Optimal estimator and artificial neural network as efficient tools for the subgrid-scale scalar flux modeling

By A. Vollant†, G. Balarac†, G. Geraci‡ AND C. Corre†

This work is devoted to exploring a new procedure to develop subgrid-scale (SGS) models in the context of large-eddy simulation (LES) of a passive scalar. Starting from the Noll's formula (Noll 1967), the concept of an optimal estimator is first used to determine an accurate set of parameters to derive a SGS model. The SGS model is then defined as a surrogate model built from this set of parameters by training an artificial neural network (ANN) on a filtered DNS database. This ANN model is next compared with the dynamic nonlinear tensorial diffusivity (DNTD) model proposed by Wang *et al.* (2007). The DNTD model is also based on Noll's formula and can be seen as a nonlinear extension of the dynamic eddy-diffusivity (DED) model proposed by Moin *et al.* (1991). The *a priori* and *a posteriori* tests performed on the ANN model demonstrate the ability of this new model to well reproduce the behavior of the exact SGS term and show an improvement in comparison with DED and DNTD models. The concept of an optimal estimator associated with a machine-learning procedure thus appears as a useful tool for SGS model development.

1. Introduction

Various applications need to solve a scalar equation simultaneously with the governing flow equations. In these applications, the scalar can represent the temperature field or the concentration of chemical species in combustion, mixing, or heat transfer studies. Owing to the large range of motion scales in turbulent flows, the direct numerical simulation (DNS) of realistic applications is not yet available because of significant computational cost. To overcome this limitation, the LES technique proposes to explicitly solve only the large scales of the flow and to model the smallest scales. This separation between resolved large scales and modeled small scales is performed by a filtering operation

$$\bar{f}(\vec{x},t) = \int f(\vec{y},t)G(\vec{x}-\vec{y})d\vec{y}, \tag{1.1}$$

to obtain the large-scale resolved field, $\bar{f}$, from the turbulent field, f, with G the filter kernel. The filtered transport equation for a passive scalar, Z, is given by

$$\frac{\partial \bar{Z}}{\partial t} + \bar{u}_i \frac{\partial \bar{Z}}{\partial x_i} = \frac{\nu}{Sc}\frac{\partial^2 \bar{Z}}{\partial x_i^2} - \frac{\partial \tau_i}{\partial x_i}, \tag{1.2}$$

where $\bar{Z}$ is the resolved passive scalar, $\bar{u}_i$ is the component of the filtered velocity in the direction x_i, ν is the kinematic viscosity, and Sc is the molecular Schmidt number. In this equation, the SGS scalar flux divergence, $\partial\tau_i/\partial x_i$, with $\tau_i = \overline{u_i Z} - \bar{u}_i\bar{Z}$, has to

† Grenoble-INP/CNRS/UJF-Grenoble 1, LEGI UMR 5519, France
‡ Inria Bordeaux, France

be modeled to perform LES. The SGS model is an algebraic expression using resolved (large-scales) quantities as input parameters, which is expected to correctly predict the SGS term and its effects on the resolved field. In the context of passive scalar LES, the usual modeling strategy is based on the definition of an eddy-diffusivity, D_T, to model the SGS scalar flux as $\tau_i = -D_T \partial \bar{Z}/\partial x_i$. Moin *et al.* (1991) introduce a dynamic model for D_T, similarly to the dynamic Smagorinsky model used to model the eddy viscosity (Germano *et al.* 1991). This dynamic eddy-diffusivity (noted DED) model is defined as

$$\tau_i^{\mathrm{DED}} = -D_T \frac{\partial \bar{Z}}{\partial x_i} = C \bar{\Delta}^2 |\bar{S}| \frac{\partial \bar{Z}}{\partial x_i}, \tag{1.3}$$

with $\bar{\Delta}$ the filter size, $|\bar{S}| = \left(2\bar{S}_{ij}\bar{S}_{ij}\right)^{1/2}$ the norm of the filtered strain rate tensor, $\bar{S}_{ij}$, and C the dynamic coefficient. More complex models can be proposed based on numerical and physical considerations. For example, Wang *et al.* (2007) have recently proposed to extend the DED model. According to the theory of tensor invariants and functions, a vector-valued function of a second-order symmetric tensor M_{ij} and a vector v_i can be represented by Noll's formula (Zheng 1994; Noll 1967). From this formula, a general form of the SGS scalar flux can be written as

