Subgrid-scale scalar flux modelling based on optimal estimation theory and machine-learning procedures

Antoine Vollant, Guillaume Balarac, Christophe Eric Corre

Journal of Turbulence

Vol. 00, No. 00, 2016, 1–37

RESEARCH ARTICLE

Subgrid-scale scalar flux modeling based on optimal estimation theory and machine-learning procedures

A. Vollant[a,b], G. Balarac [a*], and C. Corre[c]

[a] *Univ. Grenoble Alpes / CNRS, LEGI UMR 5519, Grenoble, F-38041, France*;

[b] *OptiFluides, Villeurbanne, F-69603, France*;

[c] *Ecole Centrale de Lyon, LMFA UMR 5509, Ecully, F-69134, France*

(*v3.2 released February 2009*)

This work is devoted to the exploration of new procedures for the development of subgrid-scale (SGS) models in the context of large-eddy simulation (LES) of a passive scalar. The starting idea is to combine the optimal estimator theory with machine-learning procedures. The concept of optimal estimator is then used to determine the most accurate set of input parameters (also called features in machine-learning terminology) to be used when deriving a model of the SGS scalar flux. The SGS model can be defined as a surrogate model built from this set of parameters by training an artificial neural network (ANN) on a database built by the filtering of direct numerical simulation (DNS) results. This procedure leads to a model with good structural performance. This allows to perform LES very close to the filtered DNS results, and showing an improvement in comparison with algebraic models. However, this first procedure does not control the functional performance of the SGS model and the model can fail when the flow configuration is different from the training DNS database. Another procedure is then proposed, where the functional form is imposed and the ANN used only to define the model coefficients. The training step is an optimization based on a multi-objective genetic algorithm allowing to simultaneously control the structural and functional performances of the generated model. The model obtained from this second procedure proves to be more

*Corresponding author. Email: guillaume.balarac@legi.grenoble-inp.fr

ISSN: 1468-5248 (online only)

DOI: 10.1080/14685248.YYYYxxxxxx

http://www.tandfonline.com

robust, leading to accurate LES for different mixing conditions. This work is thus a first step to establish the concept of optimal estimator associated with machine-learning procedures as a useful tool for SGS model development.

Keywords: LES, subgrid-scale scalar flux modeling, optimal estimator, surrogate function, artificial neural network (ANN), multi-objective genetic algorithm (MOGA)

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. The filtered transport equation for a passive scalar takes the form of the instantaneous advection-diffusion equation applied to the resolved passive scalar and also includes a SGS scalar flux divergence, which has to 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.

Two main strategies exist for developing SGS models: functional and structural strategies [1]. The functional modeling strategy considers the action of the subgrid term on the transported quantity and not the unknown term itself. It can introduce for instance a dissipative term, producing a similar effect while not necessarily displaying the same spatial structure. In the context of passive scalar LES, the usual functional modeling strategy is based on the definition of an eddy-diffusivity D_T

to model the SGS scalar flux. Moin *et al.* [2] introduced a dynamic model for D_T, similarly to the dynamic Smagorinsky model used to model the eddy viscosity [3]. This dynamic eddy-diffusivity model will be denoted DED from now on.

Conversely, the structural modeling strategy makes use of the known structure of the unknown SGS term to develop the best local approximation for this SGS term. A classical way to develop such a model is to rely on formal mathematical developments. For example, the gradient model [4] is based on the Taylor series expansion of a canonical class of filtering kernel. More recently, Wang et al. [5] have proposed to extend the DED model based on mathematical properties of tensor invariants and also using a dynamic procedure close to the procedure proposed by Germano et al. [3]. The dynamic nonlinear tensorial diffusivity (DNTD) model developed by these authors can be considered as a nonlinear extension of the dynamic eddy-diffusivity model, derived following a structural modeling strategy.

In the spirit of these different modeling strategies, SGS models can also be assessed in terms of functional and structural performances [6, 7]. 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). Langford and Moser [8] propose the quadratic error between the exact and the modeled term as the relevant modeling error to consider in LES to measure the structural performance. The functional performance measures the model's ability to reproduce the effect of the SGS term on the transported quantity, and not the SGS term itself. For SGS scalar flux models, the functional performance can be measured by the model's ability to well reproduce the grid-scales/subgrid-scales (GS/SGS) transfer between the resolved scalar variance and the SGS scalar variance. This transfer is controlled by the SGS scalar dissipation rate, which should therefore be correctly reproduced by the SGS model to achieve an accurate LES

[9]. Both model performance measurements, structural or functional, require the exact SGS quantities and are thus performed in the framework of *a priori* test where DNS results are filtered to obtain an exact evaluation of SGS terms.

Since structural performance measurement is based on the evaluation of a quadratic error, a possible strategy to improve SGS models can be built upon a systematic reduction of this error. Within the LES context, a modeling error decomposition can be proposed, which relies on the concept of optimal estimator [10] in the framework of the optimal estimation theory [11]. The optimal estimator concept forecasts that any model built on a given set of parameters will display a quadratic error higher than a minimal value, called the irreducible error. The total modeling error is thus split into the irreducible error and the formal error. The irreducible error is the part of the modeling error which results from the set of parameters chosen to write the model, whereas the formal error is the part of the modeling error which results from the functional form chosen to link these parameters when approximating the SGS term. This error decomposition provides valuable information on the SGS models used in LES. First, the total error 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 the irreducible error for different models. The set of parameters with the smallest irreducible error will be the best candidate to design a model. Finally, the optimal estimator theory informs to what extent a model based on a fixed set of parameters can be improved. Indeed, a quadratic error for a given model found much higher than its irreducible part means the formal error is important and a modeling improvement can be expected from a better choice of functional form while keeping the same set of parameters. This concept has already been used as an analysis tool to improve existing mod-

els [6, 7, 12, 13]. In the present work, this modeling error decomposition will be directly used to derive a new SGS model.

The starting point retained in this work to develop accurate SGS models is to take advantage of the growing available computational resources, which allow to generate a large DNS database. In the field of big data processing, the DNS database associated with explicit filtering can be used to better understand SGS model performance but also to devise an accurate SGS model by learning from this DNS database. Probably one of the first application of a machine-learning approach to the development of turbulence closures and more particularly SGS models can be found in the work of Sarghini *et al.* [14], where an artificial neural network (ANN) is trained and validated using the flowfields provided by the scale-similarity model [15] in order to model the turbulent viscosity coefficient. Milano and Koumoutsakos [16] used turbulent channel flow DNS results in order to train an ANN for reconstructing the near wall flow. Recently, Tracey *et al.* [17] used supervised learning algorithms to reproduce RANS results obtained with the one-equation Spalart-Allmaras model, retained as truth model, without knowledge of the structure, functional form and coefficients of this model. Noteworthy in the field of physics-informed RANS modeling is also the work of Wang *et al* [18] where a machine learning technique based on random forest is applied to train RANS Reynolds stresses from DNS databases. Whatever the turbulence modeling context (RANS or LES), applying machine-learning to derive improved models require to carefully select the features set processed by the algorithm as well as the target outputs. Singh and Duraisamy [19] have thus developed a data-informed approach which allows to quantify errors and uncertainties in the functional form of turbulence closure models. The information provided by this field inversion procedure can next be used as input to machine learning algorithms in lieu of deficient

modeling terms. Parish and Duraisamy [20] combine the previous field inversion and machine learning to propose a novel data-driven predictive modeling recently applied to the prediction of turbulent channel flow. In the field of RANS modeling, Ling *et al.* [21] have proposed a novel neural network architecture to embed key physical modeling properties, namely Galilean invariance, into the predicted output, namely the Reynolds stress anisotropy tensor. This methodology has been generalized in [22] to physical systems with invariance properties. As far as SGS modeling is concerned, autonomic closures have been studied by King *et al.* [23], which rely on a general nonparametric relation (a Volterra series) to represent the unclosed quantity in terms of resolved variables. This adaptative, self-optimizing approach was successfully applied to *a priori* tests but remain to be assessed on *a posteriori* tests.

After a brief review of the LES framework for the transport of a passive scalar in Section 2, two SGS modeling procedures are presented which combine optimal estimator theory and machine learning. The optimal estimator theory is used in a first step so as to identify the relevant features which are processed in a second step by the machine learning algorithm (ANN), using structural and / or functional performance as target outputs. The first modeling procedure described in Section 3 is based on the sole improvement of the structural performance. The 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 a classical ANN training. It is established that such an ANN model exhibits good performances, in comparison with the DNTD and the DED models, but the model can fail if it is used on different mixing conditions in comparison with the condition of the DNS database used for the training step. A second improved modeling procedure is then proposed in Section 4, which retains the algebraic

expression of the DNTD model but computes the model coefficient with an ANN. In order to take into account both structural and functional performance of the new model, the training procedure of the ANN relies on a multi-objective optimization algorithm. The model thus obtained leads to accurate LES for different mixing conditions but leads to an over-prediction of the mixing process when applied to a plane jet flow configuration with features far from the DNS database used for the ANN training.

2. Review of LES framework for the transport of a passive scalar

The separation between resolved large scales and modeled small scales is performed by a filtering operation, which takes the form of an integration on the overall domain D :

$$\bar{f}(\vec{x},t) = \int_{\vec{y}\in D} f(\vec{y},t)G(\vec{x}-\vec{y})d\vec{y}, \tag{1}$$

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

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

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. The SGS scalar flux divergence, $\partial T_i/\partial x_i$, with $T_i = \overline{u_i Z} - \bar{u}_i\bar{Z}$, is the SGS term which must be modeled to perform LES.

