Consistent data-driven subgrid-scale model development for large-eddy simulation

By X. Huang†, S. C. Leung‡, M. P. Whitmore, A. Elnahhas AND H. J. Bae†

Large-eddy simulations (LESs) using data-driven models suffer from a performance inconsistency between *a-priori* and *a-posteriori* testing due to a numerical gap that is dependent on the numerical method used. We study the effect of the choice of the filtering operator on such inconsistency and its contribution, with comparison to the subgrid-scale (SGS) stress term. By including the numerical gap, a data-driven eddy-viscosity model is able to achieve improved numerical stability. We further model the SGS stress term through sparse regression as a linear combination of selected invariant tensors, and a residual neural network (NN) is trained to account for the modeling error and the numerical gap. Results from both *a-priori* and *a-posteriori* testing of the model show promising improvement to have a consistent model in the forced isotropic turbulence.

1. Introduction

LES has been widely used for its computational affordability compared to direct numerical simulations (DNSs) while being able to preserve energy-containing eddies. The small-scale motions that are not directly resolved in the LES grid are then modeled through SGS models (Meneveau & Katz 2000). In recent years, data-driven techniques have shown the potential to improve the performance of SGS modeling (Park & Choi 2021).

Data-driven models often show significantly better performance in the *a-priori* tests, but these models do not outperform conventional models in the *a-posteriori* tests (Zhou *et al.* 2019). This can be partly explained by the fact that the training data for current data-driven techniques are typically obtained from filtered DNS data, which is not equivalent to LES. To account for the discrepancy in filtered DNS and LES, we introduce a so-called consistent training data extraction method for SGS terms.

The LES equations in terms of the filtered Navier-Stokes equations are expressed as

$$\frac{\partial \overline{u}_i}{\partial t} = -\frac{\partial \overline{u}_i \overline{u}_j}{\partial x_j} - \frac{1}{\rho}\frac{\partial \overline{p}}{\partial x_i} + \nu \frac{\partial^2 \overline{u}_i}{\partial x_j \partial x_j} - \frac{\partial \tau_{ij}}{\partial x_j} + \delta_i^{\text{mom}}, \quad \frac{\partial \overline{u}_i}{\partial x_i} + \delta^{\text{div}} = 0, \tag{1.1}$$

where u_i represents the velocity component in each Cartesian direction, $i \in \{1, 2, 3\}$, x_i is the Cartesian coordinate, ν is the kinematic viscosity, p is the pressure and ρ is the fluid density. We use $\overline{(\cdot)}$ to represent the filtering operator, and $\tau_{ij} = \overline{u_i u_j} - \overline{u}_i \overline{u}_j$ is the SGS stress tensor. The extra terms δ_i^{mom} and δ^{div} are the numerical gap terms that differentiate the *a-priori* tests from the *a-posteriori* tests. Their existence is due to the commutation error between the discrete filter and the differentiation operators and the coarsening of the grid resolution (Bae & Lozano-Duran 2022), making them dependent on the specific choice of filter and numerical scheme.

† Graduate Aerospace Laboratory, California Institute of Technology
‡ Mechanical and Civil Engineering, California Institute of Technology

To develop a model performing consistently in the *a-posteriori* test, one popular solution is to dampen the numerical gap. The *a-priori* comparison shows a relatively smaller contribution by the numerical term when the filter-to-grid ratio in LES increases (Chow & Moin 2003). However, LES with a larger filter size is more computationally demanding, and this only partially damps the numerical gap, for example, the aliasing error.

Additional efforts have been made to design the filtering appropriate for a consistent LES, including developing commutative filtering operators (Vasilyev *et al.* 1998) and explicit filtering (Bose *et al.* 2010). However, the numerical discrepancy is unavoidable when the resolution changes from DNS to LES (Bae & Lozano-Duran 2022), which will essentially change the formulation of the differential operators during discretization. Despite the importance of filtering, the impact of filtering on the numerical gap term that emerges in developing a consistent SGS model for the LES simulation is unclear.

Instead of damping the numerical gap, the possibility of directly modeling this term has also been explored. As one of the pioneer works, Langford & Moser (1999) developed the idea of the optimal subgrid model, though the optimal models show large errors in the energy spectrum due to the limited number of model forms. With the assistance of data-driven techniques, nonlinear models with greater complexity can be developed. However, the models need nonlocal information in either space (Beck *et al.* 2019), or time (Kurz & Beck 2020). In addition, the application of the models in *a-posteriori* simulations may still display numerical instability (Beck *et al.* 2019).

In this proceeding, we will first examine the importance of filtering and provide details of the training data set in Section 2. Then, an SGS model using an NN and sparse nonlinear models will be developed and applied in Sections 3 and 4. We then show how the filtering will change the consistency between the *a-priori* tests during training and the *a-posteriori* simulations. Finally, conclusions will be drawn in Section 5.

