Journal of Wind Engineering & Industrial Aerodynamics 228 (2022) 105105

Contents lists available at ScienceDirect

Journal of Wind Engineering & Industrial Aerodynamics

journal homepage: www.elsevier.com/locate/jweia

On the use of ensemble averaging techniques to accelerate the Uncertainty Quantification of CFD predictions in wind engineering

Riccardo Tosi [a,b,*], Marc Núñez [a,b], Jordi Pons-Prats [a,c], Javier Principe [a,d], Riccardo Rossi [a,b]

[a] Centre Internacional de Mètodes Numèrics a l'Enginyeria, C/ Gran Capitán s/n Campus Nord UPC, 08034, Barcelona, Spain
[b] Departament d'Enginyeria Civil i Ambiental, Universitat Politècnica de Catalunya, Campus Nord UPC, 08034, Barcelona, Spain
[c] Departament de Física, Divisió d'Aeronàutica, Universitat Politècnica de Catalunya, Campus PMT UPC, 08860, Castelldefels, Spain
[d] Departament de Mecànica de Fluids, Universitat Politècnica de Catalunya, Campus Diagonal Besòs, Av. Eduard Maristany 16, Edifici A (EEBE), 08019, Barcelona, Spain

ARTICLE INFO

Keywords:
Uncertainty quantification
Ensemble averaging
Monte Carlo
Statistical estimate
CFD
Wind engineering
CAARC

ABSTRACT

In this work we focus on reducing the wall clock time required to compute statistical estimators of highly chaotic incompressible flows on high performance computing systems. Our approach consists of replacing a single long-term simulation by an ensemble of multiple independent realizations, which are run in parallel. A failure probability convergence criteria must be satisfied by the statistical estimator of interest to assess convergence. The error analysis leads to the identification of two error contributions: the initialization bias and the statistical error. We propose an approach to systematically detect the *burn-in time* needed in order to minimize the initialization bias, as well as techniques to choose the *effective time* needed to keep the statistical error under control. This is accompanied by strategies to reduce simulation cost. The proposed method is assessed in application to the prediction of the drag force over high rise buildings and specifically in application to the CAARC building, a relevant benchmark for the wind engineering community.

1. Introduction

The simulation of highly turbulent flows represents a well-established challenge in computational fluid dynamics (CFD), with predictions becoming more difficult as the Reynolds number (Re) increases. This situation is explained by Kolmogorov's theory which establishes that turbulent flows are characterized by multiple temporal and spatial scales, with an energy transfer cascade from larger eddies to smaller ones (Pope, 2000). According to the theory, the ratio between the largest and smallest length scales is proportional to $Re^{3/4}$, while the ratio between the timescales is proportional to $Re^{1/2}$. This phenomenon has practical implications on the flow around large objects, for example at the building scale in wind engineering. Such simulations require dealing with $Re \approx 10^8$, thus implying that the smallest eddies in the flow will be around 10^6 times smaller than the largest ones with dynamics occurring at time scales around 10^5 times shorter. Such estimates effectively rule out the possibility of performing direct numerical simulations and show how even large eddy simulation (LES) approaches (either based on filtering or on numerics, see e.g. implicit large eddy simulations (ILESs) Grinstein et al., 2007) are challenging (Cochran and Derickson, 2011).

From a practical point of view, one has to estimate statistics of the flow, e.g. mean or variance quantities. Such estimations typically require very long simulations which include the initial transient dynamics, required for the flow to develop, followed by the effective dynamics, required for the estimator to converge. Unfortunately, despite decades of hardware improvements, such simulations require prohibitive runtimes. While the use of high performance computing (HPC) systems may reduce these runtimes, practical limits exist on the achievable speedup for a given problem size. The most important feature controlling the runtime is that time evolution in a single simulation is intrinsically sequential (Makarashvili et al., 2017).[1]

Abbreviations: CAARC, Commonwealth Advisory Aeronautical Council; CFD, computational fluid dynamics; CFL, Courant–Friedrichs–Lewy; CPU, central processing unit; FE, finite element; HPC, high performance computing; ILES, implicit large eddy simulation; Kratos, Kratos Multiphysics; LES, large eddy simulation; MC, Monte Carlo; NS, Navier–Stokes; OSS, orthogonal subgrid scales; QoI, quantity of interest; Re, Reynolds number; SC, spatially correlated; SE, statistical error; VMS, variational multiscale

* Corresponding author at: Departament d'Enginyeria Civil i Ambiental, Universitat Politècnica de Catalunya, Campus Nord UPC, 08034, Barcelona, Spain.
E-mail addresses: rtosi@cimne.upc.edu (R. Tosi), mnunez@cimne.upc.edu (M. Núñez), jpons@cimne.upc.edu (J. Pons-Prats), principe@cimne.upc.edu (J. Principe), rrossi@cimne.upc.edu (R. Rossi).

[1] A potential solution could be parallel-in-time methods (Gander, 2015), which received much attention in the last years exactly due to their potential in providing a solution for the latter problem. Unfortunately, their application does not seem to be viable in chaotic problems (Wang et al., 2013).

https://doi.org/10.1016/j.jweia.2022.105105
Received 25 November 2021; Received in revised form 17 July 2022; Accepted 17 July 2022
Available online 10 August 2022

Acknowledging such limitations, we aim at exploring an alternative strategy based on estimating statistics by averaging over numerous independent simulations, i.e. statistical ensembles. The upshot of this strategy is that each of the simulations within the ensemble can be launched independently and run in parallel, thus providing an obvious opportunity for acceleration when abundant computational resources are available. This technique has been investigated in the literature in two different settings. In the first (Krasnopolsky, 2018), the focus is on reducing the wall clock time on constant hardware resources. Approaches to this problem typically consist of solving linear systems with multiple right-hand sides (Jiang and Layton, 2014). In the second, which our work considers, the focus is on exploiting the concurrency capabilities of HPC systems (Makarashvili et al., 2017).

As we shall show the use of ensemble averaging is always beneficial in speeding up calculations. Nevertheless the "optimal" strategy appears to be different when a constant inlet is considered (deterministic or uncertain) and when turbulent fluctuations are considered. In the first two cases the best approach to accelerate the solution seems to consist in generating multiple random initial conditions for each realization of inlet conditions. The use of ensemble averaging between the simulations obtained for a same inlet but different initial conditions allows to effectively employ much shorter time windows while achieving a high level of parallelism. A similar effect can be achieved when including random fluctuations in the inlet description by directly taking ensembles over "short" realization with randomly generated inlet conditions and fluctuations.

Even though ensemble averaging has been investigated previously, the application we target (wind engineering) as well as the numerical method we employ (ILES) is significantly different from previous investigations, thus leaving the applicability of ensemble averaging unclear. The goal of our work is therefore to develop a technique to assess the efficacy of ensemble averaging when applied to any given turbulent flow problem. To this end, one of our main contributions is a statistical analysis of the approach. Complementarily, the practical question we address in this work is: How efficient is the ensemble approach in the context of under-resolved LES methods, in particular, in wind engineering applications? Although we have targeted a specific class of engineering problems, our strategies are general and can be applied to assess the ensemble average approach for other, unrelated, problems.

In order to control bias associated to the initial conditions, the estimation of statistics of a turbulent flow entails collecting data starting only at some point in time after the flow has developed (Krasnopolsky et al., 2018), i.e. once the solution has been drawn to the attractor (Sagaut, 2006). We refer to the discarded initial time interval as the *burn-in time* and the remainder as the *effective time*. When a single long simulation is performed, the burn-in time is small compared to the remaining simulation time, which contains the effective dynamics. Unfortunately, this is not the case when the same amount of simulation time is distributed across an ensemble. Indeed, the same burn-in time will be paid by all realizations in the ensemble and, as a consequence, the total effective time will be reduced. The reduction of the burn-in time is therefore key to making ensemble averaging feasible. Our statistical model provides the tools to analyze the bias associated with the initial conditions, thus allowing us to faithfully select a practical burn-in time.

One of the additional research questions we address is what to assign as a distribution for the initial conditions in the cases in which they play a crucial role (inlet profile constant in time, without fluctuations). The aim here is again to control the burn-in time. To this end we consider spatially-correlated solenoidal fields. Another approach to cost reduction which we explore is to use a less accurate but less expensive time integration procedure during the burn-in phase, e.g. by increasing the time step size.

The final question we seek to address is how long each realization should be. Increasing the number of realizations improves concurrency but also increases the aggregated burn-in time, so there is a trade-off that needs to be considered. The present study evaluates this trade-off in the case of flows around bodies. In this work, we demonstrate that very short simulations are sufficient, thus making the approach highly efficient for the class of problems we have targeted. Moreover we verify that the analytical estimates correlate well with the measured data, thus confirming the validity of the predictions.

Our approach is validated with multiple numerical examples, which are available online (Tosi et al., 2021b). For solving the physical problems considered in this work, we use Kratos Multiphysics (Kratos) (Dadvand et al., 2010, 2013) as finite element (FE) solver software, XMC (Ayoul-Guilmard et al., 2020) as hierarchical Monte Carlo (MC) library and PyCOMPSs (Badia et al., 2015; Lordan et al., 2014; Tejedor et al., 2017) as programming model for distributed computing. The wind model described in Section 2 is based on Keith et al. (2021a,b). A reference implementation is available within Mataix et al. (2021).

The remainder of the article is structured as follows. In Section 2 we describe the wind model, the system we aim at solving and the ILES method we use. In Section 3, we describe our statistical approach. Different numerical experiments evaluating the performance of the methods we propose are presented in Section 4. Concluding remarks close the work in Section 5.

2. Wind modeling

In this work, we are interested in solving problems of wind flows around high-rise buildings, whose design is of great concern for researchers and engineers to ensure correct performance in terms of serviceability and habitability (Zheng et al., 2018). Such buildings reside entirely within the atmospheric boundary layer; a layer of Earth's atmosphere, extending vertically from its surface, which is characterized by constant shear stress in the vertical direction (Kaimal and Finnigan, 1994). This region is generally recognized to be *neutrally stable* at high wind speeds. That is, the buoyancy forces due to temperature gradients are negligible in comparison to surface-driven friction forces.

As is typical in the wind engineering community, we model the natural wind effects in the atmospheric boundary layer by decomposing the incoming velocity field, $\boldsymbol{u} = \overline{\boldsymbol{u}} + \boldsymbol{u}'$, into its stationary *mean profile* $\overline{\boldsymbol{u}}$ and unsteady *turbulent fluctuations* $\boldsymbol{u}'$. The mean wind profile is a contribution to the velocity field which only changes gradually over the span of several hours or days. It is generally considered constant with respect to the scale of most numerical simulations. On the other hand, the turbulent fluctuations introduce short term wind gusts with a time span of seconds or minutes. This term induces temporary states of maximum overall wind loading and can induce resonant effects in large structures if their structural eigenfrequencies coincide with frequencies in the gust-induced wind load pattern.

Let D denote a section of the atmospheric boundary layer lying above a flat section of Earth's surface, parameterized by the Cartesian coordinates (x, y, z). Let $\boldsymbol{e} \in \mathbb{R}^3$ be a unit normal vector which denotes the mean wind direction, $\overline{\boldsymbol{u}} \parallel \boldsymbol{e}$. Under the assumptions of neutral stability, homogeneous roughness, the mean velocity $\overline{\boldsymbol{u}}$ is function of the vertical coordinate ($\overline{\boldsymbol{u}} = f(z)$, with $f(\cdot)$ being a generic function useful to highlight dependencies) and is often modeled by the following logarithmic profile (Tamura and Kareem, 2013) when $z > z_0$:

$$\overline{\boldsymbol{u}} = \frac{u_*}{\kappa} \ln\left(\frac{z}{z_0}\right)\boldsymbol{e}, \tag{1}$$

where $\kappa \approx 0.41$ is the von Karmán constant and $z_0 > 0$ the *roughness height* parameter.

Depending on the local terrain, the variety of obstacles affecting ground friction can change vastly. For instance, consider that different friction forces will arise from flow across grass, forests, open water, or urban canopies. In our studies, we consider cases in which the local terrain type can be characterized by a fixed or a varying *roughness height* parameter z_0. Intervals of validity for this parameter, for various

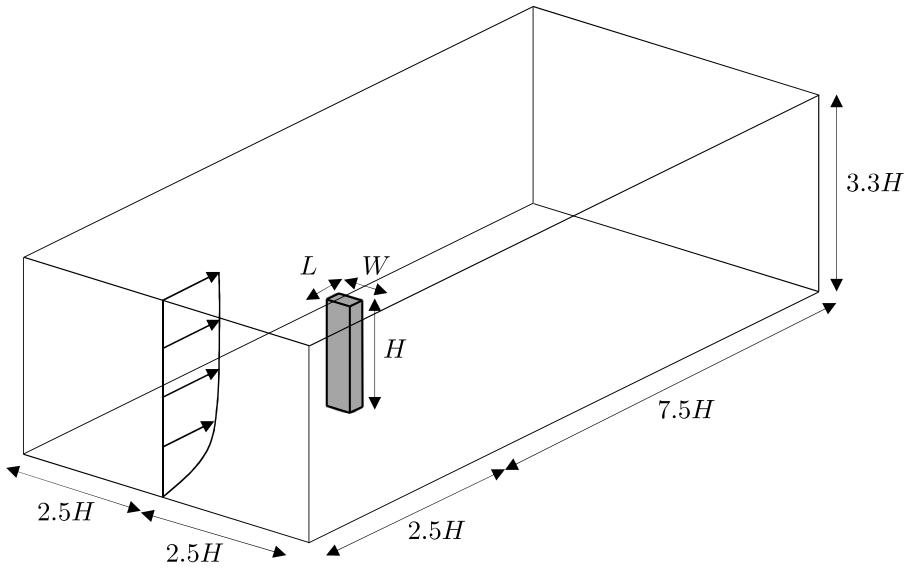

Fig. 2.1. CAARC problem domain. $H = 180\,\mathrm{m}$, $W = 45\,\mathrm{m}$ and $L = 30\,\mathrm{m}$. Wind gusts are omitted for the sake of simplicity, and only the mean wind profile is represented.

terrain categories, can be found in numerous engineering code books; e.g., Joint committee on structural safety. In addition, one must specify the *friction velocity* u_*, related to the shear stress on the ground τ_0 by $\tau_0 = \rho u_*^2$. This is done by specifying a reference velocity $\bar{\boldsymbol{u}}_H$ at a reference height H (see Section 2.1) which are inserted into Eq. (1) to obtain u_*.

In this work, we also assume that $\boldsymbol{e} = (1, 0, 0)$, that corresponds to an equal probability that the incoming wind will arrive from a fixed horizontal direction. In some environments, this is a poor assumption because geographic features may make the incoming wind arriving from any horizontal direction. In addition, we assume that the three mean profile parameters u_*, z_0, and $\boldsymbol{e}$ are independent. In specific application scenarios, this may also be a poor modeling assumption.

Modeling turbulent fluctuations with physical wind gust statistics is significantly more challenging than modeling the mean profile. Numerous techniques have been proposed in the engineering literature to tackle this issue, see, e.g., Huang et al. (2010), Guichard (2019), Ji et al. (2022) and Yang et al. (2020) for a numerical comparison. In this study, we choose to model the velocity fluctuations using the stochastic atmospheric boundary layer turbulence model proposed in Mann (1994, 1998); hereafter referred to as the *Mann model*. This model requires the parameters $\Gamma = 3.9$, $L = 0.59H$ and $\alpha\varepsilon^{2/3} = 3.2u_*/H^{2/3}$. The first controls the anisotropy due to shear, the second is a characteristic length scale defined in terms of the characteristic height of the problem H and the third is required for the definition of the energy spectrum. The fluctuations are generated in the Fourier space with a prescribed spectrum (that depends on Γ) and they have a correlation proportional to L and a variance proportional to u_*^2. The proportionality coefficients depend on Γ (see figure 4 in Mann, 1994) and their values are around 4.7 and 3.1 for $\Gamma = 3.9$. Once translated into the physical space, the Taylor's hypothesis is invoked to obtain the time dependent inlet field by evaluating at $x = \bar{\boldsymbol{u}}_H t$, see Mann (1998). Therefore, the temporal autocorrelation scale is proportional to $L/\bar{\boldsymbol{u}}_H$.

2.1. High-rise building problem

Let us introduce the high-rise Commonwealth Advisory Aeronautical Council (CAARC) building that we consider in this work, which has been widely used in literature in wind engineering benchmarks (Braun and Awruch, 2009; Huang et al., 2007; Holmes and Tse, 2014; Obasaju, 1992; Chen and Letchford, 2004; Li et al., 2020; Thordal et al., 2020a,b; Alminhana et al., 2018). The problem domain is presented in Fig. 2.1. The CAARC is a parallelepiped building with width 45 m, length 30 m and height 180 m. The domain is 1800 m long, 864 m large and 576 m high. We report in Table 1 the physical properties of the problem, where we remark that the reference mean wind velocity $\bar{\boldsymbol{u}}_H$ is defined at reference height H and that the roughness height is problem dependent. The roughness height can be fixed or can be a uniformly distributed random variable $\mathcal{U}(0.1, 0.7)$, where 0.1 and 0.7 are the minimum and maximum values, respectively. The Re is 119 millions, where a characteristic length of 45 m is considered. Under such conditions, it is clear that the problem is badly under-resolved.