In the context of passive scalar LES, the usual functional modeling strategy makes use of an eddy-diffusivity, D_T, to model the SGS scalar flux as $T_i = -D_T\partial\bar{Z}/\partial x_i$. The dynamic eddy-diffusivity (DED) model proposed by Moin *et*

al. [2] is defined as :

$$T_i^{DED} = -D_T \frac{\partial \bar{Z}}{\partial x_i} = C\bar{\Delta}^2 |\bar{S}| \frac{\partial \bar{Z}}{\partial x_i}, \tag{3}$$

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.

Following a structural modeling strategy, Wang *et al.* [5] have extended the DED model into a dynamic nonlinear tensorial diffusivity (DNTD) model. According to the theory of tensor invariants and functions, a vector-valued function T_i can be decomposed by Noll's formula [24] in a second-order symmetric tensor M_{ij} and a vector v_i. From this formula, a general form of the SGS scalar flux can be written as

$$T_i = f_1 v_i + f_2 M_{ij} v_j + f_3 M_{ik} M_{kj} v_j, \tag{4}$$

where f_1, f_2, and f_3 are coefficients. In this decomposition, there is not a unique choice to define v_i and M_{ij}. It is proposed in [5] 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.* 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. [3], yielding the DNTD model :

$$T_i^{DNTD} = \chi_1 \bar{\Delta}^2 |\bar{S}| \frac{\partial \bar{Z}}{\partial x_i} + \chi_2 \bar{\Delta}^2 \bar{S}_{ik} \frac{\partial \bar{Z}}{\partial x_k} + \chi_3 \bar{\Delta}^2 \frac{\bar{S}_{ik}\bar{S}_{kl}}{|\bar{S}|} \frac{\partial \bar{Z}}{\partial x_l}, \tag{5}$$

with χ_1, χ_2, and χ_3 the dynamic coefficients. Note that by keeping only the first term of the RHS in Eq. (5), the DED model is recovered.

The structural performance of a SGS scalar flux model is defined as the model's ability to locally describe the SGS scalar flux divergence. Meanwhile, the functional performance of a SGS scalar flux models can be measured by 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, $-T_i \partial \bar{Z}/\partial x_i$ [25, 26], a usually positive term on average but with possibly negative local value characterizing an inverse transfer (backscatter).

A possible strategy to improve models can be based on a systematic reduction of the quadratic error measuring the structural performance. Denoting h the SGS term to model and $g(\phi)$ a model for h, based on a given set of input parameters ϕ, the quadratic error writes

$$\epsilon_{\mathrm{Q}} = \langle (h - g\left(\phi\right))^2 \rangle. \tag{6}$$

In this definition, the brackets indicate a statistical average over a suitable ensemble. The optimal estimator concept forecasts that any model g, built on the set of parameters ϕ, will display a quadratic error higher than the minimal value, ϵ_{irr}, also called the irreducible error and defined by the optimal estimation theory as

$$\epsilon_{\mathrm{irr}} = \langle (h - \langle h|\phi \rangle)^2 \rangle \leq \epsilon_{\mathrm{Q}}, \tag{7}$$

where $\langle h|\phi \rangle$ is the expectation of the exact quantity h conditioned with the set of parameters ϕ used to derive the model. The quantity $\langle h|\phi \rangle$ is thus called the optimal estimator of h for the set of parameters ϕ because no model using only ϕ as set of parameters can lead to a smaller error. From the optimal estimator concept, the total modeling error ϵ_{Q} can be split into an irreducible error ϵ_{irr} and a formal error ϵ_{form} as follows :

$$\underbrace{\langle (h - g\left(\phi\right))^2 \rangle}_{\epsilon_{\mathrm{Q}}} = \underbrace{\langle (h - \langle h|\phi \rangle)^2 \rangle}_{\epsilon_{\mathrm{irr}}} + \underbrace{\langle (\langle h|\phi \rangle - g\left(\phi\right))^2 \rangle}_{\epsilon_{\mathrm{form}}}, \tag{8}$$

Identifying the set of parameters ϕ which yields the smallest irreducible error pro-

vides, in a first step, the best candidate to design a model. This approach will be followed next in paragraph 3.1. In a second step, the formal error resulting from the functional form chosen to link this selected set of parameters ϕ can be minimized using an artificial neural network, as detailed in paragraph 3.2. Both structural and functional performance of a SGS scalar flux model will be simultaneously taken into account when building another machine-learning based SGS model in Section 4.

3. SGS model built from an artificial neural network

In this section, a first strategy to develop SGS models is described. 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. A statistical steady flow is achieved by using a random forcing scheme [27]. The scalar field is initialized between 0 and 1 [28], and to achieve a steady state for the scalar, a forcing scheme is also applied to low-wave number modes in Fourier space [26, 29]. The Schmidt number is taken equal to 0.7 and the Reynolds number based on the Taylor microscale is around 95 at the stationary state. The parameters are chosen to ensure all the dissipative scales of the turbulence are simulated [30], thus such that $k_{\max}\eta \approx k_{\max}\eta_B \approx 1.5$, where $k_{\max}$ is the maximal wavenumber in the box, and η and η_B are respectively the Kolmogorov and Batchelor scales. From DNS data, LES quantities are emulated with an explicit spectral cutoff filter at several ratios $\bar{\Delta}/\Delta$, with Δ the DNS grid size. The code and the flow configuration are similar to those used in previous works [6, 7, 12, 13].

3.1. *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). The irreducible error will be low if a large set of uncorrelated parameters is used [31]. Starting from the Noll's formula, Eq. (4), various sets of parameters can be proposed, depending on the choices made to define the symmetric tensor M_{ij} and the vector v_i. In this work, the vector v_i is kept equal to $\bar{\Delta}^2|\bar{S}|\partial\bar{Z}/\partial x_i$, as proposed by Wang et al. [5], and only the definition of M_{ij} is discussed. Future works could be devoted to investigate alternative choices for v_i as well. Moreover, in Eq. (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 $II_{Mv} = v_i M_{ik} M_{kj} v_j$ [24]. Thus, neglecting the spatial derivatives of coefficients, a set of parameters, ϕ, to model the SGS scalar flux divergence, $\partial T_i/\partial x_i$, can be defined as

$$\phi = \left\{ I_M, II_M, III_M, I_v, I_{Mv}, II_{Mv}, \frac{\partial v_i}{\partial x_i}, \frac{\partial M_{ij} v_j}{\partial x_i}, \frac{\partial M_{ik} M_{kj} v_j}{\partial x_i} \right\}. \tag{9}$$

At this stage, various choices can be made to define the symmetric tensor, M^*_{ij}, based on the filtered velocity gradients [32]. A first choice can be $M^*_{1,ij} = \bar{S}_{ij}$, as proposed for the DNTD model [5]. 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 [4]. 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. (9) using $M^*_{l,ij}$ to define the symmetric tensor. To

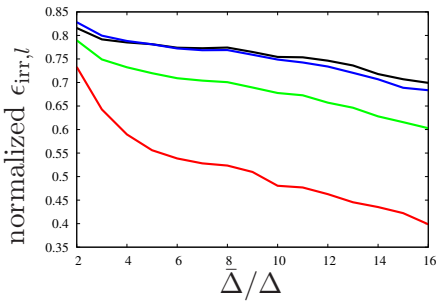


Figure 1. Evolution of the normalized irreducible errors as a function of the filter width for various set of parameters: $\epsilon_{\mathrm{irr},1}$ (red line), $\epsilon_{\mathrm{irr},2}$ (green line), $\epsilon_{\mathrm{irr},3}$ (black line), and $\epsilon_{\mathrm{irr},4}$ (blue line).

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

$$\epsilon_{\mathrm{irr},l} = \left\langle \left(\frac{\partial T_i}{\partial x_i}^{DNS} - \left\langle \frac{\partial T_i}{\partial x_i}^{DNS} \middle| \phi_l \right\rangle \right)^2 \right\rangle, \tag{10}$$

where $\partial T_i^{DNS}/\partial x_i$ 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 to the various proposed set of parameters, is shown in Fig. 1. In this figure, the irreducible errors are normalized by the statistical variance of $\partial T_i^{DNS}/\partial x_i$. 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,$\bar{S}_{ij}$, to define M_{ij}^* is much smaller than the other ones for all filter sizes. This observation leads to conclude the set of parameters ϕ_1 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. (8), so as to ensure a weak total quadratic error for the proposed model.

3.2. *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 only taken as input for developing a surrogate model to approximate one output (*i.e.* $\partial T_i/\partial x_i$). When considering the use of the machine-learning procedure from the DNS database and taking into account the amount of data to process, an artificial neural network (ANN) appears to be one of the most robust approach, as identified from the literature review provided in the Introduction. The ANN retained in the present study is described in Appendix A.

To avoid issues with dimensional consistency, the ANN procedure is applied on dimensionless inputs and output. The dimensionless inputs are built from the physical inputs by subtracting their average and normalizing with their root mean square. The dimensionless 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 process. Moreover, it is also expected that this will favor the development of a well-performing ANN model on a broader range of turbulent mixing conditions.