2. The importance of filtering

2.1. *DNS of homogeneous isotropic turbulence*

We use forced homogeneous isotropic turbulence, which allows a comprehensive study of the energy spectrum. DNS is conducted with an incompressible flow solver that uses a second-order staggered finite difference scheme in space and a third-order Runge-Kutta method in time with a fractional step method. The Taylor microscale Reynolds number is $Re_\lambda = \sqrt{2k/3}\lambda/\nu = 180$, where k is the turbulent kinetic energy, ε is the dissipation rate and $\lambda = \sqrt{10k/\varepsilon}$. The Kolmogorov length scale is $\eta = \sqrt[4]{\nu^3/\varepsilon}$, and the Kolmogorov time scale is $\tau_\eta = \sqrt{\nu/\varepsilon}$. A linear forcing will be used to provide stationary turbulence as proposed by Lundgren (2003). We use a periodic box of size $(2\pi L_x)^3$ with a resolution of 256^3, where L_x is the characteristic length for the domain, which corresponds to $\Delta_{\mathrm{DNS}} \approx 5.1\eta$. The time step is $\Delta t_{\mathrm{DNS}} = 0.013\tau_\eta$.

2.2. *Filtering in LES*

Along with the DNS, we simultaneously run an LES with SGS forcing computed from the DNS field (Bae & Lozano-Duran 2018). According to Eq. (1.1), we can compute δ_i^{mom} through a comparison between the filtered Navier-Stokes forcing on the LES grid and the LES forcing, where the SGS stress term is from the DNS field. Here, we chose a filter with zero divergence in the LES grid such that δ^{div} is negligible; we use $\delta_i \equiv \delta_i^{\mathrm{mom}}$ for brevity.

In order to achieve consistency with the filtered DNS, we first construct the initial

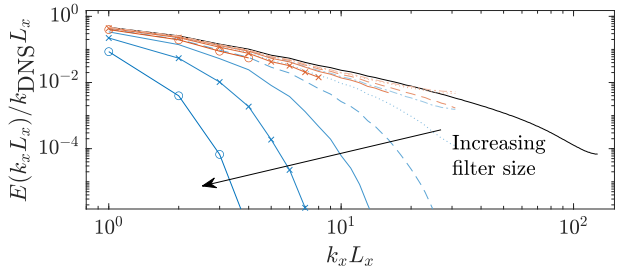

FIGURE 1. One-dimensional kinetic energy spectrum from DNS (black), LES with Gaussian filter (blue) and spectral cutoff filter (orange). Filter widths are $\sigma/\Delta_{\mathrm{LES}}$ and $\lambda_c/\Delta_{\mathrm{LES}}$ of 0.25 (dash-dotted), 0.5 (dotted), 1.0 (dashed), 2.0 (solid), 4.0 (crosses) and 8.0 (circles).

condition of the LES simulation using a filtered DNS solution $\overline{u}_i$ interpolated to the LES grid. Then, the SGS stress term is computed on the LES grid using the DNS solution as $\tau_{ij} = \overline{u_i u_j} - \overline{u}_i \overline{u}_j$. Finally, δ_i is computed by comparing the filtered DNS to the LES solution at the same time step. Note that both the SGS stress term τ_{ij} and the numerical gap term δ_i are unknown and need modeling, and they are dependent on the choice of the filtering operator.

If the filtering operator is assumed to be homogeneous and isotropic, the effect of the filtering kernel G can be seen as a modulation to the wave number $\boldsymbol{k}$ in the spectral space after the Fourier transformation $\widehat{(\cdot)}$ (Pope 2000),

$$\widehat{\overline{\phi}}(\boldsymbol{k}) = \widehat{G}(\boldsymbol{k})\widehat{\phi}(\boldsymbol{k}). \tag{2.1}$$

Filters can be sharp or smooth in the spectral space, which leads to qualitative differences in modeling unresolved terms (De Stefano & Vasilyev 2002). We consider the spectral cutoff filter as the representative of the sharp filtering, where $\widehat{G}(\boldsymbol{k}) = H(k_c - |\boldsymbol{k}|)$, and the Gaussian filter as the representative of the smooth filtering, where $\widehat{G}(\boldsymbol{k}) = \exp\left(-|\boldsymbol{k}|^2\sigma^2/2\right)$. $H(\cdot)$ is the Heaviside step function, k_c is the cutoff wave number, defining the spectral cutoff wavelength $\lambda_c = 2\pi/k_c$ and the width of the Gaussian filter σ. All filtered quantities reside on the LES grid, and the high-frequency components that are no longer resolvable are discarded (downsampled).

2.3. *Training data set*

We create multiple data sets using different filtering for the training of data-driven models utilizing the method introduced in the previous section. The LES simulations reside in the same periodic box as DNS with the resolution fixed to be 64^3 in this study. Since the velocity field is formulated on the staggered grid, the filtered DNS field is spectrally interpolated to the LES grid. We only consider the deviatoric part of the SGS stress τ_{ij}^d since the isotropic residual stress is included in the modified filtered pressure,

$$\tau_{ij}^d = \tau_{ij} - \frac{1}{3}\tau_{kk}\delta_{ij}, \tag{2.2}$$

where δ_{ij} indicates the Kronecker delta. The superscript $(\cdot)^d$ will be neglected from here on for convenience.

