Interpreting neural network models of residual scalar flux

By G. D. Portwood†, B. T. Nadiga, J. A. Saenz, D. Livescu
Los Alamos National Laboratory, Los Alamos, USA 87545

21 April 2020

We show that in addition to providing effective and competitive closures, when analysed in terms of dynamics and physically-relevant diagnostics, artificial neural networks (ANNs) will be both interpretable and provide useful insights in the on-going task of developing and improving turbulence closures. In the context of large-eddy simulations (LES) of a passive scalar in homogeneous isotropic turbulence, exact subfilter fluxes obtained by filtering direct numerical simulations (DNS) are used both to train deep ANN models as a function of filtered variables, and to optimise the coefficients of common spatio-temporally local LES closures. \textit{A-priori} analysis of the subfilter scalar variance transfer rate demonstrates that learnt ANN models out-perform optimised turbulent Prandtl number closures and non-linear gradient models. Next, \textit{a-posteriori} solutions are obtained with each model over several integral timescales. These experiments reveal, with single- and multi-point diagnostics, that ANN models temporally track exact resolved scalar variance with greater accuracy compared to other subfilter flux models for a given filter length scale. Finally, we interpret the artificial neural networks statistically with differential sensitivity analysis to show that the ANN models learn dynamics reminiscent of so-called "mixed models", where mixed models are understood as comprising both a structural and functional component. Besides enabling enhanced-accuracy LES of passive scalars henceforth, we anticipate this work to contribute to utilising neural network models as a tool in interpretability, robustness and model discovery.

1. Introduction

The application of data-driven deep learning to physical sciences has been an emergent area of research in recent years, encouraged by the success of data-driven methods in fields such as computer vision, natural language processing, and other industrial and scientific disciplines. While a large variety of data-driven applications and methodologies are currently being explored in fluid mechanics (Kutz 2017, for a review), the application of data-driven models, particularly artificial neural networks (ANNs), to turbulence closure has shown promise as means to perform calibration, augmentation or replacement of existing turbulence closure models (e.g. Sarghini et al. 2003; Beck et al. 2019; Maulik et al. 2019b; Nikolaou et al. 2019; Ling et al. 2010; Moghaddam & Sadaghiyani 2018; Salehipour & Peltier 2019; Portwood & et al. 2019).

In reduced-order frameworks such as large-eddy simulation (LES), ANNs are attractive due to their ability to (1) discover complex relationships in data, (2) effectively leverage and reduce the growing volume of high-fidelity direct numerical simulation (DNS) data and (3) be expressed algebraically such that they are tractable for mathematical analysis.

† portwood@lanl.gov
Whereas demonstrations of the first two points are valuable in assessing capabilities of ANN models, in the third point we assert that the data-driven models must be robustly interpreted before being credibly certified for engineering or scientific applications. The objective of the research presented here is to robustly interpret ANN closure models while providing statistical insight into model optimization and performance metrics. We conduct this analysis by considering a data-driven algebraic residual passive scalar flux model which out-performs common algebraic closures with respect to several a-priori and a-posteriori diagnostics.

Algebraic LES closures, the de facto standard class of approaches to LES closure, relate resolved filter scale flow parameters to subfilter scale, or residual, dynamics. These models may be derived on functional grounds, whereby the effects of unresolved quantities on the resolved quantities are modeled, thus requiring knowledge about the nature of interactions, e.g. the divergence of the residual stresses which represents the net rate of kinetic energy transfer between subfilter to resolved scales. Functional models, such as Smagorinsky-type closures [Smagorinsky 1963], are phenomenological and limited in terms of the range of dynamics they are able to model. For instance, most Smagorinsky-type implementations are incapable of reproducing backscatter of kinetic energy from sub-filter to filter scales. While this is a limitation in the physical sense, these characteristics lead to such models exhibiting desirable stability properties in a-posteriori simulation. The simplicity and numerical stability of this class of models have prompted the application of neural networks and machine learning to such functional model frameworks, for instance, in the local determination of model constants [Sarghini et al. 2003; Maulik et al. 2019a; Gamahara & Hattori 2017].