The DNS database is divided into two distinct parts : the training database and the test database, corresponding to different grid points. 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 shows both training and generalized errors as a function of the iterations of the ANN

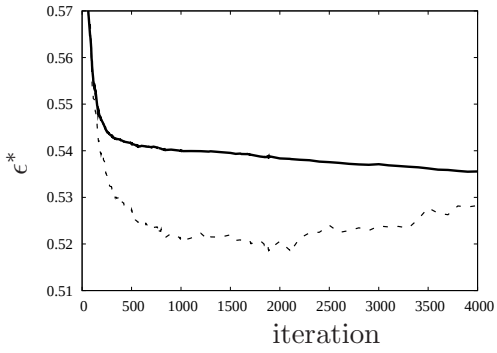
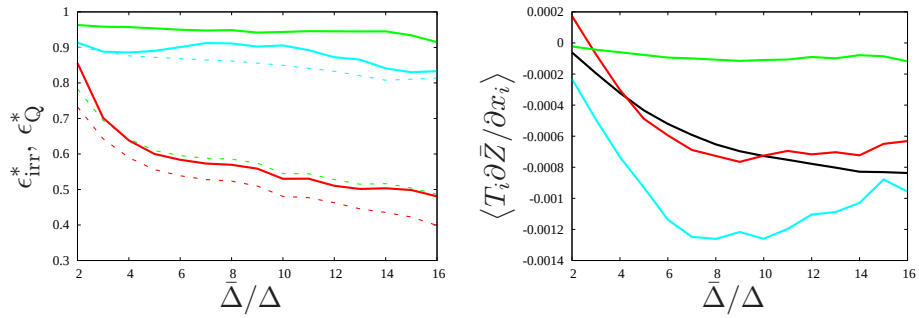


Figure 2. Evolution of the training error (solid line) and the generalized error (dashed line) as functions of the iterations of ANN training stage.

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 [33]. A surrogate model, denoted ANN model, is then generated. Some *a priori* tests are now 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.

3.3. *A priori measurement of model performance*

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. In this work, the dynamic procedure for the DED model is the classic procedure using the Germano identity [3], extended for the SGS scalar flux [2] and taking into account the modification proposed by Lilly [34]. For the DNTD model, the procedure proposed by Wang et al. [5] is used. For both models, an averaging is performed over homogeneous directions. 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



(a) (b)

Figure 3. *A priori* measurements of the structural and functional model performance for DED (cyan line), DNTD (green line) and ANN (red line) models. (a) Normalized quadratic errors as a function of the filter width. The normalized irreducible errors are also shown for comparison (dashed line). (b) Mean SGS dissipation as a function of the filter width. The black line shows the SGS dissipation given by the filtered DNS data.

algebraic relationship, as already evoked for the error decomposition, Eq. (8). A satisfactory performance of the generated ANN model is made clear because this model displays a quadratic error that remains close to the irreducible error, showing that 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, leading to conclude 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.

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

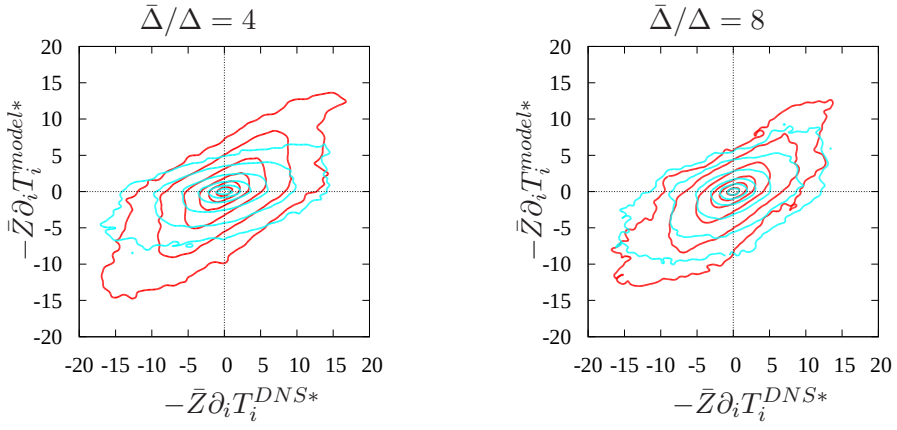


Figure 4. Joint probability density function (J-PDF) between the exact and modeled normalized SGS transfer terms, for DED (cyan line) and ANN (red line) models and for two different filter sizes. The isocontours are in the range 10^{-5} to 10^{-1} with a logarithm scale.

dissipation in good agreement with DNS data. This is an encouraging result because the SGS model development procedure does not directly take into account the functional performance. However, the model tends to underpredict the SGS dissipation magnitude for some filter widths. This can lead to unstable LES with an accumulation of the scalar variance at the smallest resolved scale. Figure 4 compares the joint probability density function (J-PDF) between the exact and modeled normalized SGS transfer terms (including SGS dissipation and SGS diffusion) for the DED and ANN models. 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 GS/SGS transfers are better localized with the ANN model than with the DED model.

3.4. *A posteriori tests*

A posteriori tests are now performed to validate the overall model development procedure leading to the ANN model. The flow configuration is a forced homogeneous isotropic turbulence, similar to the DNS database previously described. 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 [7]. Two mixing conditions are considered. The first test corresponds to a restart of the DNS database after a spectral interpolation of the scalar field on the LES grid, and keeping the scalar forcing scheme. Then, the ANN model is compared with exactly the same mixing condition as the DNS database used for its generation. The second test corresponds to a random initialization of the scalar field [28] and the scalar is not forced. This permits testing the ANN model in another mixing process, where no global equilibrium is enforced and where a decay of the scalar variance is thus expected.

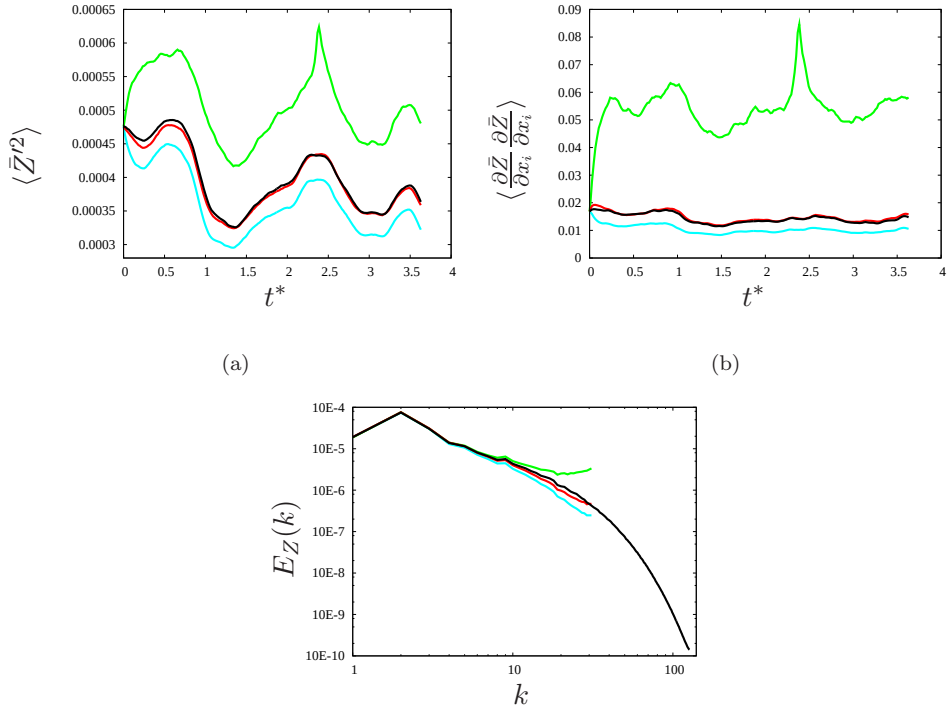
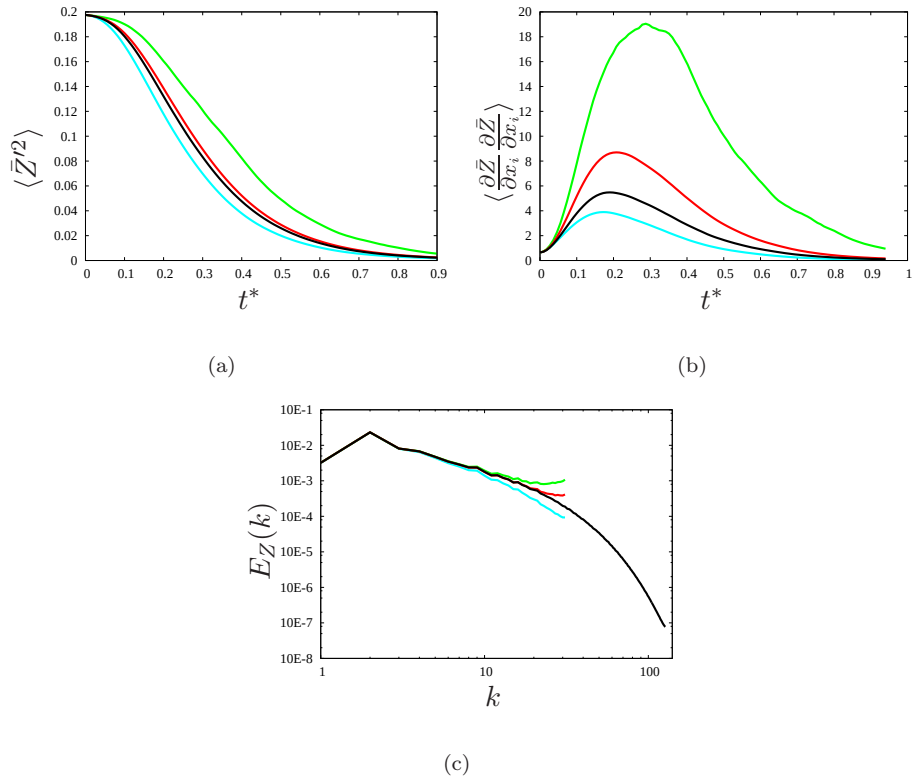


Figure 5. Forced scalar case: evolution of the resolved scalar variance (a) and scalar enstrophy (b) with time, and scalar variance spectrum (c). The models are compared with filtered DNS (a,b) or full DNS (c). DNS (black line), DED model (cyan line), DNTD model (green line) and ANN model (red line).