We examine the Gaussian filter with filter widths $\sigma/\Delta_{\mathrm{LES}}$ of 0.25, 0.5, 1.0, 2.0, 4.0 and 8.0, and the spectral cutoff filter with cutoff wavelengths $\lambda_c/\Delta_{\mathrm{LES}}$ of 0.25, 0.5, 1.0, 2.0, 4.0 and 8.0, where Δ_{LES} is the LES grid size. In Figure 1, the kinetic energy spectra for the examined filters are compared with the DNS spectrum. We see damping in the high-

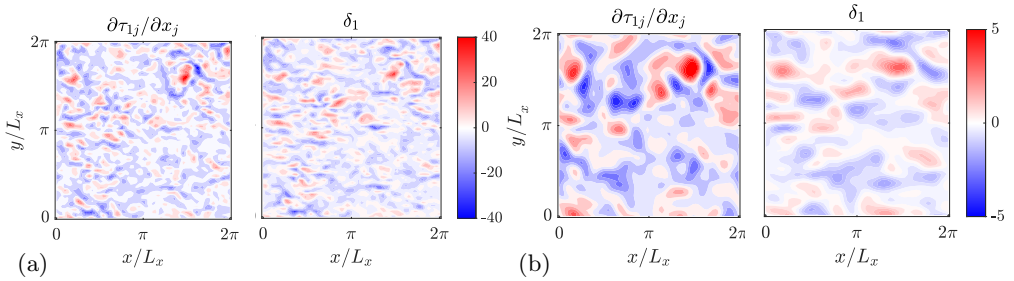

FIGURE 2. Contribution of the SGS term $\partial\tau_{1j}/\partial x_j$ (left) and the numerical gap term δ_1 (right) in the x-direction for the Gaussian filter with (a) $\sigma/\Delta_{\mathrm{LES}} = 0.5$ and (b) 2.0. All components normalized by $\varepsilon/\sqrt{2k/3}$.

frequency components for a smooth filter as the Gaussian filter, and an abrupt cutoff in the high-frequency components for a sharp filter as the spectral cutoff filter.

The LES simulation assumes that the unresolved quantities, τ_{ij} and δ_i, can be modeled as functions of resolved quantities. We use the strain rate and rotation rate tensors,

$$\boldsymbol{S} \equiv S_{ij} = \frac{1}{2}\left(\frac{\partial \overline{u}_i}{\partial x_j} + \frac{\partial \overline{u}_j}{\partial x_i}\right), \quad \boldsymbol{R} \equiv R_{ij} = \frac{1}{2}\left(\frac{\partial \overline{u}_i}{\partial x_j} - \frac{\partial \overline{u}_j}{\partial x_i}\right), \tag{2.3}$$

to formulate a model for τ_{ij} that allows for Galilean invariance (Pope 2000). Similarly, we utilize the finite difference scheme to calculate the quantities for model formulation to keep the training process consistent with the LES simulation.

2.4. *Flow statistics*

In Figure 2, we compare δ_1 and the contribution of the SGS stress term in the x-momentum forcing, $\partial\tau_{1j}/\partial x_j$ for different filters. In Figure 2(a), the importance of δ_i is comparable to the SGS stress term when the filter size is small for a Gaussian filter. Although it is not shown, this also occurs with the spectral cutoff filter, which indicates that locally δ_i is playing an important role. However, the relative importance is qualitatively reduced as the filter size increases. In Figure 2(b), δ_i is small compared to the SGS stress term. Though increasing the filter size helps to alleviate the impact of the numerical scheme, δ_i cannot be fully neglected locally.

To get a quantitative comparison between δ_i and $\partial\tau_{ij}/\partial x_j$, we examine the budget of these unresolved quantities. We obtain the equation for the kinetic energy $k_{\mathrm{LES}} = \overline{u}_i\overline{u}_i/2$ when neglecting the higher-order numerical terms due to commutation error, that is,

$$\frac{\partial k_{\mathrm{LES}}}{\partial t} = \mathcal{P} + \mathcal{T} - \underbrace{\nu\frac{\partial \overline{u}_i}{\partial x_j}\frac{\partial \overline{u}_i}{\partial x_j}}_{\varepsilon_{\mathrm{visc}}} - \underbrace{(-\tau_{ij}S_{ij})}_{\varepsilon_{\mathrm{SGS}}} - \underbrace{(-\overline{u}_i\delta_i)}_{\varepsilon_{\mathrm{num}}}. \tag{2.4}$$

