Downsampling Strategies for LES Data from High-Fidelity CFD Solvers

Raffaella Sapio[1,2(✉)], Carmine De Lucia[2], Andrea Mastellone[2], Alfonso Matrone[2], Gerardo Toraldo[1], and Ferdinando Zullo[1]

[1] University of Campania Luigi Vanvitelli, Viale Lincoln 5, 81100 Caserta, Italy
{raffaella.sapio,gerardo.toraldo,ferdinando.zullo}@unicampania.it
[2] Centro Italiano Ricerche Aerospaziali (CIRA), Via Maiorise, 81043 Capua, Italy
{c.delucia,a.mastellone,a.matrone}@cira.it

Abstract. Downsampling is a fundamental operation in the analysis of high-fidelity Computational Fluid Dynamics (CFD) data, particularly in the context of Large-Eddy Simulation (LES), where flow fields are already represented as filtered, cell-averaged quantities. Recently, machine-learning-integrated approaches have been increasingly employed to accelerate the convergence of CFD solutions. Training neural networks in this setting requires highly accurate data, and one effective strategy is to downsample CFD solutions computed on finer grids. However, designing downsampling procedures that remain consistent with the physical and numerical structure of finite-volume solvers remains a challenging task. In this work, we investigate grid-to-grid downsampling strategies for LES data on structured meshes. We consider restriction operators that map fine-grid fields to coarser resolutions by aggregating values over blocks of cells, including arithmetic averaging, distance-weighted averaging, and extrema-based operators (minimum and maximum).

The methods are assessed through qualitative comparisons and quantitative metrics, including conservation error, blockwise discrepancies, and interface-based errors relevant to flux computations. The results show that arithmetic averaging provides the most consistent and physically meaningful coarse representation, while distance-weighted averaging offers comparable accuracy but lacks strict conservation. Extrema-based operators introduce significant distortions.

Keywords: Computational Fluid Dynamics (CFD) · Large-Eddy Simulation (LES) · Downsampling

1 Introduction

Computational Fluid Dynamics (CFD) plays a central role in the numerical analysis of fluid flows arising in engineering and physical sciences. The governing equations, typically the Navier–Stokes equations, describe the conservation of mass, momentum, and energy in a fluid and give rise to highly nonlinear and multi-scale dynamics. In turbulent regimes, the flow exhibits a wide range

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2027
O. Gervasi et al. (Eds.): ICCSA 2026 Workshops, LNCS 16763, pp. 221–238, 2027.
https://doi.org/10.1007/978-3-032-30530-5_14

of interacting spatial and temporal scales, making direct numerical simulation (DNS) prohibitively expensive for most practical applications. As a consequence, various modeling strategies have been developed to balance accuracy and computational cost, including Reynolds-Averaged Navier–Stokes (RANS) methods and Large-Eddy Simulation (LES) [11,16]. Among these approaches, LES represents a widely adopted compromise, in which the large-scale turbulent structures are explicitly resolved while the effect of the unresolved small scales is modeled. From a mathematical viewpoint, LES can be interpreted as a filtering operation applied to the Navier–Stokes equations, leading to a coarse-grained representation of the flow field [7]. In practical implementations, the computational grid and the numerical discretization act as an implicit filter, so that the stored solution corresponds to a filtered, cell-averaged field defined on the mesh. In recent years, the increasing availability of high-fidelity CFD data and advances in machine learning (ML) have opened new perspectives for the development of hybrid modeling approaches. In particular, ML techniques have been successfully employed to enhance numerical solvers, accelerate simulations, and reconstruct fine-scale features from coarse data [6,10,12]. A key aspect of these approaches is the ability to transfer information across resolutions, for instance when generating training datasets, designing surrogate models, or coupling numerical solvers with data-driven components. In this context, downsampling plays a fundamental role. Although often treated as a simple preprocessing step, downsampling in CFD is a non-trivial operation that involves reducing the spatial resolution of the data while preserving relevant physical and numerical properties. Different strategies—ranging from geometric coarsening and averaging to more complex filtering or data-driven approaches—can lead to significantly different coarse representations of the same flow field [9,14,17]. In particular, the choice of the downsampling operator directly affects the preservation of conservation laws, the representation of turbulent structures, and the consistency with the underlying numerical scheme. Despite its importance, the problem of defining suitable downsampling strategies for CFD data remains relatively underexplored, especially in the context of finite-volume LES solvers. In such settings, the solution is naturally expressed in terms of cell-averaged quantities, and numerical fluxes depend critically on interface values reconstructed from these averages. Therefore, a meaningful downsampling procedure should not only approximate the fine-grid field but also remain consistent with the conservative structure of the discretization and with the quantities entering the flux computation.

The objective of this work is to investigate and compare different downsampling strategies for CFD data generated by high-fidelity LES solvers. We focus on structured grids and consider restriction operators that map fine-grid data to coarser resolutions by aggregating values over blocks of cells. In particular, we analyze four strategies: arithmetic averaging, distance-weighted averaging, and extrema-based operators (minimum and maximum). These methods are evaluated in terms of their ability to preserve conservation properties, represent fine-scale variability, and maintain consistency with interface quantities relevant to finite-volume schemes. The analysis is carried out on a set of LES snapshots

obtained from a representative test case, and a series of quantitative metrics are introduced to assess the performance of each strategy. The results provide insight into the trade-offs between physical consistency, numerical accuracy, and feature preservation, and allow us to identify the most suitable approaches for coarse-grid representation of LES data.

The remainder of the paper is organized as follows. In Sect. 2, we provide an overview of downsampling strategies for CFD data. Section 3 describes the characteristics of the considered LES datasets. In Sect. 4, we introduce the downsampling operators and analyze their properties. Section 5 presents the numerical tests and comparative results. Finally, conclusions and perspectives for future work are discussed in Sect. 6.