$$\tau_i = f_1 v_i + f_2 M_{ij} v_j + f_3 M_{ik} M_{kj} v_j, \tag{1.4}$$

where f_1, f_2, and f_3 are coefficients. Wang *et al.* (2007) propose to define v_i equal to $\bar{\Delta}^2 |\bar{S}| \partial \bar{Z}/\partial x_i$. With this definition, the coefficients and the symmetric tensor, M_{ij}, have to be dimensionless. M_{ij} can be generally defined as $M_{ij} = M^*_{ij}/|M^*|$, with M^*_{ij} a dimensional symmetric tensor. Wang *et al.* (2007) propose to take M^*_{ij} equal to $\bar{S}_{ij}$ and to compute coefficients with a dynamic procedure close to the procedure proposed by Germano *et al.* (1991). This enables defining the dynamic nonlinear tensorial diffusivity (DNTD) model as

$$\tau_i^{DNTD} = \chi_1 \bar{\Delta}^2 |\bar{S}| \frac{\partial \bar{Z}}{\partial x_i} + \chi_2 \bar{\Delta}^2 \bar{S}_{ik} \frac{\partial \bar{Z}}{\partial x_k} + \chi_3 \bar{\Delta}^2 \frac{\bar{S}_{ik}\bar{S}_{kl}}{|\bar{S}|} \frac{\partial \bar{Z}}{\partial x_l}, \tag{1.5}$$

with χ_1, χ_2, and χ_3 the dynamic coefficients. Note that by keeping only the first term of the RHS in Eq. (1.5), the DED model is recovered. The DNTD model can thus be considered as a nonlinear extension of the dynamic eddy-diffusivity model. Note also that Lund & Novikov (1992) have already used such a decomposition to model the SGS stress tensor.

The goal of the present work is to investigate a new procedure to develop accurate SGS models. The starting point of the proposed procedure is to take advantage of the growing available computational resources, which now can generate a large DNS database. In the field of big data processing, this DNS database associated with explicit filtering can be used to better understand SGS model performance. Moreover, with the recent development of machine-learning procedures, it is expected that an accurate SGS model will be directly learned from this DNS database. Such a procedure would be a useful tool for model development, allowing for a means to better measure the performance of classic algebraic SGS model in comparison with the best available model derived from a machine-learning procedure. Building on a modeling error decomposition, the new model development procedure is made of two main steps. First, the concept of an optimal estimator is used to determine an appropriate set of input parameters. The model is then derived by building a surrogate model based on this set of parameters, using an artificial neural network (ANN) training. This procedure is applied to the modeling of

the SGS scalar flux divergence. It will be hereafter established that such an ANN model exhibits good performances, in comparison with the DNTD and the DED models.

2. SGS model-development procedure

In this section, the proposed strategy to develop a SGS model is described in two steps. This strategy is based on a DNS database, used to extract exact filtered quantities. In this work, the flow configuration of the DNS database consists of a forced scalar field in a forced homogeneous isotropic turbulence. The DNS is generated from a standard pseudo-spectral code, and the simulation domain is discretized using 256^3 grid points on a domain of length 2π. The forcing scheme used to obtain a statistical steady flow follows the one proposed by Alvelius (1999). The scalar field is initialized between 0 and 1 according to the procedure proposed by Eswaran & Pope (1988), and to achieve a steady state for the scalar, a forcing scheme is also applied to low-wave number modes in Fourier space, similarly to the velocity forcing (da Silva & Pereira 2007). The Schmidt number is taken equal to 1 and the Reynolds number based on the Taylor microscale is around 95 at the stationary state. From DNS data, LES quantities are emulated with an explicit spectral cutoff filter at several ratio $\bar{\Delta}/\Delta$, with Δ the DNS grid size.