Here, $\mathcal{P}$ is the production; $\mathcal{T}$ is the transport between neighboring grids; and the flow field is dissipated by the viscous term $\varepsilon_{\mathrm{visc}}$, the SGS stress $\varepsilon_{\mathrm{SGS}}$ and the numerical dissipation $\varepsilon_{\mathrm{num}}$. We compute the ratio of the local dissipation rate $r_\varepsilon = \varepsilon_{\mathrm{num}}/\varepsilon_{\mathrm{SGS}} = |\overline{u}_i\delta_i|/|\tau_{ij}S_{ij}|$ for three filter sizes using the Gaussian filter in Figure 3. As the filter size increases, the local numerical gap dissipation rate becomes relatively negligible as $\varepsilon_{\mathrm{num}}/\varepsilon_{\mathrm{SGS}} \ll 1$. Though not shown, the same conclusion is also seen for the spectral cutoff filter.

A global dissipation rate can be computed by averaging the local dissipation rate within the field due to homogeneity. In Figure 4(a), we compare the relative contribution of the numerical gap dissipation $r_\varepsilon = -\int \overline{u}_i\delta_i dV / -\int \tau_{ij}S_{ij}dV$ for all the filter types and

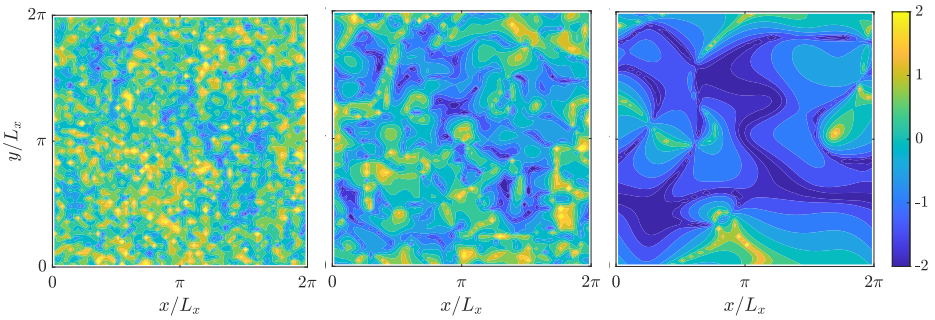

FIGURE 3. Local relative dissipation ratio $\log(|u_i\delta_i|/|\tau_{ij}S_{ij}|)$ for the Gaussian filter with $\sigma/\Delta_{\mathrm{LES}}$ of 0.5, 2.0 and 8.0.

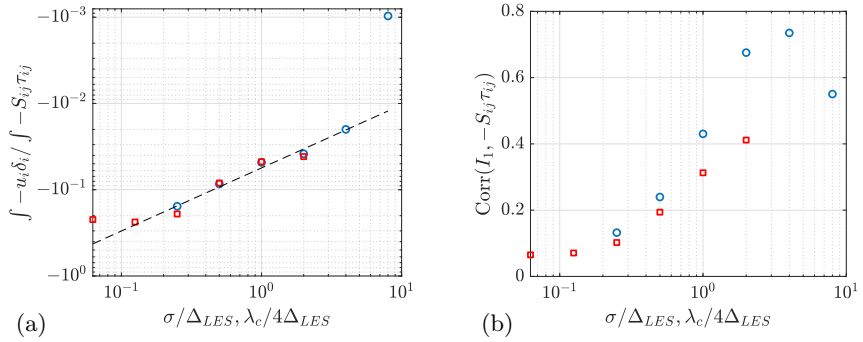

FIGURE 4. (a) Global relative dissipation ratio between $\int u_i\delta_i dV$ and $\int \tau_{ij}S_{ij}dV$ as a function of filter widths $\sigma/\Delta_{\mathrm{LES}}$ and $\lambda_c/4\Delta_{\mathrm{LES}}$ for the Gaussian (blue) and spectral cutoff (red) filters. The Black dashed line indicates the fitted power law. (b) Correlation coefficient between the SGS dissipation rate and $I_1 = \mathrm{tr}(\boldsymbol{S}^2)$ for the Gaussian (blue) and spectral cutoff (red) filters.

filter sizes we examined. As expected, the global relative dissipation rate $|r_\varepsilon|$ decreases with the filter size, similar to the local relative dissipation rate. Note that the global numerical gap dissipation rate is not necessarily positive, and for our data set using spectral interpolation, it will be negative. In this way, we are able to find an equivalence between the two filters under consideration based on r_ε. Between the spectral cutoff filter with the cutoff wavelength λ_c and the Gaussian filter with the filter width σ, they are usually equivalent when $\sigma \approx \lambda_c/4$. For a filter with a moderate filter size, we can fit a power law between r_ε and the filter size. Specifically, for the two filter types examined, the relationship is

$$r_\varepsilon^{\mathrm{Gaussian}}(\sigma/\Delta_{\mathrm{LES}}),\ r_\varepsilon^{\mathrm{spectral}}(\lambda_c/4\Delta_{\mathrm{LES}}) \approx -0.05591 r_L^{-0.7310}. \tag{2.5}$$

Here, the relative length is $r_L = \sigma/\Delta_{\mathrm{LES}}$ for the Gaussian filter, and $r_L = \lambda_c/4\Delta_{\mathrm{LES}}$ for the spectral cutoff filter. This is an empirical law for a moderate filter size and needs to be further examined with other filter types. For both filters, increasing the filter size leads to decreasing r_ε, indicating that the numerical gap term has a lower impact in developing SGS models.