Alternatively, algebraic models may be developed on structural grounds, whereby models attempt to reconstruct dynamical quantities (e.g. the residual stress, instead of its divergence as in functional modelling), representing a broader range of residual dynamics from mathematically- or theoretically-rigorous bases [Sagaut 2006]. This approach relies on the following assumptions: the structure of residual quantities are (i) universal, independent of the resolved scales, and (ii) can be determined from the resolved quantities. For example, a class of gradient-type models are derived from the assumption of asymptotically-small filter scales [Clark et al. 1979] and a class of scale-similarity models are developed by the imposition of scale symmetries [Bardina et al. 1980]. Model development from asymptotic or statistical symmetry assumptions is unreliable in practice, where such assumptions are, at best, approximate. Frequently, these imperfect assumptions often lead to issues, such as under-dissipation of filter scale kinetic energy in gradient-type models [Leonard 1974], such that structural models are commonly linearly combined with functional closures in so-called “mixed” models [Balarac et al. 2013]. A significant contribution of neural networks in the past decade has been in optimising the balance between functional and structural components in mixed models (e.g. Sarghini et al. 2003; Maulik et al. 2019a; Beck et al. 2019).

By virtue of the universal approximation theorem [Cybenko 1989], ANNs are expected to be adept at estimating closures without having to rely on further assumptions such as those of statistical symmetry or asymptotically-small filter scales. However, taking advantage of this capability of ANNs to develop closures has typically led to the characterisation of such closures as “black-box” models. We contend that appropriately designed and trained ANNs will learn the correct structure of the closure, and that it remains to be revealed and interpreted through proper analysis. That is, we take the point of view that in addition to providing effective and competitive closures, when analysed in terms of dynamics and physically-relevant diagnostics, a learnt ANN model will be both inter-
Interpreting neural network models of residual scalar flux

pretable and provide useful insights in the on-going task of developing and improving turbulence closures.

We explore the above point of view in the context of statistically-stationary homogeneous isotropic turbulence coupled to a passive scalar with a mean gradient (c.f. Overholt & Pope [1996]). After demonstrating the capability of ANNs to learn algebraic LES closures when trained with filtered DNS data, the performance of the learnt closures are evaluated against that obtained with optimised canonical algebraic models in both a-priori and a-posteriori settings. To the best of our knowledge, we show for the first time that a data-driven SGS model without an assumed form, is not only stable in a-posteriori testing, but also out-performs canonical structural and functional models, even after the coefficients of the canonical models are optimised for the flows considered. Finally, towards interpretability of the learnt model, we perform differential sensitivity analysis of the modelled flux with respect to input parameters. Such an analysis permits us to (a) demonstrate that the ANN learns a closure that is a combination of structural and functional LES closures (cf. mixed models) and in effect (b) highlight the potential of machine learning in accelerating model discovery and closure development.

2. Experimental Configuration

2.1. Equations of motion

We consider flows governed by the incompressible Navier-Stokes equations,

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla) \mathbf{u} = -\frac{1}{\rho_0} \nabla p + \nu \nabla \cdot \nabla \mathbf{u} + A\mathbf{u}; \quad \nabla \cdot \mathbf{u} = 0$$ (2.1)

where $\mathbf{u} = (u_x, u_y, u_z)$ is the velocity vector on the coordinate system $(x, y, z)$, $p$ is the pressure, $\rho_0$ is the constant reference density, $\nu$ is the kinematic fluid viscosity and $A$ is a dynamic coefficient modulated by a forcing scheme, here enforcing statistical stationarity of the kinetic energy. Parametric dependence of the fluctuating quantities has been omitted for simplicity of notation. We introduce a passive scalar with an imposed mean gradient in $z$, which is decomposed as $\phi_t = \Phi + \phi$ where $\phi_t$ is the total scalar concentration and $\phi$ is the turbulent, spatio-temporally fluctuating quantity with respect to the mean scalar concentration $\Phi = zd\Phi/dz$. The turbulent scalar concentration follows

$$\frac{\partial \phi}{\partial t} + (\mathbf{u} \cdot \nabla) \phi = u_z \frac{d\Phi}{dz} + D\nabla \cdot \nabla \phi$$ (2.2)

