Learning Closed-form Equations for Subgrid-scale Closures from High-fidelity Data: Promises and Challenges

Karan Jakhar[1], Yifei Guan[1], Rambod Mojgani[1], Ashesh Chattopadhyay[1], Pedram Hassanzadeh[1,2], and Laure Zanna[3]

[1]Department of Mechanical Engineering, Rice University, Houston, TX, USA
[2]Department of Earth, Environmental and Planetary Sciences, Rice University, Houston, TX, USA
[3]Courant Institute of Mathematical Sciences, New York University, New York City, NY, USA

Key Points:

- Subgrid-scale momentum/heat flux closures discovered using common algorithms are the analytically derivable nonlinear gradient model (NGM)
- In 2D turbulence/convection, NGM leads to unstable online simulations due to its inability to fully capture key inter-scale energy transfers
- We suggest that physics-informed loss functions, libraries, and sparsity-level/metrics are needed to discover accurate and stable closure

Corresponding author: Pedram Hassanzadeh, pedram@rice.edu

Abstract

There is growing interest in discovering interpretable, closed-form equations for subgrid-scale (SGS) closures/parameterizations of complex processes in Earth system. Here, we apply a common equation-discovery technique with expansive libraries to learn closures from filtered direct numerical simulations of 2D forced turbulence and Rayleigh-Bénard convection (RBC). Across common filters, we robustly discover closures of the same form for momentum and heat fluxes. These closures depend on nonlinear combinations of gradients of filtered variables (velocity, temperature), with constants that are independent of the fluid/flow properties and only depend on filter type/size. We show that these closures are the nonlinear gradient model (NGM), which is derivable analytically using Taylor-series expansions. In fact, we suggest that with common (physics-free) equation-discovery algorithms, regardless of the system/physics, discovered closures are always consistent with the Taylor-series. Like previous studies, we find that large-eddy simulations with NGM closures are unstable, despite significant similarities between the true and NGM-predicted fluxes (pattern correlations > 0.95). We identify two shortcomings as reasons for these instabilities: in 2D, NGM produces zero kinetic energy transfer between resolved and subgrid scales, lacking both diffusion and backscattering. In RBC, backscattering of potential energy is poorly predicted. Moreover, we show that SGS fluxes diagnosed from data, presumed the "truth" for discovery, depend on filtering procedures and are not unique. Accordingly, to learn accurate, stable closures from high-fidelity data in future work, we propose several ideas around using physics-informed libraries, loss functions, and metrics. These findings are relevant beyond turbulence to closure modeling of any multi-scale system.

Plain Language Summary

Even in state-of-the-art climate models, the effects of many important small-scale processes cannot be directly simulated due to limited computing power. Thus, these effects are represented using functions called parameterizations. However, many of the current physics-based parameterizations have major shortcomings, leading to biases and uncertainties in the models' predictions. Recently, there has been substantial interest in learning such parameterizations directly from short but very high-resolution simulations. Most studies have focused on using deep neural networks, which while leading to successful parameterizations in some cases, are hard to interpret and explain. A few more recent studies have focused on another class of machine-learning methods that discover equations. This approach has resulted in fully interpretable but unsuccessful parameterizations that produce unphysical results. Here, using widely-used test cases, we 1) explain the reasons for these unphysical results, 2) connect the discovered equations to well-known mathematically derived parameterizations, and 3) present ideas for learning successful parameterizations using equation-discovery methods. Our main finding is that the common loss functions that match patterns representing effects of small-scale processes are not enough, as important physical phenomena are not properly learned. Based on this, we have proposed a number of physics-aware metrics and loss functions for future work.