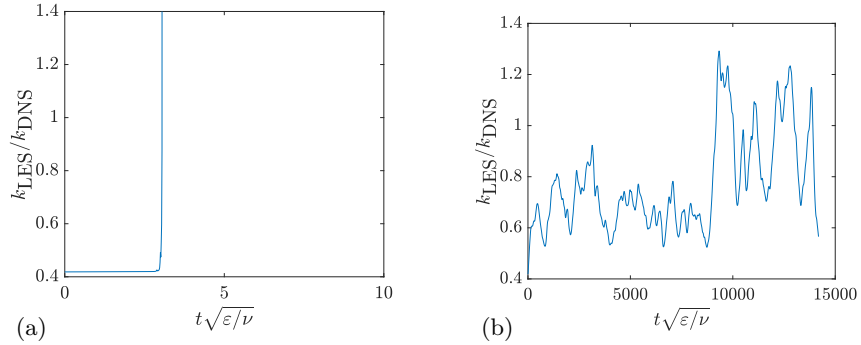

FIGURE 5. Evolution of kinetic energy in the *a-posteriori* simulations for the eddy-viscosity model that considers (a) only the SGS stress term and (b) both the SGS stress term and δ_i.

3. A data-driven eddy-viscosity model

According to Section 2, the existence of δ_i is comparatively important to the SGS model as the SGS stress term, and it is dependent on the choice of the filter and the specific numerical scheme. Here, we try to develop the eddy-viscosity model for only the SGS stress term and for both the SGS stress term and δ_i to examine how δ_i will change the development of the SGS model and its performance. To be exact, the modeled SGS stress term will be $\tau_{M,ij} = -2\nu_T S_{ij}$.

The eddy viscosity ν_T is considered a function of the scalar invariants

$$\begin{aligned} &I_1 = \mathrm{tr}(\boldsymbol{S}^2), \quad I_2 = \mathrm{tr}(\boldsymbol{R}^2), \quad I_3 = \mathrm{tr}(\boldsymbol{S}^3), \\ &I_4 = \mathrm{tr}(\boldsymbol{S}\,\boldsymbol{R}^2), \quad I_5 = \mathrm{tr}(\boldsymbol{S}^2\boldsymbol{R}^2), \quad I_6 = \mathrm{tr}(\boldsymbol{S}^2\boldsymbol{R}^2\boldsymbol{S}\,\boldsymbol{R}). \end{aligned} \tag{3.1}$$

We train a feed-forward NN with the six scalar invariants as the input features and the local eddy viscosity as the output features, and the size is fixed to be three hidden layers with 20 neurons in each layer according to the balance between the training efficiency and the training accuracy. Both the input and output features will be nondimensionalized by the characteristic velocity scale $u_{\mathrm{char}} = 10(\sqrt{2S_{ij}S_{ij}}\nu)^{1/2}$ and the characteristic length scale $l_{\mathrm{char}} = 0.2\sqrt{3}\Delta_{\mathrm{LES}}$, which are both local information available even in cases of complex geometry present in real-world applications. The coefficients in the characteristic scales will normalize the features to a similar order of magnitude so that the training can be numerically stabilized. If the training includes δ_i, we further consider $\sqrt{2k/3}$ as an extra input feature to inform the magnitude of δ_i as an extra feature for the specific numerical scheme. The loss function will be the error in the dissipation rate $\tau_{ij}S_{ij}$. In Figure 4(b), we examine the correlation coefficient between the SGS dissipation rate and I_1. The correlation level increases with the filter size, which indicates the possibility of higher instantaneous accuracy of the SGS modeling. The data set for training uses a Gaussian filter with the moderately high filter width $\sigma/\Delta_{\mathrm{LES}} = 2.0$ to obtain a reasonable SGS stress modeling.

In Figure 5, we show the development of the kinetic energy in the *a-posteriori* simulations for both models. As we can see, the model that only considers the SGS stress term is very sensitive to numerical instabilities, and blows up very fast after the initiation. In contrast, the model that considers both the SGS stress term and the numerical gap term in the training performs much better, with a stabilized kinetic energy in the simulation. The inclusion of δ_i is necessary for stabilizing the LES simulation.

4. Sparse nonlinear model

Sparse regression has been widely adopted to identify nonlinear dynamic systems. Initially proposed by Lund & Novikov (1992), it applies Lasso regression to select the most contributing terms from a set of given variables and assign corresponding weights. Here, we extend its application to derive a simple SGS model, represented as a linear combination of the selected invariant tensors. Additionally, an NN residual model is incorporated alongside the linear model to correct the initial predictions of the SGS stress, enhancing the overall accuracy.