Table 1
Physical parameters problem. The velocity $\bar{\boldsymbol{u}}_H$ is expressed in $\mathrm{m\,s^{-1}}$, the height H in m, the roughness height z_0 in m, the density ρ in $\mathrm{kg\,m^{-3}}$, the density μ in $\mathrm{kg\,m^{-1}\,s^{-1}}$ and the Re number is dimensionless.

$\bar{\boldsymbol{u}}_H$	H	z_0	ρ	μ	Re
40	180	2 or 0.1 – 0.7	1.225	1.846e−5	119447453

The fluid flow is modeled with the incompressible Navier–Stokes (NS) equations, and the system reads

$$\begin{aligned} \frac{\partial \boldsymbol{u}}{\partial t} + \boldsymbol{u}\cdot\nabla\boldsymbol{u} - \nu\Delta\boldsymbol{u} + \nabla p &= \boldsymbol{f} && \text{in } [0,T]\times D \\ \nabla\cdot\boldsymbol{u} &= 0 && \text{in } [0,T]\times D \\ \boldsymbol{u} &= \boldsymbol{u}_0(\boldsymbol{x}) && \text{in } t = 0\times D \\ \boldsymbol{u} &= f(z, z_0) && \text{on } [0,T]\times\Gamma_{\text{in}}, \end{aligned} \tag{2}$$

where Γ_{in} refers to the inlet boundary of the domain D, $[0, T]$ is the time window and $f(\cdot)$ is a generic function useful to express dependencies. Slip boundary conditions are applied on the walls and the ceiling, no-slip boundary conditions on the building and on the floor.

Three different boundary conditions of the problem are considered and therefore different configurations of the system are solved.

- The first scenario considers a constant in time fixed mean wind profile as boundary condition. The problem is solved with the standard time averaging method and the ensemble averaging method.
- The second scenario considers a constant in time stochastic mean wind profile as boundary condition. The problem is solved with the ensemble-based MC method.
- The third scenario considers turbulent fluctuations around a stochastic mean wind profile as boundary condition. The problem is solved with the ensemble-based MC method.

In next section, we present the statistical framework that we develop for solving the high-rise building problem, under the three scenarios described above.

2.2. Implicit large eddy simulation

In this section we briefly describe the ILES model we use for the simulation of turbulent flows around obstacles. Traditional LES methods are based on a filtering of the NS equations, which results in an extra term that requires modeling. Many different types of models are available (Sagaut, 2006) but the problem is considered from the continuous point of view, the discretization is performed after physical approximations have been made. However, the influence of the numerical scheme is very important, as described in Sagaut (2006, Chapter 7). Numerical experiments suggest that numerical errors are larger than the contribution of extra terms introduced by physical models when second order methods are used. The solution to this problem is to increase the accuracy of the numerical scheme or to use "pre-filtering" which consists of keeping the filter size constant while reducing the mesh size.

An alternative approach is the ILES method, pioneered in Boris et al. (1992). The idea is to consider the development of LES methods as discretization problems, taking into account the effect of missing scales through purely numerical artifacts. This line was further developed in Grinstein et al. (2007) but it became very popular after the introduction of the variational multiscale (VMS) method in Hughes et al. (1998). The VMS method was originally introduced as a framework to develop stable and accurate numerical methods of some class of problems as a generalization of older techniques like upwinding and artificial viscosities. However, scale separation is at the heart of VMS and it was soon recognized that it could be used to perform LES (Hughes et al., 2000). A distinctive feature of this approach is that scale separation is obtained by a projection operator. The solution space V is decomposed as $V = V_h \oplus \widetilde{V}$, into a discrete space V_h, e.g. a FE space built on top of a $\mathcal{T}_h$ of the domain, and a space of subgrid or unsolved scales $\widetilde{V}$ whose effect is modeled to obtain a closed stable and accurate problem in V_h. In this way, applying the VMS decomposition to the NS equations, the velocity $\boldsymbol{u} \in V$ is decomposed as $\boldsymbol{u} = \boldsymbol{u}_h + \widetilde{\boldsymbol{u}}$; a similar decomposition for the pressure can be introduced although it is not essential to have stability.

There are many VMS methods and a complete review of them is out of the scope of this article, see e.g. Codina et al. (2017). Some of these models have been successfully used for the LES of turbulent flows (Bazilevs et al., 2007; Nogueira et al., 2010; Colomés et al., 2015).

In all cases an algebraic approximation of the differential operator is required, e.g.

$$\boldsymbol{u} \cdot \nabla \widetilde{\boldsymbol{u}} - \nu \nabla^2 \boldsymbol{u} \approx \tau^{-1} \widetilde{\boldsymbol{u}}, \tag{3}$$

where τ is a piecewise constant function, computed within each element $K \in \mathcal{T}_h$ as

$$\tau_K^{-1} = \frac{c_1 \nu}{h_K^2} + \frac{c_2 \|\boldsymbol{u}_h + \widetilde{\boldsymbol{u}}\|_K}{h_K}. \tag{4}$$

Here, h_K is a characteristic length of K, c_1 and c_2 are algorithmic constants that depend only on the degree of the finite element approximation being used, and $\|\cdot\|_K$ is some norm defined on each element, e.g. the $L^2(K)$-norm.

Some further modeling choices lead to different VMS models. These choices include:

- *Static/Dynamic subscales:* from the VMS decomposition it follows that $\partial_t \boldsymbol{u} = \partial_t \boldsymbol{u}_h + \partial_t \widetilde{\boldsymbol{u}}$. Considering *dynamic subscales*, introduced in Codina et al. (2007), has some advantages like a correct behavior of time integration schemes and better accuracy. In particular, stability and convergence for the Stokes problem can be proved without any restriction on the time step size and the stabilization parameters on which the formulation depends. The typical approach, however, is the use of *quasistatic subscales* to neglect $\partial_t \widetilde{\boldsymbol{u}}$.
- *Linear/Nonlinear subscales:* applying the VMS decomposition to the nonlinear convective term, four different contributions are obtained on each equation (fine and coarse), that is, $\boldsymbol{u} \cdot \nabla \boldsymbol{u} = (\boldsymbol{u}_h + \widetilde{\boldsymbol{u}}) \cdot \nabla (\boldsymbol{u}_h + \widetilde{\boldsymbol{u}})$. After the approximation in Eq. (3) it is possible to keep all the contributions, as proposed in Codina et al. (2007). A simpler alternative is to perform the approximation $\boldsymbol{u} \cdot \nabla \boldsymbol{u} \approx \boldsymbol{u}_h \cdot \nabla \boldsymbol{u} + \widetilde{\boldsymbol{u}}$ (thus neglecting $\widetilde{\boldsymbol{u}}$ in Eq. (4) and the quadratic term in $\widetilde{\boldsymbol{u}}$ in Eq. (3)) which is enough to have numerical stability.
- *The space of subscales:* the choice of a space for the approximation of the subscales defines a projector $\mathcal{P}$ to be used in the fine scale equation. One option is to choose $\widetilde{V}$ as the space of the residual, that is to simply take $\mathcal{P} = \mathbb{I}$ (the identity). We refer to this space of subscales as the *algebraic subscales*. Another possibility is to consider the space of the subscales orthogonal to the FE space, that is, to take $\mathcal{P} := \Pi_h^{\perp} = \mathbf{I} - \Pi_h$, where Π_h is the projection onto the FE space (Codina et al., 2007).

A complete assessment of these modeling choices for the LES of turbulent flows can be found in Colomés et al. (2015). In this work we use static, linear, orthogonal subscales. Using nonlinear and/or dynamic subscales requires tracking them along the iterative and time integration loops, with the consequent increase in memory demands and computational cost (the simplest option is to store the subscales at the integration points). Although using dynamic, nonlinear orthogonal subscales provides a better accuracy, these subscales also imply a higher computational cost. The evaluation of this problem-dependent trade-off is outside the scope of this article.

However, even if it is simpler to consider algebraic subscales, orthogonal subgrid scales (OSS) enjoy a number of important properties that are worth having, such as stability without restrictions on the time step size (Codina et al., 2007), a clear scale separation in the energy transfers and the possibility of predicting backscatter with a stable numerical method (Principe et al., 2010). It is worth mentioning that, after the introduction of OSS a number of projection-based method appeared in which only some terms involving fine scales are kept in the resolved scale equation, see section 4.1 in Codina et al. (2017). The method we use in this article belongs to this class. It retains only the advective and pressure terms and projects them separately, so it is usually referred to as term-by-term OSS.

3. Statistical analysis

In this section, we introduce our statistical framework. In Section 3.1 we define the statistical operators and the convergence probability criterion. Different sources of error are identified and analyzed in Section 3.2. We discuss the generation of initial velocity fields, which provides independent ensemble realizations in Section 3.3. Finally, considerations on how to determine the length of the burn-in time are provided in Section 3.4.

3.1. Problem outline

We start introducing some preliminary notation. Given a random variable X, the expected value, variance, standard deviation and covariance (with respect to another random variable Y) of X are denoted with $\mathbb{E}[X]$, $\mathbb{V}[X]$, $\sigma[X] = \sqrt{\mathbb{V}[X]}$ and $\mathrm{co}\mathbb{V}[X, Y]$, respectively. Given J realizations of X, such statistical estimators can be approximated by sample average $\mathrm{E}^J[X]$, sample variance $\mathrm{V}^J[X]$, sample standard deviation $\sigma^J[X]$ and sample covariance, respectively.

Let us consider the NS problem of Eq. (2). We have $\boldsymbol{u}_0(\boldsymbol{x}, w_1)$ that is a random field, but also $z_0(w_2)$ is a random variable.

Remark 3.1. If $\boldsymbol{u}' \neq 0$, wind gusts dominate over initial conditions $\boldsymbol{u}_0(\boldsymbol{x}, w_1)$ and we can therefore consider $\boldsymbol{u}'(w_1)$ as random variable.

We describe this situation with $\Omega = \Omega_1 \times \Omega_2$, where $w_1 \in \Omega_1$ and $w_2 \in \Omega_2$ are independent uncertainties and Ω_1 and Ω_2 are their sample spaces (Ross, 2010, chapter 6). Then, $\mathbb{P} = \mathbb{P}_1 \times \mathbb{P}_2$ and the expected value of a generic joint random variable $X(w_1, w_2)$ is

$$\begin{aligned}\mathbb{E}[X] &= \int_{\Omega_1\times\Omega_2} X(w_1,w_2)\,\mathrm{d}\Omega_1\times\Omega_2(w_1,w_2)\\ &= \int_{\Omega_1}\int_{\Omega_2} X(w_1,w_2)\,\mathrm{d}\Omega_2(w_2)\,\mathrm{d}\Omega_1(w_1),\end{aligned} \tag{5}$$

where $\mathrm{d}\Omega_i(w_i) := \rho_i\,\mathrm{d}w_i$ and ρ_i is the probability density function of w_i, with $i = \{1,2\}$. We can introduce the operators $\mathbb{E}_1$ and $\mathbb{E}_2$ as

$$\mathbb{E}_i[X] = \int_{\Omega_i} X(w_1,w_2)\,\mathrm{d}\Omega_i(w_i), \tag{6}$$

where $i = \{1,2\}$. We remark that $\mathbb{E}_i$ is a random variable in Ω_j, where $i \neq j$. Moreover,

$$\mathbb{E}[X] = \mathbb{E}_1[\mathbb{E}_2[X]] = \mathbb{E}_2[\mathbb{E}_1[X]] = \mathbb{E}_1 \circ \mathbb{E}_2[X]. \tag{7}$$

Let $u(t,\boldsymbol{x},w_1,w_2) := (\boldsymbol{u}(t,\boldsymbol{x},w_1,w_2), p(t,\boldsymbol{x},w_1,w_2))$ denote the solution of the NS problem. From the solution field we compute the quantity of interest (QoI), that is a process, and reads $\mathrm{Q}(t,w_1,w_2) := \mathrm{Q}(u(t,\boldsymbol{x},w_1,w_2))$. The estimator we aim at computing is the long-term expected value of $\mathrm{Q}(t,w_1,w_2)$, which reads

$$\overline{\mathrm{Q}} = \lim_{t\to\infty} \mathbb{E}[\mathrm{Q}](t). \tag{8}$$

Let us now consider to have *ergodicity* in Ω_1 and not in Ω_2. If we think, for example, about $z_0(w_2)$, it is easy to see that different realizations will not converge to the same value, no matter how long the simulation is.

Now $\mathbb{E}_1$ can be approximated in two ways, since

$$\mathbb{E}_1[\mathrm{Q}] = \langle \mathrm{Q}(t,w_1,w_2)\rangle = \lim_{T\to\infty}\frac{1}{T}\int_0^T \mathrm{Q}(t,w_1,w_2)\,\mathrm{d}t. \tag{9}$$

We can take a finite time interval

$$\mathbb{E}_1[\mathrm{Q}] \approx \langle \mathrm{Q}(t,w_1,w_2)\rangle_{T_0,T} = \frac{1}{T-T_0}\int_{T_0}^T \mathrm{Q}(t,w_1,w_2)\,\mathrm{d}t \tag{10}$$

or an ensemble of them

$$\mathbb{E}_1[\mathrm{Q}] \approx \sum_{m=1}^M \langle \mathrm{Q}_m\rangle_{T_0,T} = \frac{1}{M}\sum_{m=1}^M \frac{1}{T-T_0}\int_{T_0}^T \mathrm{Q}_m\,\mathrm{d}t, \tag{11}$$

where $\mathrm{Q}_m := \mathrm{Q}(t, w_1^{(m)}, w_2)$ are different realizations in w_1 and T_0 is the burn-in time.

Now, the expected value operator *without ergodicity* can be approximated as

$$\mathbb{E}_2[\mathrm{Q}] \approx \frac{1}{N}\sum_{n=1}^N \mathrm{Q}_n, \tag{12}$$

where Q_n are different realizations in w_2, i.e. $\mathrm{Q}_n := \mathrm{Q}\left(t, w_1, w_2^{(n)}\right)$.

Finally, the long term expected value can be estimated as

$$\overline{\mathrm{Q}} \approx \mathrm{E}^{N,M}[\langle \mathrm{Q}_{n,m}\rangle_{T_0,T}] = \frac{1}{NM}\sum_{n=1}^N\sum_{m=1}^M \langle \mathrm{Q}_{n,m}\rangle_{T_0,T}, \tag{13}$$

where $\mathrm{Q}_{n,m} := \mathrm{Q}\left(t, w_1^{(m)}, w_2^{(n)}\right)$.

Remark 3.2. The estimation provided by Eq. (13) comes from considering $\boldsymbol{u}_0(\boldsymbol{x}, w_1)$ (or $\boldsymbol{u}\prime(w_1)$) and $z_0(w_2)$, but it is not limited to such random variables. It can be generalized to any random variable satisfying ergodicity in Ω_1 and any random variable not satisfying ergodicity in Ω_2.

Using this notation, the statistical problem we face is to find the optimal values of N, M, T_0, T (those that minimize the computational cost and/or the time to solution) while satisfying the probability convergence criterion

$$\mathbb{P}\left[\left|\frac{1}{NM}\sum_{n=1}^N\sum_{m=1}^M \langle \mathrm{Q}_{n,m}\rangle_{T_0,T} - \overline{\mathrm{Q}}\right| > \varepsilon\right] \leq \phi\,,\ \varepsilon > 0\,,\ \phi \ll 1\,, \tag{14}$$

where ε denotes the absolute tolerance and $1-\phi$ is the confidence on the final estimation. Such a condition requires that the probability of the error exceeding ε should not be greater than ϕ.

3.2. Error analysis

There are two sources of error in Eq. (13). First, the choice of the random initial condition may not be compatible with the long-term statistically stationary, thus triggering a transient perturbation with a nonvanishing contribution to the mean, sometimes called initialization bias (Banks et al., 2004). The use of a finite number of samples of finite duration is a second source of error. Increasing the number of samples, as well as the length of their effective time intervals, will also reduce the influence of the initialization bias.