1 Introduction

Turbulent flows are ubiquitous in many geophysical systems, including atmospheric and oceanic circulations, and play an important role, e.g., greatly enhancing mixing and transport. Direct numerical simulation (DNS) of high-dimensional turbulent flows often becomes computationally intractable. Therefore, numerical simulations of most geophysical turbulent flows cannot resolve all the relevant scales (Fox-Kemper et al., 2019; Palmer, 2001; Schneider, Teixeira, et al., 2017). Large-eddy simulation (LES) is a practical approach to balance computational cost and accuracy: the large scales of the flow are explicitly resolved, while the effects of the small-scale features which cannot be resolved by the given grid resolution, called subgrid-scale (SGS) features, are parameterized as a function of the

resolved flow (Pope, 2000; Sagaut, 2006; Smagorinsky, 1963). However, the performance of the LES models strongly depends on the accuracy of the employed SGS closure. Over years, there have been extensive efforts focused on formulating physics-based and semi-empirical SGS closures using various techniques in many turbulent flows (Meneveau & Katz, 2000; Moser et al., 2021; Pope, 2000; Sagaut, 2006), including geophysical flows (Alexander & Dunkerton, 1999; Anstey & Zanna, 2017; Berner et al., 2017; Cessi, 2008; Gallet & Ferrari, 2020; Herman & Kuang, 2013; Jansen & Held, 2014; O'Kane & Frederiksen, 2008; Khodkar et al., 2019; Sadourny & Basdevant, 1985; Schneider, Teixeira, et al., 2017; Sridhar et al., 2022; Sullivan et al., 1994; Tan et al., 2018; Zanna et al., 2017).

The challenge of modeling SGS closures lies in faithfully representing the two-way interactions between the SGS processes and the resolved, large-scale dynamics. There are two general approaches to SGS modeling: (a) functional and (b) structural (Sagaut, 2006). The functional SGS closures are developed by considering the inter-scale interactions (e.g., energy transfers). This is often achieved by introducing a dissipative term. Hence, functional SGS closures generally take an eddy-viscosity form to mimic the average function of the SGS eddies. Among the first and most-used functional closures is the Smagorinsky model (Smagorinsky, 1963). Later, dynamic formulations of this model were proposed, in which the key coefficient is dynamically adjusted to the local structures of the flow (Germano, 1992; Lilly, 1992; Ghosal et al., 1993; Chai & Mahesh, 2012). Existing functional closures, most of which are the eddy-viscosity type, can be excessively dissipative (Vreman et al., 1996; Guan et al., 2022a). Furthermore, they cannot capture the structure of the SGS terms, leading to a low correlation coefficient (CC<0.5) with the true SGS terms, i.e., those diagnosed from the DNS data (Carati et al., 2001; Guan et al., 2022a; Moser et al., 2021).

On the contrary, structural closures tend to have much higher CC with the true SGS terms. Structural closures approximate the SGS terms by constructing it from an evaluation of large-scale motions or a formal series expansion. One of the most common structural closures is the nonlinear gradient model (Leonard, 1975; Clark et al., 1979), referred to as NGM hereafter (it is also known as the tensor diffusivity model). *The NGM can be derived analytically*: the SGS term is approximated using a first-order truncated Taylor-series expansion of the SGS stress' convolution integral (details discussed later). However, despite CC>0.9, LES with NGM closure has been found to be unstable in many studies of two-dimensional (2D) and three-dimensional (3D) turbulence. These instabilities are often attributed to insufficient dissipation and more importantly, to the presence of too-strong backscattering in NGM (Leonard, 1997, 2016; Liu et al., 1994; Fabre & Balarac, 2011; Lu & Porté-Agel, 2010; Meneveau & Katz, 2000; Prakash et al., 2021; S. Chen et al., 2003, 2006; Vollant et al., 2016; Moser et al., 2021). As a result, while backscattering (basically anti-diffusion or up-gradient flux) is an important process to represent in closure models (Grooms et al., 2015; Guan et al., 2022a; Hewitt et al., 2020; Nadiga, 2010; Shutts, 2005), it is ignored in most practical SGS closures in favor of stability (though there has been some new exciting progress; see, e.g., Jansen et al. (2015) and Juricke et al. (2020)). In fact, currently operational climate models do not accounts for backscattering in their ocean parameterizations (Hewitt et al., 2020). Consequently, a framework for developing SGS closures with the right amount of diffusion and backscattering, that can capture both the structure and function of the SGS terms, has remained elusive (Moser et al., 2021; Pope, 2000; Sagaut, 2006).