2 Overview on Downsampling Strategies for CFD Data

Although often informally associated with simple subsampling, downsampling encompasses fundamentally different operations depending on the context, including mesh coarsening, filtering or coarse graining, adaptive resolution reduction, and data-driven compression. These approaches are not equivalent: each induces a distinct transformation of the underlying flow field and therefore preserves different invariants, such as conservation laws, spectral content, or turbulence statistics. Consequently, the choice of downsampling strategy must be aligned with the intended application, whether it is solver acceleration, turbulence modeling, data storage, or machine-learning-based reconstruction. A first major class of methods is geometric downsampling via mesh coarsening, in which the computational grid is explicitly reduced. This approach is central to multigrid methods and hierarchical discretizations, where coarse representations are constructed either geometrically (e.g., structured coarsening or agglomeration of elements) or algebraically through operator-based techniques. In this setting, restriction and prolongation operators define the transfer of solution variables between grid levels, and careful design is required to maintain stability and conservation properties. While mesh coarsening directly reduces the number of degrees of freedom and thus computational cost, it may degrade the representation of localized flow features such as boundary layers, shocks, or vortical structures if the coarsening is too aggressive or not adapted to the flow physics [5,17]. A second class consists of field-based downsampling through filtering or volume averaging, which is particularly relevant in the context of turbulence modeling. In large-eddy simulation, the coarse representation of the flow is interpreted as a filtered field, where unresolved scales are removed and their effects are modeled. This perspective, formalized by Germano's filtering framework, emphasizes that a physically consistent coarse field should be obtained via a low-pass filtering operation rather than pointwise subsampling [7,16]. Common filtering strategies include box filters, Gaussian filters, and spectral cut-off filters, each defining a characteristic filter width that separates resolved and unresolved scales. In contrast to implicit filtering—where the grid and numerical discretization act as the effective filter—explicit filtering allows greater control over the

scale separation and has been shown to improve robustness and consistency in certain LES formulations [8,15]. An important distinction within field-based approaches is that between decimation and averaging. Pure decimation, such as selecting every k-th grid point or cell value, is computationally inexpensive but is generally inconsistent with finite-volume data, since it discards most of the cell-averaged information within each coarse control volume. As a consequence, it may not preserve integral quantities over the coarse cells and may lead to non-physical representations of the flow. Averaging-based methods, by contrast, act as discrete low-pass filters and, on uniform grids, correspond to volume averaging, making them consistent with the finite-volume interpretation of the solution as a filtered field. This distinction has become particularly relevant in recent data-driven CFD workflows, where the generation of low-resolution training data significantly influences the performance and physical consistency of learned models [6]. A third category includes adaptive and feature-aware downsampling strategies, in which resolution is reduced non-uniformly across the domain. These methods aim to retain high resolution in dynamically important regions—such as shear layers, vortices, or discontinuities—while coarsening regions of smooth flow. Adaptive mesh refinement, multiresolution wavelet methods, and subsampling-based compression techniques fall within this class. Such approaches are especially valuable in large-scale simulations and in situ data processing, where storage and I/O constraints necessitate selective data reduction. However, adaptive strategies introduce additional complexity in terms of error estimation, refinement criteria, and the preservation of conservation properties across interfaces [9,18]. A fourth class is spectral or modal downsampling, where the flow field is represented in a reduced basis rather than on a coarser physical grid. Examples include Fourier truncation, proper orthogonal decomposition, and wavelet-based representations. These approaches are particularly effective for capturing dominant energetic structures and for reduced-order modeling, as they concentrate information in a small number of modes. However, the resulting coarse representation may not directly correspond to a standard CFD discretization, and additional reconstruction or projection steps are typically required to recover physical-space fields [1,14]. More recently, data-driven downsampling and reconstruction strategies have gained significant attention. In this setting, low-resolution flow fields are generated—typically via filtering, averaging, or coarsening—and used as inputs to machine learning models trained to reconstruct high-resolution fields or to correct coarse-grid simulations. Approaches include supervised and unsupervised super-resolution, physics-informed learning, and hybrid solvers that embed learned closures into coarse CFD computations. A key insight from this literature is that the downsampling operator is not merely a preprocessing step but a critical component that shapes the statistical structure of the data and the learnability of fine-scale features [10,12]. Overall, downsampling in CFD should be viewed as a design choice rather than a purely technical operation. Mesh-based, filtering-based, adaptive, spectral, and learned strategies each provide different trade-offs between computational efficiency, physical fidelity, and reconstructability. In particular, it is generally not

possible to simultaneously preserve conservation laws, spectral accuracy, and small-scale intermittency under aggressive resolution reduction. Therefore, the selection of a downsampling strategy must be guided by the specific objectives of the study, whether they involve numerical efficiency, turbulence-consistent modeling, data compression, or data-driven inference.

3 Considered CFD Data

The data considered in this work are obtained from high-fidelity simulations performed with a Large-Eddy Simulation solver, specifically SPARK-LES [2–4]. This solver computes the evolution of fluid flows by numerically solving the compressible Navier-Stokes equations using a finite-volume formulation on structured grids. As a consequence, the resulting datasets consist of discrete representations of physical fields defined over a set of control volumes, where conservation laws are enforced at the discrete level through flux balances across cell interfaces. More precisely, the computational domain is partitioned into a structured Cartesian mesh, and physical quantities are stored as cell-centered variables. Each cell contains averaged values of the primary flow quantities, including velocity components, pressure, density, and, depending on the physical model, additional variables such as temperature or species concentrations. These quantities represent volume-averaged states, consistent with the finite-volume discretization, rather than pointwise values. The evolution of these variables is governed by the balance of convective and diffusive fluxes computed at cell faces using high-order interpolation schemes. Since SPARK-LES operates within the LES framework, the data inherently correspond to a filtered representation of the flow field. In particular, only the large-scale turbulent structures are explicitly resolved on the computational grid, while the effects of unresolved small scales are modeled. Therefore, the grid spacing defines an effective filter width, and the stored solution can be interpreted as a coarse-grained version of the true physical flow. This intrinsic filtering plays a crucial role when further downsampling is applied, as it implies that the data are already the result of a scale-reduction process. From a numerical standpoint, the datasets exhibit several important properties that must be preserved during downsampling. First, the finite-volume formulation guarantees discrete conservation of mass, momentum, and energy, meaning that any coarsening strategy should ideally maintain consistency with these conservation laws. Second, the solution is strongly influenced by flux evaluations at cell interfaces, which depend on interpolation procedures and local flow gradients. Inaccuracies in these interface fluxes can significantly affect the overall solution quality, especially in turbulent regimes. Consequently, downsampling methods that ignore inter-cell flux consistency may introduce non-physical artifacts. Additionally, the data are typically high-dimensional and multi-scale. Even in relatively simple test cases such as the Taylor–Green vortex, the flow exhibits evolving vortical structures and a cascade of energy across scales, making it sensitive to resolution changes. In more complex configurations, such as compressible or reacting flows, the coupling between turbulence, thermodynamics,