The results of the first test case, corresponding to the mixing condition of the

training DNS database, are displayed in Fig. 5. 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 is compared with the filtered DNS result. As expected, the under-dissipative behavior of the DNTD model leads quickly to a large over-prediction of the resolved scalar variance in comparison with the filtered scalar variance extracted from the DNS. The weak under-estimation of scalar variance is found for the DED model in agreement with the over-dissipation observed for this model in *a priori* test. As expected, the ANN model leads to an evolution of the resolved scalar variance very close to the filtered DNS result. To characterize the smallest resolved scales behavior, figure 5(b) shows the evolution of the resolved scalar enstrophy [35], $\langle \frac{\partial \bar{Z}}{\partial x_i} \frac{\partial \bar{Z}}{\partial x_i} \rangle$. The good behavior of the ANN model is confirmed. Note that the gap between the DNTD model and the filtered DNS result is more important for the scalar enstrophy, showing that the DNTD model fails mainly at the smallest resolved scales. This is confirmed by the scalar variance spectra, Fig. 5(c), which shows an unphysical accumulation at these scales for the DNTD model. The over-dissipation of the DED model leads to an under-estimation of the scalar variance spectrum at the smallest scales. Finally, the good performance of the ANN model is confirmed, with a spectrum in good agreement with the DNS data for all the resolved scales.

Figure 6 shows the same statistics as Fig. 5 for the second test case: unforced scalar evolution starting from a random initialization. Without forcing, the scalar variance decays in time since no production term is present to compensate dissipation term, Fig. 6(a). Moreover, because the initial scalar field is only composed of large scales, the first stage of the scalar mixing consists of a growth of the scalar enstrophy. This stage corresponds to the generation of smaller mixing scales due



(a) (b)

(c)

Figure 6. Unforced scalar with random scalar initialization case: evolution of the resolved scalar variance (a) and scalar enstrophy (b) with time, and scalar variance spectrum (c). The models are compared with filtered DNS (a,b) or full DNS (c). DNS (black line), DED model (cyan line), DNTD model (green line) and ANN model (red line).

to the transport by the velocity field. After this stage, the scalar enstrophy decreases due to the dissipation process, Fig. 6(b). For DNTD and DED models, the conclusions are to the ones made for the first test case. A large over-prediction of the scalar variance and of the scalar enstrophy is found for the DNTD model due to an under-prediction of the dissipation at the smallest scales. Conversely, a weak over-dissipation of the DED model leads to a weak under-prediction of the scalar variance and enstrophy in comparison with filtered DNS results. However, the good agreement between the ANN model and the filtered DNS results observed in the first test case, is no longer found in this second test case. Indeed, the ANN model leads to a behavior similar to the DNTD model, even though the negative effects are less pronounced. The over-estimation of the statistical values (variance and enstrophy) shows that the model eventually fails, leading to unphysical accu-

mulation of scalar variances at the smallest resolved scales, Fig. 6(c). This seems to indicate that the ANN model developed from a given DNS database can not be successfully used for other mixing conditions.

4. SGS model built from multi-objective optimization

To develop a more robust model, *i.e.* a model able to deal with flow conditions that differ from the training DNS database, a second (improved) strategy is proposed, which takes into account the functional performance during the model development process.

4.1. *Determination of model coefficients based on multi-objective optimization*

In the previous section, the ANN model has been developed as a surrogate model without any assumption on the relationship between the set of input parameters. However, the Noll's decomposition, Eq. (4), already fixed an algebraic relation (which has been ignored in the previous section). This relation is the starting point of the DNTD model [5]. On the other hand, the DNTD model has displayed very weak structural performance with a total quadratic error much larger than its irreducible error (Fig. 3). This probably means that the dynamic procedure is not efficient to compute model coefficients. The new procedure conserves the algebraic expression but evaluates the model coefficients from a machine-learning procedure.

As already stated in the introduction, Noll's formula applied to the modeling of the SGS scalar flux allows to write a complete and irreducible nonlinear tensor diffusivity model [5], as

$$T_i^{AC} = g_1 \bar{\Delta}^2 |\bar{S}| \frac{\partial \bar{Z}}{\partial x_i} + g_2 \bar{\Delta}^2 \bar{S}_{ik} \frac{\partial \bar{Z}}{\partial x_k} + g_3 \bar{\Delta}^2 \frac{\bar{S}_{ik}\bar{S}_{kl}}{|\bar{S}|} \frac{\partial \bar{Z}}{\partial x_l}. \tag{11}$$

From now on, this model is denoted AC for 'adaptative coefficients' because the coefficients g_1, g_2 and g_3 are defined from a machine-learning procedure, instead of a dynamic procedure as involved by the DNTD model. As already explained, the model coefficients depend on a set of parameters defined from the principal invariants of $M_{ij} = \bar{S}_{ij}/|\bar{S}|$ and $v_i = \bar{\Delta}^2|\bar{S}|\frac{\partial \bar{Z}}{\partial x_i}$. However, due to the flow configuration, the first invariant of M_{ij} is zero and the second one is a constant, the set of parameters is thus defined as, $\phi = \{III_M, I_v, I_{Mv}, II_{Mv}\}$, with $III_M = M_{ik}M_{kl}M_{li}$, $I_v = v_i v_i$, $I_{Mv} = v_i M_{ik} v_k$, and $II_{Mv} = v_i M_{ik} M_{kj} v_j$. A surrogate model based on an ANN is then computed to define the vectorial relation between the vector of the model coefficients $\boldsymbol{g} = (g_1, g_2, g_3)$ and the set of input parameter ϕ. The ANN is composed of 4 input variables (the input parameters, ϕ) and 3 output variables (the model coefficient, $\boldsymbol{g}$). Its topology is a two-layer perceptron composed of two hidden layers with 8 and 5 neurons, respectively. The training stage determines the best set of ANN parameters (noted $\omega_{ij,n}$ and $b_{k,n}$ on Fig. A1). Note that with this ANN topology, there are 103 parameters to determine. For convenience a set of ANN parameters is noted $\boldsymbol{p}$, and the vector of model coefficients computed from this set is noted $\boldsymbol{g}|_{\boldsymbol{p}}$.

The new determination process of the ANN parameters is performed to guarantee simultaneously the structural and the functional performances of the AC model. The structural performance is measured by defining the normalized quadratic error as first criterion to minimize,

$$c_1(\bar{\Delta}/\Delta, \boldsymbol{p}) = \frac{\left\langle \left(\frac{\partial T_i}{\partial x_i}^{DNS} - \frac{\partial T_i}{\partial x_i}^{AC} \right)^2 \right\rangle}{\left\langle \left(\frac{\partial T_i}{\partial x_i}^{DNS} \right)^2 \right\rangle - \left\langle \frac{\partial T_i}{\partial x_i}^{DNS} \right\rangle^2}. \tag{12}$$

A second criterion is defined for the functional performance, as the relative error between the mean exact SGS scalar dissipation and the mean SGS scalar dissipation

provided by the model,

$$c_2(\bar{\Delta}/\Delta, \boldsymbol{p}) = \left| \frac{\left\langle T_i^{DNS} \frac{\partial \bar{Z}}{\partial x_i} \right\rangle - \left\langle T_i^{AC} \frac{\partial \bar{Z}}{\partial x_i} \right\rangle}{\left\langle T_i^{DNS} \frac{\partial \bar{Z}}{\partial x_i} \right\rangle} \right|. \tag{13}$$