We then apply the standard LES decomposition by applying a Gaussian filter,

$$G(r) = \left(\frac{6}{\pi \Delta^2}\right)^{1/2} e^{-6r^2/\Delta^2}$$ (2.3)

where $\Delta$ is the isotropic filter length scale, to (2.2) such that the filtered scalar concentration follows

$$\frac{\partial \bar{\phi}}{\partial t} + (\bar{\mathbf{u}} \cdot \nabla) \bar{\phi} = \bar{u}_z \frac{d\Phi}{dz} + D\nabla \cdot \nabla \bar{\phi} + \nabla \cdot \mathbf{q}$$ (2.4)

where $\bar{\cdot}$ is the linear filter operator ($\phi = \bar{\phi} + \phi'$, $\mathbf{u} = \bar{\mathbf{u}} + \mathbf{u}'$) and $\mathbf{q}$ is the residual scalar flux. The residual flux is defined exactly as

$$\mathbf{q}_{DNS} \equiv \bar{\mathbf{u}} - \bar{\phi}\mathbf{u}$$ (2.5)

2.2. Closure models

We consider two canonical local closure models, a “functional”, Prandtl-Smagorinsky (PRS) model, and a “structural”, scalar asymptotic gradient (SAG) model. In the PRS
model, the turbulent diffusivity is related to a turbulent viscosity through a turbulent Prandtl number,

\[ q_{\text{PRS}} = -\frac{\nu_t}{Pr_t} \nabla \bar{\phi} \]  

(2.6)

and \( \nu_t \) itself is determined using the Smagorinsky model \( \nu_t = (C_s \Delta)^2 ||S||_2 \). Here \( C_s \) is the Smagorinsky constant and \( ||S||_2 \) is the \( L_2 \) norm of the resolved strain rate tensor.

While the residual flux in this model is constrained to be parallel to the resolved gradient, stability properties resulting from further limiting the residual flux to a down-gradient direction (forward-scatter) makes this a common modelling choice.

In the second SAG model, a truncated series expansion about the mean gradient that uses the filter scale as a small parameter \( (\Delta)^2 \) \( \nabla \bar{\phi} \cdot \nabla \bar{u} \), (2.7)

leaves the orientation of the residual flux unconstrained. However, since the SAG model has been found to be under-dissipative at filter scales sufficiently larger than dissipation scales \( \text{[Leonard 1974]} \), further ad-hoc fixes of the SAG model, such as combining it with an eddy diffusivity model \( \text{[Clark et al. 1979; Balarac et al. 2013]} \) or using an artificial ‘clipping’ procedure \( \text{[La & Porté-Agel 2013]} \) has been found to be necessary to use this model form in LES.

As an alternative, we model the residual flux using a deep feed-forward ANN, where such a network consists of \( L \) directed and fully connected layers with \( N^{[l]} \) neurons in the \( l \text{-th} \) layer, leading to a bias vector \( b^{[l]} \in \mathbb{R}^{N^{[l]}} \) and a weight tensor \( W^{[l]} \in \mathbb{R}^{N^{[l]} \times N^{[l-1]}} \).