and possibly chemical kinetics further increases the complexity of the data. This multi-scale nature implies that naive resolution reduction (e.g., simple subsampling) can lead to loss of physically relevant information, including energy spectra and coherent structures. Finally, the data are generated through explicit time integration schemes, often based on multi-stage Runge-Kutta methods, resulting in time-resolved sequences of flow fields. This temporal dimension introduces additional constraints: downsampling strategies must not only preserve spatial consistency but also avoid introducing spurious temporal dynamics when applied across successive time steps.

4 Downsampling Strategies

In this work, we consider downsampling procedures aimed at transferring solution fields from a fine computational grid to a coarser grid while preserving, as much as possible, the physical and numerical properties of the original data. The data originate from Large-Eddy Simulation computations performed with a finite-volume solver such as SPARK-LES, where the solution is represented as cell-averaged quantities over a structured mesh.

4.1 Grid Hierarchy and Notation

Let Ω^f denote the fine grid and Ω^c the corresponding coarse grid. We assume a structured hierarchy such that each coarse control volume $C \in \Omega^c$ is defined as the union of a set of non-overlapping fine-grid cells:

$$C = \bigcup_{c_f \in \mathcal{F}(C)} c_f, \tag{1}$$

where $\mathcal{F}(C) \subset \Omega^f$ denotes the set of fine cells contained in C.

Let $u^f : \Omega^f \to \mathbb{R}$ be a scalar field defined on the fine grid (e.g., pressure, density, or a component of velocity). The goal of the downsampling procedure is to construct a coarse field $u^c : \Omega^c \to \mathbb{R}$ by applying an aggregation operator over the fine cells in $\mathcal{F}(C)$.

The extension to vector-valued fields (e.g., velocity $\mathbf{u}$) is performed component-wise.

4.2 Arithmetic Averaging

The first and most natural strategy is based on arithmetic averaging. For each coarse cell $C \in \Omega^c$, we define

$$u^c(C) = \frac{1}{|\mathcal{F}(C)|} \sum_{c_f \in \mathcal{F}(C)} u^f(c_f), \tag{2}$$

where $|\mathcal{F}(C)|$ denotes the number of fine cells in C.

On uniform grids, this operation is equivalent to a discrete volume average and is therefore consistent with the finite-volume interpretation of the solution. In particular, if u^f represents a conserved quantity, the arithmetic mean preserves integral quantities across scales up to discretization error. Moreover, this operator can be interpreted as a low-pass filter, removing small-scale fluctuations while retaining the large-scale structure of the field.

For these reasons, arithmetic averaging is the most physically consistent downsampling strategy in the context of LES data.

4.3 Distance-Weighted Averaging

In addition to uniform arithmetic averaging, we consider a weighted averaging strategy in which fine-grid values closer to the center of a coarse cell are assigned a larger contribution. For each coarse cell $C \in \Omega^c$, the downsampled value is defined as

$$u^c(C) = \sum_{c_f \in \mathcal{F}(C)} w(c_f; C)\, u^f(c_f), \tag{3}$$

where the weights satisfy

$$w(c_f; C) \geq 0, \qquad \sum_{c_f \in \mathcal{F}(C)} w(c_f; C) = 1. \tag{4}$$

In the implementation considered here, the weights are chosen to be inversely proportional to the distance between the fine-cell location and the geometric center of the coarse cell. Denoting by $\mathbf{x}_{c_f}$ the position of the fine cell c_f and by $\mathbf{x}_C$ the center of the coarse cell C, we define

$$w(c_f; C) = \frac{\dfrac{1}{\|\mathbf{x}_{c_f} - \mathbf{x}_C\| + \varepsilon}}{\displaystyle\sum_{c'_f \in \mathcal{F}(C)} \frac{1}{\|\mathbf{x}_{c'_f} - \mathbf{x}_C\| + \varepsilon}}, \tag{5}$$

where $\varepsilon > 0$ is a small regularization parameter introduced to avoid singularities when a fine-cell center coincides with the coarse-cell center.

This construction yields a convex combination of the fine-cell values, so that the weighted coarse value remains bounded between the local minimum and maximum:

$$\min_{c_f \in \mathcal{F}(C)} u^f(c_f) \leq u^c(C) \leq \max_{c_f \in \mathcal{F}(C)} u^f(c_f). \tag{6}$$

Compared with arithmetic averaging, this operator gives greater importance to the central part of the coarse cell and reduced influence to values located near its boundaries. As a consequence, it may provide a representation that is more localized and less diffusive, especially when the field exhibits strong intra-cell gradients. However, unlike uniform averaging, distance-weighted averaging is generally not conservative with respect to the integral of the field over the coarse

control volume. Indeed, conservation would require weights proportional to fine-cell volumes, whereas the present choice depends on the geometric distance from the cell center. Therefore, while the weighted strategy may better preserve local features around the coarse-cell centroid, it does not in general preserve integral quantities across scales.

4.4 Maximum-Based Downsampling

The second strategy assigns to each coarse cell the maximum value among the corresponding fine cells:

$$u^c(C) = \max_{c_f \in \mathcal{F}(C)} u^f(c_f). \tag{7}$$

This operator emphasizes the most intense local features within each coarse region. It can be particularly useful for detecting strong gradients, localized peaks, or intermittent structures, which are common in turbulent flows.

However, this approach is non-conservative and introduces a systematic bias toward larger values. As a result, it does not preserve integral quantities and may distort statistical properties such as mean values or energy distributions when applied repeatedly across multiple coarsening levels.