3.2.1. Bias of initial conditions

We analyze now the error in approximating Eq. (13). We assume each realization $\mathrm{Q}_{n,m}$ can be described as the sum of a transient perturbation $A_{n,m}$ and of a statistical steady-state process $S_{n,m}$. Then,

$$\mathrm{Q}_{n,m} = A_{n,m} + S_{n,m}\,. \tag{15}$$

Assuming Eq. (15) implies that $\mathbb{E}_1[A_{n,m}](t) \xrightarrow[t\to\infty]{} 0$, from which it follows that

$$\mathbb{E}_1[\mathrm{Q}_{n,m}](t) \xrightarrow[t\to\infty]{} \mathbb{E}_1[S_{n,m}]\,. \tag{16}$$

By applying the $\mathbb{E}_2$ operator to the above equation, from Eq. (7) we obtain

$$\mathbb{E}[\mathrm{Q}_{n,m}](t) \xrightarrow[t\to\infty]{} \mathbb{E}[S_{n,m}] = \overline{\mathrm{Q}}\,. \tag{17}$$

We remark that $\mathbb{E}[A_{n,m}](t)$ and $\mathbb{E}[\mathrm{Q}_{n,m}](t)$ are the expected values of the transient perturbation $A_{n,m}$ and of the QoI $\mathrm{Q}_{n,m}$ at time t.

In this setting,

$$\begin{aligned}\mathbb{E}\left[\frac{1}{NM}\sum_{n=1}^N\sum_{m=1}^M \langle \mathrm{Q}_{n,m}\rangle_{T_0,T}\right] &= \frac{1}{NM}\sum_{n=1}^N\sum_{m=1}^M \mathbb{E}\left[\langle \mathrm{Q}_{n,m}\rangle_{T_0,T}\right]\\ &= \frac{1}{NM}\sum_{n=1}^N\sum_{m=1}^M \left(\mathbb{E}\left[\langle A_{n,m}\rangle_{T_0,T}\right] + \mathbb{E}\left[\langle S_{n,m}\rangle_{T_0,T}\right]\right).\end{aligned} \tag{18}$$

Since $S_{n,m}$ is statistically steady-state, then

$$\mathbb{E}[\langle S_{n,m}\rangle_{T_0,T}] = \mathbb{E}\left[\frac{\int_{T_0}^T S_{n,m}(t)\,\mathrm{d}t}{T-T_0}\right] = \frac{1}{T-T_0}\int_{T_0}^T \mathbb{E}[S_{n,m}]\,\mathrm{d}t = \mathbb{E}[S_{n,m}] = \overline{\mathrm{Q}}\,. \tag{19}$$

Therefore,

$$\mathbb{E}\left[\frac{1}{NM}\sum_{n=1}^N\sum_{m=1}^M \langle \mathrm{Q}_{n,m}\rangle_{T_0,T}\right] = \mathrm{B} + \overline{\mathrm{Q}}\,, \tag{20}$$

where

$$\mathrm{B} = \left|\frac{1}{NM}\sum_{n=1}^N\sum_{m=1}^M \mathbb{E}\left[\langle A_{n,m}\rangle_{T_0,T}\right]\right| \tag{21}$$

is the bias of the initial condition, which can be mitigated if $\mathbb{E}[\langle A_{n,m}\rangle_{T_0,T}]$ decays sufficiently fast. If

$$\int_0^\infty \left|\mathbb{E}[A_{n,m}](t)\right|\,\mathrm{d}t < \infty, \tag{22}$$

then we have that

$$\left|\mathbb{E}[\langle A_{n,m}\rangle_{T_0,T}]\right| = \frac{1}{T-T_0}\left|\int_{T_0}^{T}\mathbb{E}[A_{n,m}](t)\,\mathrm{d}t\right| \le \frac{1}{T-T_0}\int_{T_0}^{T}\left|\mathbb{E}[A_{n,m}](t)\right|\,\mathrm{d}t \le \frac{1}{T-T_0}\int_{0}^{\infty}\left|\mathbb{E}[A_{n,m}](t)\right|\,\mathrm{d}t \xrightarrow[T-T_0\to\infty]{} 0. \tag{23}$$

In other words, if Eq. (22) holds, then increasing $T - T_0$ eventually decreases $|\mathrm{B}|$.

An estimation of the decay rate can be made under stronger assumptions on the transient perturbation. For illustration purposes we consider a fast decay of the form

$$A^f_{n,m}(t) = A^f_{0,n,m}\mathrm{e}^{-\frac{t}{\tau}}, \tag{24}$$

which is an example of exponentially ergodic processes. For a wide class of stochastic processes satisfying a dissipativity condition, it can be proved that the transient perturbation decays exponentially (Meyn and Tweedie, 1993, theorem 6.1), i.e. it satisfies

$$|\mathbb{E}[A^f]| = |\mathbb{E}[A^f_0]|\mathrm{e}^{-\frac{t}{\tau}}, \tag{25}$$

as in Fang and Giles (2019, equation (3)).

If we now consider its time average we get

$$\mathbb{E}[\langle A^f_{n,m}\rangle_{T_0,T}] = \mathbb{E}[A^f_{0,n,m}]\langle \mathrm{e}^{-\frac{t}{\tau}}\rangle_{T_0,T} = \tau\mathbb{E}[A^f_{0,n,m}]\left(\frac{\mathrm{e}^{-T_0/\tau}-\mathrm{e}^{-T/\tau}}{T-T_0}\right), \tag{26}$$

from where we see a decay of the form

$$\mathbb{E}[\langle A^f_{n,m}\rangle_{T_0,T}] = \mathcal{O}((T-T_0)^{-1}). \tag{27}$$

Therefore, $\mathbb{E}[\langle A^f_{n,m}\rangle_{T_0,T}]$ is a decreasing function of $T - T_0$. However, as it can be seen in Eq. (26), it is also decreasing when $T - T_0$ is kept constant while T and T_0 separately increase. In practice, T is fixed so increasing T_0 decreases $T - T_0$; this is the trade-off we analyze in the examples of Section 4.

3.2.2. Statistical error

The previous analysis makes it clear that the bias can be reduced by increasing $T-T_0$ and, for specific transient perturbations, increasing T_0. However, to assess statistical accuracy, Eq. (14) needs to be evaluated. Given a bound $\left|\mathbb{E}\left[\frac{1}{NM}\sum_{n=1}^{N}\sum_{m=1}^{M}\langle \mathrm{Q}_{n,m}\rangle_{T_0,T}\right] - \frac{1}{NM}\sum_{n=1}^{N}\sum_{m=1}^{M}\langle \mathrm{Q}_{n,m}\rangle_{T_0,T}\right| \le \varepsilon$, the asymptotic normality of the estimator $\frac{1}{NM}\sum_{n=1}^{N}\sum_{m=1}^{M}\langle \mathrm{Q}_{n,m}\rangle_{T_0,T}$, in the limit $N, M \to \infty$, implies that (Durrett, 2019, chapter 3)

$$\left|\mathbb{E}\left[\frac{1}{NM}\sum_{n=1}^{N}\sum_{m=1}^{M}\langle \mathrm{Q}_{n,m}\rangle_{T_0,T}\right] - \frac{1}{NM}\sum_{n=1}^{N}\sum_{m=1}^{M}\langle \mathrm{Q}_{n,m}\rangle_{T_0,T}\right| \le C_\phi\sqrt{\mathbb{V}\left[\frac{1}{NM}\sum_{n=1}^{N}\sum_{m=1}^{M}\langle \mathrm{Q}_{n,m}\rangle_{T_0,T}\right]} \le \varepsilon, \tag{28}$$

with probability $1 - \phi$ as the tolerance $\varepsilon \to 0$. C_ϕ is the confidence coefficient defined as $C_\phi = \Phi^{-1}(1-\frac{\phi}{2})$, where Φ is the cumulative distribution function of a standard normal distribution. The total error in Eq. (14) can then be bounded with confidence $1 - \phi$, as follows,

$$\begin{aligned}\left|\frac{1}{NM}\sum_{n=1}^{N}\sum_{m=1}^{M}\langle \mathrm{Q}_{n,m}\rangle_{T_0,T} - \overline{\mathrm{Q}}\right| &\le \left|\mathbb{E}\left[\frac{1}{NM}\sum_{n=1}^{N}\sum_{m=1}^{M}\langle \mathrm{Q}_{n,m}\rangle_{T_0,T}\right] - \overline{\mathrm{Q}}\right| \\ &\quad + \left|\mathbb{E}\left[\frac{1}{NM}\sum_{n=1}^{N}\sum_{m=1}^{M}\langle \mathrm{Q}_{n,m}\rangle_{T_0,T}\right] - \frac{1}{NM}\sum_{n=1}^{N}\sum_{m=1}^{M}\langle \mathrm{Q}_{n,m}\rangle_{T_0,T}\right| \\ &\le \mathrm{B} + C_\phi\sqrt{\mathbb{V}\left[\frac{1}{NM}\sum_{n=1}^{N}\sum_{m=1}^{M}\langle \mathrm{Q}_{n,m}\rangle_{T_0,T}\right]}.\end{aligned} \tag{29}$$

We define the statistical error (SE) to be

$$\mathrm{SE} := \sqrt{\mathbb{V}\left[\frac{1}{NM}\sum_{n=1}^{N}\sum_{m=1}^{M}\langle \mathrm{Q}_{n,m}\rangle_{T_0,T}\right]}. \tag{30}$$

Thus, using Eq. (20), we get

$$\left|\frac{1}{NM}\sum_{n=1}^{N}\sum_{m=1}^{M}\langle \mathrm{Q}_{n,m}\rangle_{T_0,T} - \overline{\mathrm{Q}}\right| \le |\mathrm{B}| + C_\phi\mathrm{SE}, \tag{31}$$

where B is the initialization bias defined in Eq. (21). For a given confidence $1 - \phi$, the probability convergence criterion then reads

$$|\mathrm{B}| + C_\phi\mathrm{SE} \le \varepsilon\,. \tag{32}$$

The bias error $|\mathrm{B}|$ was analyzed above, let us now focus on the SE term. We can write

$$\begin{aligned}&\mathbb{V}\left[\frac{1}{NM}\sum_{n=1}^{N}\sum_{m=1}^{M}\langle \mathrm{Q}_{n,m}\rangle_{T_0,T}\right] \\ &= \frac{1}{N^2}\sum_{n=1}^{N}\mathbb{V}\left[\frac{\sum_{m=1}^{M}\langle A_{n,m}\rangle_{T_0,T}}{M}\right] + \frac{1}{N^2}\sum_{n=1}^{N}\mathbb{V}\left[\frac{\sum_{m=1}^{M}\langle S_{n,m}\rangle_{T_0,T}}{M}\right] \\ &\quad + \frac{1}{N^2}\sum_{\substack{n,k=1\\ n\neq k}}^{N}\mathrm{co}\mathbb{V}\left[\frac{\sum_{m=1}^{M}\langle \mathrm{Q}_{n,m}\rangle_{T_0,T}}{M}, \frac{\sum_{m=1}^{M}\langle \mathrm{Q}_{k,m}\rangle_{T_0,T}}{M}\right],\end{aligned} \tag{33}$$

where we assume $A_{n,m}$ and $S_{n,m}$ to be independent. The second term of the right hand side of Eq. (33) can be written as

$$\begin{aligned}\mathbb{V}\left[\frac{\sum_{m=1}^{M}\langle S_{n,m}\rangle_{T_0,T}}{M}\right] &= \frac{1}{M^2}\sum_{m=1}^{M}\mathbb{V}\left[\langle S_{n,m}\rangle_{T_0,T}\right] \\ &\quad + \frac{1}{M^2}\sum_{\substack{m,k=1\\ m\neq k}}^{M}\mathrm{co}\mathbb{V}\left[\langle S_{n,m}\rangle_{T_0,T}, \langle S_{n,k}\rangle_{T_0,T}\right] \\ &= \frac{\mathbb{V}\left[S_{n,m}\right]}{M(T-T_0)^2}\int_{T_0}^{T}\int_{T_0}^{T}\rho(s-t)\,\mathrm{d}s\,\mathrm{d}t \\ &\quad + \frac{1}{M^2}\sum_{\substack{m,k=1\\ m\neq k}}^{M}\mathrm{co}\mathbb{V}\left[\langle S_{n,m}\rangle_{T_0,T}, \langle S_{n,k}\rangle_{T_0,T}\right],\end{aligned} \tag{34}$$

where ρ is the autocorrelation function, see Pope (2000, section 3.6). As demonstrated in Pope (2000, problem 3.37), the long-time limit of the first term of the right hand side of Eq. (34) is $2\frac{\mathbb{V}[S_{n,m}]\overline{\tau}}{M(T-T_0)}$, where $\overline{\tau}$ is the integral timescale of the process (Pope, 2000, section 3.6), which is a correlation constant associated to the QoI.

The last term in Eq. (34) depends on the correlation between realizations with the same $\omega_2^{(n)}$. An example of the effects of the correlation between realizations of turbulent flow in a channel is presented in Makarashvili et al. (2017). If these realizations are independent, the final term in Eq. (34) is negligible. To this end, we discuss an initial condition strategy which helps to provide independent realizations in Section 3.3.

The first term on the right-hand side of Eq. (33) cannot be estimated without making assumptions on the behavior of the transient perturbation A. If we consider the same fast decay of the previous subsection, Eq. (24), a straightforward computation shows that $\mathbb{V}[\langle A^f_{n,m}\rangle_{T_0,T}] = \mathcal{O}((T-T_0)^{-2})$.

If the decay of the transient perturbation is slower than Eq. (24), the decay of the statistical error will be dominated by the first and third terms in Eq. (33). On the other hand, if the decay of the transient perturbation is fast (Eq. (24) holds[2]), the overall error is dominated by

[2] Or any other scenario such that $\mathbb{E}[\langle A_{n,m}\rangle_{T_0,T}] = \mathcal{O}((T-T_0)^{-p})$ and $\mathbb{V}[\langle A_{n,m}\rangle_{T_0,T}] = \mathcal{O}((T-T_0)^{-q})$, $p, q > 1$.

the second and third terms in Eq. (33), and such an equation becomes

$$\begin{aligned}\mathbb{V}\left[\frac{1}{NM}\sum_{n=1}^{N}\sum_{m=1}^{M}\langle \mathrm{Q}_{n,m}\rangle_{T_0,T}\right] &= \frac{1}{N^2}\sum_{n=1}^{N} 2\frac{\mathbb{V}[S_{n,m}]\overline{\tau}}{M(T-T_0)} \\ &+ \frac{1}{N^2}\sum_{\substack{n,k=1\\ n\neq k}}^{N} \mathrm{co}\mathbb{V}\left[\left[\sum_{m=1}^{M}\frac{\langle \mathrm{Q}_{n,m}\rangle_{T_0,T}}{M}\right],\left[\sum_{m=1}^{M}\frac{\langle \mathrm{Q}_{k,m}\rangle_{T_0,T}}{M}\right]\right].\end{aligned} \tag{35}$$

The first term of the right hand side of Eq. (35) decays as $\frac{c_0^2}{NM(T-T_0)}$, for some constant c_0. The covariance term of the right hand side of Eq. (35) behaves as $\frac{c^2}{N}$, for some constant c. The SE therefore decays as $\sqrt{\frac{c_0^2}{NM(T-T_0)}+\frac{c^2}{N}}$, that is bounded by

$$\frac{\frac{c_0}{M^{0.5}(T-T_0)^{0.5}}+c}{N^{0.5}}. \tag{36}$$

Such a result suggests that we cannot exactly estimate $\overline{\mathrm{Q}}$ by only increasing M or $T-T_0$, but we need $N\to\infty$ to have null SE. This is intuitively true, since we cannot describe all possible scenarios by only increasing M or $T-T_0$ (that is related with Ω_1), but we need to explore different stochastic (and non-ergodic) scenarios by sampling on Ω_2, that is increasing N.

The dominant term in Eq. (32) depends on the decay rate of the transient perturbations $A_{n,m}$. If the transient perturbation decays fast enough, Eq. (32) simplifies to

$$C_\phi \mathrm{SE} \leq \varepsilon, \tag{37}$$

which can be estimated as

$$C_\phi\sqrt{\mathrm{V}^K\left[\frac{1}{NM}\sum_{n=1}^{N}\sum_{m=1}^{M}\langle \mathrm{Q}_{n,m}\rangle_{T_0,T}\right]} \leq \varepsilon, \tag{38}$$

with K sufficiently large.[3] On the other hand, if the decay of the transient perturbation is slower than Eq. (24), the left-hand side of Eq. (32) will decay like $(T-T_0)^{-q}$ for some $q<1$.

In the numerical experiments of Section 4, we verify that the initialization bias decays fast enough with respect to the burn-in time T_0 we set, and therefore that Eqs. (36) and (37) hold. Insights for choosing T_0 are given in Section 3.4.