2.1. *Concept of optimal estimator*

The proposed strategy is based on a systematic reduction of the SGS modeling error. Considering f as the SGS term to model and $g(\phi)$ as a model of f based on a given set of input parameters ϕ, Langford & Moser (1999) define the quadratic error

$$\epsilon_{\mathrm{Q}} = \langle (f - g(\phi))^2 \rangle \tag{2.1}$$

as the relevant modeling error to consider in LES. In this definition, the brackets indicate a statistical average over a suitable ensemble. Within the LES context, a modeling error decomposition can be proposed by using the concept of optimal estimator introduced by Moreau *et al.* (2006) in the framework of the optimal estimation theory (Deutsch 1965). The concept of an optimal estimator forecasts that any model g, built on the set of parameters ϕ, will display a quadratic error higher than a minimal value, ϵ_{irr}, called an irreducible error. Moreover, this minimal value is defined by the optimal estimation theory as

$$\epsilon_{\mathrm{irr}} = \langle (f - \langle f|\phi \rangle)^2 \rangle \leq \epsilon_{\mathrm{Q}}, \tag{2.2}$$

where $\langle f|\phi \rangle$ is the expectation of the exact quantity f conditioned with the set of parameters ϕ used to write the model. The quantity $\langle f|\phi \rangle$ is thus called the optimal estimator of f for the set of parameters ϕ because no model using only ϕ as set of parameters can lead to a smaller error.

From the concept of optimal estimator, a modeling error decomposition can be written as

$$\underbrace{\langle (f - g(\phi))^2 \rangle}_{\epsilon_{\mathrm{Q}}} = \underbrace{\langle (f - \langle f|\phi \rangle)^2 \rangle}_{\epsilon_{\mathrm{irr}}} + \underbrace{\langle (\langle f|\phi \rangle - g(\phi))^2 \rangle}_{\epsilon_{\mathrm{form}}}, \tag{2.3}$$

where the total modeling error, ϵ_{Q}, is split into the irreducible error, ϵ_{irr}, and the formal error, ϵ_{form}. The irreducible error is the part of the modeling error from the set of parameters chosen to write the model, whereas the formal error is the part of the modeling error from the functional form chosen to link these parameters to approximate the SGS term f. This error decomposition provides various information on the SGS models used

in LES. First, the total error ϵ_{Q} can be assessed for each model to see which one yields the best results as far as the modeling of the unknown SGS term is concerned. The most suitable set of parameters to model the SGS term can also be determined by comparing ϵ_{irr} for different models. The set of parameters with the smallest irreducible error will be the best candidate to design a model. Finally, this theory informs to what extent a model based on a fixed set of parameters can be improved. Indeed, a quadratic error of a given model much higher than its irreducible part means the formal error, ϵ_{form}, is important and an improvement can be probably expected. This concept has already been used as an analysis tool to improve existing models. For example, in the 2012 CTR Summer Program, this concept was used to evaluate a regularized version of the gradient model (Balarac *et al.* 2012, 2013). In the present work, this modeling error decomposition is directly used to derive a new SGS model in two steps. In a first step, the appropriate set of input parameters will be determined by considering irreducible errors. In a second step, the formal error will be reduced by using a machine-learning procedure on the DNS database.

2.2. *Determination of the set of input parameters*

The optimal estimator theory is first used to determine an appropriate set of parameters to develop the model (i.e., the set of parameters leading to the smallest irreducible error). Papoulis (1965) demonstrates the irreducible error will be low if a large set of uncorrelated parameters is used. Starting from the Noll's formula, Eq. (1.4), various sets of parameters can be proposed. In this work, the vector v_i is kept equal to $\bar{\Delta}^2|\bar{S}|\partial\bar{Z}/\partial x_i$. Moreover, in Eq. (1.4), f_1, f_2, and f_3 are coefficients depending on principal invariants of M_{ij} and v_i defined as $I_M = M_{ii}$, $II_M = M_{ij}M_{ji}$, $III_M = M_{ik}M_{kl}M_{li}$, $I_v = v_i v_i$, $I_{Mv} = v_i M_{ik} v_k$, and $II_{Mv} = v_i M_{ik} M_{kj} v_j$ (Zheng 1994; Noll 1967). Thus, neglecting the spatial derivation of coefficients, a set of parameters, ϕ, to model the SGS scalar flux divergence, $\partial\tau_i/\partial x_i$, can be defined as

$$\phi = \left\{ I_M, II_M, III_M, I_v, I_{Mv}, II_{Mv}, \frac{\partial v_i}{\partial x_i}, \frac{\partial M_{ij} v_j}{\partial x_i}, \frac{\partial M_{ik} M_{kj} v_j}{\partial x_i} \right\}. \tag{2.4}$$