4.5 Minimum-Based Downsampling

Similarly, a minimum-based strategy can be defined as

$$u^c(C) = \min_{c_f \in \mathcal{F}(C)} u^f(c_f). \tag{8}$$

This operator highlights the lowest values within each coarse cell and may be useful in identifying low-intensity regions or preserving lower bounds of physical quantities. As in the case of the maximum operator, this method is non-conservative and may lead to systematic underestimation of the field.

Remark 1. We underline that the same strategy is applied for the data arising from interface values.

4.6 Properties of the Downsampling Operators

The four strategies described above exhibit different mathematical and physical properties.

Conservation. Arithmetic averaging is consistent with the conservation principles underlying finite-volume methods, as it approximates volume integration over the coarse cells on uniform grids. Distance-weighted averaging, while still based on a convex combination of fine-cell values, is generally not conservative because the weights are not proportional to cell volumes. Maximum and minimum operators are also non-conservative.

Linearity. Both arithmetic averaging and distance-weighted averaging are linear operators, i.e.,

$$\mathcal{R}(\alpha u + \beta v) = \alpha \mathcal{R}(u) + \beta \mathcal{R}(v), \tag{9}$$

where $\mathcal{R}$ denotes the restriction operator. This property is important for consistency with the linear structure of the governing equations. Maximum and minimum operators are nonlinear and therefore do not satisfy this property.

Boundedness. Averaging and distance-weighted averaging satisfy

$$\min_{c_f \in \mathcal{F}(C)} u^f(c_f) \leq u^c(C) \leq \max_{c_f \in \mathcal{F}(C)} u^f(c_f), \tag{10}$$

since both are convex combinations of the fine-cell values. The maximum and minimum operators attain the upper and lower bounds, respectively.

Spectral Behavior. Arithmetic averaging acts as a discrete low-pass filter and damps high-frequency components of the field. Distance-weighted averaging also has a smoothing effect, but with a stronger emphasis on the central region of the coarse cell, thus producing a filtered field that is less uniform than the arithmetic mean. Maximum and minimum operators do not correspond to classical linear filters and may introduce distortions in the spectral content.

4.7 Implications for LES Data

Since the input data are obtained from LES simulations, they already represent a filtered version of the underlying flow. Therefore, additional downsampling introduces a second level of filtering or transformation. In this context, arithmetic averaging is consistent with the LES framework, as it preserves the interpretation of the solution as a coarse-grained field.

On the other hand, extrema-based operators (maximum and minimum) modify the physical meaning of the variables and may alter the balance between resolved and unresolved scales. Their use should therefore be carefully evaluated depending on the application, particularly in relation to turbulence statistics and energy transfer across scales.

5 Numerical Tests

5.1 Comparison Metrics

To assess the quality of the different downsampling strategies, we introduce a set of comparison metrics designed to quantify both volumetric consistency and interface accuracy. Since the original high-fidelity data are defined as cell-centered quantities on the fine grid, while flux computations in the finite-volume solver depend on interface values, it is important to evaluate the coarse representation from both perspectives.

Let u^f denote the fine-grid field and u^c the corresponding downsampled coarse-grid field. For a given coarse cell $C \in \Omega^c$, let $\mathcal{F}(C)$ be the set of fine cells contained in C, and let $N_C = |\mathcal{F}(C)|$ be their number.

Global Conservation Error. A first metric measures the discrepancy between the total fine-grid quantity and the total coarse-grid quantity:

$$Q^f = \sum_{c_f \in \Omega^f} u^f(c_f)\,\Delta V_f, \qquad Q^c = \sum_{C \in \Omega^c} u^c(C)\,\Delta V_C, \tag{11}$$

where ΔV_f and ΔV_C denote the fine- and coarse-cell volumes, respectively. The relative conservation error is then defined as

$$\varepsilon_{\text{cons}} = \frac{|Q^c - Q^f|}{|Q^f|}, \qquad \text{if } Q^f \neq 0. \tag{12}$$

If $Q^f = 0$, the absolute discrepancy $|Q^c - Q^f|$ is used instead. This metric quantifies the extent to which a downsampling operator preserves the global integral of the field. It is particularly relevant for conserved variables such as density, momentum, or total energy.

Intra-Block Cell-Centered Error. A second family of metrics evaluates how well the coarse value $u^c(C)$ represents the distribution of fine-cell values inside the corresponding coarse control volume. For each coarse cell C, we define the blockwise root-mean-square error (RMSE) as

$$\text{RMSE}_{\text{blk}}(C) = \left(\frac{1}{N_C} \sum_{c_f \in \mathcal{F}(C)} \left(u^f(c_f) - u^c(C) \right)^2 \right)^{1/2}, \tag{13}$$

and the blockwise mean absolute error (MAE) as

$$\text{MAE}_{\text{blk}}(C) = \frac{1}{N_C} \sum_{c_f \in \mathcal{F}(C)} \left| u^f(c_f) - u^c(C) \right|. \tag{14}$$

The corresponding global indicators are obtained by averaging over all coarse cells:

$$\overline{\text{RMSE}}_{\text{blk}} = \frac{1}{|\Omega^c|} \sum_{C \in \Omega^c} \text{RMSE}_{\text{blk}}(C), \tag{15}$$

$$\overline{\text{MAE}}_{\text{blk}} = \frac{1}{|\Omega^c|} \sum_{C \in \Omega^c} \text{MAE}_{\text{blk}}(C). \tag{16}$$

These metrics measure the representativeness of the coarse cell value with respect to the unresolved fine-scale variability inside each block. The RMSE is more sensitive to large deviations, whereas the MAE provides a more robust measure of the average discrepancy.

Interface-Based Error. Because the finite-volume discretization relies on numerical fluxes through cell faces, it is also necessary to evaluate how accurately the downsampled field represents interface quantities. Let $u_{\text{int}}^{f,\alpha}$ denote the fine-grid interface value in direction α, reconstructed from the fine cell-centered field,

and let $u_{\mathrm{int}}^{c,\alpha}(C)$ denote the corresponding coarse interface value associated with coarse cell C.

