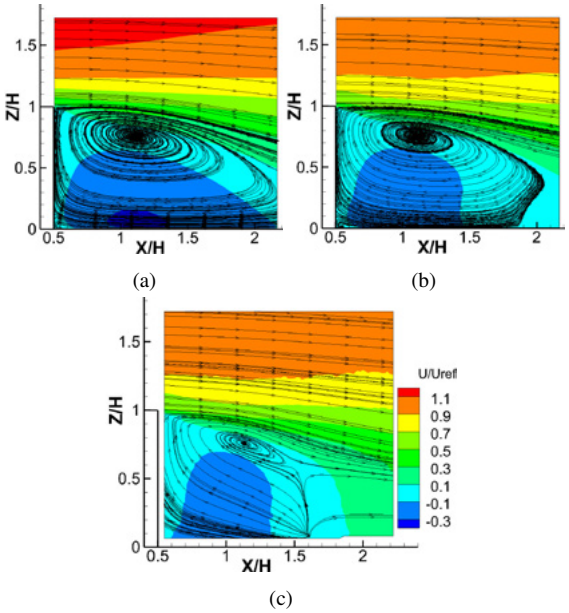

Figure 5: Contours of the streamwise velocity component for: (a) the CFD run, (b) the analysis solution (after DA) and (c) the PIV data for a (symmetry) plane covering the near-wake region. The flow is from left to right and the cube center is at $X/H = Z/H = 0$. Streamlines are also included.

The influence of the number of measurements is briefly discussed here. The initial PIV data are uniformly distributed in both X and Z directions while $\Delta X = \Delta Z$. In this way, fractions of (uniformly distributed) measurements $f = 1, 11$ and 100% can be tested. Streamwise and vertical velocity profiles are given in Figure 7 at $X/H = 1.1$. Even $f = 1\%$ ($\sim$ 150 measurements) gives a correction which is comparable to that of $f = 100\%$ (full PIV dataset) for the streamwise velocity. Hence, less time-consuming and expensive (e.g. point-wise) measurement techniques, e.g. Laser Doppler Velocimetry, could yield similar correction to that of the PIV method. For the vertical velocity component no correction is obtained at $X/H = 1.1$, presumably for the already mentioned reason of comparable CFD and PIV uncertainties. Another reason could be the choice of $r = 10$ closest measurements to be assimilated. This might not be the optimal choice, especially for the vertical velocity.

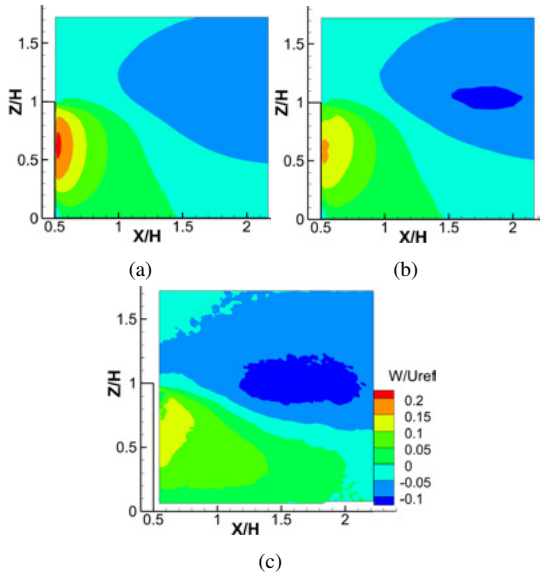

Figure 6: Contours of the vertical velocity component for: (a) the CFD run, (b) the analysis solution (after DA) and (c) the PIV data for a (symmetry) plane covering the near-wake region. The flow is from left to right and the cube center is at $X/H = Z/H = 0$.

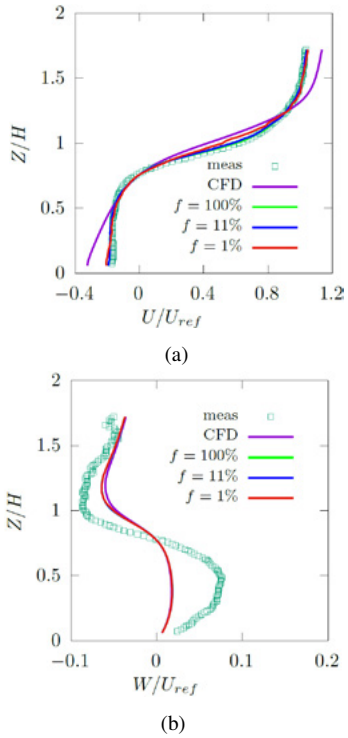

Figure 7: (a) Vertical profiles of (a) streamwise velocity and (b) vertical velocity for $f = 1, 11$ and 100% (after DA) at $X/H = 1.1$, compared to the CFD solution and the PIV data.