4.1. *Sparse nonlinear training*

Utilizing sparse regression, we can model the SGS stress as a linear combination of the products between the tensor invariants

$$\begin{array}{lll} \boldsymbol{m}_1 = \boldsymbol{S}, & \boldsymbol{m}_2 = \boldsymbol{S}^2, & \boldsymbol{m}_3 = \boldsymbol{R}^2, \\ \boldsymbol{m}_4 = \boldsymbol{S}\boldsymbol{R} - \boldsymbol{R}\boldsymbol{S}, & \boldsymbol{m}_5 = \boldsymbol{S}^2\boldsymbol{R} - \boldsymbol{R}\boldsymbol{S}^2, & \boldsymbol{m}_6 = \boldsymbol{I}, \\ \boldsymbol{m}_7 = \boldsymbol{S}\boldsymbol{R}^2 + \boldsymbol{R}^2\boldsymbol{S}, & \boldsymbol{m}_8 = \boldsymbol{R}\boldsymbol{S}\boldsymbol{R}^2 - \boldsymbol{R}^2\boldsymbol{S}\boldsymbol{R}, & \boldsymbol{m}_9 = \boldsymbol{S}\boldsymbol{R}\boldsymbol{S}^2 - \boldsymbol{S}^2\boldsymbol{R}, \\ \boldsymbol{m}_{10} = \boldsymbol{S}^2\boldsymbol{R}^2 + \boldsymbol{R}^2\boldsymbol{S}^2, & \boldsymbol{m}_{11} = \boldsymbol{R}\boldsymbol{S}^2\boldsymbol{R}^2 - \boldsymbol{R}^2\boldsymbol{S}^2\boldsymbol{R} & \end{array} \tag{4.1}$$

and the scalar invariants $I_1 \dots I_6$ established by Lund & Novikov (1992). A regression-based SGS model takes the form of $\boldsymbol{\tau}_M = \sum_k C_{k,\tau}\boldsymbol{m}_k$, where $C_{k,\tau}$ are the constant coefficients determined through regression.

Data for the invariants as well as $\boldsymbol{\tau}_M$ is locally computed and scaled by the characteristic velocity and length scale of $u_{\text{char}}/10$ and $50l_{\text{char}}$, respectively, where u_{char} and l_{char} are defined as in Section 3. The trace is removed from the invariant tensors to maintain incompressibility. The training set contains data from two consecutive time snapshots, and the testing set from another snapshot.

The linear regression algorithm is conducted via PyTorch with a batch size of 64, and the mean-squared-error (MSE) loss function. As in Lasso regression, sparsity in the regression models is enforced by adding L_1 regularization, expressed by $\lambda_{L_1}\sum_{k=1}^{N}|C_{k,\tau}|$ where N is the number of terms, onto the loss function. In our study, the regularization strength of $\lambda_{L1} = 0.2$ is found to best maintain the balance between model simplicity and accuracy. To evaluate model accuracy, the relative root–MSE (rRMSE) is computed for the quantities of interest. For example,

$$\text{rRMSE}(\boldsymbol{\tau}_M) = \sqrt{\frac{\sum(\boldsymbol{\tau}_M - \boldsymbol{\tau})^2}{\sum \boldsymbol{\tau}^2}} \tag{4.2}$$

is the rRMSE for $\boldsymbol{\tau}_M$. With the converged coefficients, the invariant terms are ranked based on the magnitude of each coefficient multiplied by the average norm of its corresponding term. Models with varying numbers of terms are developed according to this ranking and then evaluated individually. The optimal number of invariant terms is determined by examining the variation in rRMSE as the number of terms increases (Figure 6).

The three-term model obtained from sparse regression for LES, using a Gaussian filter with a filter size of $2\Delta_{\text{LES}}$, is given by

$$\frac{\boldsymbol{\tau}_M}{u_{\text{char}}^2} = -\frac{166.58}{S_{\text{char}}^2}\boldsymbol{m}_4 - \frac{67.16}{S_{\text{char}}}\boldsymbol{m}_1 - \frac{166.82}{S_{\text{char}}^2}\boldsymbol{m}_3, \tag{4.3}$$

where $S_{\text{char}} = u_{\text{char}}/(5l_{\text{char}})$. The $\boldsymbol{m}_4$, $\boldsymbol{m}_3$ and $\boldsymbol{m}_1$ are considered the contributing invariant terms for predicting the SGS stress. In our *a-priori* tests, this model yields an rRMSE($\boldsymbol{\tau}_M$) of 0.704. For reference, the classical Smagorinsky model produces an