At this stage, various choices can be made to define the symmetric tensor, M^*_{ij}, based on the filtered velocity gradients (Lund & Novikov 1992). A first choice can be $M^*_{1,ij} = \bar{S}_{ij}$, as proposed by Wang *et al.* (2007) with the DNTD model. A second one can be $M^*_{2,ij} = \partial\bar{u}_i/\partial x_k \partial\bar{u}_j/\partial x_k$, considering the gradient model (Clark *et al.* 1979). Other choices can also be $M^*_{3,ij} = \bar{S}_{ik}\bar{\Omega}_{jk} + \bar{\Omega}_{ik}\bar{S}_{jk}$ and $M^*_{4,ij} = \bar{\Omega}_{ik}\bar{\Omega}_{jk}$. The two last propositions come from the decomposition of $M^*_{2,ij}$, using the filtered strain rate tensor, $\bar{S}_{ij}$, and the filtered rotation rate tensor, $\bar{\Omega}_{ij}$.

These propositions yield various sets of input parameters to write a model for the SGS scalar flux divergence. The set of parameters ϕ_l is defined as the set of parameters given by Eq. (2.4) using $M^*_{l,ij}$ to define the symmetric tensor. To determine the most appropriate set, the irreducible error of each set of parameters is now computed on the DNS database. The irreducible error of the set of parameters ϕ_l is defined as

$$\epsilon_{\mathrm{irr},l} = \left\langle \left(\frac{\partial \tau_i}{\partial x_i}^{DNS} - \left\langle \frac{\partial \tau_i}{\partial x_i}^{DNS} \middle| \phi_l \right\rangle \right)^2 \right\rangle, \tag{2.5}$$

where $\partial\tau_i^{DNS}/\partial x_i$ represents the exact divergence of the SGS scalar flux extracted from the filtered DNS database, and a spatial averaging is used owing to the flow configuration. The evolution with the filter width of various irreducible errors, corresponding

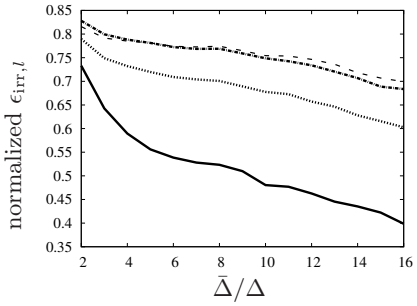

FIGURE 1. Evolution of the normalized irreducible errors as a function of the filter width for various set of parameters: $\epsilon_{\mathrm{irr},1}$ (solid line), $\epsilon_{\mathrm{irr},2}$ (dotted line), $\epsilon_{\mathrm{irr},3}$ (dashed line), and $\epsilon_{\mathrm{irr},4}$ (dashed-dotted line).

to the various proposed set of parameters, are shown in Figure 1. In this figure, the irreducible errors are normalized by the statistical variance of $\partial\tau_i^{DNS}/\partial x_i$. All the normalized irreducible errors decrease with the increase of the filter width. However, the irreducible error of the set of parameters ϕ_1, i.e., using only the filtered strain rate tensor to define M_{ij}^* as proposed by Wang *et al.* (2007), is much smaller than the other ones for all filter sizes. This observation leads to conclude this set of parameters is the best candidate to develop a SGS model. The next step is to determine an appropriate link between the parameters of this set, leading to a weak formal error in Eq. (2.3), so as to ensure a weak total quadratic error for the proposed model.

2.3. *Formal error reduction using an artificial neural network*

In this second step, only the set of parameters ϕ_1 is considered. Owing to the divergence-free condition, the first invariant of M_{ij} is equal to zero, and because of the dimensionless form of the symmetric tensor M_{ij}, the second invariant is constant. The set of parameters ϕ_1 is thus made of seven parameters. A multi-dimensional interpolation on scattered data is performed, to develop a surrogate model based on these seven inputs to approximate one output (i.e., $\partial\tau_i/\partial x_i$). When considering the use of the machine-learning procedure from the DNS database and taking into account the amount of data to process (256^3 grid points and various filter sizes), the artificial neural network (ANN) approach appears to be one of the most robust. In the LES context, Sarghini *et al.* (2003) have already used ANN to show the ability of this approach to approximate the scale-similarity model.