We consider now the simplified situation in which there are not non-ergodic uncertainties, that is $\Omega_2\equiv\emptyset$. It follows that the covariance term of the right hand side of Eq. (35) is null. If the decay of the transient perturbation is slower than Eq. (24), the left-hand side of Eq. (32) decays like $(T-T_0)^{-q}$ for some $q<1$. If the decay of the transient perturbation is fast (Eq. (24) holds), the SE decays like $M^{-0.5}$ and $(T-T_0)^{-0.5}$ and can be estimated as

$$C_\phi\sqrt{\mathrm{V}^K\left[\frac{1}{M}\sum_{m=1}^{M}\langle \mathrm{Q}_{m}\rangle_{T_0,T}\right]} \leq \varepsilon, \tag{39}$$

where we omit the dependencies on the non-ergodic random variable.

3.3. On the generation of initial fields

Ensemble averaging benefits from independent initial conditions to generate uncorrelated flow evolutions. It is known that different turbulent flows will diverge with a rate determined by the Lyapunov exponent (Nikitin, 2009; Nastac et al., 2017), and that this is the case of our target problems. We decide then to generate perturbed initial conditions, and to let the system evolve for a defined burn-in time T_0 to arrive at uncorrelated solutions.

Our approach consists of adding nonlocal *spatially-correlated* and *divergence-free* solenoidal noise to the averaged velocity field $\bar{\boldsymbol{u}}$; we refer to this as the spatially correlated (SC) approach. Exploiting solenoidal fluctuations in the initial conditions is not new; we refer for example to Krasnopolsky (2018), where the author used uncorrelated divergence-free initial conditions to ensure independence of different realizations. Our novelty is that we propose to generate spatially-correlated fluctuations $\boldsymbol{w}(\boldsymbol{x})$, which arise from a well-established synthetic turbulence model.

Our approach is inspired by the work of Hunt in Hunt (1973) (see also Keith et al., 2021a). The underlying assumption is that the inhomogeneous contributions to fully developed turbulence fluctuations in the inviscid source layer above a solid body have negligible vorticity. From this assumption, one arrives at the following inhomogeneous turbulent fluctuation model: $\boldsymbol{w}(\boldsymbol{x})=\boldsymbol{w}^{(\mathrm{H})}(\boldsymbol{x})-\nabla\phi(\boldsymbol{x})$, where $\boldsymbol{w}^{(\mathrm{H})}(\boldsymbol{x})$ is a homogeneous turbulent velocity field and $\phi(\boldsymbol{x})$ satisfies

$$\Delta\phi=\nabla\cdot\boldsymbol{w}^{(\mathrm{H})} \quad \text{in } D, \quad (\nabla\phi-\boldsymbol{w}^{(\mathrm{H})})\cdot\boldsymbol{n}=0 \quad \text{on } \partial D. \tag{40}$$

In this work, we adopt the classical von Kármán model (Von Kármán, 1948) for the homogeneous random field $\boldsymbol{w}^{(\mathrm{H})}(\boldsymbol{x})$.

Realizations of this type of nonlocal spatially-correlated field can be generated using a Fourier transform on a Cartesian grid containing D; see, e.g., Mann (1998). Once a realization $\boldsymbol{w}^{(\mathrm{H})}(\boldsymbol{x})$ is generated, we may interpolate the boundary conditions so that the solution to Eq. (40) can be solved with the same finite element spaces used to solve the NS problem. After interpolating the sum $\boldsymbol{w}^{(\mathrm{H})}(\boldsymbol{x})-\nabla\phi$, we arrive at the nonlocal spatially-correlated perturbation $\boldsymbol{w}(\boldsymbol{x})$. We finally scale the perturbation to ensure that its norm is 1% of the norm of the mean velocity, which is taken as the average velocity obtained in preliminary simulations. Then, the SC initial condition used is $\boldsymbol{u}_0(\boldsymbol{x})=\bar{\boldsymbol{u}}(\boldsymbol{x})+\boldsymbol{w}(\boldsymbol{x})$.

In Tosi et al. (2021c), both SC and spatially-uncorrelated initial conditions are developed and considered. It is demonstrated that the two strategies give very close results. In this work, we prefer to use SC initial conditions because they are more physically meaningful.

3.4. On the optimal choice of the burn-in time

Given the full time interval $[0,T]$, we split it into a burn-in time interval $[0,T_0]$ and an effective time interval $[T_0,T]$. In this subsection, we focus on how to optimally choose T_0.

First, a single simulation is executed for a time long enough to reach a statistically stationary turbulent state, which is saved. Thereafter, N realizations are run with SC (or other strategies) initial conditions to ensure independent flow evolution. Once the required transient time T_0 is passed, statistical data are collected and updated on the fly, until the end of the effective time window.

We propose a systematic manner to minimize T_0, which makes use of the SE defined above. Given N realizations and a quantity of interest Q, our idea is to analyze how the statistical estimates of the QoI change for different burn-in times. We can observe this plotting the mean $\mathrm{E}^N[\langle \mathrm{Q}\rangle_{T_0,T}]$ as function of T_0, together with its confidence intervals. The time interval $T-T_0$ is kept constant, while varying T_0. The confidence intervals are computed as $C_\phi\mathrm{SE}$, with confidence $1-\phi$. By looking at the plot, we can detect a starting point after which the statistical result is effectively insensitive to T_0 variations. In addition to the statistical checks, we decide to apply a physical constraint, which in our case is the time the flow needs to go from the inlet to the obstacle. Therefore, T_0 will be the *maximum* of these two time values. In order to further reduce the computational cost of the transient phase, we also explore the possibility of using larger time steps in $[0,T_0]$.

Another way to estimate T_0 is analyzed in Beyhaghi et al. (2018), where the authors choose a burn-in time which minimizes the estimated variance of the sample average estimator of the time average for a given signal. To do so, we average at each time step over all realizations, for different numbers of realizations N, and apply the procedure to the resultant time signal. As we will see in Section 4, both procedures give similar results.

[3] Notice that K is independent of N and M. Each repetition K requires a total of NM samples.

4. Numerical experiments

We consider the problem of wind flowing around the three-dimensional CAARC building, introduced in Section 2.1. As commented above, three different situations are considered and are summarized next.

- The first scenario considers a constant in time fixed mean wind profile as boundary condition. The problem is solved with the standard time averaging method and the ensemble averaging method. The case is presented in Section 4.1.
- The second scenario considers a constant in time stochastic mean wind profile as boundary condition. The problem is solved with the ensemble-based MC method. The case is presented in Section 4.2.
- The third scenario considers turbulent fluctuations around a stochastic mean wind profile as boundary condition. The problem is solved with the ensemble-based MC method. The case is presented in Section 4.3.

In all cases, a comparison against literature is made, to verify the implementation. The quantities of interest are the drag force F_d on the body, the base moment M_b on the body and the pressure field $p(x)$ on all nodes of the body surface. The quantity of interest we choose to analyze is the drag force. Therefore we set $\mathrm{Q} \equiv F_d$. We omit the subscript T_0, T if there is no risk of misunderstanding. We remark that international units are used to measure physical quantities.

We shall remark that the mesh remains relatively coarse with respect to the resolution that would be needed to resolve the flow at the Reynolds number of interest. This situation is often encountered in the field of wind engineering (Tamura, 2008). The VMS stabilization, which is used as basis of our solver, represents an alternative to classical LES approaches and provides a simple yet effective turbulence model for the applications of interest, as explained in Section 2.2.

The computational efficiency of the joint use of XMC, Kratos and PyCOMPSs has already been demonstrated, and optimal strong scalability was ensured up to 128 nodes (6144 central processing units (CPUs)) (Tosi et al., 2021a).

4.1. CAARC system with constant in time deterministic boundary conditions

The analyses of this subsection are run on MareNostrum 4. This supercomputer has 11.15 Petaflops of peak performance, which consists of 3456 compute nodes equipped with two Intel R Xeon Platinum 8160 (24 cores at 2.1 GHz each) processors.

4.1.1. Problem formulation

The problem is the wind flow around the CAARC building, introduced and modeled in Section 2.1. In this subsection we use constant in time fixed mean wind field boundary conditions. Therefore, there are not non-ergodic random variables and for simplicity we write $\mathrm{E}^M[\mathrm{Q}] \equiv \mathrm{E}^{N,M}[\mathrm{Q}]$. In Table 1 we present the physical properties of the problem, and we set $z_0 = 2\,\mathrm{m}$ for this system. We point out that such a value is typical of centers of large cities (Joint committee on structural safety).

The selected tolerance and confidence for solving the stochastic problem in Section 4.1.5 are $\varepsilon = 50000$ and $1 - \phi = 99\%$. The relative value of the tolerance, with respect to the drag force mean value, is 0.55%.

The mesh considered to solve the problem has approximately 312000 nodes, and a minimal size, close to the body, of 0.2 m. The Courant–Friedrichs–Lewy (CFL) number, is 100. The considered mesh is adaptive with respect to a metric built on top of velocity and pressure fields. We refer for example to Alauzet and Loseille (2016), Dapogny et al. (2014), Tosi et al. (2021c) for details about metric-based adaptive mesh refinement.

Table 2
Normalized force and moment coefficients mean values of current problem, compared with literature. Time averages of our simulation are estimated for an effective time window $T - T_0 = 600\,\mathrm{s}$. The coefficients are dimensionless quantities.

References	$\langle C_{F_X} \rangle$	$\langle C_{F_Y} \rangle$	$\langle C_{M_X} \rangle$	$\langle C_{M_Y} \rangle$
Problem 4.1	1.563	0.004	−0.002	0.607
(Braun and Awruch, 2009)	1.660	0.008	0.004	0.570

4.1.2. Validation

To ensure correctness of the solver and of the mesh, we compute the following normalized coefficients (Braun and Awruch, 2009)

$$C_{F_X} = \frac{F_X}{1/2\rho W \int_0^H \bar{\boldsymbol{u}}^2 \,\mathrm{d}Z} \quad C_{F_Y} = \frac{F_Y}{1/2\rho W \int_0^H \bar{\boldsymbol{u}}^2 \,\mathrm{d}Z}$$
$$C_{M_X} = \frac{M_X}{1/2\rho \quad \bar{\boldsymbol{u}}_H^2 W H^2} \quad C_{M_Y} = \frac{M_Y}{1/2\rho \bar{\boldsymbol{u}}_H^2 W H^2}, \tag{41}$$

where ρ is the density of the fluid, $\bar{\boldsymbol{u}}$ the velocity of the mean profile, $\bar{\boldsymbol{u}}_H$ the velocity at height $H = 180\,\mathrm{m}$, W the building width and F_X, F_Y, M_X and M_Y the forces and moments computed on directions X and Y, respectively. X is the direction parallel to the ground and orthogonal to the inlet, while Y is parallel to both the ground and the inlet. Fig. 4.1 shows the normalized coefficients for the drag force and base moment for a 600 s simulation of the CAARC building with a steady inlet. The signals are compared against the results from Braun and Awruch (2009). Also, a comparison of the time-averaged normalized coefficients is shown in Table 2. This quantitative comparison of the mean of the normalized coefficients show a good agreement between the current problem and literature.

4.1.3. Comparison ensemble average and time average

We know from Tosi et al. (2021c) that perturbing initial conditions with spatially uncorrelated or SC noise is equivalent, from both computational and statistical points of view. For this reason, we prefer to use the latter, since more consistent from a physical point of view.

First, we analyze in Fig. 4.2 which are the dominant terms of Eq. (32) by plotting $\left(\mathrm{V}^K[\mathrm{E}^M[\langle \mathrm{Q} \rangle_{T_0,T}]]\right)^{-1}$ as function of $T - T_0$. We consider different K and M and $\mathrm{Q} = F_d$. A linear decay of the variance with respect to both M and $(T - T_0)$ can be observed. In fact, $\left(\mathrm{V}^K[\mathrm{E}^M[\langle \mathrm{Q} \rangle_{T_0,T}]]\right)^{-1}$ grows linearly as M or $(T - T_0)$ increase. If we compare the left and the right plots of Fig. 4.2, we see that $\left(\mathrm{V}^K[\mathrm{E}^M[\langle \mathrm{Q} \rangle_{T_0,T}]]\right)^{-1}$ grows by a factor M, for the same $T - T_0$ value. The linear decay of the variance with respect to the effective time can be observed on each plot independently from the other. The linear decay of the variance estimation with respect to M and $(T - T_0)$ suggests that the dominant term of the total error is $C_\phi \sqrt{\mathrm{V}^K[\mathrm{E}^M[\langle \mathrm{Q} \rangle_{T_0,T}]]}$, which implies simplifying Eq. (32) to $C_\phi \mathrm{SE} \approx C_\phi \sqrt{\mathrm{V}^K[\mathrm{E}^M[\langle \mathrm{Q} \rangle_{T_0,T}]]} \leq \varepsilon$.

The statistical error SE as a function of the size of the ensemble (for a constant effective time) is shown in Table 3. On the contrary, Table 4 gives the evolution of the SE as a function of the effective time. As expected, the SE decreases as more realizations or larger time windows are considered. We can observe that the ensemble average approach drastically reduces the time to solution, for the same statistical error (provided that more computational resources are available). For example, the case $M = 4$, $T - T_0 = 210\,\mathrm{s}$ of the ensemble average approach, compared to $M = 1$, $T - T_0 = 840\,\mathrm{s}$ of standard time average, reduces the time to solution by almost a factor 4, while guaranteeing a similar SE (consistent between the two approaches).

4.1.4. On the reduction of burn-in time computational cost

We analyze now if it is statistically consistent to reduce the burn-in time. As before, we consider 128 realizations and we keep constant $T - T_0 = 110\,\mathrm{s}$. Fig. 4.3 shows that, above a minimal threshold of the order of 30 s, the expected value estimation is largely independent of

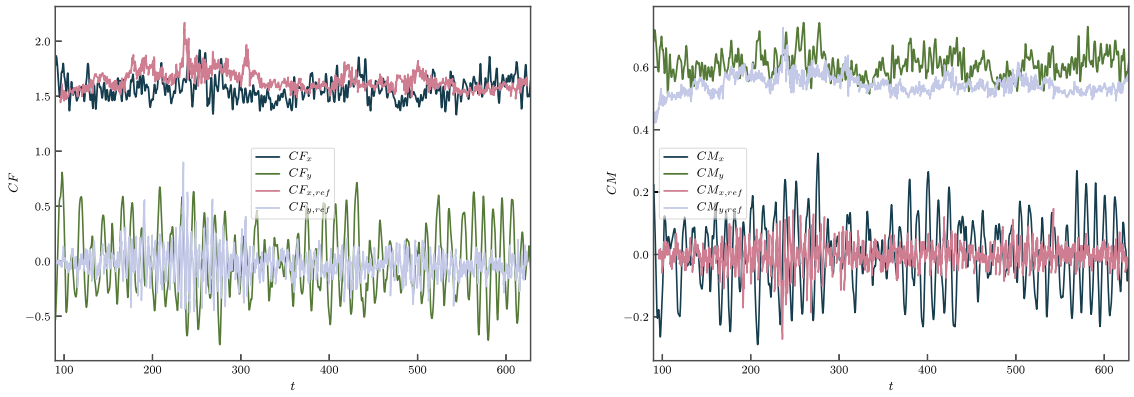

Fig. 4.1. Drag (C_F) and moment (C_M) coefficients comparison between our work and Braun and Awruch (2009). The plot is done using the web based tool https://github.com/ankitrohatgi/WebPlotDigitizer to extract numerical data from plot images. The coefficients are dimensionless quantities.

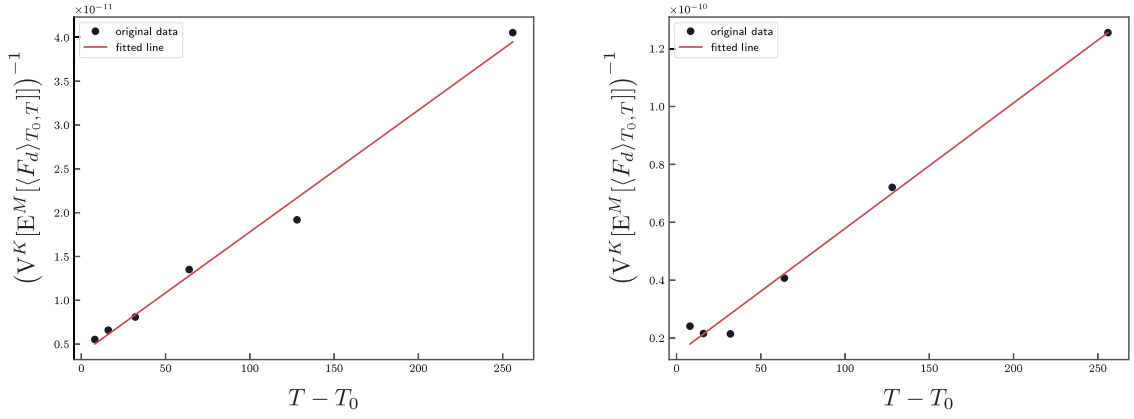

Fig. 4.2. Computation of $\mathrm{V}^K[\mathrm{E}^M[\langle \mathrm{Q}\rangle_{T_0,T}]]$ as function of the effective time $T-T_0$ for $\mathrm{Q} = F_d$. The left plot presents $(K, M, T_0) = (128, 1, 30\,\mathrm{s})$ and the right plot $(K, M, T_0) = (32, 4, 30\,\mathrm{s})$. Time is measured in seconds and the drag force in Newtons.