The output of the \( n \text{-th} \) neuron in the \( l \text{-th} \) layer, \( a^{[l]}_n \), is given by

\[ a^{[l]}_n = f^{[l]} \left( \sum_k W^{[l]}_{nk} a^{[l-1]}_k + b^{[l]}_n \right) \]  

(2.8)

where \( f \) is an activation function and \( a^{[0]}, a^{[L]} \) are the input and output objects, respectively. We optimise for \( a^{[L]} = q_{\text{ANN}} \) with inputs \( a^{[0]} = \{ \nabla \bar{\phi}, \nabla \bar{u} \} \), selected to be consistent with the fundamental assumptions of structural models outlined in the introduction and to ensure required symmetries of the sub-filter flux such as those of Galilean-invariance and homogeneity \( \text{[Speziale 1985]} \).

### 2.3. Numerics and optimization

Numerical solution of the relevant equations are obtained by using a standard Fourier pseudo-spectral discretisation of the equations over a triply periodic spatial domain in conjunction with exact pressure projection, a third-order Adams-Bashforth time integration, and a forcing scheme as described in \( \text{[Overholt & Pope 1998]} \). Similar numerics to the system defined by (2.1) and (2.2) have been extensively studied in literature \( \text{[Overholt & Pope 1998; Daniel et al. 2018; Shete & de Bruyn Kops 2019]} \) such that we do not find it necessary to discuss the numerical method in detail.

Reference data for training and testing is obtained from DNS of (2.2) and (2.1) using \( N = 512^3 \) collocation points and at \( Re = 170 \) over 25 large-eddy times. For \textit{a-posteriori} model evaluation, which is described in §4, the filtered scalar equation \( \text{(2.4)} \) is solved with explicitly filtered velocity and the three residual flux models \( \text{(c.f. Vollaant et al. 2016)} \).

For reference, time series of kinetic energy, scalar variance and their respective dissipation rates

\[ E_k = \tfrac{1}{2} (u \cdot u), \quad E_\phi = \tfrac{1}{2} (\phi^2), \quad \epsilon \equiv \nu \langle \nabla u : \nabla u \rangle, \quad \chi \equiv \kappa \langle \nabla \phi \cdot \nabla \phi \rangle, \]
Interpreting neural network models of residual scalar flux

where the notation \( \langle \cdot \rangle \) denotes a spatial average for the DNS solution are shown in figure 1. The timescales considered ensure sufficient sampling of the temporal deviations from mean statistics (cf. Rao & de Bruyn Kops 2011; Portwood et al. 2019).

We use the Adam optimiser to optimise the weights and biases with respect to the mean-squared-error loss function

\[
L = E[(q_{\text{ANN}} - q_{\text{DNS}})^2].
\]

The optimiser iteratively modifies the weights and biases by using gradient descent with added momentum and dampening heuristics (Kingma & Ba 2014). Training data was obtained from the reference DNS solution by using (2.5) at multiple filter scales as summarised in table 1, and where the filter scale is relative to the DNS grid spacing as \( \Delta^* \equiv \Delta / \Delta_{\text{DNS}} \). With spatio-temporal sampling at large eddy scales, the reference DNS data yielded a total of 450,000 samples, with 20% of the data held out for out-of-sample testing. The deep feed-forward ANN consisted of 8 non-linear layers with ‘relu’ activation, 64 neurons per layer and a final linear layer with 512 neurons before the output layer \( q \) to yield a total of 65,000 trainable parameters. This architecture was observed to perform well with respect to different \( a\text{-priori} \) diagnostics in a hyper-parameter study. However, we do not dwell on other architectures since finding optimal architecture is not the objective of this study.

Training the ANN, first over 2000 iterations with a learning rate \( \alpha = 10^{-3} \) and then in a second stage over a similar number of iterations but with the learning rate decimated by an order of magnitude, was found to be robust. The loss function, normalised by \( ||q_{\text{DNS}}||_2^2 \), is shown for the training and testing sets in figure 2. We observe the loss function remaining approximately flat for each case for the first 700 iterations. We show

![Figure 1 & Table 1](image)

**Figure 1 & Table 1:** Left, the time series of kinetic energy, scalar variance and their respective dissipation rates for the training dataset. Right, configuration of LES filter scales. A value of \( \Delta^* = 1 \) indicates DNS resolution. Forcing is applied to the velocity field from largest scales in the domain until the smallest scale \( L_f \).