For each direction α and each coarse cell C, let $\mathcal{I}_\alpha(C)$ denote the set of fine interfaces associated with the corresponding coarse interface, and let $N_{I,\alpha}(C) = |\mathcal{I}_\alpha(C)|$. We define

$$\mathrm{RMSE}_{\mathrm{face}}^{(\alpha)}(C) = \left(\frac{1}{N_{I,\alpha}(C)} \sum_{i_f \in \mathcal{I}_\alpha(C)} \left(u_{\mathrm{int}}^{f,\alpha}(i_f) - u_{\mathrm{int}}^{c,\alpha}(C) \right)^2 \right)^{1/2}, \tag{17}$$

and

$$\mathrm{MAE}_{\mathrm{face}}^{(\alpha)}(C) = \frac{1}{N_{I,\alpha}(C)} \sum_{i_f \in \mathcal{I}_\alpha(C)} \left| u_{\mathrm{int}}^{f,\alpha}(i_f) - u_{\mathrm{int}}^{c,\alpha}(C) \right|. \tag{18}$$

The global interface indicators are then obtained by averaging over all coarse cells and all coordinate directions:

$$\overline{\mathrm{RMSE}}_{\mathrm{face}} = \frac{1}{N_{\mathrm{dir}} \, |\Omega^c|} \sum_{\alpha=1}^{N_{\mathrm{dir}}} \sum_{C \in \Omega^c} \mathrm{RMSE}_{\mathrm{face}}^{(\alpha)}(C), \tag{19}$$

$$\overline{\mathrm{MAE}}_{\mathrm{face}} = \frac{1}{N_{\mathrm{dir}} \, |\Omega^c|} \sum_{\alpha=1}^{N_{\mathrm{dir}}} \sum_{C \in \Omega^c} \mathrm{MAE}_{\mathrm{face}}^{(\alpha)}(C), \tag{20}$$

where $N_{\mathrm{dir}} = 2$ in two-dimensional configurations and $N_{\mathrm{dir}} = 3$ in three-dimensional configurations.

These interface-based metrics are especially important in CFD applications because they quantify the extent to which a downsampled field preserves the quantities entering the flux evaluation. A method may perform well in terms of cell-centered averages while still introducing large errors at interfaces, which can significantly affect the numerical solution when the coarse field is used for further computation or modeling.

Interpretation of the Metrics. The conservation error measures the preservation of integral quantities and therefore reflects the physical consistency of the downsampling operator with the finite-volume formulation. The blockwise RMSE and MAE quantify how well a single coarse value represents the unresolved fine-scale variability within each coarse cell. The face-based RMSE and MAE instead evaluate the consistency of the downsampled field with the interface quantities that govern convective and diffusive fluxes. Taken together, these metrics provide a complementary assessment of conservation, representativeness, and numerical relevance.

5.2 Computational Tests

In this section, we assess the different downsampling strategies on a sequence of LES snapshots from the COVO test case. COVO (COnvection of a VOrtex)

is a basic test used to verify the correct functioning of convective schemes. In this test, a vortex is transported through a periodic grid without deformation or dissipation, and the scheme must ensure that the transport occurs at space and time scales suitable for simulating other cases of interest. The analysis is carried out on structured two-dimensional data, with fine-grid resolution 160×160 and coarse-grid resolution 40×40, corresponding to a downsampling factor equal to 4 in each spatial direction.

The variables considered are the conservative fields

$$\rho, \qquad \rho u, \qquad \rho v, \qquad \rho w, \qquad \rho E,$$

together with the temperature T. Since the underlying solver adopts a finite-volume formulation, the stored fine-grid variables are interpreted as cell-averaged quantities. A meaningful restriction operator should therefore preserve, as much as possible, both the discrete integral of conservative quantities and the local structure relevant to the evaluation of numerical fluxes.

The four downsampling operators compared in this study are:

- arithmetic averaging (`mean`),
- distance-weighted averaging (`weight`),
- minimum-based restriction (`min`),
- maximum-based restriction (`max`).

For each variable and each snapshot, we compute the following indicators:

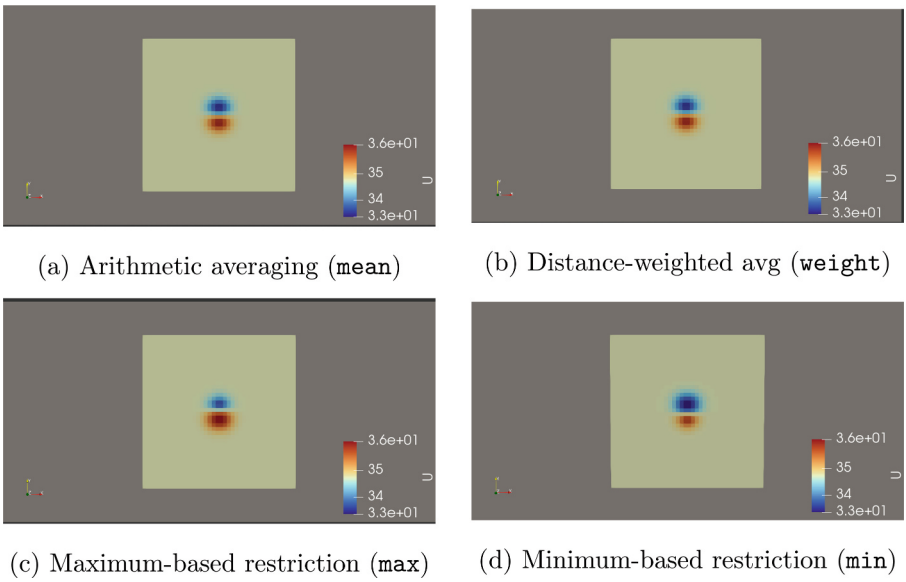
- the global conservation error $\varepsilon_{\text{cons}}$,
- the blockwise cell-centered errors $\overline{\text{RMSE}}_{\text{blk}}$ and $\overline{\text{MAE}}_{\text{blk}}$,
- the face-based errors $\overline{\text{RMSE}}_{\text{face}}$ and $\overline{\text{MAE}}_{\text{face}}$.

5.3 Qualitative Comparison of the Coarse Fields

Figures 1, 2 and 3 show a qualitative comparison of the downsampled streamwise velocity field u at the initial, intermediate and final stage of the flow evolution for the COVO test case. The final snapshot is selected as the representative visual example because it corresponds to a more evolved state of the flow, in which the relevant structures are more clearly developed and the differences among the restriction operators are more evident.