Table 3
The table reports the mean estimation and its associated SE of ensemble averaging for the estimation of the drag force mean (measured in Newtons). M and $T - T_0$ refer to the number of realizations and the effective time window of the simulation (measured in seconds), respectively. C is the computational cost and is expressed in CPU hours, while time to solution is expressed in hours.

$\mathrm{E}^M[\langle F_d\rangle]$	SE	M	$T - T_0$	C	Time to solution
8982493	84767	4	210	1666	17.36
8932223	59939	8	210	3333	17.36
8973444	42383	16	210	6666	17.36
8986913	29969	32	210	13332	17.36
8927955	21191	64	210	26664	17.36
8930547	14984	128	210	53329	17.36

Table 4
The table reports the mean estimation and its associated SE of standard time averaging for the estimation of the drag force mean (measured in Newtons). M and $T-T_0$ refer to the number of realizations and the effective time window of the simulation (measured in seconds), respectively. C is the computational cost and is expressed in CPU hours, while time to solution is expressed in hours.

$\mathrm{E}^M[\langle F_d\rangle]$	SE	M	$T - T_0$	C	Time to solution
8658884	324256	1	52.5	215.58	8.98
8879404	229284	1	105	300.19	12.50
9109719	162128	1	210	468.56	19.52
9003445	114642	1	420	853.28	35.55
8950303	93604	1	630	1189.42	49.55
8956216	81064	1	840	1524.88	63.53

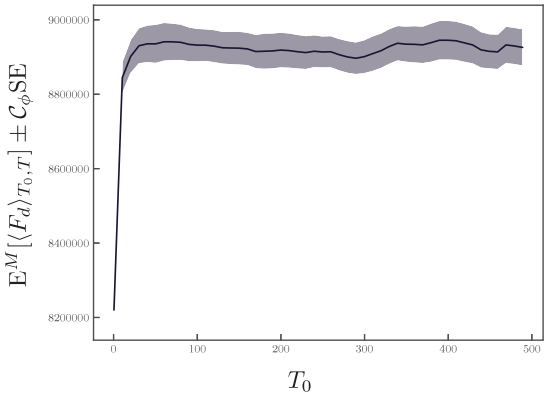

Fig. 4.3. Expected value estimation and associated SE for a confidence of 99% as a function of the burn-in time. The burn-in time is measured in seconds and the drag force in Newtons. The plot is computed with $M = 128$, $T - T_0 = 110\,\mathrm{s}$.

the burn-in time. Fig. 4.4 suggests that a similar result holds when the ensemble average approach is employed. The autocorrelation of the computed drag force is of the order of 20 s.

To ensure robustness of our strategy, we also consider the more restrictive "heuristic" constraint that T_0 should be larger than the time needed to travel from the inlet to the body. Since for an average velocity of $40\,\mathrm{m\,s^{-1}}$ such time is 11.625 s, it is statistically consistent to use $T_0 = 30\,\mathrm{s}$.

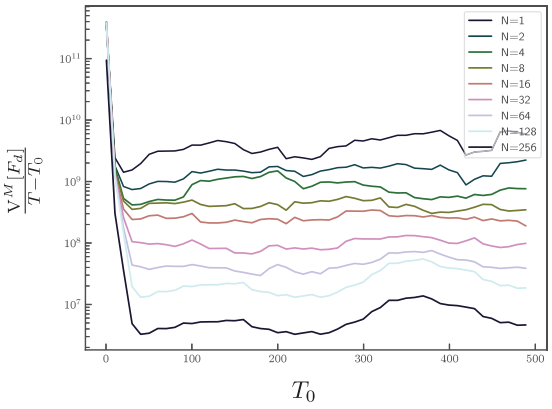

Fig. 4.4. Ratio between drag force variance (F_d measured in Newtons) and effective time window, for different realizations and $T - T_0 = 110\,\mathrm{s}$. The burn-in time T_0 is measured in seconds.

Table 5
Expected value and associated SE with 99% confidence for different time steps during burn-in time. F_d is measured in Newtons, $T - T_0$ in seconds and $\frac{\Delta t_0}{\Delta t}$ is a dimensionless quantity.

$\mathrm{E}^M[\langle F_d\rangle_{T_0,T}]$	C_ϕSE	$T - T_0$	$\frac{\Delta t_0}{\Delta t}$
8922399	39013	170	1.0
8907406	39534	170	2.5

Table 6
The table reports the expected value and the statistical error values of the time-averaged drag force $\langle F_d\rangle_{T_0,T}$ (measured in Newtons), with a 99% confidence. M, T and T_0 refer to the number of ensemble realizations, the time window $[0,T]$ upper bound of the simulation and the burn-in time, respectively. These last two are measured in seconds. $\frac{\Delta t_0}{\Delta t}$ is the ratio between the time steps of T_0 and of the effective time window $T - T_0$. C is the computational cost, expressed in CPU hours, and time to solution is the real time we need to wait for the solution and is expressed in hours.

$\mathrm{E}^M[\langle F_d\rangle_{T_0,T}]$	$\frac{C_\phi \mathrm{SE}}{\mathrm{E}^M[\langle F_d\rangle]}$	M	$T - T_0$	T_0	$\frac{\Delta t_0}{\Delta t}$	C	Time to solution
8972727	0.545%	142	110	30	2.5	35749	10.34
8946768	0.511%	100	170	30	2.5	38041	15.54
8943515	0.455%	76	230	30	2.5	38566	20.60

Finally we explore the possibility of employing larger time steps during the burn-in time, in order to reduce its computational cost. Table 5 shows that running with a larger time step during T_0 is statistically equivalent to employing a constant time step, where C_ϕ is computed for a 99% confidence. This proves that such an approach is convenient whenever possible.

4.1.5. Results

Finally, we run the problem exploiting ensemble average, larger time step Δt_0 and smaller burn-in time $T_0 = 30\,\mathrm{s}$. Convergence is checked with Eq. (32), which is simplified to $C_\phi \mathrm{SE} \leq \varepsilon$. The chosen confidence is 99%, and the relative tolerance with respect to the time-averaged drag force mean estimator is around 0.5%. We run the problem for different configurations, keeping the overall cost constant, that is maintaining approximately constant the product between time window and the number of realizations TM. Results are shown in Table 6.

We report in Fig. 4.5 the instantaneous velocity field at $t = 200\,\mathrm{s}$ for one realization. We refer to Tosi et al. (2021c, figure 8.11) to observe the pressure field for the same problem.

4.1.6. Other observables

We select the case with minimal statistical error of Table 6 to show the statistical results for other quantities of interest. Table 7 shows the expected value and the standard deviation estimators of the drag force and the base moment. Fig. 4.6 shows the estimation of the expected value and the standard deviation of the pressure field. We refer to Tosi et al. (2021c, figure 8.12) to observe the estimation of the expected value and the standard deviation of the time-averaged pressure field. Moreover, we refer to Tosi et al. (2021c, table 8.13) to observe the conditional value at risk results for the time-averaged drag force and the drag force of the case with minimal statistical error of Table 6.

Table 7
Statistical analysis of drag force (measured in Newtons) and base moment (measured in Newton meters).

Q	$\mathrm{E}^M[\langle \mathrm{Q}\rangle_{T_0,T}]$	$\sigma^M[\langle \mathrm{Q}\rangle_{T_0,T}]$	$\sigma^M[\mathrm{Q}]$
F_d	8943515	152726	662129
M_b	−14943	436540	7332992

4.1.7. Best configuration

The objective of this section is to find the configuration which minimizes the wall clock time to solve the fluid flow past the CAARC building problem with constant in time deterministic boundary conditions. By analyzing Tables 3 and 4 we observe that exploiting an ensemble of multiple independent realizations is more convenient than a single long-term time averaging. In Table 6 we see that we should increase as much as possible the number of ensemble realizations M, so that we can reduce the effective time of the simulation $T - T_0$. Table 5 shows that using a larger time step during the burn-in time reduces the computational cost of the simulation (since the time integration is cheaper) and does not alter the outcome of the problem. Therefore, the best method to solve the fluid flow past the CAARC building problem with constant in time deterministic boundary conditions is ensemble averaging, together with the described strategies which reduce the burn-in phase computational cost.

4.2. CAARC system with constant in time stochastic boundary conditions

The analyses of this subsection are run on Salomon. This cluster presents 2 Petaflops of peak performance and is made by 1008 compute nodes. Each node is equipped with an x86-64 24 cores CPU (two 2.5 GHz twelve-core Intel Xeon processors) and 128 GB RAM.

4.2.1. Problem formulation

We solve the wind flow past the CAARC building, that is introduced and modeled in Section 2.1. In this subsection we use constant in time stochastic mean wind field boundary conditions, which means we consider null fluctuations $\boldsymbol{u}' \equiv 0$ and a stochastic roughness height, modeled as $z_0 \sim \mathcal{U}(0.1, 0.7)$. This scenario is typical of sparsely built-up urban areas, suburbs and wooded areas (Joint committee on structural safety). Apart from the roughness height, other physical properties are reported in Table 1.

As above, the QoI for which we assess the failure probability criterion of Eq. (14) is the drag force F_d. The selected tolerance and confidence for solving the stochastic problem are $\varepsilon = 110000$ and $1 - \phi = 99\%$. The relative value of the tolerance, with respect to the drag force mean value, is 1.15%.

The mesh we use is adaptive with respect to a metric built on top of velocity and pressure fields and has approximately 312000 nodes. We refer for example to Alauzet and Loseille (2016), Dapogny et al. (2014), Tosi et al. (2021c) for details about metric-based adaptive mesh refinement. The minimal size, close to the body, is 0.2 m and the CFL number is 100. Since in this subsection we consider a constant in time wind inflow, we refer to Section 4.1.2 for the validation.

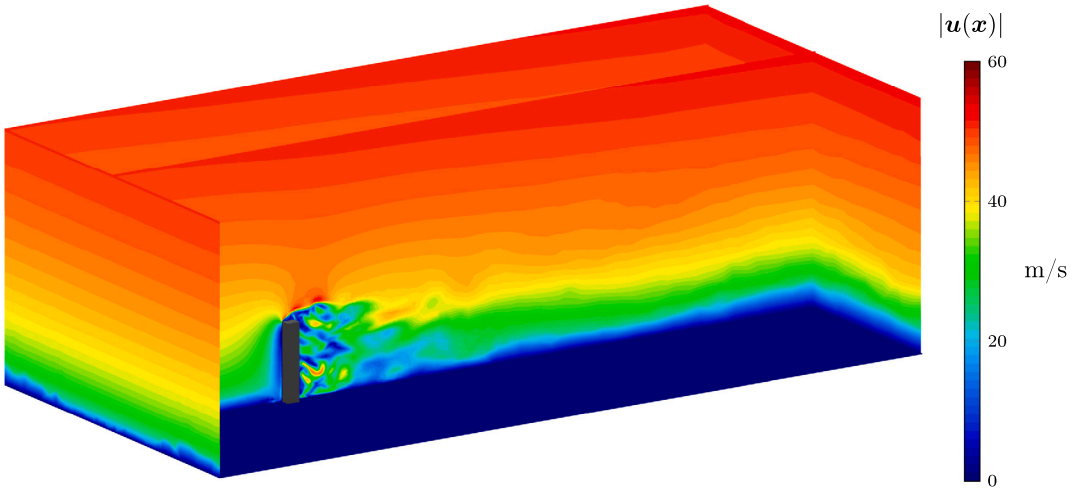

Fig. 4.5. Velocity field snapshot at $t = 200\,\mathrm{s}$, post-processed using the https://www.gidsimulation.com/ software.

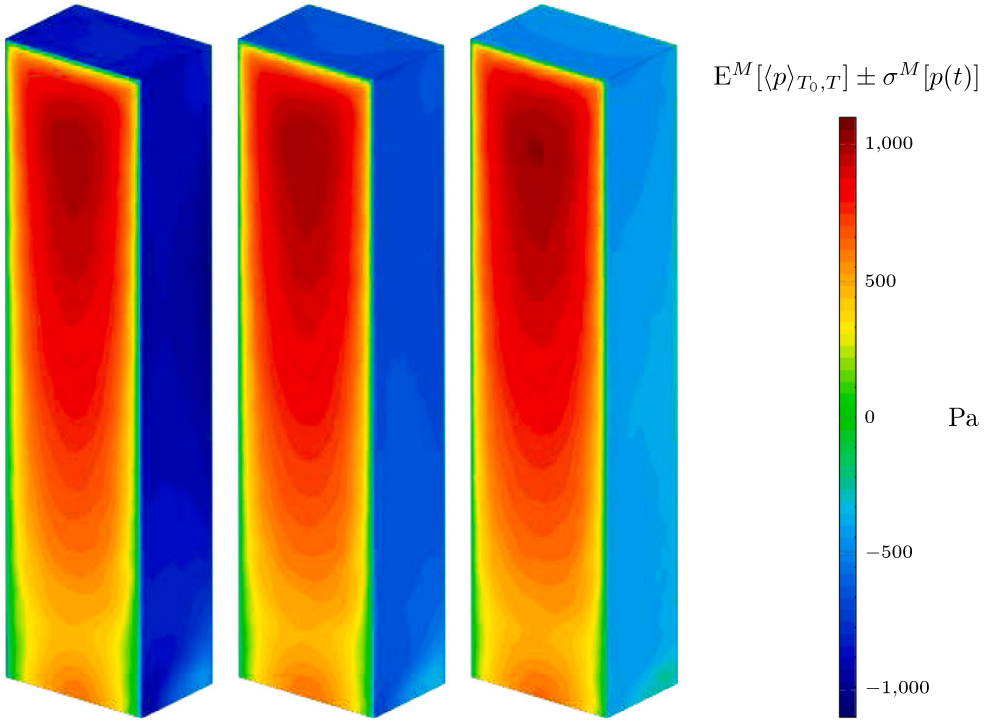

Fig. 4.6. Statistical result of the pressure field $p(\boldsymbol{x})$. From left to right, $\mathrm{E}^M[\langle p(\boldsymbol{x})\rangle_{T_0,T}] - \sigma^M[p(\boldsymbol{x})]$, $\mathrm{E}^M[\langle p(\boldsymbol{x})\rangle_{T_0,T}]$ and $\mathrm{E}^M[\langle p(\boldsymbol{x})\rangle_{T_0,T}] + \sigma^M[p(\boldsymbol{x})]$.

4.2.2. Results

First, we verify that for $T_0 = 30\,\mathrm{s}$ the initialization bias is negligible, where $T_0 = 30\,\mathrm{s}$ is the optimal burn-in time we obtain in Section 4.1. Figs. 4.7–4.9 show the decay of the statistical error in terms of the wind realizations N, the effective time of the simulation $T - T_0$, and the number of ensemble realizations M, respectively. The agreement with the fitted curve indicates that Eq. (36) holds, which allows simplifying Eq. (32) to Eq. (37).

Then, we solve the problem for a fixed total time NMT and we report results in Table 8. We remark that the product $M(T - T_0)$ is approximately equal to $600\,\mathrm{s}$ for each wind scenario. Looking at the results, we observe that exploiting multiple ensembles ($M > 1$) reduces the wall clock time, at the price of a slightly larger SE, for the same computational cost. Slightly larger SE values appear due to the fact that the effective time is smaller as M grows.

We present as well results where only the product NMT is kept constant and the constraint $M(T - T_0) \approx 600\,\mathrm{s}$ is relaxed in Table 9.

We report in Fig. 4.10 the instantaneous velocity field at $t = 200\,\mathrm{s}$ for one realization. We refer to Tosi et al. (2021c, figure 9.9) to observe the pressure field of the current problem. We point out that the main difference with respect to Fig. 4.5 is the roughness height, that here is equal to $0.3\,\mathrm{m}$.