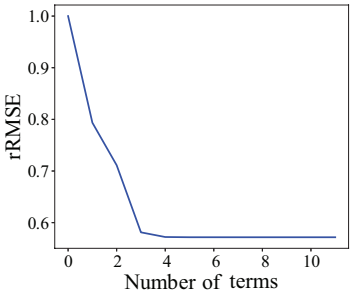

FIGURE 6. rRMSE of $\boldsymbol{\tau}_M$ for the number of invariant terms to be included for the SGS model.

rRMSE($\boldsymbol{\tau}_M$) of 0.928 [Figure 7(a)]. To better assess the model's *a-posteriori* performance when applied to the momentum equation, we compute and evaluate the divergence of $\boldsymbol{\tau}_M$ or $\partial\tau_{M,ij}/\partial x_j$. On the testing set, the model results in an rRMSE($\partial\tau_{M,ij}/\partial x_j$) of 0.817. We also explore higher-order models that incorporate products of tensor invariants and scalar invariants. However, they do not yield significant improvements in accuracy, at least for the cases tested, and are therefore not reported here.

To improve accuracy, we utilize an NN to capture the residuals from the regression models and the numerical gap. With the added contribution from the residual model, the divergence of the SGS stress can now be represented as

$$\left[\frac{\partial\tau_{M,ij}}{\partial x_j}\right]_{\text{corr}} = \frac{\partial\tau_{M,ij}}{\partial x_j} + \varepsilon_{\text{NN},i} + \delta_{\text{NN},i}, \tag{4.4}$$

where $[\partial\tau_{M,ij}/\partial x_j]_{\text{corr}}$ is the corrected SGS term, $\varepsilon_{\text{NN},i}$ is the predicted residual, and $\delta_{\text{NN},i}$ is the predicted numerical gap. The NN takes as inputs the divergence of the invariant tensors $\nabla\cdot(m_{k,ij})$, the convective term $\partial u_i u_j/\partial x_j$, the viscous term $\nu\left(\partial^2 u_i/\partial x_j\partial x_j\right)$ and the SGS term modeled by the regression $\partial\tau_{M,ij}/\partial x_j$, all of which are locally scaled by the characteristic quantities. It then outputs $\varepsilon_{\text{NN},i}+\delta_{\text{NN},i}$. The NN is a multilayer perceptron consisting of two fully connected hidden layers of layer size 32. All layers are followed by a Rectified Linear Unit activation, except for the last one, which uses a sigmoid activation to help constrain the output. The standard MSE loss function and the Adam optimizer with variable learning rates depending on the filter are utilized for training. During *a-priori* testing, the combination of the residual model and the regression model given in Eq. (4.3) achieves a 10% reduction in the rRMSE($\partial\tau_{M,ij}/\partial x_j$), improving the accuracy of the model.

4.2. *The impact of filtering*

The filter type and size greatly impact the modeling results, especially regarding the trainability of the SGS stress and the model formulation. As shown in Figure 7(a), the rRMSE($\boldsymbol{\tau}_M$) of the model decreases to levels indicating a reasonable learning outcome as the filter size increases. This leads to a trade-off between the model's trainability and the retention of small-scale information. With varying filter types and sizes, the sparse regression method generates different coefficients for the invariant terms. As displayed in Figure 7(b), the regression coefficients for all variables follow a near power-law behavior with respect to the filter size, with an exponent approximately equal to 2. An order of magnitude discrepancy in the regression coefficients is observed between the Gaussian and the spectral cutoff filters.

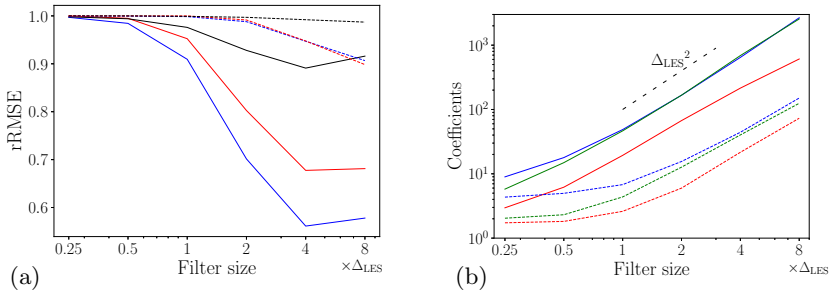

FIGURE 7. (a) rRMSE of $\tau_{M,ij}$ (blue), $\partial\tau_{M,ij}/\partial x_j$ (red) and the classical Smagorinsky model (black) for a Gaussian (solid) and spectral cut-off (dashed) filter. (b) Magnitude of the regression coefficients of $\boldsymbol{m}_4$ (blue), $\boldsymbol{m}_1$ (red), and $\boldsymbol{m}_3$ (green) for Gaussian (solid) and spectral cut-off (dashed) filter. Black dotted line indicates $O(\Delta_{\mathrm{LES}}{}^2)$.