The ANN is a nonlinear surrogate function between inputs and output. The linear relation in Eq. (1.4) is no longer taken into account. The most efficient ANN topology in this work is a two-layers perceptron with a back-propagation training algorithm composed of two hidden layers and fifteen neurons per hidden layer (Figure 2(a)). The activation functions are sigmoid type. The j^{th} neuron of the first layer, noted $N_{j,1}$, is defined as

$$N_{j,1} = \tanh\left(\sum_{l=1}^{7} \omega_{jl,0}\phi_{1,l} + b_{j,0}\right), \tag{2.6}$$

with $\phi_{1,l}$ the l^{th} parameter of the set of parameters, ϕ_1. The i^{th} neuron of the second layer, noted $N_{j,2}$, is then defined as

$$N_{i,2} = \tanh\left(\sum_{j=1}^{15} \omega_{ij,1}N_{j,1} + b_{i,1}\right), \tag{2.7}$$

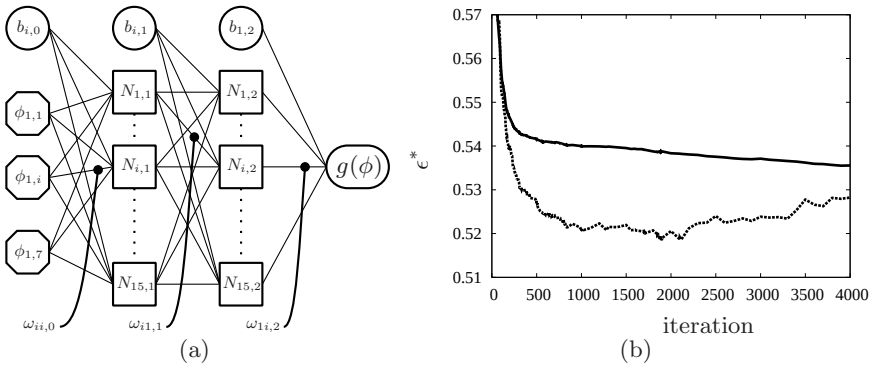

FIGURE 2. (a) ANN topology used for training; (b) evolution of the training error (solid line) and the generalized error (dotted line) as functions of the iterations of ANN training stage.

allowing to be the analytical form of the nonlinear function for the output, $g(\phi_1)$

$$g(\phi_1) = \sum_{i=1}^{15} \omega_{1i,2} N_{i,2} + b_{1,2}. \tag{2.8}$$

In these definitions, $\omega_{ij,n}$ and $b_{k,n}$ are the ANN parameters. $\omega_{ij,n}$ represents the weights linking the neuron $N_{j,n}$ (or input) to neuron $N_{i,n+1}$ (or output) in the n^{th} layer and $b_{k,n}$ is the bias of the n^{th} layer to the k^{th} neuron of this layer. The learning procedure is then an optimization of the ANN parameter, $\omega_{ij,n}$ and $b_{k,n}$, in order to minimize the quadratic error of the ANN model.

To avoid issues with dimensional consistency, the ANN procedure is applied on dimensionless inputs and output. The dimensionless inputs are built by subtracting their average and normalizing with their root mean square. The output is normalized by the root mean square of $\partial v_i/\partial x_i$, to generate a dimensional quantity *a posteriori*. The use of dimensionless parameters allows for a more efficient training stage. Moreover, we can expect that this will also favor the development of a well-performing ANN model on a broader range of turbulent flow configurations.

The database is divided into a training database and a test database. The optimization procedure of the ANN parameters is performed on the training database with the objective to decrease the training error, defined as the quadratic error of the ANN model on this database. Moreover, at each optimization step, a generalized error is also defined as the quadratic error of the ANN on the test database. Figure 2(b) shows both training and generalized errors as a function of the iterations of the ANN training stage. The generalized error presents a minimum error while the training error is still decreasing. The selected result corresponds to the iteration leading to this minimal generalized error. Beyond this iteration, the over-fitting phenomenon occurs and the ANN model is specifically linked with the training database (Lodwich *et al.* 2009). A surrogate model, denoted ANN model, is then generated. In the next section, some *a priori* tests are performed to check the training accuracy, and some *a posteriori* tests are also carried out to validate the overall procedure. Comparisons of the ANN model performance with other models will be presented below.