![Figure 2](image)

**Figure 2:** Training and testing loss for each filter scale as a function of gradient descent iteration is shown in panel (a), where curves with decreasing losses correspond the finer filter scales. Central moments of the output parameter \( q_z \), normalised by ground-truth data, for \( \Delta^* = 10 \) as a function of gradient descent iteration is shown in panel (b). Note that the mean of \( q_z \) develops much faster than higher-order moments.
Figure 3: Panels (a, b) demonstrate a-priori comparison of the p.d.f.s of modelled and ground-truth production for coarsest (a) and finest filter scales (b) for the testing dataset, logscales inset. Note that forward- and back-scatter are more accurately captured by the ANN models.

3. a-priori analysis

A useful a-priori characterisation of an LES model is its ability to reproduce the probability distribution function (p.d.f) of the subfilter scale production of scalar variance, $P_\phi$ as seen in the fully resolved computations, where the production is given by

$$P_\phi = q \cdot \nabla \phi. \quad (3.1)$$

The net downscale cascade of scalar variance in the setting considered, that is, a transfer from resolved scales to unresolved scales, leads to its mean value being positive.

The p.d.f of the production of scalar variance is shown for the DNS data and the three closures considered at two filter sizes in figures 3a, b, with the computation using test data. In this figure, the down-gradient nature of the residual flux in the PRS model constrains the production to be positive-definite whereas production in the DNS is not seen to be constrained in such a fashion. Indeed, the SAG model is seen to be able to produce counter-gradient fluxes (negative production) like in the reference DNS data. The ANN model is seen to similarly produce both counter-gradient and down-gradient fluxes. However, as mentioned earlier, that the SAG model tends to be under-dissipative is seen from the fact that the integrated value $E[P_\phi]$ has a relative error of -56% at the coarsest filter scale, and -22% at the finest filter scale considered. The corresponding numbers for the PRS and ANN models are (0.19%, 4.2%) and (-1.5%, -0.12%) respectively. The ANN model is thus seen to exhibit advantageous characteristics by accurately capturing both mean dynamics of the subfilter scalar variance transfer rate, which in unenforced in optimisation procedure, and also the distribution of the transfer rate.

4. a-posteriori analysis

In an LES setting that models subfilter stresses in the momentum equation as well (that is, in addition to the subfilter scalar flux), dynamics of the passive scalar are affected
Interpreting neural network models of residual scalar flux

not only by the model for the residual scalar flux, but also by the model for the residual momentum stress. It is easy to imagine parameter regimes where the indirect effects of the latter dominates the direct effects of the former, as far as the dynamics of the scalar are concerned. Therefore, in an effort to isolate the fundamental issue of scalar closure, we simulate the filtered passive scalar with exact advective coupling via explicit filtering of DNS solutions (2.1) (c.f. Vollant et al. 2016).

The three closure models are implemented in an a-posteriori simulation fashion and run for 5 large-eddy times using explicitly filtered DNS solution, not used in testing or training datasets, as initial conditions. The scalar variance for each model, \( \langle \bar{\phi}^2 \rangle / 2 \), is shown in figure 4 at four filter scales, each normalized by scalar variances obtained by explicitly filtering DNS solutions. For the coarsest filter, shown in figure 4a, the SAG model characteristically under-dissipates and scalar variance diverges from the filtered DNS solution at early time. The optimal PRS model initially tracks the filtered DNS solution before \( t/\tau_{LE} \approx 0.1 \). For both the PRS and SAG models, the these the relative errors in resolved scalar variance soon begins to asymptote at large values. Whereas the large timescale trends in the evolution of filtered scalar variance is follow by all models, the SAG and PRS model over-predict scalar variance by 25% and 16%, respectively, after 5 large-eddy turnover times. The ANN model, however, tracks the filtered DNS solution accurately where the scalar variance after 5 large-eddy turnover times after is within 4.7% of the filtered DNS solution. The same general trend are observed in figure 4b where \( \Delta^* = 10 \). The accuracy of the ANN model for \( \Delta^* = 18 \) is only matched in PRS and SAG models for \( \Delta^* = 5.7 \) as shown in figure 4c. When the filter length is small, all models track the filtered DNS solution well in time as shown in figure 4d as would be expected.