4.2.3. Other observables

We select the case with $N = 20$, $M = 1$, $T - T_0 = 600\,\mathrm{s}$ to compute drag force, base moment (computed around the center of the bottom of the building) and pressure field around the building statistical estimators. Table 10 shows expected value and standard deviation estimations of the drag force and of the base moment. Fig. 4.11 shows the expected value and the standard deviation estimations of the pressure field. We

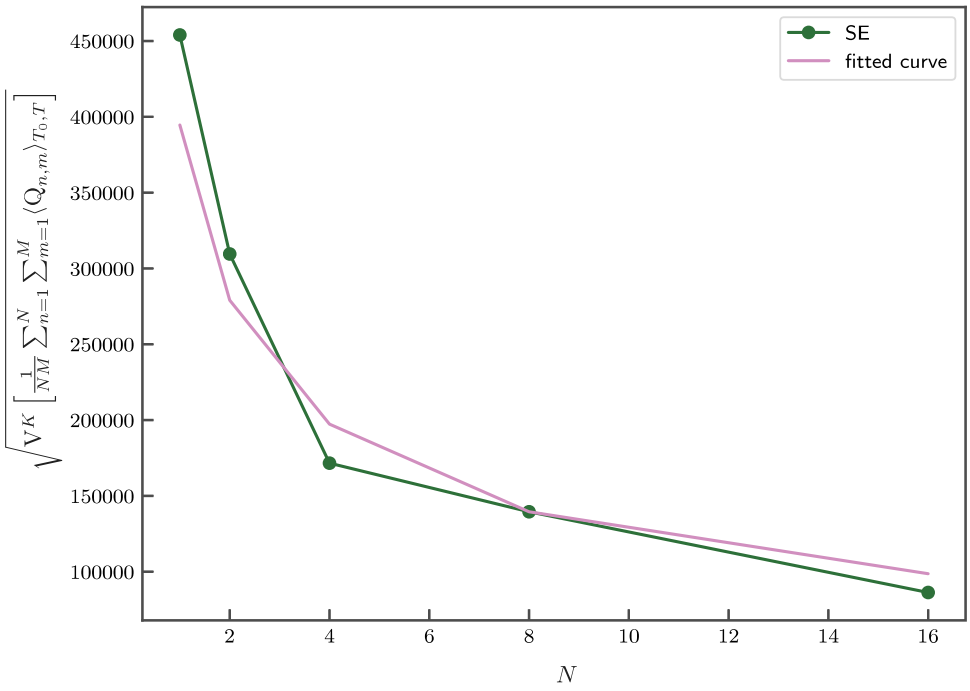

Fig. 4.7. Computation of SE as function of N for the drag force $Q \equiv F_d$ (measured in Newtons). $M = 1$ and $T - T_0 = 48.75$ s. The fitted curve is Eq. (36).

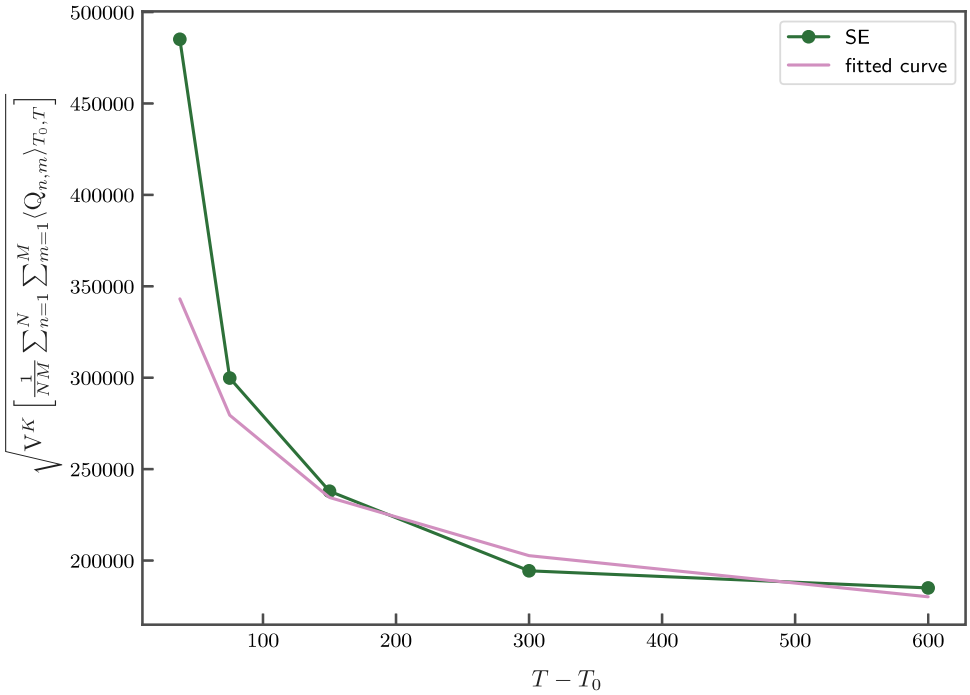

Fig. 4.8. Computation of SE as function of $T - T_0$ (measured in seconds) for the drag force $Q \equiv F_d$ (measured in Newtons). $K = 20$, $N = 1$ and $M = 1$. The fitted curve is Eq. (36).

refer to Tosi et al. (2021c, figure 9.10) to observe the estimation of the expected value and the standard deviation of the time-averaged pressure field. Moreover, we refer to Tosi et al. (2021c, table 9.8) to observe the conditional value at risk results for the time-averaged drag force and the drag force of the current problem.

4.2.4. Best configuration

The objective of this section is to find the configuration which minimizes the wall clock time to solve the fluid flow past the CAARC building problem with constant in time stochastic boundary conditions. By analyzing Table 8 we observe that the best method is using an ensemble of realizations M for each event N since we reach a wall clock time reduction up to a factor 10.

4.3. CAARC system with general boundary conditions

The analyses of this subsection are run on Salomon. This cluster presents 2 Petaflops of peak performance and is made by 1008 compute

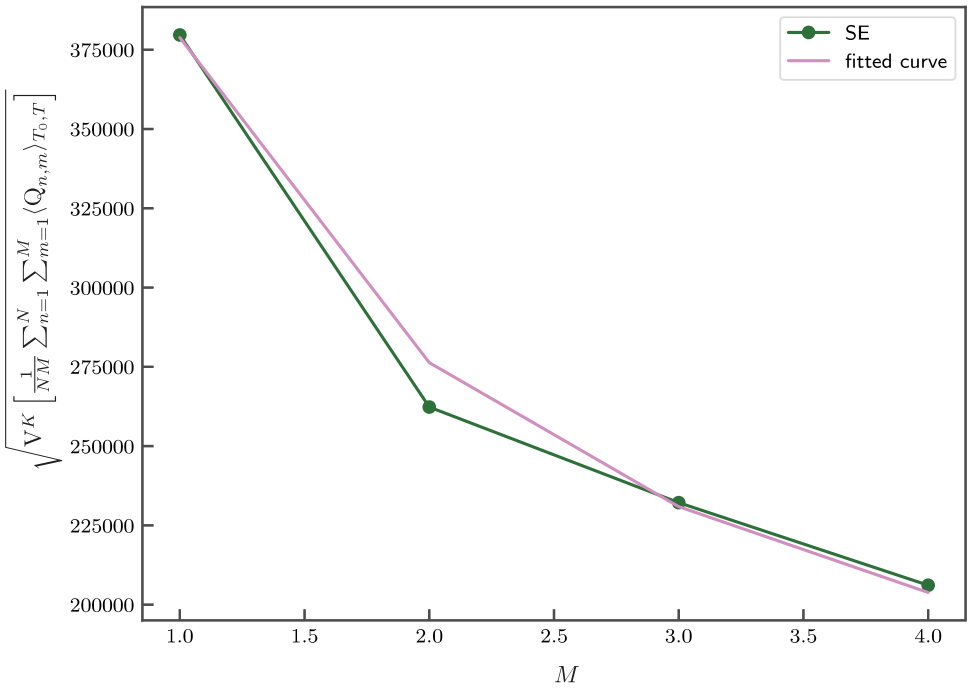

Fig. 4.9. Computation of SE as function of M for the drag force $Q \equiv F_d$ (measured in Newtons). $K = 40$, $N = 1$ and $T - T_0 = 75$ s. The fitted curve is Eq. (36).

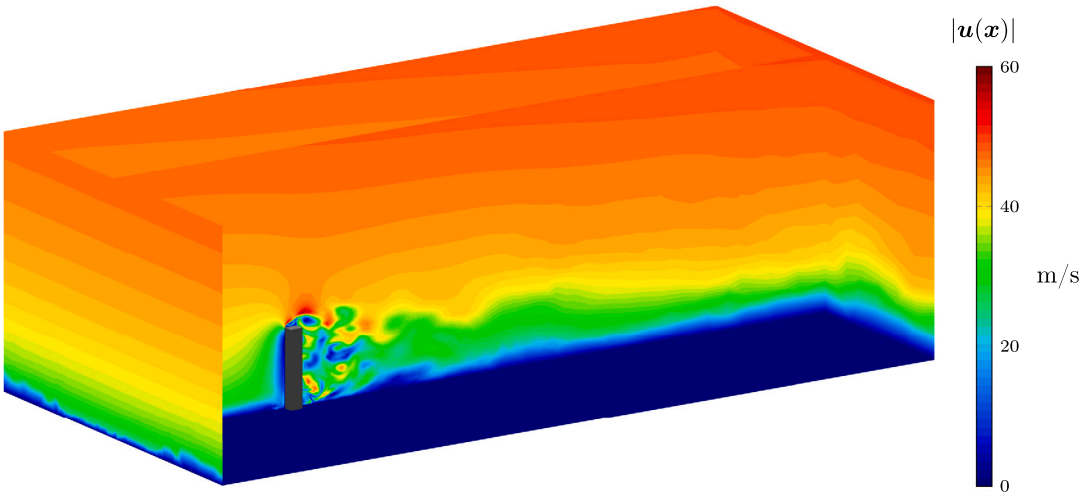

Fig. 4.10. Velocity field snapshot at $t = 200$ s.

Table 8
The table reports the drag force F_d expected value estimation (measured in Newtons) and its associated SE, with a 99% confidence. N, M, $T - T_0$ and T_0 refer to the number of wind realizations, the number of ensembles per wind scenario, the effective time window and the burn-in time, respectively. Times are expressed in seconds. C is the computational cost, expressed in CPU hours. Time to solution is the real time we need to wait for solving the problem, and it is expressed in hours. Results are sorted for decreasing time to solution.

$\mathrm{E}^{N,M}[\langle F_d \rangle]$	$\frac{C_\phi \mathrm{SE}}{\mathrm{E}^{N,M}[\langle F_d \rangle]}$	N	M	$T - T_0$	T_0	C	Time to solution
9632478	0.97%	20	1	600	30	10468	20.77
9748082	1.26%	20	2	285	30	10087	10.25
9652424	1.00%	20	4	127.5	30	10263	5.27
9598793	1.12%	20	8	48.75	30	10364	2.68

nodes. Each node is a powerful x86-64 computer equipped with 24 cores (two 2.5 GHz twelve-core Intel Xeon processors) and 128 GB RAM.

Table 9
The table reports the drag force F_d expected value estimation (measured in Newtons) and its associated SE, with a 99% confidence. N, M, $T - T_0$ and T_0 refer to the number of wind realizations, the number of ensembles per wind scenario, the effective time window and the burn-in time, respectively. Times are expressed in seconds. C is the computational cost, expressed in CPU hours. Time to solution is the real time we need to wait for solving the problem, and it is expressed in hours. Results are sorted for decreasing time to solution.

$\mathrm{E}^{N,M}[\langle F_d \rangle]$	$\frac{C_\phi \mathrm{SE}}{\mathrm{E}^{N,M}[\langle F_d \rangle]}$	N	M	$T - T_0$	T_0	C	Time to solution
9632478	0.97%	20	1	600	30	10468	20.77
9576676	0.82%	160	1	48.75	30	10319	2.67

4.3.1. Problem formulation

We solve the wind flow past the CAARC building, that is introduced and modeled in Section 2.1. In this subsection we use turbulent fluctuations around a stochastic mean wind field boundary conditions, which

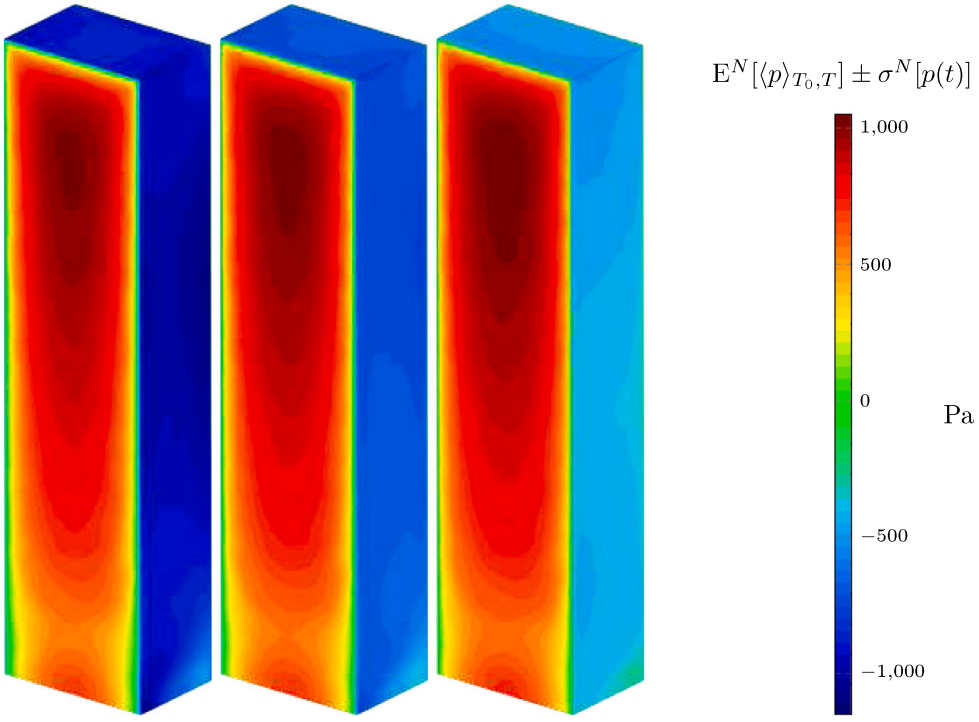

Fig. 4.11. Statistical analysis of the pressure field $p(x)$. From left to right, $\mathrm{E}^N[\langle p(x)\rangle_{T_0,T}] - \sigma^N[p(x)]$, $\mathrm{E}^N[\langle p(x)\rangle_{T_0,T}]$ and $\mathrm{E}^N[\langle p(x)\rangle_{T_0,T}] + \sigma^N[p(x)]$.

Table 10
Statistical analysis of the drag force F_d (measured in Newtons) and of the base moment M_b (measured in Newton meters), computed around the center of the CAARC building base.

Q	$\mathrm{E}^N[\langle \mathrm{Q}\rangle_{T_0,T}]$	$\sigma^N[\langle \mathrm{Q}\rangle_{T_0,T}]$	$\sigma^N[\mathrm{Q}]$
F_d	9632478	180333	783596
M_b	−3603	293732	7990526

Table 11
Normalized force and moment coefficients mean values of current problem, compared with literature. Time averages of our simulation are estimated for an effective time window $T - T_0 = 600$ s. The coefficients are dimensionless quantities.

References	$\langle C_{F_x}\rangle$	$\langle C_{F_y}\rangle$	$\langle C_{M_x}\rangle$	$\langle C_{M_y}\rangle$
Problem 4.3	1.818	0.001	−0.001	0.694
(Obasaju, 1992)	1.490	−0.039	0.000	0.640
(Huang et al., 2007)	1.830	0.006	–	–
(Braun and Awruch, 2009)	1.660	0.008	0.004	0.570

means that $\boldsymbol{u}'$ is not null and such fluctuations are modeled by the Mann model. Moreover, the roughness height is a uniformly distributed random variable $z_0 \sim \mathcal{U}(0.1, 0.7)$; a scenario typical of sparsely built-up urban areas, suburbs and wooded areas (Joint committee on structural safety). Other physical quantities are reported in Table 1.

The QoI for which we assess the failure probability criterion is the drag force F_d. The tolerance and confidence of the convergence criterion are $\varepsilon = 110000$ and $1 - \phi = 99\%$, respectively. The relative value of the tolerance, with respect to the drag force mean value, is around 1.10%.

The mesh we use is adaptively refined with respect to pressure and velocity fields. We refer for example to Alauzet and Loseille (2016), Dapogny et al. (2014), Tosi et al. (2021c) for details about metric-based adaptive mesh refinement. The mesh presents around 283 thousands nodes and a minimal size of 0.2 m close to the building. The chosen time step is 0.2375 s, which gives a CFL number of 100. Such a configuration is validated next in Section 4.3.2.

4.3.2. Validation

To ensure correctness of the solver and of the mesh, we compute the normalized coefficients of Eq. (41). We compare the time-averaged normalized coefficients with Obasaju (1992), Huang et al. (2007), Braun and Awruch (2009) in Table 11.

4.3.3. Results

We estimate the burn-in time following the approach presented in Section 3.4, which estimates T_0 on top of statistical and physical constraints. Fig. 4.12 shows that statistical results of the QoI are insensitive for $T_0 > 30$ s, that is larger than the time required by the wind to go from the inlet to the building.

Similarly to Section 4.2.2, we show in Figs. 4.13–4.15 the decay of the statistical error in terms of the wind realizations N, the effective time of the simulation $T - T_0$, and the number of ensemble realizations M, respectively. The agreement between the SE and the fitted curve indicates that Eq. (36) holds, which allows simplifying Eq. (32) to Eq. (37).