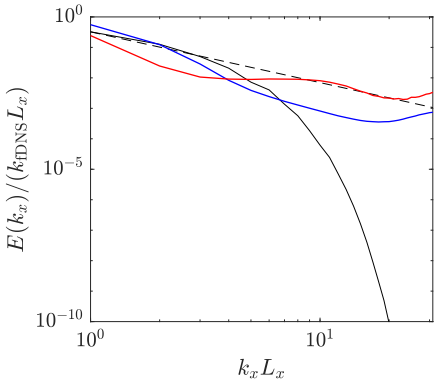

FIGURE 8. *A-posteriori* energy spectra of the regression model with (blue) and without (red) the residual correction provided by the NN. These are compared against the filtered DNS (black solid) and the $-5/3$ law (black dashed).

4.3. *A-posteriori testing*

As demonstrated by the energy spectra of *a-posteriori* tests in Figure 8, the residual correction step, which compensates for the errors from the regression model and numerical operators, helps stabilize the simulation. The spectra with the residual model exhibit not only significant suppression of energy accumulation in the smaller scales but also improved alignment with the filtered DNS spectrum in the larger scales.

5. Conclusion

In this work, we focus on the performance gap between the *a-priori* and *a-posteriori* tests for a data-driven model, which is highly determined by the specific numerical scheme. We examine this numerical gap, *i.e.*, δ_i, and compare it to the SGS stress τ_{ij} which is traditionally the only modeled term. The comparison of the contribution of the two terms using the DNS flow field shows that δ_i is locally important, and this importance can be reduced by using a filter with a larger filter size. The global relative dissipation ratio between the numerical gap and the SGS stress can find the equivalence in different filter types, and a power law can be fit for the two types of filter we examined, the Gaussian filter and the spectral cutoff filter.

Additionally, we try to improve the consistency of data-driven models by including

the numerical gap. A data-driven eddy-viscosity model is developed and including the numerical gap in the training will stabilize the simulation. To further understand the importance of filtering to modeling, we further develop sparse nonlinear models for the SGS stress. The SGS tensor is thus a linear combination of multiple nonlinear tensor invariants. We can thus achieve good accuracy when using only three terms in the model through the analysis of sparse regression. Our model error shows a significant decrease with increasing filter size, and we find the relationship of the coefficients with the filter size. With the residual NN model correction for both the error in the regression-based SGS model and the numerical gap term, the sparse nonlinear model shows promising *a-posteriori* test results, as evidenced by the stabilization of the simulation. Further investigation will be done to analyze the impact of different numerical schemes.

REFERENCES

Bae, H. J. & Lozano-Duran, A. 2018 DNS-aided explicitly filtered LES of channel flow. *Annual Research Briefs*, Center for Turbulence Research, Stanford University, pp. 197–207.

Bae, H. J. & Lozano-Duran, A. 2022 Numerical and modeling error assessment of large-eddy simulation using direct-numerical-simulation-aided large-eddy simulation. *arXiv preprint arXiv:2208.02354*.

Beck, A., Flad, D. & Munz, C.-D. 2019 Deep neural networks for data-driven LES closure models. *J. Comput. Phys.* **398**, 108910.

Bose, S. T., Moin, P. & You, D. 2010 Grid-independent large-eddy simulation using explicit filtering. *Phys. Fluids* **22** (10), 105103.

Chow, F. K. & Moin, P. 2003 A further study of numerical errors in large-eddy simulations. *J. Comput. Phys.* **184** (2), 366–380.

De Stefano, G. & Vasilyev, O. V. 2002 Sharp cutoff versus smooth filtering in large eddy simulation. *Phys. Fluids* **14** (1), 362–369.

Kurz, M. & Beck, A. 2020 A machine learning framework for LES closure terms. *arXiv preprint arXiv:2010.03030*.

Langford, J. A. & Moser, R. D. 1999 Optimal LES formulations for isotropic turbulence. *J. Fluid Mech.* **398**, 321–346.

Lund, T. S. & Novikov, E. 1992 Parameterization of subgrid-scale stress by the velocity gradient tensor. *Annual Research Briefs*, Center for Turbulence Research, Stanford University, pp. 27–43.

Lundgren, T. S. 2003 Linearly forced isotropic turbulence. *Annual Research Briefs*, Center for Turbulence Research, Stanford University, pp. 461–473.

Meneveau, C. & Katz, J. 2000 Scale-invariance and turbulence models for large-eddy simulation. *Annu. Rev. Fluid Mech.* **32** (1), 1–32.

Park, J. & Choi, H. 2021 Toward neural-network-based large eddy simulation: Application to turbulent channel flow. *J. Fluid Mech.* **914**, A16.

Pope, S. 2000 *Turbulent Flows*. Cambridge University Press.

Vasilyev, O. V., Lund, T. S. & Moin, P. 1998 A general class of commutative filters for LES in complex geometries. *J. Comput. Phys.* **146** (1), 82–104.

Zhou, Z., He, G., Wang, S. & Jin, G. 2019 Subgrid-scale model for large-eddy simulation of isotropic turbulent flows using an artificial neural network. *Comput. Fluids* **195**, 104319.