The arithmetic and distance-weighted averages produce very similar coarse fields and both preserve the overall dipolar structure of the fine solution. Nevertheless, the arithmetic average yields a slightly more balanced representation of the positive and negative lobes. By contrast, the extrema-based operators visibly distort the field: the maximum-based restriction tends to overemphasize the positive lobe, whereas the minimum-based restriction enhances the negative one. This behavior is consistent with the mathematical definition of the operators, since `max` and `min` do not compute representative coarse values but instead select local extrema within each coarse block.

Overall, the qualitative comparison already suggests that arithmetic averaging provides the most physically meaningful coarse representation, while the weighted average remains a close alternative and the extrema-based operators are less suitable for consistent state transfer across resolutions.



(a) Arithmetic averaging (`mean`)

(b) Distance-weighted avg (`weight`)

(c) Maximum-based restriction (`max`)

(d) Minimum-based restriction (`min`)

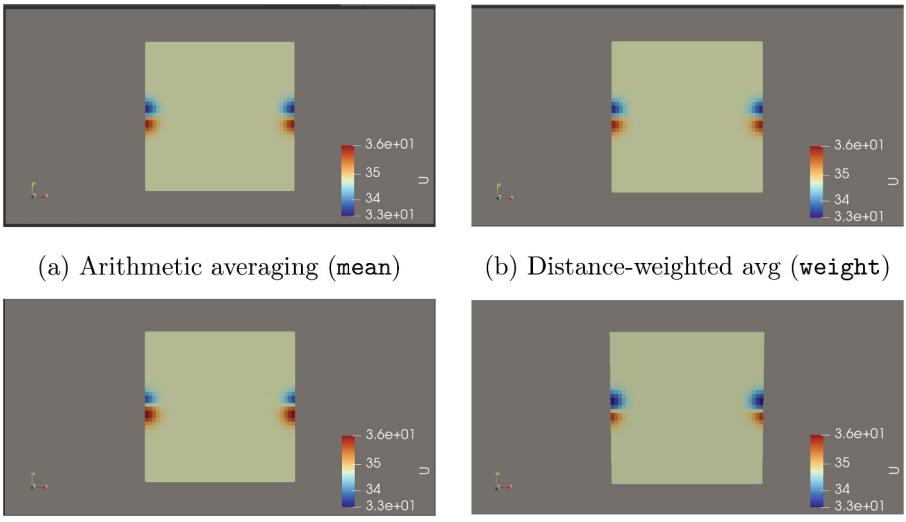
Fig. 1. Downsampling comparison at the **initial time step**.

5.4 Representative Quantitative Results

Table 1 reports the quantitative error metrics for the final snapshot, which is taken as the representative reference case. This choice is motivated by the fact that the final state corresponds to a more developed flow configuration, while still exhibiting the same ranking among the downsampling operators as observed at the initial and intermediate times.

The quantitative results consistently indicate that arithmetic averaging provides the best overall performance. For the conservative variables ρ, ρu, and ρE, the arithmetic average preserves the discrete integral up to machine precision and also yields the smallest blockwise and face-based errors. The distance-weighted strategy produces errors that are often very close to those of the arithmetic mean, but remains systematically less conservative. By contrast, the minimum- and maximum-based operators lead to significantly larger discrepancies, both in the cell-centered and in the interface-based metrics.

A particularly relevant observation concerns the face-based errors. Since numerical fluxes in the finite-volume solver are constructed from interface quantities, the quality of the face representation is directly related to the suitability of a



(a) Arithmetic averaging (`mean`) (b) Distance-weighted avg (`weight`)

(c) Maximum-based restriction (`max`) (d) Minimum-based restriction (`min`)

Fig. 2. Downsampling comparison at an **intermediate time step**.

downsampling operator for subsequent numerical use. In this respect, arithmetic averaging again emerges as the most reliable option, with the distance-weighted strategy remaining competitive but consistently slightly inferior, whereas the extrema-based operators exhibit substantially larger errors.

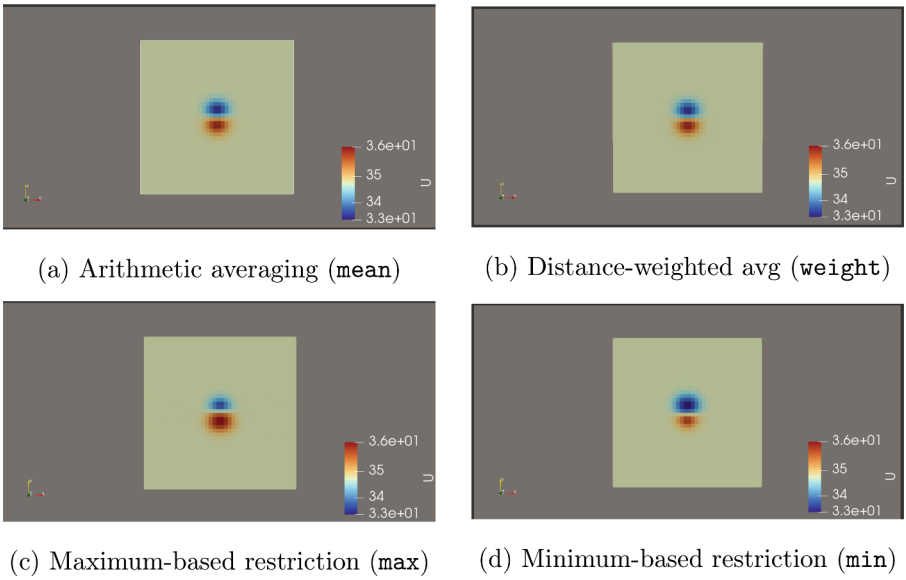
5.5 Consistency Across the Flow Evolution

The same analysis was repeated for the initial, intermediate, and final snapshots of the COVO test case. Although the absolute values of the errors vary as the flow evolves, the qualitative ranking of the downsampling operators remains unchanged. In particular, arithmetic averaging always provides the best or nearly best overall performance, distance-weighted averaging remains a close but non-conservative alternative, and the extrema-based operators systematically yield larger errors.