The temporal tracking of scalar variance is an interesting metric for evaluation of the ANN model because the model is developed without temporal correlations. The model also only depends on spatially-local quantities such that reconstruction of multi-point statistics may only be accurate if the structure of the ANN model accurately predicts the residual flux \( q \), i.e. it does not mimic dynamics in a purely functional sense. We evaluate two physically relevant multi-point diagnostics in figure 5. The second-order structure function of the scalar in the direction of the mean scalar gradient, normalised by the filtered DNS solution, after 5 large-eddy turnover times is shown in figures 5a,b for \( \Delta^* = 10, 18 \), respectively. For \( \Delta^* = 10 \), the SAG model notably misses small-scale behaviour of the scalar, where the PRS and ANN models exhibit more similarity with the filtered DNS results. At the coarser filter width, as shown in figure 5a, the PRS model begins to diverge at larger scales. The ANN model performs comparably for both filter

\[ \langle \bar{\phi}^2 \rangle, \text{rel. error} \]

\[ -0.2 \quad 0 \quad 0.2 \quad 0.4 \]

\[ 0 \quad 1 \quad 2 \quad 3 \quad 4 \quad 5 \]

**Figure 4:** (a-d) Time series of relative error, with respect to filtered DNS solutions, of scalar variance \( \langle \bar{\phi}^2 \rangle \) at multiple filter scales. The finest filter is shown is panel (d), which demonstrates good performance for all models.
scales presented and the diverging behaviour of the SAG at small scales is exaggerated with respect to the case with $\Delta^* = 10$.

Higher order structure functions are also instructive. In passive scalar turbulence forced with a mean gradient, the third-order vertical structure function of the scalar is a physically-relevant diagnostic which should be preserved in a physically-accurate subfilter model (Warhaft 2000). Third-order scalar structure functions are shown in figure 5c,d for $\Delta^* = 10, 18$. First, we note that whereas the PRS model exhibits behaviour similar to that of the ANN model for $\Delta^* = 10$ in the second-order structure functions, the same trends are not observed in figure 5c, especially at larger scales. The SAG model exhibits poorer consistency with large and small scales in 5c. Whereas the ANN model has some inconsistency with the DNS data at small scales, these errors are smaller than as observed in other models, and it closely follows the DNS results at larger scales. Similar behaviour is observed for $\Delta^* = 18$ in figure 5d, except that errors may be more exaggerated at smaller scales for SAG and PRS models. The ANN model for this case exhibits some positive bias, almost uniformly for all $r/\Delta$.

5. Interpretation of data-driven models

We propose that it is instructive to consider the sensitivity of the residual flux to the input fields of scalar and velocity gradients as a means to get further insight into the learnt closure. Indeed, the automatic differentiation capability of the computational frameworks for developing ANNs, which are necessary for the optimisation of the networks via backpropagation, can be leveraged for such a study of the sensitivity of the residual flux to the input parameters.