Results for a constant total time NMT and $M \approx (T - T_0) \approx 600$ s are presented in Table 12. We observe that the last case of the table drastically reduces the time to solution, at the price of a slightly larger SE, for the same computational cost. Therefore, $M > 1$ is the most promising configuration.

We consider also the scenario where the constraint $M(T - T_0) \approx 600$ s is relaxed and only NMT is kept constant. Results are reported in Table 13. We observe that both SE and time to solution are drastically reduced by increasing N and reducing M and $T - T_0$.

We report in Fig. 4.16 the instantaneous velocity field at $t = 200$ s for one realization and we refer to Tosi et al. (2021c, figure 9.17) to observe the instantaneous pressure field of this problem.

4.3.4. Other observables

We select the case with $N = 40$, $M = 1$, $T - T_0 = 600$ s to show statistical results of the physical quantities we compute, that are the drag force, the base moment (computed around the base of the building) and the pressure field around the building. Table 14 shows

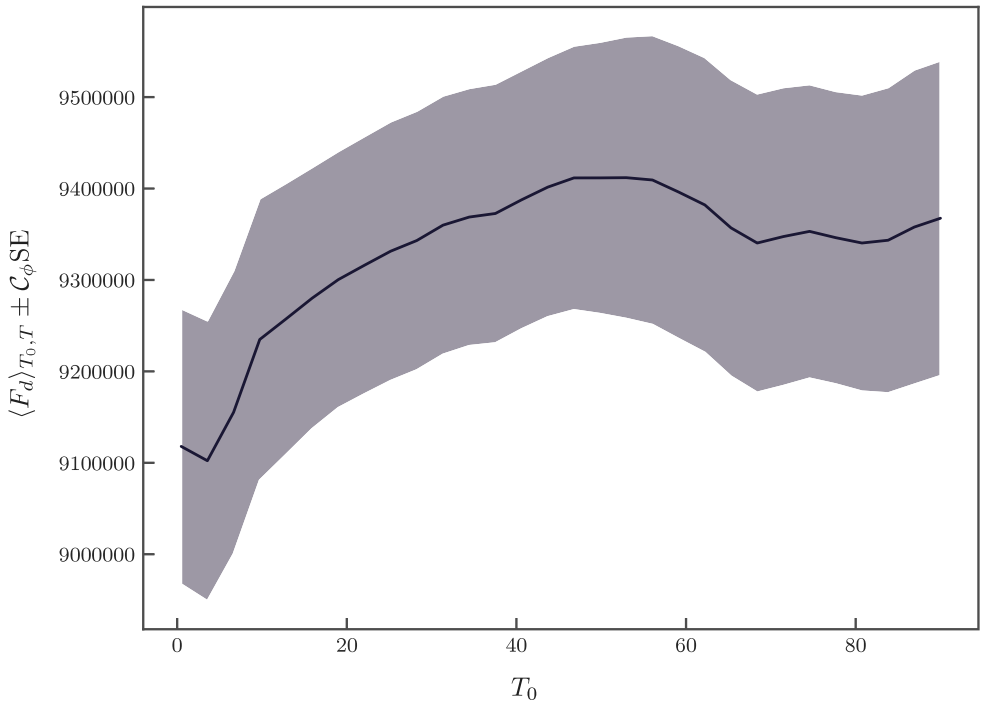

Fig. 4.12. Expected value estimation and associated SE with 99% confidence plotted as function of the burn-in time, with $M = 128$. The effective time $T - T_0 = 60\,\mathrm{s}$ is fixed among all realizations and all burn-in times. The drag force F_d is measured in Newtons and the burn-in time T_0 in seconds.

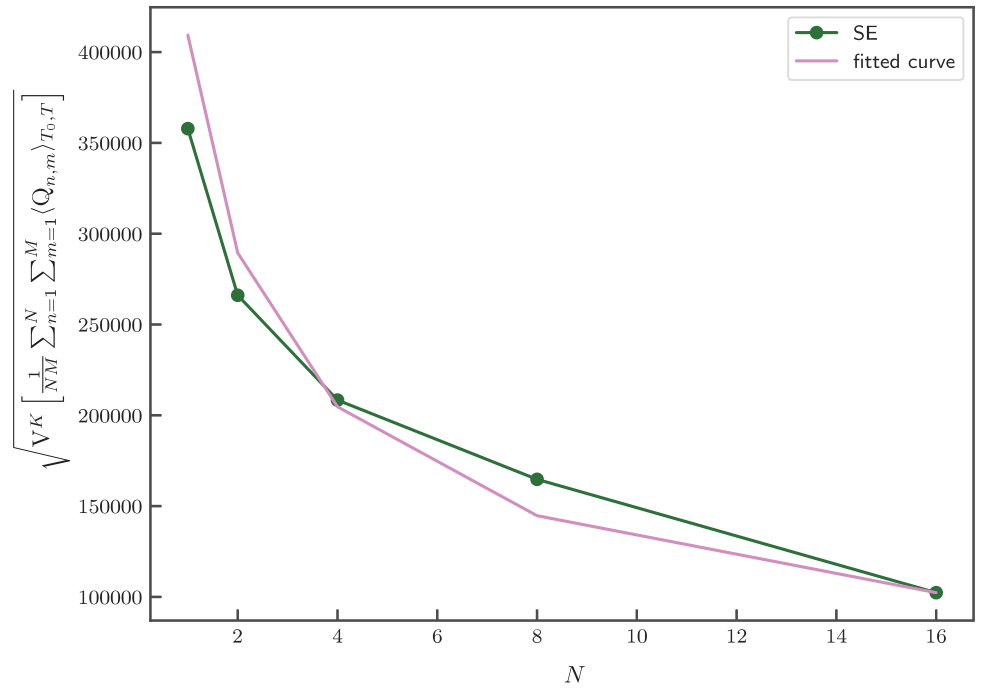

Fig. 4.13. Computation of SE as function of N for the drag force $\mathrm{Q} \equiv F_d$ (measured in Newtons). $M = 1$ and $T - T_0 = 75\,\mathrm{s}$. The fitted curve is Eq. (36).

expected value and standard deviation estimations of the drag force and the base moment. Fig. 4.17 shows the estimation of the expected value and the standard deviation of the pressure field. We refer to Tosi et al. (2021c, figure 9.18) to observe the estimation of the expected value and the standard deviation of the time-averaged pressure field. Moreover, we refer to Tosi et al. (2021c, table 9.12) to observe the conditional value at risk results for the time-averaged drag force and the drag force.

4.3.5. Best configuration

The objective of this section is to find the configuration which minimizes the wall clock time to solve the fluid flow past the CAARC

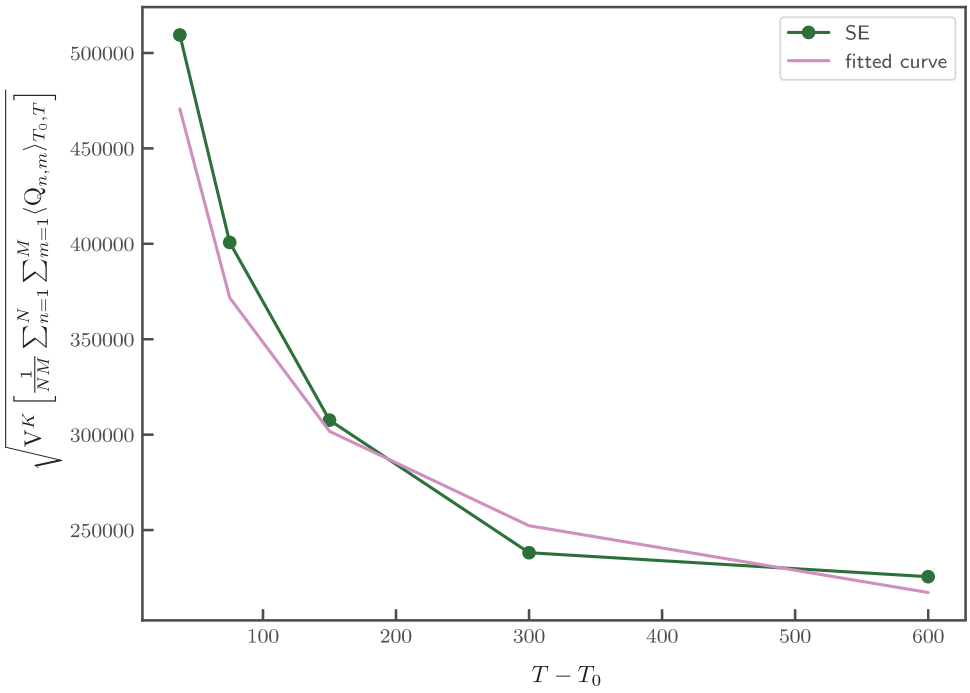

Fig. 4.14. Computation of SE as function of $T - T_0$ (measured in seconds) for the drag force $Q \equiv F_d$ (measured in Newtons). $K = 40$, $N = 1$ and $M = 1$. The fitted curve is Eq. (36).

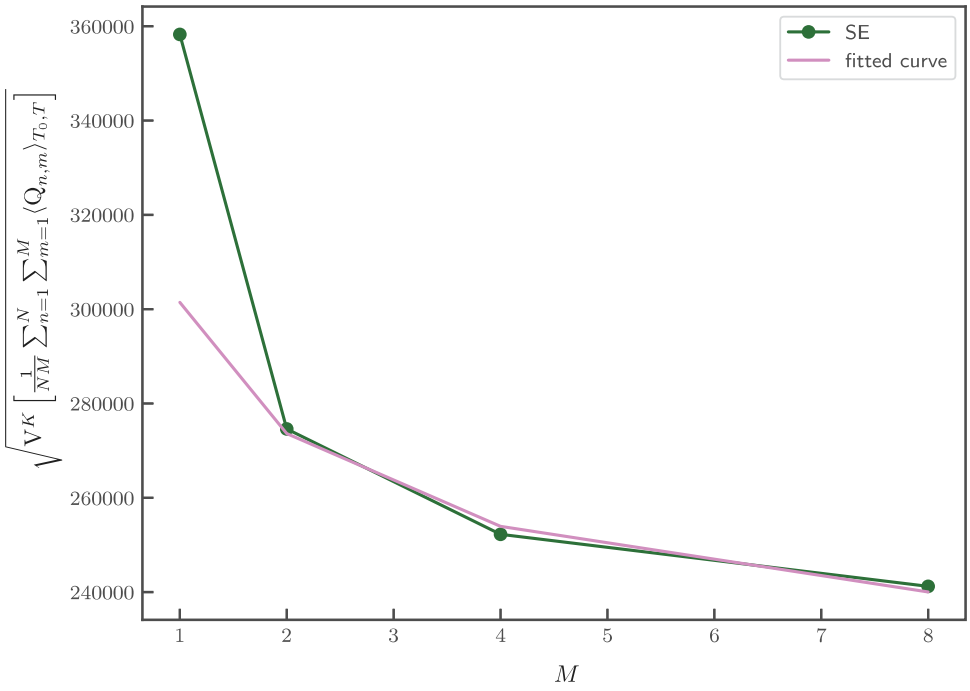

Fig. 4.15. Computation of SE as function of M for the drag force $Q \equiv F_d$ (measured in Newtons). $K = 40$, $N = 1$ and $T - T_0 = 75$ s. The fitted curve is Eq. (36).

building problem with general boundary conditions. By analyzing Table 12 we observe that the best method is using an ensemble of realizations M for each event N, since we reach a wall clock time reduction up to a factor 7.5.

5. Conclusions

As shown in the work, the statistical analysis of the use of ensemble averaging in the estimation of the expected value leads to the identification of two error components: an initialization bias, related to the transient perturbation of the flow, and a statistical error, related to finite sampling. Convergence rates of both error contributions are analyzed and discussed.

Decay rates are estimated for the target test cases to verify that the initialization bias is effectively negligible. The cost of the burn-in phase is minimized by following a statistical-based approach combined with the use of a less expensive time integration procedure. Multiple observables (drag force, base moment and pressure field) are computed.

We show that ensemble averaging can be successfully applied to highly chaotic incompressible flows and we propose strategies to minimize the total error and the computational cost of the simulation.

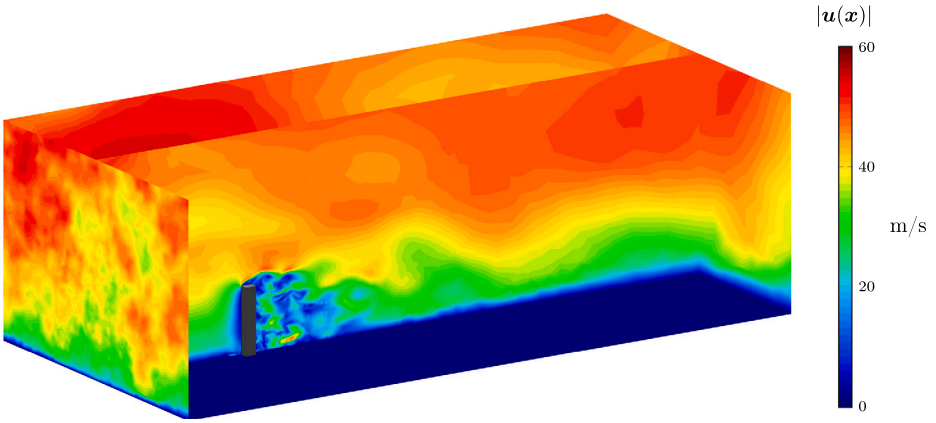

Fig. 4.16. Velocity field snapshot at $t = 200\,\mathrm{s}$.

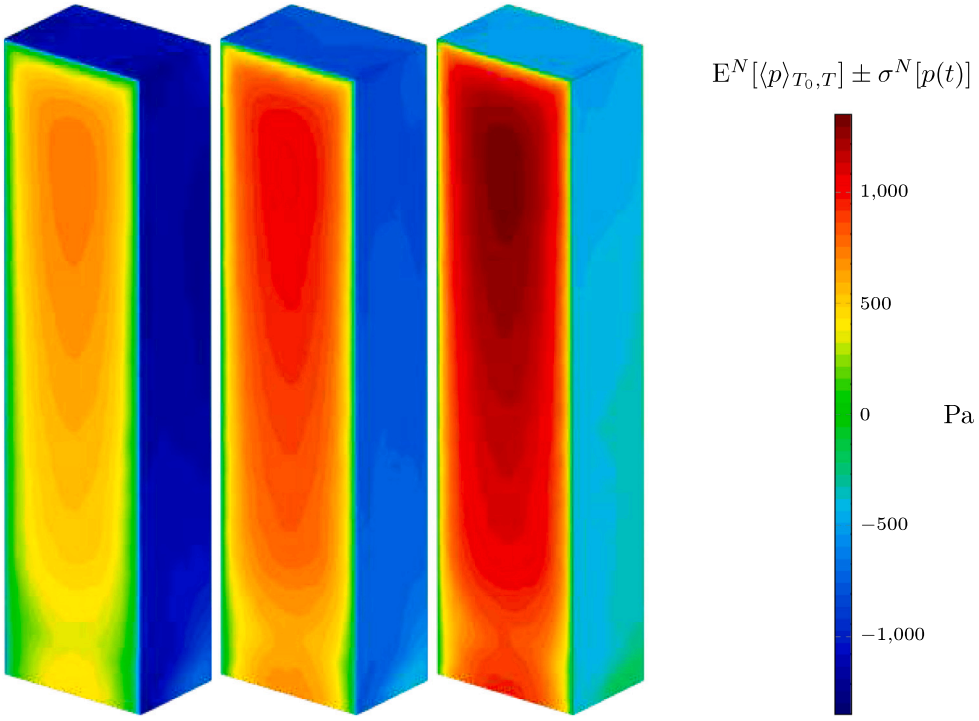

Fig. 4.17. Statistical analysis of the pressure field $p(x)$. From left to right, $\mathrm{E}^N[\langle p(x)\rangle_{T_0,T}] - \sigma^N[p(x)]$, $\mathrm{E}^N[\langle p(x)\rangle_{T_0,T}]$ and $\mathrm{E}^N[\langle p(x)\rangle_{T_0,T}] + \sigma^N[p(x)]$.

Table 12
The table reports the drag force F_d expected value estimation (measured in Newtons) and its associated SE, with a 99% confidence. N, M, $T-T_0$ and T_0 refer to the number of wind realizations, the number of ensembles per wind scenario, the effective time window and the burn-in time, respectively. Times are expressed in seconds. C is the computational cost, expressed in CPU hours. Time to solution is the wall clock time we need to wait for solving the problem, and it is expressed in hours. Results are sorted for decreasing time to solution.