Both criteria are defined at a given filter size. The objectives of the training stage (noted Ob_1 and Ob_2, respectively) are then respectively defined from these criteria as the minimization of the maximum value of c_1 and c_2 over the filter size. The training stage is then a bi-objective optimization yielding a set of Pareto-optimal solutions $\boldsymbol{p}^*$ such that $\boldsymbol{p}^* = \min_{\boldsymbol{p}} [Ob_1(\boldsymbol{p}), Ob_2(\boldsymbol{p})]$, with $Ob_1(\boldsymbol{p}) = \max_{\bar{\Delta}}(c_1)$ and $Ob_2(\boldsymbol{p}) = \max_{\bar{\Delta}}(c_2)$. For the bi-objective problem under consideration, a set $\boldsymbol{p}^*$ of ANN parameters will be Pareto optimal if there is not another set $\boldsymbol{p}$ such that $Ob_i(\boldsymbol{p}) \leq Ob_i(\boldsymbol{p}^*)$ for $i = 1, 2$ and $Ob_i(\boldsymbol{p}) < Ob_i(\boldsymbol{p}^*)$ for at least one value of i. In other words, $\boldsymbol{p}^*$ will be Pareto-optimal if it is not dominated by any other parameter set in the solution space. The objectives Ob_1 and Ob_2 are conflicting, in the sense it is not possible to improve (decrease) one of these objectives without degrading (increasing) the other one. As a consequence, the Pareto-optimal set for the simultaneous minimization of Ob_1 and Ob_2 will include an infinite number of trade-off solutions, which do not dominate each other but dominate all the other parameter sets. When plotted in the objective space, namely the (Ob_1, Ob_2) plane, the trade-off optimal solutions form a Pareto front (see Fig. 8). The Non-dominated Sorting Genetic Algorithm (NSGA) (see [36] for more details on this well-established multi-objective genetic algorithm) is used to efficiently explore the solution space and provide a set of well-spread optimal solutions along the Pareto front. Figure 7 summarizes the global optimization algorithm. A random population of individuals $\boldsymbol{p}$ is first generated. For each individual $\boldsymbol{p}$, the model coefficient $\boldsymbol{g}|_{\boldsymbol{p}}$ and the model T_i^{AC} are computed for $\bar{\Delta}/\Delta = \{4, 8, 12, 16\}$ and both c_1 and c_2 criteria defined by (12) and (13) are computed for each filter size

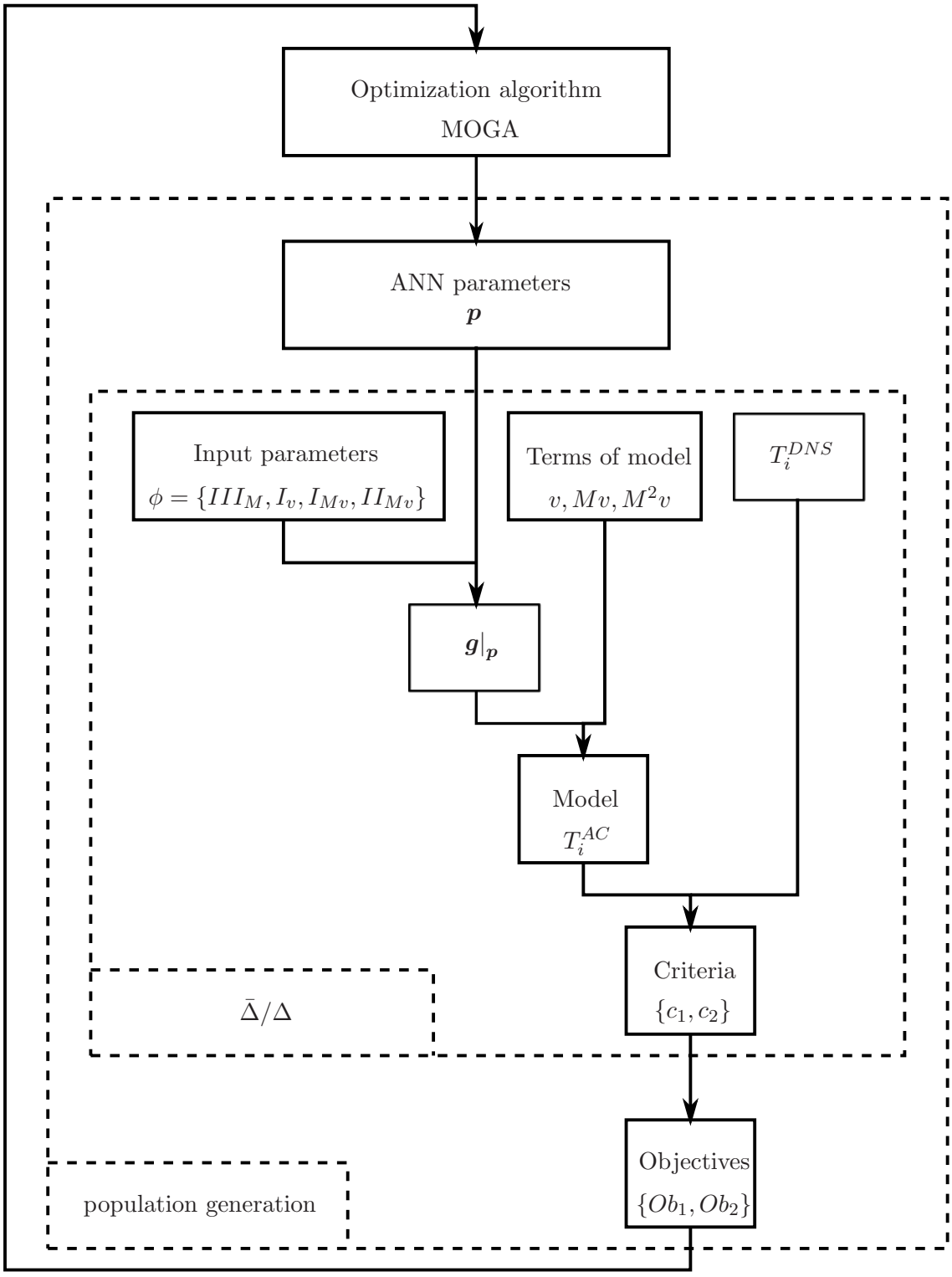


Figure 7. Schematic view of the modeling procedure based on the multi-objective optimisation of the ANN parameters

to yield the corresponding objective functions Ob_1 and Ob_2. Once a population of individuals $\boldsymbol{p}$ has been computed, the available values of the objective functions and the parameter values are used by the multi-objective genetic algorithm to efficiently explore the design space (of dimension 103). Individuals are sorted based on their rank of dominance, which allows to take simultaneously into account both objectives Ob_1 and Ob_2. A new population of designs is generated by applying selection, crossover and mutation operators to the current population. The process is iterated until no new Pareto-optimal designs are found by the algorithm.

4.2. *A priori tests: optimization results*

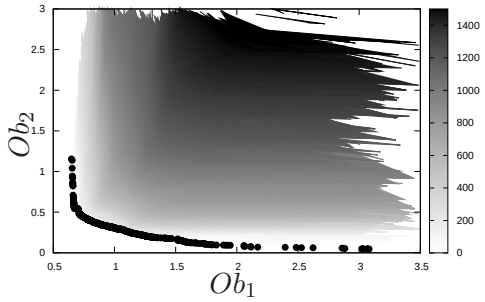
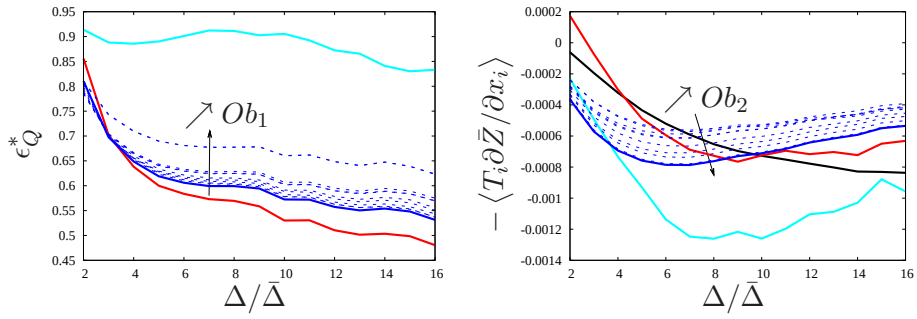


Figure 8. Representation of the individuals of the optimization procedure in the 2D objectives plan. The points represent the individuals of the optimal Pareto front and the grey level represent the rank of the Pareto fronts.

The described procedure is applied on the previously presented DNS database. The genetic algorithm optimization of the 103 parameters of the ANN has been performed by evolving a population of 1000 individuals during 1304 generations, yielding over 10^6 distinct designs. The large population size has been selected because of the large number (103) of design parameters. Figure 8 shows the results of the optimization procedure in the objective plane. The grey level shows the rank of Pareto fronts and the points represent the 230 individuals of the optimal Pareto front (rank equal to 1). It is interesting to note that the existence of the optimal

Pareto front clearly shows that the structural and functional performances can appear as antagonist objectives for the development of SGS models.



(a) (b)

Figure 9. Measurements of the structural (a) and functional (b) models performances for individuals selected from the Pareto front (blue lines). The performances are compared with the ANN model (red line) and the DED model (cyan line). The SGS scalar dissipation given by filtered DNS is also shown for comparison (black line). The individual used to define the AC model is identified by the blue solid line.

Additional processing has been performed to select the individual in the Pareto front, which will be used to define the final AC model. First, all the individuals leading to structural performance worst than the DED model are neglected. In order to further avoid Pareto-optimal individuals too specifically linked with the training database, the remaining optimal individuals, with structural performance better than the DED model, are also computed on a test database derived from the training database but with velocity and scalar fields uncorrelated with those of the training database. Only the designs that are still Pareto-optimal for the test database are eventually considered. This refined selection process eventually leaves 11 Pareto-optimal designs, the structural and functional performance of which are displayed in Fig. 9 These performances are compared with the performances of the ANN model developed in section 2 and the DED model. Note that the DNTD model is no longer considered in this section, because the ANN model has been found with better performances of the DNTD model in previous section. As expected, the ANN model has the best structural performance because only this performance was taken

into account during its training stage. In other words, the ANN model corresponds to a design point located on the upper left end of the Pareto front displayed in Fig. 8. However, the new models computed from the 11 selected individuals of the Pareto front also allow a significant improvement of structural performance in comparison with the DED model. In term of functional performance, these new models lead to a weak over-estimation of the SGS dissipation in the range of filter size used for the *a posteriori* tests, $\bar{\Delta}/\Delta < 8$. This over-estimation is less important than for the DED model, and it can avoid the unstable behavior observed with the ANN model. Finally, the one individual with the best structural performance is selected to define the AC model, which is the next used in *a posteriori* tests.