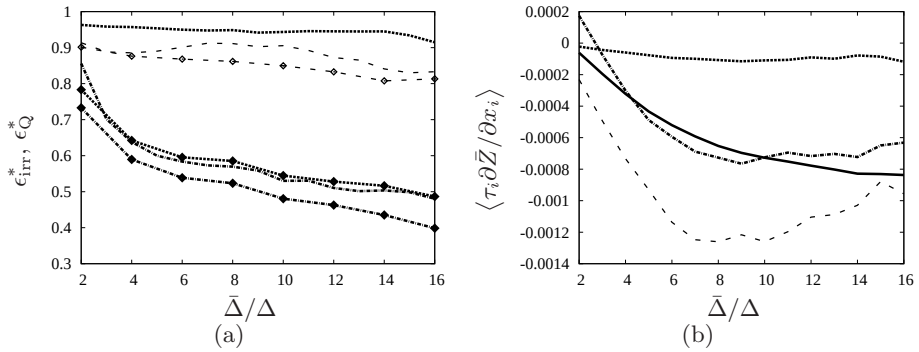

FIGURE 3. *A priori* measurements of the structural and functional models performances for DED (dashed line), DNTD (dotted line), and ANN (dotted-dashed line) models. (a) Normalized quadratic errors as a function of the filter width. The normalized irreducible errors are also shown for comparison (symbols). (b) Mean SGS dissipation as a function of the filter width. The solid line shows the SGS dissipation given by the filtered DNS data.

3. Performance of the ANN model

3.1. *A priori tests*

The previously generated ANN model is compared with the DNTD and DED models through *a priori* tests relying on the use of the DNS database. For SGS models, performances can be measured in terms of functional and structural performances (Sagaut 2006). The structural performance is defined as the model's ability to locally describe the SGS unknown term appearing in the resolved equation (here, the SGS scalar flux divergence). This performance is measured by computing the quadratic error, defined by Eq. (2.1). The functional performance measures the model's ability to reproduce the averaged effect of the SGS term on the transported quantity, and not the term itself. For SGS scalar flux models, the functional performance is defined as the model's ability to well reproduce the grid-scales/subgrid-scales (GS/SGS) transfer between the resolved scalar variance, $\bar{Z}^2$, and the SGS scalar variance, $\overline{Z^2} - \bar{Z}^2$. This transfer is controlled by the SGS scalar dissipation rate, $-\tau_i \partial \bar{Z}/\partial x_i$ (Jiménez *et al.* 2001; da Silva & Pereira 2007), a usually positive term on average but with possibly negative local value characterizing an inverse transfer (backscatter). For an accurate LES, a SGS model should correctly reproduce the average SGS scalar dissipation rate (Kang & Meneveau 2002).

The models quadratic errors are displayed in Figure 3(a) for various filter sizes, along with the irreducible errors, to assess whether the errors are mainly caused by the set of variables retained to derive the model or from the algebraic relationship, as already evoked for the error decomposition, Eq. (2.3). A satisfactory performance of the generated ANN model is made clear because this model leads to the smallest irreducible error (the set of the input parameters is relevant) and displays a quadratic error that remains close to the irreducible error. The convergence of the ANN training is good. Meanwhile, the DED model also yields a quadratic error close to the irreducible error, thus hinting that the dynamic procedure is well adapted for this case; however, the high level of quadratic error observed for the DED model hints that the number of input parameters for this model is too limited. Conversely, the DNTD model displays a large gap between its (low) quadratic error and its irreducible error, concluding the dynamic procedure is probably not efficient in this case. Overall the ANN model appears to serve an accurate surrogate function for the optimal estimator.

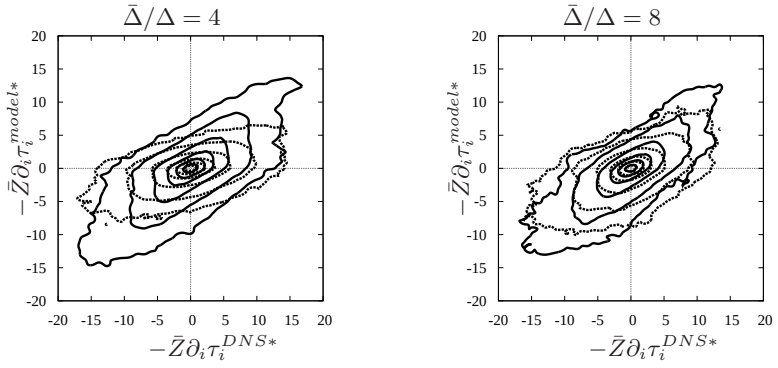

FIGURE 4. Joint probability density function (J-PDF) between the exact and modeled normalized SGS transfer terms, for DED (dotted line), and ANN (solid line) models and for both filter sizes. Level lines represent density of probability from 0.0001% to 10% with a geometric sequence with a ratio of 10.