This temporal consistency indicates that the conclusions are not tied to a specific flow realization, but rather reflect an intrinsic property of the restriction operators themselves. The final state is therefore used in the main quantitative presentation because it offers the most developed and visually informative flow structures, while remaining fully representative of the trends observed throughout the simulation.

6 Conclusions

In this work, we investigated several downsampling strategies for LES data generated by high-fidelity finite-volume CFD solvers, with particular attention to



(a) Arithmetic averaging (`mean`) (b) Distance-weighted avg (`weight`)

(c) Maximum-based restriction (`max`) (d) Minimum-based restriction (`min`)

Fig. 3. Downsampling comparison at the **final time step**.

their suitability for cell-centered fields arising from SPARK-LES-type computations. The study was motivated by the need to transfer information from fine to coarse grids in a way that remains consistent with the numerical structure of the solver and with the physical interpretation of LES data as already filtered flow fields.

We first framed the problem in the broader context of downsampling for CFD, highlighting that restriction operators are not interchangeable: different choices preserve different properties, including conservation, boundedness, and spectral content. Starting from this perspective, we analyzed four grid-to-grid strategies on structured meshes, namely arithmetic averaging, distance-weighted averaging, minimum-based restriction, and maximum-based restriction, under the assumption that each coarse cell is the union of a set of fine cells.

The numerical results on the two-dimensional COVO test case show a clear and stable hierarchy among the considered operators. Arithmetic averaging consistently provides the best overall behavior. In particular, it preserves the integral of conservative quantities up to machine precision on uniform grids and yields the lowest blockwise and face-based errors for the variables considered. This confirms that arithmetic averaging is the most coherent choice when the objective is to obtain a coarse representation compatible with the conservative finite-volume structure of the underlying solver. Distance-weighted averaging emerges as a reasonable alternative when a more localized representation is desired. Its performance remains close to that of the arithmetic mean in many of the tested quantities, both qualitatively and quantitatively. By contrast, the extrema-based operators, i.e. minimum- and maximum-based restriction, sys-

Table 1. Representative comparison of downsampling strategies for the final flow state of the COVO test case. The columns correspond to arithmetic averaging (`mean`), distance-weighted averaging (`weight`), minimum-based restriction (`min`), and maximum-based restriction (`max`). For ρv, the relative conservation error may become poorly informative because the domain-integrated quantity is close to zero. In this case, the blockwise and face-based errors provide a more robust basis for comparison. We will use the notation $\varepsilon_{\text{cons}}(\cdot)$, $\overline{\text{RMSE}}_{\text{blk}}(\cdot)$, $\overline{\text{MAE}}_{\text{blk}}(\cdot)$, $\overline{\text{RMSE}}_{\text{face}}(\cdot)$ and $\overline{\text{MAE}}_{\text{face}}(\cdot)$, where $\cdot$ can be $\rho, \rho u, \rho v$ or ρE, to denote the corresponding error for the considered variable.

Metric(Variable)	`mean`	`weight`	`min`	`max`
$\varepsilon_{\text{cons}}(\rho)$	8.456792×10^{-16}	6.898326×10^{-14}	1.491120×10^{-7}	1.278181×10^{-7}
$\overline{\text{RMSE}}_{\text{blk}}(\rho)$	9.625817×10^{-8}	9.634005×10^{-8}	2.005217×10^{-7}	1.789093×10^{-7}
$\overline{\text{MAE}}_{\text{blk}}(\rho)$	8.179648×10^{-8}	8.176986×10^{-8}	1.753982×10^{-7}	1.503505×10^{-7}
$\overline{\text{RMSE}}_{\text{face}}(\rho)$	6.123171×10^{-8}	6.141042×10^{-8}	9.988677×10^{-8}	9.989998×10^{-8}
$\overline{\text{MAE}}_{\text{face}}(\rho)$	5.558632×10^{-8}	5.562534×10^{-8}	7.869585×10^{-8}	7.874203×10^{-8}
$\varepsilon_{\text{cons}}(\rho u)$	1.302656×10^{-14}	6.506816×10^{-10}	3.145298×10^{-4}	3.145655×10^{-4}
$\overline{\text{RMSE}}_{\text{blk}}(\rho u)$	7.669575×10^{-3}	7.673540×10^{-3}	1.499306×10^{-2}	1.499423×10^{-2}
$\overline{\text{MAE}}_{\text{blk}}(\rho u)$	6.547496×10^{-3}	6.540503×10^{-3}	1.284659×10^{-2}	1.284805×10^{-2}
$\overline{\text{RMSE}}_{\text{face}}(\rho u)$	4.795590×10^{-3}	4.803121×10^{-3}	7.996384×10^{-3}	7.996010×10^{-3}
$\overline{\text{MAE}}_{\text{face}}(\rho u)$	4.270746×10^{-3}	4.270813×10^{-3}	6.391113×10^{-3}	6.390626×10^{-3}
$\varepsilon_{\text{cons}}(\rho v)$	1.081697×10^{-3}	1.839160×10^{7}	2.815816×10^{12}	2.813619×10^{12}
$\overline{\text{RMSE}}_{\text{blk}}(\rho v)$	7.683832×10^{-3}	7.688001×10^{-3}	1.502378×10^{-2}	1.501530×10^{-2}
$\overline{\text{MAE}}_{\text{blk}}(\rho v)$	6.560065×10^{-3}	6.553028×10^{-3}	1.287379×10^{-2}	1.286374×10^{-2}
$\overline{\text{RMSE}}_{\text{face}}(\rho v)$	4.805584×10^{-3}	4.813296×10^{-3}	8.012499×10^{-3}	8.012188×10^{-3}
$\overline{\text{MAE}}_{\text{face}}(\rho v)$	4.279609×10^{-3}	4.279862×10^{-3}	6.403712×10^{-3}	6.403241×10^{-3}
$\varepsilon_{\text{cons}}(\rho E)$	2.390365×10^{-15}	5.255124×10^{-13}	1.636917×10^{-7}	1.425787×10^{-7}
$\overline{\text{RMSE}}_{\text{blk}}(\rho E)$	9.205312×10^{-3}	9.213652×10^{-3}	1.901099×10^{-2}	1.716526×10^{-2}
$\overline{\text{MAE}}_{\text{blk}}(\rho E)$	7.827272×10^{-3}	7.825163×10^{-3}	1.658196×10^{-2}	1.444322×10^{-2}
$\overline{\text{RMSE}}_{\text{face}}(\rho E)$	5.832684×10^{-3}	5.851590×10^{-3}	9.459973×10^{-3}	9.462631×10^{-3}
$\overline{\text{MAE}}_{\text{face}}(\rho E)$	5.323460×10^{-3}	5.328789×10^{-3}	7.425071×10^{-3}	7.431006×10^{-3}