We compute such sensitivities with the test data and compare them against those for the ANN, PRS and SAG models. In particular, we present an analysis that considers

$$D_1 \equiv \frac{\partial q_z}{\partial (\partial \phi/\partial z)} \text{ and } D_2 \equiv \frac{\partial q_z}{\partial (\partial u_z/\partial x)}.$$  

(5.1)

Isosurfaces of the joint pdf of (5.1) are shown in figure 6a for $\Delta^* = 18$. At large $D_1$, the ANN model exhibits more similarity to the SAG model in terms of $D_2$. However, at moderate negative $D_1$, the model behaves more similar to the PRS model, particularly in the regions near $D_2 \approx 0$. For the next finer filter scale with $\Delta^* = 10$, as shown in figure 6b, isosurfaces of the gradients of the ANN model appear more similar to the SAG model. The interpretation of the ANN model being ‘intermediate’ of the SAG and PRS models suggests that perhaps the ANN model may be approximated by a mixed gradient modes, wherein functional gradient diffusion is added to the structural gradient model.
Interpreting neural network models of residual scalar flux

Figure 6: (a,b) joint p.d.f.s of the derivative of the vertical residual flux with respect to the vertical scalar gradient ($D_1$) and with respect to the horizontal derivative of vertical velocity ($D_2$). Note that the distribution of the ANN and SAG models become more similar with decreasing filter scale.

Furthermore, the ANN model increasingly mimics the SAG model with decreasing $\Delta^*$. The trend with $\Delta^*$ is consistent with the assumption of asymptotically small filter scale with the the SAG model, and suggests that the neural network learns to compensate with gradient-diffusion type dynamics as the filter scale is increased.

6. Concluding remarks

Encouraged by recent advancements of deep learning to industrial and technological applications, data-driven deep learning models have emerged as a promising route to calibrate, augment or replace existing models in the context of fluid turbulence. A major criticism of such learnt turbulence models is that they act as black-boxes, often impeding robust model certification. Responses to this criticism have ranged from attempts to incorporate physical or theoretical constraints, to imposing statistical symmetries, etc., either directly in the architecture of ANNs or through the loss function (Raissi et al. 2017). Surprisingly, a point of view that has not received much attention is that appropriately designed and trained ANNs will learn the correct structure of the closure, and that it remains to be interpreted and understood through proper analysis.

We have approached this point of view by the consideration of spatiotemporally local algebraic residual flux models. By reconstructing the residual fluxes from resolved parameters, we demonstrate the capability of neural networks to discover structural relations within the data without typical assumptions of scale similarity or asymptotically small filter scale. In training these ANN models, we interpret the training process from a statistical perspective by demonstrating that higher-order moments of model outputs require significantly more gradient-descent iterations to converge compared to the means.

We are able to show that in addition to providing effective and competitive closures, when analysed in terms of dynamics and physically-relevant diagnostics, the learnt ANN model can indeed be interpreted in a physically and dynamically meaningful fashion. Indeed, such an analysis permits us to demonstrate that the ANN learns a closure that is a combination of structural and functional LES closures (cf. mixed models). This in effect highlights the potential of machine learning in not only providing useful insights in the on-going task of developing and improving closures, but accelerating the process of model discovery.
REFERENCES

Balarac, G., Le Sommer, J., Meunier, X. & Vollant, A. 2013 A dynamic regularized gradient model of the subgrid-scale scalar flux for large eddy simulations. *Physics of Fluids* **25** (7), 075107.

Bardina, J., Ferziger, J. H. & Reynolds, W. C. 1980 Improved subgrid scale models for large eddy simulations. Paper. AIAA.

Beck, A., Flad, D. & Munz, C.-D. 2019 Deep neural networks for data-driven large eddy simulations. *Journal of Computational Physics* **398**, 108910.

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.

Cybenko, G. 1989 Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems* **2** (4), 303–314.

Daniel, D., Livescu, D. & Ryu, J. 2018 Reaction analogy based forcing for incompressible scalar turbulence. *Phys. Rev. Fluids* **3**, 094602.

Gamahara, M. & Hattori, Y. 2017 Searching for turbulence models by artificial neural network. *Phys. Rev. Fluids* **2**, 054604.