4.3. *A posteriori tests*

A posteriori tests are finally performed to validate the second procedure based on ANN trained by using a MOGA, and leading to the AC model. The tests are the ones previously described in section 2.4 when assessing the ANN model. Passive scalar LES on 64^3 grid points are thus performed for two different conditions. The first case is exactly the case of the training data base: forced scalar and established scalar field, whereas the second case is a unforced scalar with random initial scalar field. Note that the AC model is compared with the DED model only for clarity. Results provided by the ANN model are available in Fig. 5 and 6.

Figure 10 shows the results for the first test case, which corresponds exactly to the mixing configuration of the training database. As expected, the results are similar to the ANN model results presented in section 2.4. The AC model also leads to a temporal evolution of the scalar variance and enstrophy very close to the filtered DNS. This confirms the ability of machine-learning procedure to reproduce the exact SGS scalar flux in the same condition as the training stage. This allows

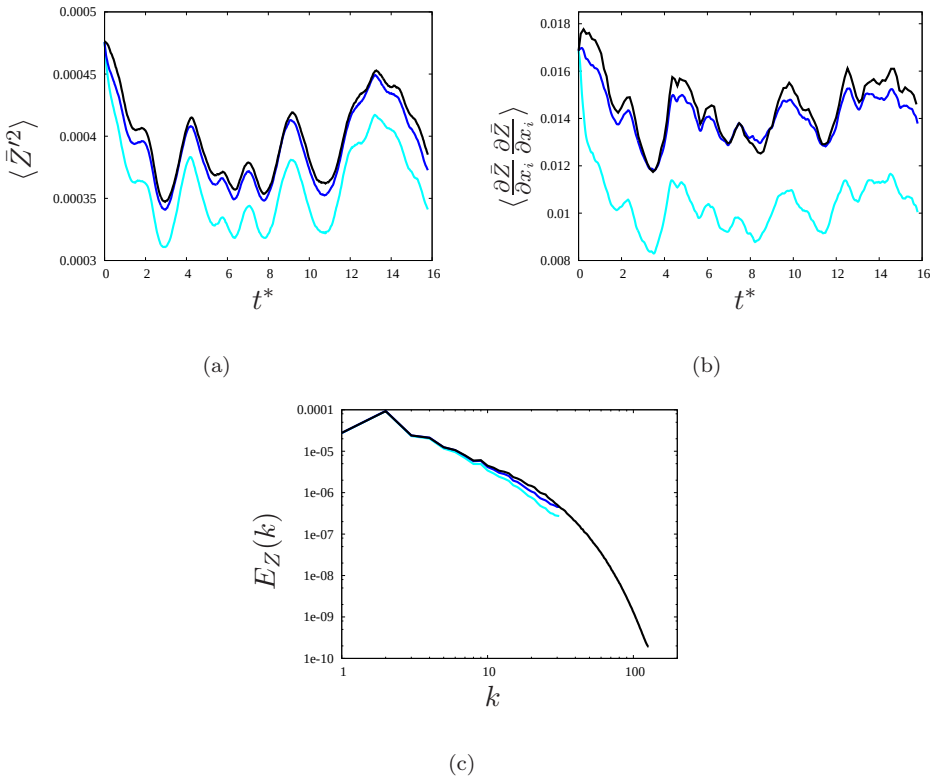
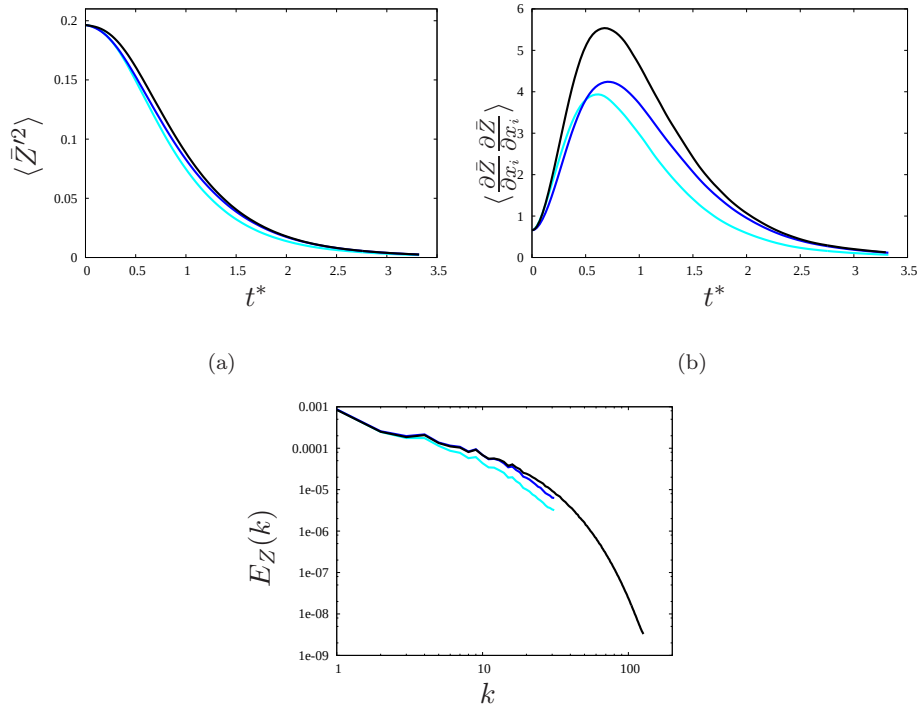


Figure 10. Forced scalar case: evolution of the resolved scalar variance (a) and scalar enstrophy (b) with time, and scalar variance spectrum (c). The models are compared with filtered DNS (a,b) or full DNS (c). DNS (black line), DED model (cyan line) and AC model (blue line).

for example to correct the unphysical behavior observed with the DNTD model (see Fig. 5) or the weak over-dissipation due to the DED model.

Figure 11 shows the scalar statistics for the second test case corresponding to the unforced scalar with random initial field. In section 2.4, it has been seen that the ANN model generated unphysical behavior at the smallest scales for this flow configuration. The better control of the functional performance offered by the AC model allows to correct this behavior of the ANN model. No accumulation at the smallest resolved scales is observed on the scalar variance spectrum computed for the AC model. This avoids the over-estimation of the scalar variance and enstrophy observed with the ANN model. Note that since the AC model has been defined so as to favor a weak over-estimation of magnitude of the SGS scalar dissipation, this allows to avoid unphysical accumulation, but leads in the same time to a weak

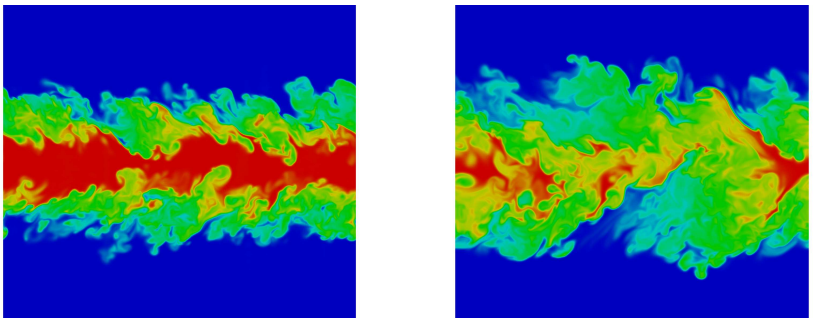


(a) (b)

(c)

Figure 11. Unforced scalar with random scalar initialization case: evolution of the resolved scalar variance (a) and scalar enstrophy (b) with time, and scalar variance spectrum (c). The models are compared with filtered DNS (a,b) or full DNS (c). DNS (black line), DED model (cyan line) and AC model (blue line).

under-estimation at the smallest resolved scales as shown by the under-estimation of the scalar enstrophy. Nevertheless, the AC model appears as the best compromise for this configuration, allowing to correct the unphysical behavior of the ANN and DNTD models, and being less dissipative than the DED model.



(a) (b)

Figure 12. Contour of scalar from DNS results during the plane jet transition toward a turbulent state at $tH/\Delta U \approx 8$ (a) and $tH/\Delta U \approx 14$ (b).The scalar is between 0 (blue) and 1 (red).