tematically produce larger discrepancies. Although these operators may still be useful for feature-detection purposes or for bounding analyses, they do not provide representative coarse values in the finite-volume sense and significantly distort both the cell-centered and interface-based description of the flow. This makes them less appropriate for consistent state transfer across resolutions or for subsequent use inside a numerical solver. An important outcome of the study is that the ranking of the operators remains unchanged throughout the flow evolution. The same qualitative and quantitative trends are observed at the initial, intermediate, and final stages of the simulation, indicating that the conclusions are not tied to a particular snapshot but rather reflect intrinsic properties of the restriction procedures themselves. This temporal consistency strengthens the

robustness of the analysis and supports the use of arithmetic averaging as the reference downsampling strategy for this class of data.

The present work constitutes a first step toward the definition of reliable restriction operators for LES-based CFD workflows and, more generally, for future integration with data-driven models acting on coarse–grid representations. The proposed restriction operators are defined in a dimension-independent manner, as they rely on the aggregation of cell-averaged quantities over control volumes, and are therefore directly applicable to fully three–dimensional structured grids. As a next step, we are extending the analysis to the three-dimensional Taylor–Green vortex (see [13]), which represents a more demanding benchmark due to its fully three-dimensional vortex dynamics and richer inter–scale interactions. Preliminary results indicate that the relative performance of the considered downsampling operators remains consistent with the two-dimensional case, with arithmetic and distance-weighted averaging outperforming the extrema-based strategies. For brevity, the corresponding three-dimensional results are not reported here, but we underline that they further support the robustness of the observed trends. Although the study is partly motivated by data-driven CFD applications, its primary objective is to identify downsampling strategies that preserve the physical and numerical consistency of the underlying finite-volume formulation. This aspect is particularly relevant for machine–learning–based approaches, where the quality and physical coherence of the training data play a crucial role in ensuring the reliability of the resulting models.

Acknowledgments. We thank Professor Francesco Capuano (Polytechnic University of Catalonia), Professor Raffaele D'Ambrosio (University of L'Aquila) and Professor Anna Esposito (University of Campania) for the fruitful discussions and useful comments. We also thank the reviewers for their valuable comments. The research of Gerardo Toraldo is partially supported by INdAM-GNCS and the research of Ferdinando Zullo is partially supported by INdAM-GNSAGA.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Berkooz, G., Holmes, P., Lumley, J.L.: The proper orthogonal decomposition in the analysis of turbulent flows. Annu. Rev. Fluid Mech. **25**, 539–575 (1993)
2. Capuano, F., Catalano, P., Mastellone, A.: Large-eddy simulation of a spatially-evolving turbulent mixing layer, APS Division of Fluid Dynamics Meeting Abstracts, R03-005 (2015)
3. Capuano, F., Mastellone, A., De Angelis, E.M.: A conservative overlap method for multi-block parallelization of compact finite-volume schemes. Comput. Fluids **159**, 327 (2017)
4. Capuano, F., Mastellone, A., Benedetto, S.D., Cutrone, L., Schettino, A.: Preliminary developments towards a high-order and efficient LES code for propulsion applications. In: Proceedings of 6th European Conference on Computational Fluid Dynamics (ECFD VI), Barcelona, Spain (2014)

5. Deconinck, H., Koren, B.: Algorithm development in CFD: Multidimensional upwinding and multiple semi-coarsening multigrid. In: Computational methods in applied sciences' 96, Paris, 9-13 September 1996 (1996)
6. Fukami, K., Fukagata, K., Taira, K.: Super-resolution reconstruction of turbulent flows with machine learning. J. Fluid Mech. **870**, 106–120 (2019)
7. Germano, M.: Turbulence: the filtering approach. J. Fluid Mech. **238**, 325–336 (1992)
8. Gullbrand, J., Chow, F.K.: The effect of numerical errors and turbulence models in large-eddy simulations of channel flow. J. Fluid Mech. **495**, 323–341 (2003)
9. Harten, A.: Multiresolution algorithms for the numerical solution of hyperbolic conservation laws. Commun. Pure Appl. Math. **48**(12), 1305–1342 (1995)
10. Kim, H., et al.: Unsupervised deep learning for super-resolution reconstruction of turbulence. J. Fluid Mech. **910**, A29 (2021)
11. Pope, S.B.: Turbulent flows. Meas. Sci. Technol. **12**(11), 2020–2021 (2001)
12. Kochkov, D., et al.: Machine learning–accelerated computational fluid dynamics. Proceedings of the National Academy of Sciences, vol. 118, no. 21 (2021)
13. Mastellone, A., Capuano, F., Benedetto, S.D., Cutrone, L.: Direct numerical simulation of the Taylor-Green Vortex at Re = 1600. In: Proceedings of 3rd International Workshop on High-Order CFD Methods, Kissimmee, USA (2015)
14. Pulido, A., et al.: Survey and analysis of multiresolution methods for turbulence data. Comput. Fluids **125**, 39–58 (2016)
15. Radhakrishnan, S., Bellan, J.: Explicit filtering to obtain grid-spacing-independent and discretization-order-independent large-eddy simulation of compressible single-phase flow. J. Fluid Mech. **697**, 399–435 (2012)
16. Sagaut, P.: Large Eddy Simulation for Incompressible Flows: an introduction. Springer, Heidelberg (2006)
17. Stüben, K.: A review of algebraic multigrid. In: Numerical Analysis: Historical Developments in the 20th Century, pp. 331–359 (2001)
18. Unat, D., Hromadka, T., Baden, S.B.: An adaptive sub-sampling method for in-memory compression of scientific data. In: Data Compression Conference. IEEE (2009)