To assess the functional performance, the prediction of the models for the SGS scalar dissipation rate is now analyzed. Figure 3(b) compares the mean SGS scalar dissipation rate of the models with the exact evaluation extracted from DNS database. The DNTD model displays a strong under-prediction of the SGS dissipation magnitude, which can lead to unstable simulations. Conversely, the DED model appears as weakly over-dissipative. The ANN model predicts a global SGS dissipation in good agreement with DNS data. This is an encouraging result because it confirms that, even though the SGS model development procedure does not directly take into account the functional performance, it cannot sufficiently decrease of the quadratic error, i.e., a good enough structural performance will ensure a correct functional performance. The GS/SGS local transfers are also studied locally by considering both SGS dissipation and SGS diffusion. Figure 4 compares the joint probability density function (J-PDF) between the exact and modeled normalized SGS transfer terms for the DED and ANN models. Note that all the transfers are normalized by the root mean square of the exact SGS transfers. A better local correlation with the exact term is found for the ANN model, showing that the ANN model predicts a correct level of dissipation and that the energy transfers are better localized than with the DED model.

3.2. *A posteriori tests*

A posteriori tests were performed to validate the overall model development procedure leading to the ANN model. The flow configuration is similar to the DNS database previously described but the scalar is not forced in this case. This permits testing of the ANN model in another mixing process, where no global equilibrium is enforced by a stationary forcing, unlike the DNS database used to generate the ANN model. The *a posteriori* tests consist of LES of passive scalars on 64^3 grid points, using DED, DNTD, and ANN models. The results are compared with filtered DNS still performed on 256^3 grid points. To avoid modeling errors interaction, the velocity field is solved by DNS in all cases (Balarac *et al.* 2013). The LES results are in agreement with the *a priori* analysis. Figure 5(a) shows the time evolution of the predicted resolved scalar variance during the mixing process for various LES models. The resolved scalar variance behavior

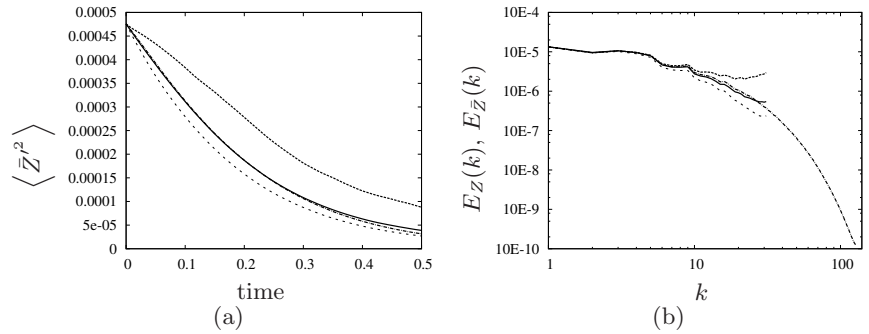

FIGURE 5. *A posteriori* test results for DED (dashed line), DNTD (dotted line), and ANN (solid line) models. (a) Comparison of the time evolution of the LES resolved scalar variance with the filtered DNS scalar variance (dotted-dashed line). (b) Comparison of resolved scalar variance spectra with the DNS scalar variance spectrum (dotted-dashed line) at time = 0.2.

is compared with the filtered DNS result. As expected, the under-dissipative behavior of the DNTD model leads to an under-prediction of the decay of the scalar variance. This is because of a development of unphysical fluctuations at the smallest resolved scales, as shown by the scalar variance spectra, Figure 5(b), with a unphysical accumulation at these scales. The weak over-dissipation of scalar variance is found for the DED model, with an under-estimation of the scalar variance at the smallest scales. Finally, the performance of ANN model is confirmed, with a prediction of the resolved scalar variance decay close to the filtered DNS data results and with a spectrum in good agreement with the DNS data for all the resolved scales.

4. Conclusion

An original SGS model development procedure has been proposed in this work. This procedure is based on decomposition of the modeling error, splitting this error between the error form the parameters used to write the model and the error from the algebraic relation used to link these input parameters on the other hand. The optimal set of parameters is first determined thanks to the concept of the optimal estimator. The ANN technique is then used to design a surrogate model with this selected set of parameters, by minimizing the formal error. This procedure leads to an accurate ANN SGS model, as shown by comparison with classic SGS models in *a posteriori* test.