Kingma, D. P. & Ba, J. 2014 Adam: A method for stochastic optimization.

Kutz, J. N. 2017 Deep learning in fluid dynamics. *Journal of Fluid Mechanics* **814**, 1–4.

Leonard, A. 1974 Energy cascade in LES of turbulent fluid flows. *Ag 18A*, 237–248.

Ling, J., Kurzawski, A. & Templeton, J. 2016 Reynolds averaged turbulence modelling using deep neural networks with embedded invariance. *Journal of Fluid Mechanics* **807**.

Lu, H. & Porté-Agel, F. 2013 A modulated gradient model for scalar transport in large-eddy simulation of the atmospheric boundary layer. *Physics of Fluids* **25** (1), 015110.

Maulik, R., San, O., Jacob, J. D. & Crick, C. 2019a Sub-grid scale model classification and blending through deep learning. *Journal of Fluid Mechanics* **870**, 784812.

Maulik, R., San, O., Rasheed, A. & Vedula, P. 2019b Subgrid modelling for two-dimensional turbulence using neural networks. *Journal of Fluid Mechanics* **858**, 122–144.

Moghaddam, A. A. & Sadaghiyani, A. 2018 A deep learning framework for turbulence modeling using data assimilation and feature extraction. *arXiv preprint arXiv:1802.06106*.

Nikolaou, Z. M., Chrysostomou, C., Minamoto, Y. & Vervisch, L. 2019 Neural network-based modelling of unresolved stresses in a turbulent reacting flow with mean shear.

Overholt, M. R. & Pope, S. B. 1996 Direct numerical simulation of a passive scalar with imposed mean gradient in isotropic turbulence. *Phys. Fluids* **8**, 3128–3148.

Overholt, M. R. & Pope, S. B. 1998 A deterministic forcing scheme for direct numerical simulations of turbulence. *Computers & Fluids* **27**, 11–28.

Portwood, G., de Bruyn Kops, S. & Caulfield, C. 2019 Asymptotic dynamics of high dynamic range stratified turbulence. *Physical review letters* **122** (19), 194504.

Portwood, G. D. & et al. 2019 Turbulence forecasting via Neural ODE. *arXiv:1911.05180*.

Raissi, M., Perdikaris, P. & Karniadakis, G. E. 2017 Physics informed deep learning (part i): Data-driven solutions of nonlinear partial differential equations.

Rao, K. J. & de Bruyn Kops, S. M. 2011 A mathematical framework for forcing turbulence applied to horizontally homogeneous stratified flow. *Phys. Fluids* **23**, 065110.

Sagaut, P. 2006 *Large Eddy Simulation for Incompressible Flows*, 3rd edn. Berlin: Springer.

Salehipour, H. & Peltier, W. R. 2019 Deep learning of mixing by two atoms of stratified turbulence. *Journal of Fluid Mechanics* **861**, R4.

Sarghini, F., De Felice, G. & Santini, S. 2003 Neural networks based subgrid scale modeling in large eddy simulations. *Computers & fluids* **32** (1), 97–108.

Shete, K. P. & de Bruyn Kops, S. M. 2019 Area of scalar isosurfaces in homogeneous isotropic turbulence as a function of reynolds and schmidt numbers. *Journal of Fluid Mechanics* **883**.

Smagorinsky, J. 1963 General circulation experiments with the primitive equations. I. The basic experiment. *Mon. Weather Rev.* **91**, 99–164.

Speziale, C. G. 1985 Galilean invariance of subgrid-scale stress models in the large-eddy simulation of turbulence. *J. Fluid Mech.* **156**, 5562.

Vollant, A., Balarac, G. & Corre, C. 2016 A dynamic regularized gradient model of the subgrid-scale stress tensor for large-eddy simulation. *Physics of Fluids* **28**, 025114.

Warhaft, Z. 2000 Passive scalar in turbulent flows. *Annu. Rev. Fluid Mech.* **32**, 203–240.