To further test the SGS model built from the multi-objective optimization process, a temporal turbulent plane jet flow configuration is also considered. This flow configuration includes transition stages and mean shear regions and is thus very far from the DNS database used to build the model. The flow problem is similar to the one studied by Silva and Pereira [37], with a computational domain periodic in the three spatial directions. The temporal evolution of the flow generated by an initial plane jet velocity profile is studied. The initial velocity and scalar profiles are described by a classic hyperbolic-tangent profile [37]. The molecular viscosity, ν, is defined to yield a Reynolds number $Re_H = \Delta U H/\nu$ equal to $10,000$, where H is the plane jet inlet slot width and $\Delta U = U_1 - U_2$ is the velocity difference between the initial jet velocity, U_1, and the co-flow velocity, U_2. The scalar value is initially 1 in the jet and 0 in the co-flow and the molecular Schmidt number is 0.7. The computational box size is defined such that $(L_x, L_y, L_z) = (4H, 6H, 4H)$, with x, y and z, respectively the streamwise, normal and spanwise directions. For the DNS, the grid size is selected in such a way the numbers of grid points in the 3 space directions are $N_x^{\mathrm{DNS}} \times N_y^{\mathrm{DNS}} \times N_z^{\mathrm{DNS}} = 1024 \times 1536 \times 1024$ grid points. The LES of the passive scalar is performed on a coarser mesh made of $N_x^{\mathrm{LES}} \times N_y^{\mathrm{LES}} \times N_z^{\mathrm{LES}} = 128 \times 192 \times 128$ grid points (1 out of 8 points is retained in the DNS grid along each direction). Note that, as for the previous tests, the velocity field is still solved by DNS and the filtered velocity is used for scalar transport in order to avoid modeling errors interaction between velocity and scalar fields. Contours of the scalar field at two different times of the transition process are displayed in Fig. 12. Scalar statistics are reported in Fig. 13 at similar times. Both SGS models lead to stable LES with mean scalar profiles in good agreement with the filtered DNS data (see Fig. 13 (a)). However, the resolved scalar variance profiles for the AC model are significantly under-predicted in comparison with fil-

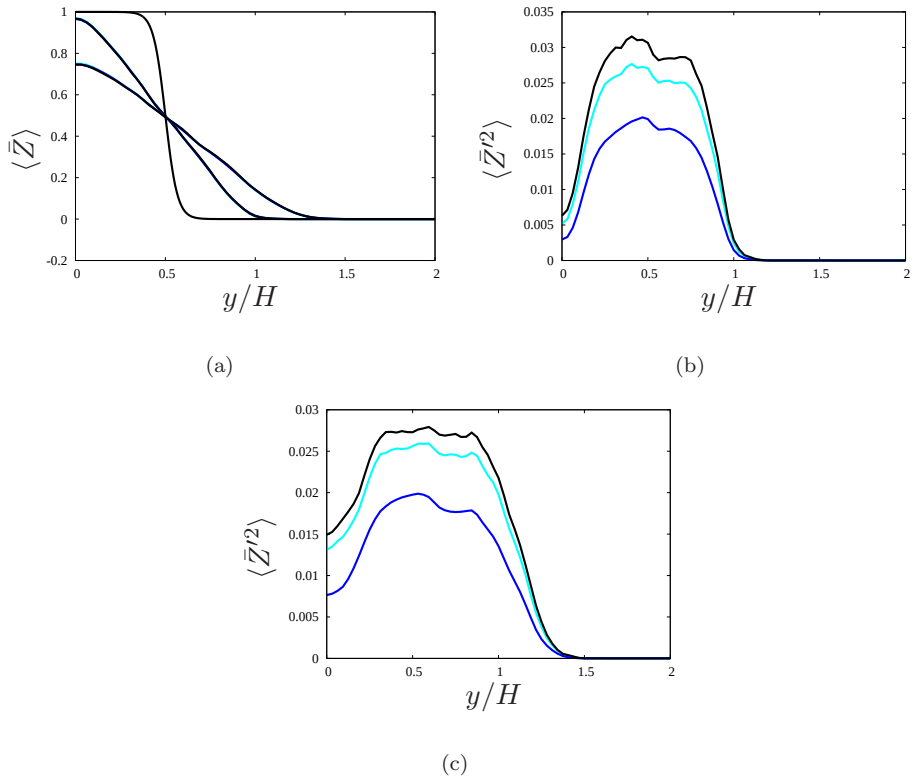


Figure 13. Mean scalar profile of the temporal jet at $tH/\Delta U = 0$, at $tH/\Delta U \approx 8$ and at $tH/\Delta U \approx 14$ (a). Profile of the resolved scalar variance, $\langle \bar{Z}'^2 \rangle$, at $tH/\Delta U \approx 8$ (b) and at $tH/\Delta U \approx 14$ (c). The DED model (cyan line) and AC model (blue line) are compared with filtered DNS results (black line).

tered DNS results (see Fig. 13 (b) and (c)). An under-prediction is also found with the DED model but remains much less pronounced. This observation means that the AC model over-predicts the SGS scalar dissipation in this configuration. The over-estimation of the SGS dissipation leads to an over-prediction of the mixing process. This is shown by the probability density function (PDF) of the resolved scalar computed for different locations toward the jet shear layer, from the center of the jet to the co-flow region (see Fig. 14). Indeed, for all locations, the peak of the PDF is larger at the most probable value of $\bar{Z}$, whereas the probability of the extrema is smaller for the AC model in comparison with the filtered DNS. Thus, even though stable LES are obtained with the AC model for this flow configuration, modeling improvement appears necessary. Note that machine-learning-assisted turbulence modeling is still in a rather early stage of development. Improvement can

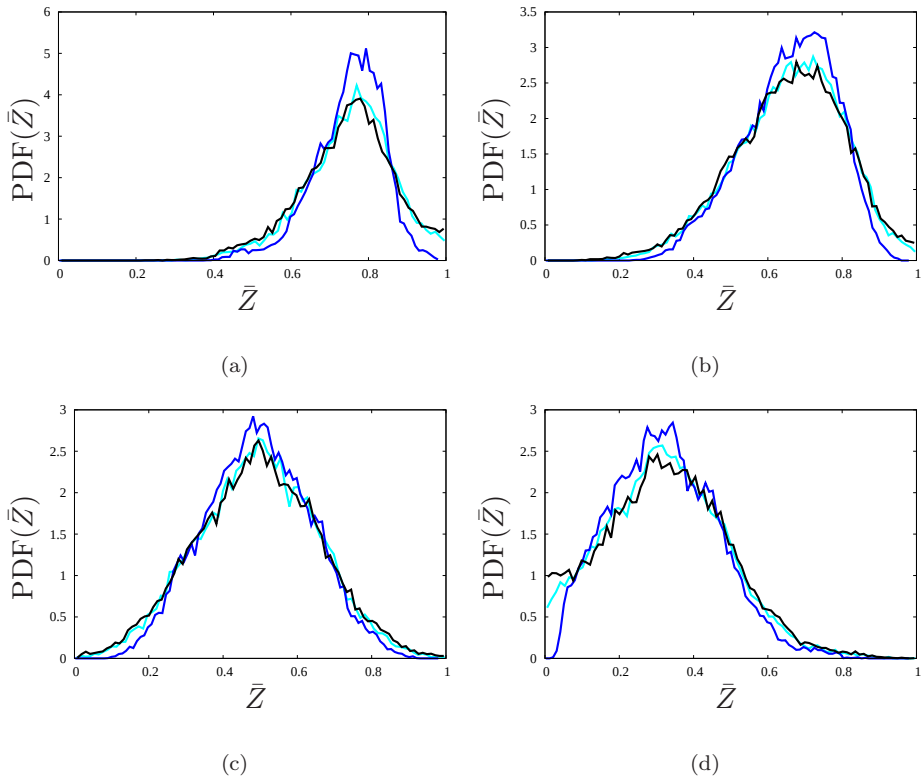


Figure 14. Probability density function (PDF) of the resolved scalar computed at $tH/\Delta U \approx 14$, for different locations toward the jet shear layer: $y/H = 0$ (a), $y/H = 0.25$ (b), $y/H = 0.5$ (c) and $y/H = 0.75$ (d). The models, DED model (cyan line) and AC model (blue line), are compared with filtered DNS (black line).

be expected from using more complex ANN topology (see for instance the work of Ling *et al.* [21]) and/or from using training database built from a wider range of flow configurations. It is still an open question to know whether such procedure can successfully model SGS terms for any flow configuration. Future works will be devoted to answer this question which is ultimately linked with the assumed universality of SGS models.

5. Conclusion

This work has been devoted to the formulation and the assessment of new strategies to develop SGS models. The proposed modeling procedures are first based on an improvement of the structural performance of the SGS model, measured by the

quadratic error between the exact (filtered DNS) SGS term and the model. In the framework of the optimal estimation theory, this modeling error can be split into the error resulting from the parameters used to write the model on one hand and the error resulting from the algebraic relation used to link these input parameters on the other hand. The optimal set of parameters is determined in a preliminary step thanks to the optimal estimator concept. Then, two strategies have been proposed to derive optimal SGS models using this selected set of parameters. The first strategy has been to directly use the artificial neural network technique to design a surrogate model, by minimizing the modeling error, *i.e.* improving only the structural performance. The second strategy has been to conserve the functional form given by Noll's formula and to determine coefficients by using the artificial neural network technique combined with a bi-objective optimisation technique to develop a model improving both structural and functional performances.

Both proposed strategies have been applied in the context of LES of turbulent mixing, to the modeling of the SGS scalar flux. The models developed from these procedures have been compared with classic algebraic SGS models. The first procedure based on ANN technique only leads to result very close to the reference results in comparison with classic SGS models. However, this first procedure fails for mixing conditions different from the mixing condition occurring in the training database used to generate the ANN. This eventually leads to unphysical behavior of the scalar field. The second procedure allows to correct this behavior. The unphysical behavior observed with the first strategy is avoided and the new model is found both more robust and leading to an improvement in comparison with classic SGS model. The second strategy differs from the first one by two main factors: (i) the functional form was imposed and (ii) the functional performance was taken into account in the optimization process. Future works will be devoted to better

understand which factor is preponderant to improve the model capability.