Acknowledgments

This work is supported by the Agence Nationale pour la Recherche (ANR) under Contract No. ANR-2010-JCJC-091601. Part of computing resources were provided by IDRIS-CNRS and CIMENT infrastructure, which is supported by the Rhone-Alpes region (GRANT CPER07_13 CIRA). The authors acknowledge use of computational resources from the Certainty cluster awarded by the National Science Foundation to CTR. The authors have benefited from fruitful discussions with CTR Summer Program participants.

REFERENCES

ALVELIUS, K. 1999 Random forcing of three-dimensional homogeneous turbulence. *Phys. Fluids* **11**, 1880–1889.

Balarac, G., Le Sommer, J., Meunier, X. & Vollant, A. 2013 A dynamic regularized gradient model of the subgrid-scale scalar flux. *Phys. Fluids* **25**, 075107.

Balarac, G., Le Sommer, J. & Vollant, A. 2012 A dynamic regularized gradient model of the subgrid-scale scalar flux. *Proceedings of the Summer Program*, Center for Turbulence Research, Stanford University, pp. 97–106.

Clark, R. A., Ferziger, J. H. & Reynolds, W. C. 1979 Evaluation of subgrid-scale models using an accurately simulated turbulent flow. *J. Fluid Mech.* **91**, 1–16.

Deutsch, R. 1965 *Estimation Theory*. Prentice-Hall.

Eswaran, V. & Pope, S. 1988 Direct numerical simulations of the turbulent mixing of a passive scalar. *Phys. Fluids* **31**, 506–520.

Germano, M., Piomelli, U., Moin, P. & Cabot, W. H. 1991 A dynamic subgrid-scale eddy viscosity model. *Phys. Fluids* **3**, 1761–1765.

Jiménez, C., Ducros, F., Cuenot, B. & Bédat, B. 2001 Subgrid scale variance and dissipation of a scalar field in large eddy simulations. *Phys. Fluids* **13**, 1748–1754.

Kang, H. S. & Meneveau, C. 2002 Universality of large eddy simulation model parameters across a turbulent wake behind a heated cylinder. *J. Turbul.* **3**, 1–27.

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

Lodwich, A., Rangoni, Y. & Breuel, T. 2009 Evaluation of robustness and performance of Early Stopping Rules with Multi Layer Perceptrons. *Int. Joint Conf. Neural Networks* pp. 1877–1884.

Lund, T. S. & Novikov, E. A. 1992 Parametrization of subgrid-scale stress by the velocity gradient tensor. *Annual Research Briefs*, Center for Turbulence Research, Stanford University, pp. 27–41.

Moin, P., Squires, K., Cabot, W. & Lee, S. 1991 A dynamic subgrid-scale model for compressible turbulence and scalar transport. *Phys. Fluids A* **3**, 2746–2757.

Moreau, A., Teytaud, O. & Bertoglio, J. P. 2006 Optimal estimation for large-eddy simulation of turbulence and application to the analysis of subgrid models. *Phys. Fluids* **18**, 1–10.

Noll, W. 1967 Representations of certain isotropic tensor functions. *Archiv der Mathematik* **21**, 87–90.

Papoulis, A. 1965 *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill.

Sagaut, P. 2006 *Large Eddy Simulation for Incompressible Flows: An Introduction*. Springer.

Sarghini, F., de Felice, G. & Santini, S. 2003 Neural networks based subgrid scale modeling in large eddy simulations. *Comput. Fluids* **32**, 97–108.

da Silva, C. B. & Pereira, J. C. F. 2007 Analysis of the gradient-diffusion hypothesis in large-eddy simulations based on transport equations. *Phys. Fluids* **19**, 1–20.

Wang, B.-C., Yin, J., Yee, E. & Bergstrom, D. J. 2007 A complete and irreducible dynamic SGS heat-flux modelling based on the strain rate tensor for large-eddy simulation of thermal convection. *Int. J. Heat Fluid Flow* **28**, 1227–1243.

Zheng, Q.-S. 1994 Theory of representations for tensor functions - A unified invariant approach to constitutive equations. *Appl. Mech. Rev.* **47**, 545–587.