$\mathrm{E}^{N,M}[\langle F_d\rangle]$	$\frac{C_\phi \mathrm{SE}}{\mathrm{E}^{N,M}[\langle F_d\rangle]}$	N	M	$T-T_0$	T_0	C	Time to solution
10219325	0.80%	40	1	600	30	19114	37.27
9987287	0.81%	40	2	285	30	18607	18.91
10266800	0.75%	40	4	127.5	30	18625	9.58
10115719	1.07%	40	8	48.75	30	18743	4.85

Table 13
The table reports the drag force F_d expected value estimation (measured in Newtons) and its associated SE, with a 99% confidence. N, M, $T-T_0$ and T_0 refer to the number of wind realizations, the number of ensembles per wind scenario, the effective time window and the burn-in time, respectively. Times are expressed in seconds. C is the computational cost, expressed in CPU hours. Time to solution is the wall clock time we need to wait for solving the problem, and it is expressed in hours. Results are sorted for decreasing time to solution.

$\mathrm{E}^{N,M}[\langle F_d\rangle]$	$\frac{C_\phi \mathrm{SE}}{\mathrm{E}^{N,M}[\langle F_d\rangle]}$	N	M	$T-T_0$	T_0	C	Time to solution
10219325	0.80%	40	1	600	30	19114	37.27
10131701	0.45%	320	1	48.75	30	19133	4.95

Table 14
Statistical analysis of the drag force F_d (measured in Newtons) and of the base moment M_b (measured in Newton meters).

Q	$\mathrm{E}^N[\langle \mathrm{Q} \rangle_{T_0,T}]$	$\sigma^N[\langle \mathrm{Q} \rangle_{T_0,T}]$	$\sigma^N[\mathrm{Q}]$
F_d	10219325	222721	1742534
M_b	41210	321223	9397070

The proposed strategies are shown to be effective in taking advantage of HPC systems.

CRediT authorship contribution statement

Riccardo Tosi: Conceptualization, Methodology, Software, Investigation, Writing – original draft. **Marc Núñez:** Software, Investigation. **Jordi Pons-Prats:** Writing – review & editing, Supervision. **Javier Principe:** Conceptualization, Methodology, Writing – review & editing, Supervision, Funding. **Riccardo Rossi:** Conceptualization, Methodology, Reviewing, Writing – review & editing, Supervision, Funding.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors thank Prof. Barbara Wohlmuth and Dr. Brendan Keith for their valuable discussions and contributions to the subject. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 800898 and from project RTI2018-096898-B-I00 from the FEDER/Ministerio de Ciencia e Innovación - Agencia Estatal de Investigación. The authors also acknowledge the Severo Ochoa Centre of Excellence (2019–2023), which financially supported this work under the grant CEX2018-000797-S funded by MCIN/AEI/10.13039/501100011033. The authors thankfully acknowledge the computer resources at MareNostrum and the technical support provided by Barcelona Supercomputing Center (IM-2020-2-0030). This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic through the e-INFRA CZ (ID:90140). Dr. Jordi Pons-Prats acknowledges the support of the Serra Hunter programme by the Catalan Government.

References

Alauzet, F., Loseille, A., 2016. A decade of progress on anisotropic mesh adaptation for computational fluid dynamics. Comput. Aided Des. 72, 13–39.

Alminhana, G., Braun, A., Loredo-Souza, A., 2018. A numerical-experimental investigation on the aerodynamic performance of CAARC building models with geometric modifications. J. Wind Eng. Ind. Aerodyn. 180, 34–48.

Ayoul-Guilmard, Q., Ganesh, S., Nobile, F., Rossi, R., Tosi, R., Badia, R.M., Amela, R., 2020. XMC. Zenodo.

Badia, R.M., Conejero, J., Diaz, C., Ejarque, J., Lezzi, D., Lordan, F., Ramon-Cortes, C., Sirvent, R., 2015. COMP superscalar, an interoperable programming framework. SoftwareX 3–4.

Banks, J., Carson, J., Nelson, B.L., Nicol, D., 2004. Discrete-Event System Simulation (4th Edition), 4 Prentice Hall.

Bazilevs, Y., Calo, V., Cottrell, J., Hughes, T., Reali, A., Scovazzi, G., 2007. Variational multiscale residual-based turbulence modeling for large eddy simulation of incompressible flows. Comput. Methods Appl. Mech. Engrg. 197 (1–4), 173–201.

Beyhaghi, P., Alimohammadi, S., Bewley, T., 2018. A multiscale, asymptotically unbiased approach to uncertainty quantification in the numerical approximation of infinite time-averaged statistics. arXiv:1802.01056.

Boris, J.P., Grinstein, F.F., Oran, E.S., Kolbe, R.L., 1992. New insights into large eddy simulation. Fluid Dyn. Res. 10 (4–6), 199, URL http://iopscience.iop.org/1873-7005/10/4-6/A01.

Braun, A.L., Awruch, A.M., 2009. Aerodynamic and aeroelastic analyses on the CAARC standard tall building model using numerical simulation. Comput. Struct. 87 (9–10), 564–581.

Chen, L., Letchford, C., 2004. Parametric study on the along-wind response of the CAARC building to downbursts in the time domain. J. Wind Eng. Ind. Aerodyn. 92 (9), 703–724.

Cochran, L., Derickson, R., 2011. A physical modeler's view of computational wind engineering. J. Wind Eng. Ind. Aerodyn. 99 (4), 139–153, URL https://www.sciencedirect.com/science/article/pii/S0167610511000171 The Fifth International Symposium on Computational Wind Engineering.

Codina, R., Badia, S., Baiges, J., Principe, J., 2017. Variational multiscale methods in computational fluid dynamics. In: Encyclopedia of Computational Mechanics Second Edition. John Wiley & Sons, Ltd, Chichester, UK, pp. 1–28, https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119176817.ecm2117 http://doi.wiley.com/10.1002/9781119176817.ecm2117.

Codina, R., Principe, J., Guasch, O., Badia, S., 2007. Time dependent subscales in the stabilized finite element approximation of incompressible flow problems. Comput. Methods Appl. Mech. Engrg. 196 (21–24), 2413–2430.

Colomés, O., Badia, S., Codina, R., Principe, J., 2015. Assessment of variational multiscale models for the large eddy simulation of turbulent incompressible flows. Comput. Methods Appl. Mech. Engrg. 285, 32–63.

Dadvand, P., Rossi, R., Gil, M., Martorell, X., Cotela, J., Juanpere, E., Idelsohn, S.R., Oñate, E., 2013. Migration of a generic multi-physics framework to HPC environments. Comput. & Fluids 80 (1), 301–309.

Dadvand, P., Rossi, R., Oñate, E., 2010. An object-oriented environment for developing finite element codes for multi-disciplinary applications. Arch. Comput. Methods Eng. 17 (3), 253–297.

Dapogny, C., Dobrzynski, C., Frey, P., 2014. Three-dimensional adaptive domain remeshing, implicit domain meshing, and applications to free and moving boundary problems. J. Comput. Phys. 262, 358–378.

Durrett, R., 2019. Probability: Theory and Examples. pp. 1–420.

Fang, W., Giles, M.B., 2019. Multilevel Monte Carlo method for ergodic SDEs without contractivity. J. Math. Anal. Appl. arXiv:1803.05932.

Gander, M.J., 2015. 50 Years of Time Parallel Time Integration. Springer, Cham, pp. 69–113.

Grinstein, F.F., Margolin, L.G., Rider, W.J., 2007. Implicit Large Eddy Simulation: Computing Turbulent Fluid Dynamics. Cambridge University Press.

Guichard, R., 2019. Assessment of an improved Random Flow Generation method to predict unsteady wind pressures on an isolated building using Large-Eddy simulation. J. Wind Eng. Ind. Aerodyn. 189, 304–313, URL https://www.sciencedirect.com/science/article/pii/S016761051830953X.

Holmes, J.D., Tse, T.K., 2014. International high-frequency base balance benchmark study. Wind Struct. Int. J. 18 (4), 457–471.

Huang, S., Li, Q., Wu, J., 2010. A general inflow turbulence generator for large eddy simulation. J. Wind Eng. Ind. Aerodyn. 98 (10), 600–617, URL https://www.sciencedirect.com/science/article/pii/S0167610510000644.

Huang, S., Li, Q.S., Xu, S., 2007. Numerical evaluation of wind effects on a tall steel building by CFD. J. Construct. Steel Res..

Hughes, T.J., Feijóo, G.R., Mazzei, L., Quincy, J.-B., 1998. The variational multiscale method - a paradigm for computational mechanics. Comput. Methods Appl. Mech. Engrg. 166 (1–2), 3–24, URL http://www.sciencedirect.com/science/article/pii/S0045782598000796.

Hughes, T.J.R., Mazzei, L., Jansen, K.E., 2000. Large eddy simulation and the variational multiscale method. Comput. Vis. Sci. 3, 47–59.

Hunt, J.C., 1973. A theory of turbulent flow round two-dimensional bluff bodies. J. Fluid Mech. 61 (4), 625–706.

Ji, B., Lei, W., Xiong, Q., 2022. An inflow turbulence generation method for large eddy simulation and its application on a standard high-rise building. J. Wind Eng. Ind. Aerodyn. 226, 105048, URL https://www.sciencedirect.com/science/article/pii/S0167610522001519.

Jiang, N., Layton, W., 2014. An algorithm for fast calculation of flow ensembles. Int. J. Uncertain. Quantif. 4 (4), 273–301.

Joint committee on structural safety, Wind load. In: Probabilistic Model Code, vol. 2. URL https://www.jcss-lc.org/jcss-probabilistic-model-code/.

Kaimal, J.C., Finnigan, J.J., 1994. Atmospheric Boundary Layer Flows: Their Structure and Measurement. Oxford University Press.

Keith, B., Khristenko, U., Wohlmuth, B., 2021a. A fractional PDE model for turbulent velocity fields near solid walls. J. Fluid Mech. 916.

Keith, B., Khristenko, U., Wohlmuth, B., 2021b. Learning the structure of wind: A data-driven nonlocal turbulence model for the atmospheric boundary layer. Phys. Fluids 33 (9), 095110, URL https://aip.scitation.org/doi/10.1063/5.0064394.

Krasnopolsky, B.I., 2018. An approach for accelerating incompressible turbulent flow simulations based on simultaneous modelling of multiple ensembles. Comput. Phys. Comm. 229, 8–19.

Krasnopolsky, B., Nikitin, N., Lukyanov, A., 2018. Optimizing generation of multiple turbulent flow states. J. Phys. Conf. Ser. 1129 (1), 12020.

Li, Y., Li, C., Li, Q.-S., Song, Q., Huang, X., Li, Y.-G., 2020. Aerodynamic performance of CAARC standard tall building model by various corner chamfers. J. Wind Eng. Ind. Aerodyn. 202, 104197.

Lordan, F., Tejedor, E., Ejarque, J., Rafanell, R., Álvarez, J., Marozzo, F., Lezzi, D., Sirvent, R., Talia, D., Badia, R.M., 2014. ServiceSs: An interoperable programming framework for the cloud. J. Grid Comput. 12 (1), 67–91.

Makarashvili, V., Merzari, E., Obabko, A., Siegel, A., Fischer, P., 2017. A performance analysis of ensemble averaging for high fidelity turbulence simulations at the strong scaling limit. Comput. Phys. Comm. 219, 236–245.

Mann, J., 1994. The spatial structure of neutral atmospheric surface-layer turbulence. J. Fluid Mech. 273, 141–168.

Mann, J., 1998. Wind field simulation. Probab. Eng. Mech. 13 (4), 269–282.

Mataix, V., Bucher, P., Zorrilla, R., Rossi, R., Cotela, J., Carbonell, J.M., Celigueta, M.A., Teschemacher, T., Cornejo, A., Roig, C., Casas, G., Masó, M., Warnakulasuriya, S., Núñez, M., Dadvand, P., Latorre, S., Pouplana, I., Irazábal, J., Arrufat, F., Tosi, R., Ghantasala, A., Wilson, P., Baumgaertner, D., Chandra, B., Franci, A., Geiser, A., Bernd, K., Lopez, I., Gárate, J., 2021. KratosMultiphysics/Kratos: KratosMultiphysics 9.0. http://dx.doi.org/10.5281/zenodo.3234644.

Meyn, S.P., Tweedie, R.L., 1993. Stability of Markovian processes III: Foster-Lyapunov criteria for continuous-time processes. In: Applied Probability, vol. 25 no. 3. pp. 518–548.

Nastac, G., Labahn, J.W., Magri, L., Ihme, M., 2017. Lyapunov exponent as a metric for assessing the dynamic content and predictability of large-eddy simulations. Phys. Rev. Fluids 2 (9).

Nikitin, N.V., 2009. Disturbance growth rate in turbulent wall flows. Fluid Dyn. 44 (5).

Nogueira, X., Cueto-Felgueroso, L., Colominas, I., Gómez, H., 2010. Implicit large eddy simulation of non-wall-bounded turbulent flows based on the multiscale properties of a high-order finite volume method. Comput. Methods Appl. Mech. Engrg. 199 (9–12), 615–624.

Obasaju, E.D., 1992. Measurement of forces and base overturning moments on the CAARC tall building model in a simulated atmospheric boundary layer. J. Wind Eng. Ind. Aerodyn..

Pope, S.B., 2000. Turbulent Flows. Cambridge University Press, URL https://www.cambridge.org/core/product/identifier/9780511840531/type/book.

Principe, J., Codina, R., Henke, F., 2010. The dissipative structure of variational multiscale methods for incompressible flows. Comput. Methods Appl. Mech. Engrg. 199 (13–16), 791–801.

Ross, S., 2010. a First Course in Probability, eighth ed. Prentice Hall, Upper Saddle River, N.J..

Sagaut, P., 2006. Large eddy simulation for incompressible flows. Scientific Computing, Springer.

Tamura, T., 2008. Towards practical use of LES in wind engineering. J. Wind Eng. Ind. Aerodyn. 96 (10–11).

Tamura, Y., Kareem, A., 2013. Advanced Structural Wind Engineering.

Tejedor, E., Becerra, Y., Alomar, G., Queralt, A., Badia, R.M., Torres, J., Cortes, T., Labarta, J., 2017. PyCOMPSs: Parallel computational workflows in Python. Int. J. High Perform. Comput. Appl. 31 (1), 66–82.

Thordal, M.S., Bennetsen, J.C., Capra, S., Koss, H.H.H., 2020a. Towards a standard CFD setup for wind load assessment of high-rise buildings: Part 1–benchmark of the CAARC building. J. Wind Eng. Ind. Aerodyn. 205, 104283.

Thordal, M.S., Bennetsen, J.C., Capra, S., Kragh, A.K., Koss, H.H.H., 2020b. Towards a standard CFD setup for wind load assessment of high-rise buildings: Part 2–blind test of chamfered and rounded corner high-rise buildings. J. Wind Eng. Ind. Aerodyn. 205, 104282.

Tosi, R., Amela, R., Badia, R., Rossi, R., 2021a. A parallel dynamic asynchronous framework for uncertainty quantification by hierarchical Monte Carlo algorithms. J. Sci. Comput. 89 (28), 25.

Tosi, R., Mataix Ferrándiz, V., de Pouplana, I., Zorrilla, R., Gracia, L., Warnakulasuriya, S., Núñez, M., Bucher, P., Chandra, B., Ghantasala, A., Cotela, J., Franci, A., Bravo, R., Arrufat, F., Lázaro, C., Sautter, K.B., Latorre, S., González-Usúa, J., Geiser, A., Dadvand, P., Maso, M., Baumgaertner, D., Celigueta, M.A., Kikkeri Nagaraja, R., Wenczowski, S., Saridar, B., Roig, C., Casas, G., 2021b. Kratos-Multiphysics/Examples: Kratos examples 9.0. http://dx.doi.org/10.5281/zenodo.4293799.

Tosi, R., Rossi, R., Pons-Prats, J., 2021c. Towards Stochastic Methods in CFD for Engineering Applications (Ph.D. thesis). Universitat Politècnica de Catalunya, p. 206, URL https://upcommons.upc.edu/handle/2117/361637.

Von Kármán, T., 1948. Progress in the statistical theory of turbulence. Proc. Natl. Acad. Sci. USA 34 (11), 530.

Wang, Q., Gomez, S.A., Blonigan, P.J., Gregory, A.L., Qian, E.Y., 2013. Towards scalable parallel-in-time turbulent flow simulations. Phys. Fluids 25 (11), arXiv:1211.2437.

Yang, Q., Zhou, T., Yan, B., Van Phuc, P., Hu, W., 2020. LES study of turbulent flow fields over hilly terrains — Comparisons of inflow turbulence generation methods and SGS models. J. Wind Eng. Ind. Aerodyn. 204, 104230, URL https://www.sciencedirect.com/science/article/pii/S0167610520301409.

Zheng, C., Xie, Y., Khan, M., Wu, Y., Liu, J., 2018. Wind-induced responses of tall buildings under combined aerodynamic control. Eng. Struct. 175, 86–100.