This contribution appears as a first step to establish the optimal estimator concept associated with machine-learning procedures as useful tools for SGS model development. In this first step, a simple flow configuration (forced HIT) has been considered, allowing to use spectral method. This leads to a clear definition of the filter kernel and guarantees the same filter kernel is applied during the training stage and *a posteriori* (LES) test. Moreover, this also allows to neglect interaction between modeling and numerical errors. Future works will be devoted to extend the proposed approach for SGS model developments to more complex flows devoid of these simplifications.

Acknowledgements

This work was supported by the Agence Nationale pour la Recherche (ANR) under Contract No. ANR-2010-JCJC-091601. It was performed using HPC resources from GENCI-IDRIS (Grant 2014-020611) and from CIMENT infrastructure (supported by the Rhône-Alpes region and the Equip@Meso project). LEGI is part of Labex OSUG@2020 (ANR10LABX56) and Labex Tec21 (ANR11LABX30). The authors thankfully acknowledge the hospitality of the Center for Turbulence Research, NASA-Ames and Stanford University, where a part of this work has been done during the Summer Program 2014.

Appendix A. Description of the Artificial Neural Network used in the study

The ANN is a nonlinear surrogate function between inputs and output, meaning that the linear relation in Eq. (4) is no longer taken into account. The most efficient ANN topology in this work is a two-layer perceptron with a back-propagation training algorithm composed of two hidden layers and fifteen neurons per hidden layer (Fig. A1). The activation functions are of sigmoid type. The j^{th} neuron of the first layer, denoted $N_{j,1}$, is defined as

$$N_{j,1} = \tanh\left(\sum_{l=1}^{7} \omega_{jl,0}\phi_{1,l} + b_{j,0}\right), \tag{A1}$$

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

$$N_{i,2} = \tanh\left(\sum_{j=1}^{15} \omega_{ij,1}N_{j,1} + b_{i,1}\right), \tag{A2}$$

yielding the following non-linear expression for the output, $g(\phi_1)$,

$$g(\phi_1) = \sum_{i=1}^{15} \omega_{1i,2}N_{i,2} + b_{1,2}. \tag{A3}$$

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 the 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. Given this surrogate function, the ANN technique consists of *a priori* learning the model from the DNS database. The learning procedure is a 'training', to optimize the ANN parameter, $\omega_{ij,n}$ and $b_{k,n}$, in order to minimize the quadratic error of the ANN model. The optimization is based on the RPROP algorithm [38].

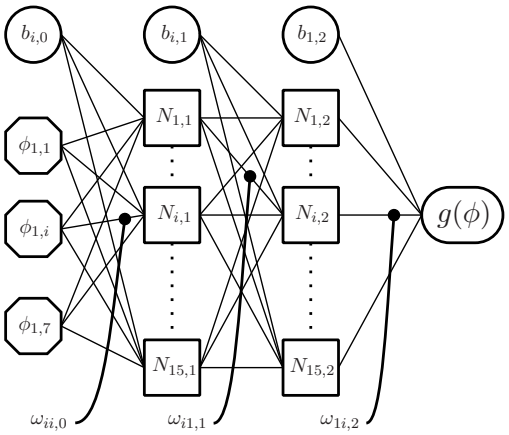


Figure A1. ANN topology used for training.

References

[1] P. Sagaut *Large eddy simulation for incompressible flows : an introduction (Third Edition)*, Springer, 2005.

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

[3] M. Germano, U. Piomelli, P. Moin, and W.H. Cabot, *A dynamic subgrid-scale eddy viscosity model*, Phys. Fluids A 3 (1991), pp. 1760–1765.

[4] R.A. Clark, J.H. Ferziger, and W.C. Reynolds, *Evaluation of subgrid-scale models using an accurately simulated turbulent flow*, J. Fluid Mech. 91 (1979), pp. 1–16.

[5] B.C. Wang, J. Yin, E. Yee, and D.J. Bergstrom, *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 (2007), pp. 1227–1243.

[6] Y. Fabre, and G. Balarac, *Development of a new dynamic procedure for the Clark model of the subgrid-scale scalar flux using the concept of optimal estimator*, Phys. Fluids 23 (2011), pp. 1–11.

[7] G. Balarac, J. Le Sommer, X. Meunier, and A. Vollant, *A dynamic regularized gradient model of the subgrid-scale scalar flux.*, Phys. Fluids 25 (2013).

[8] J.A. Langford, and R.D. Moser, *Optimal LES formulations for isotropic turbulence*, J. Fluid Mech. 398 (1999), pp. 321–346.

[9] H.S. Kang, and C. Meneveau, *Universality of large eddy simulation model parameters across a turbulent wake behind a heated cylinder*, J. Turbul. 3 (2002).

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

[11] R. Deutsch *Estimation Theory*, Prentice-ttall, Englewood Cliffs, N. J., 1965.

[12] G. Balarac, H. Pitsch, and V. Raman, *Development of a dynamic model for the subfilter scalar variance using the concept of optimal estimators*, Phys. Fluids 20 (2008), pp. 1–8.

[13] A. Vollant, G. Balarac, and C. Corre, *A dynamic regularized gradient model of the subgrid-scale stress tensor for large-eddy simulation*, Physics of Fluids (1994-present) 28 (2016), p. 025114.

[14] F. Sarghini, G. de Felice, and S. Santini, *Neural networks based subgrid scale modeling in large eddy simulations*, Comput. Fluids 32 (2003), pp. 97–108.

[15] J. Bardina, J. Ferziger, and W. Reynolds, *Improved subgrid-scale models for large-eddy simulation*, AIAA (1980).

[16] M. Milano, and P. Koumoutsakos, *Neural network modeling for near wall turbulent flow*, Journal of Computational Physics 182 (2002), pp. 1–26.

[17] B. Tracey, K. Duraisamy, and J. Alonso, *A machine learning strategy to assist turbulence model development*, AIAA 2015-1287 (2015).

[18] J.X. Wang, J.L. Wu, and H. Xiao, *Physics-informed machine learning approach for reconstructing Reynolds stress modeling discrepancies based on DNS data*, Physical Review Fluids 2 (2017), p. 034603.

[19] A.P. Singh, and K. Duraisamy, *Using field inversion to quantify functional errors in turbulence closures*, Physics of Fluids 28 (2016).

[20] E.J. Parish, and K. Duraisamy, *A paradigm for data-driven predictive modeling using field inversion and machine learning*, Journal of Computational Physics 305 (2016), pp. 758–774.

[21] J. Ling, A. Kurzawski, and J. Templeton, *Reynolds averaged turbulence modelling using deep neural networks with embedded invariance*, J. Fluid Mech 807 (2016), pp. 155–166.

[22] J. Ling, R. Jones, and J. Templeton, *Machine learning strategies for systems with invariance properties*, Journal of Computational Physics 318 (2016), pp. 22–35.

[23] R.N. King, P.E. Hamlington, and W.J.A. Dahm, *Autonomic closure for turbulence simulations*, Physical Review E - Statistical, Nonlinear, and Soft Matter Physics 93 (2016), pp. 1–6.

[24] W. Noll, *Representations of certain isotropic tensor functions.*, Archiv der Mathematik 21 (1967), pp. 87–90.

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

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

[27] K. Alvelius, *Random forcing of three-dimensional homogeneous turbulence*, Phys. Fluids 11 (1999), pp. 1880–1889.

[28] V. Eswaran, and S. Pope, *Direct numerical simulations of the turbulent mixing of a passive scalar*, Phys. Fluids 31 (1988), pp. 506–520.

[29] J.B. Lagaert, G. Balarac, and C. G.-H., *Hybrid spectral-particle method for the turbulent transport of a passive scalar.*, Journal of Computational Physics (2014).

[30] S.B. Pope *Turbulent Flows*, Cambridge Univ. Press, 2000.

[31] A. Papoulis *Probability, random variables, and stochastic processes*, McGraw-Hill, 1965.

[32] B. Lund, and E. Novikov, *Parameterization of subgrid-scale stress by the velocity gradient tensor*,

Center for Turbulence Research. Annual Research Briefs (1992), pp. 27–43.

[33] A. Lodwich, Y. Rangoni, and T. Breuel, *Evaluation of robustness and performance of Early Stopping Rules with Multi Layer Perceptrons*, 2009 International Joint Conference on Neural Networks (2009), pp. 1877–1884.

[34] D. Lilly, *A proposed modification of the Germano subgrid-scale closure method.*, Phys. Fluids A 4 (1992), pp. 633–635.

[35] M. Lesieur *Turbulence in fluids*, Springer, 2008.

[36] K. Deb *Multi-Objective Optimization using Evolutionary Algorithms*, Wiley, 2001.

[37] C. da Silva, and J. Pereira, *Invariants of the velocity-gradient, rate-of-strain, and rate-of-rotation tensors across the turbulent/nonturbulent interface in jets*, Phys. Fluids 20 (2008).

[38] M. Riedmiller, and H. Braun, *A Direct Adaptive Method for Faster Backpropagation Learning : The RPROP Algorithm*, in , Vol. 1 IEEE, San Fransisco, CA, 1993, pp. 586–591.