Similar results as that for the near-wake region (Figure 5) are shown in Figure 8 for a PIV measurement plane located above the roof of the cube and at $Y/H = 0$ (plane of symmetry). A significant correction of the CFD results is observed in terms of both

 doi: 10.5281/zenodo.17280519

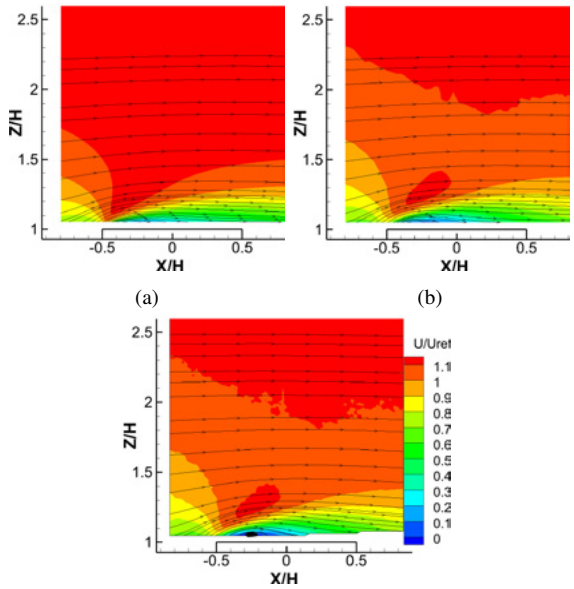

Figure 8: Contours of the streamwise velocity component for: (a) the CFD run, (b) the analysis solution (after DA) and (c) the PIV data for a (symmetry) plane covering the roof region. The flow is from left to right and the cube center is at $X/H = Z/H = 0$. Streamlines are also included.

streamwise velocity contours and streamlines. Although vertical velocity results are not shown here, it should be noted that, similarly to the wake region, the correction was not as prominent as that of the streamwise velocity. However, it was more pronounced than the one observed in Figures 6 and 7a. Finally, even a measurement fraction of $f = 1\%$ (~ 150 measurements) gave a comparable correction to that of the case with the whole measurement set being assimilated ($f = 100\%$), as observed for the near-wake area.

5 Conclusions

A localised, ensemble-based DA method, derived from the BLUE equations is presented here and applied to the case of 3D, turbulent and time-averaged flow past a surface mounted cube. The main novelty is that additional CFD runs for the generation of the ensemble, required to characterise model uncertainty, are completely avoided yielding significant reduction of the computational cost.

Significant correction was observed even for very few assimilated PIV measurements ($f = 1\%$ or about ~ 150 measurement points) for the near-wake as well as the roof region of the cube and for the streamwise velocity. For the vertical velocity component, less significant correction is observed, presumably because of comparable CFD and PIV uncertainty.

Acknowledgments

This work has been funded in part by the European Union, under the research project (PR 101079125) TWEET-IE: Twin Wind tunnels for Energy and the EnvironmenT - Innovations and Excellence.

References

Cook, N. J. (1978), Determination of the model scale factor in wind-tunnel simulations of the adiabatic atmospheric boundary layer, *Journal of Wind Engineering and Industrial Aerodynamics*, 2(4), 311-321.

Evensen, G. (2009), Data assimilation: the ensemble Kalman filter, *Springer*.

Franke, J., Hellsten, A., Schlünzen, H. and Carissimo, B. (2007), Best practice guideline for the CFD simulation of flows in the urban environment: quality assurance and improvement of microscale meteorological models. COST European Cooperation in Science and Technology.

Hunt, B. R., Kostelich, E. J., and Szunyogh, I. (2007), Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter, *Physica D: Nonlinear Phenomena* 230, 112-126.

Launder, B. E., and Spalding, D. B. (1983), The numerical computation of turbulent flows. Numerical prediction of flow, heat transfer, turbulence and combustion, *Pergamon*.

Moldovan, G., Lehnasch, G., Cordier, L., and Meldi, M. (2021), A multigrid/ensemble Kalman filter strategy for assimilation of unsteady flows, *Journal of Computational Physics* 443, 110481.

Mons, V., Vervynck, A., and Marquet, O. (2024), Data assimilation and linear analysis with turbulence modelling: application to airfoil stall flows with PIV measurements, *Theoretical and Computational Fluid Dynamics*, 38(3), 403-429.

Pallas, N. P., and Bouris, D. (2022), Calculation of the Pressure Field for Turbulent Flow around a Surface-Mounted Cube Using the SIMPLE Algorithm and PIV Data, *Fluids* 7, 140.

Pappa, V., Gromke, C., and Bouris, D. (2023), Twin test 1: Effect of vegetation on urban flows. PIV data from NTUA WT experiment and LDV from KIT WT experiment (1.0.0) [Data set], *Zenodo*, https://doi.org/10.5281/zenodo.10019002

Resseguier, V., Li, L., Jouan, G., Dérian, P., Mémin, E., and Chapron, B. (2021), New trends in ensemble forecast strategy: uncertainty quantification for coarse-grid computational fluid dynamics. *Archives of Computational Methods in Engineering* 28, 215–261.

Richards, P. J., and Hoxey, R. P. (1993), Appropriate boundary conditions for computational wind engineering models using the $k-\epsilon$ turbulence model, *Journal of Wind Engineering and Industrial Aerodynamics* 46, 145-153.

Shepard, D. (1968), A two-dimensional interpolation function for irregularly-spaced data. *In: Proceedings of the 1968 23rd ACM national conference* (pp. 517-524).

Symon, S., Dovetta, N., McKeon, B. J., Sipp, D., and Schmid, P. J. (2017), Data assimilation of mean velocity from 2D PIV measurements of flow over an idealized airfoil, *Experiments In Fluids* 58, 1-17.

VDI 3783 Part 12, Environmental Meteorology Physical Modelling of Flow and Dispersion Processes in the Atmospheric Boundary Layer. Applications of Wind Tunnels, Beuth Verlag, Berlin, 2000.