Before moving forward, it should be pointed out that while the discussion so far has been focused on closure for geophysical turbulence, many other critical processes in the Earth system (in atmosphere, ocean, land, cryosphere, biosphere and at their interfaces) require parameterizations in Earth system models (Stensrud, 2009; Schneider, Jeevanjee, & Socolow, 2021). Thus, the discussion below and as clarified later, the findings of this paper, are broadly relevant to parameterization efforts in Earth science.

Recently, machine learning (ML) has brought new tools into SGS closure modeling (Schneider, Lan, et al., 2017; Zanna & Bolton, 2021; Brunton et al., 2020; Duraisamy, 2021;

Gentine et al., 2021; Balaji, 2021). The strength of ML techniques is their ability to handle high-dimensional data and learn strongly nonlinear relationships. Therefore, ML techniques are attractive tools that might be able to extract more hidden knowledge from data, potentially providing better SGS closures and even new insights into SGS physics. Data-driven SGS closures, e.g., based on deep neural networks trained on high-fidelity simulation data such as DNS data, have been developed for canonical geophysical flows such as 2D and quasi-geostrophic turbulence (Bolton & Zanna, 2019; Frezat et al., 2022; Guan et al., 2022a, 2023; Pawar et al., 2020; Maulik et al., 2018; Srinivasan et al., 2023) and oceanic and atmospheric circulations (Beucler et al., 2021; Brenowitz & Bretherton, 2018; Cheng et al., 2022; Guillaumin & Zanna, 2021; Rasp et al., 2018; Yuval & O'Gorman, 2020; X. Zhang et al., 2022). While some of these studies found the learned data-driven SGS closures to lead to stable and accurate LES (Yuval & O'Gorman, 2020; Guan et al., 2022a, 2023; Frezat et al., 2022), a number of major challenges remain (Schneider, Jeevanjee, & Socolow, 2021; Balaji, 2021). Perhaps the most important one is *interpretability*, which is difficult for neural networks, despite some recent advances in explainable ML for climate-related applications (Clare et al., 2022; Mamalakis et al., 2022), including for SGS modeling (Subel et al., 2023). The black-box nature of neural network-based closures aside, there are also challenges related to generalizability, computational cost, and even implementation (Balaji, 2021; Chattopadhyay et al., 2020; Guan et al., 2022a; Kurz & Beck, 2020; Maulik et al., 2019; Subel et al., 2021; Xie et al., 2019; Zhou et al., 2019), limiting the broad application of such closures in operational climate and weather models, at least for now.

An alternative approach that is rapidly growing in popularity involves using ML techniques that provide interpretable, closed-form equations, e.g. using sparse linear regression. The underlying idea of this *equation-discovery* approach is that given spatial, temporal, or spatio-temporal data from a system, one can discover the governing (algebraic or differential) equations of that system (Brunton et al., 2016; Y. Chen et al., 2022; Goyal & Benner, 2022; Mojgani et al., 2022b; Schneider et al., 2020; Rudy et al., 2017; Schaeffer, 2017; Schmidt & Lipson, 2009; Schneider, Stuart, & Wu, 2021; Schneider et al., 2022; Udrescu & Tegmark, 2020; S. Zhang & Lin, 2018). Most of the aforementioned studies are focused on discovering the entire governing equations from data, though few recent studies have used this approach to discover SGS closures (see below). This approach has the following advantages over more complex methods such as neural networks in the context of SGS modeling: 1) the learned closure is significantly easier to interpret based on physics (Zanna & Bolton, 2020), 2) the number of required training samples and the training costs are often considerably lower (Brunton et al., 2020; Mojgani et al., 2022b), and 3) the computational cost of implementation in conventional solvers is lower, as the discovered closures often involve traditional operations, e.g., gradients and Laplacians (Udrescu & Tegmark, 2020; Ross et al., 2023).

A number of equation-discovery techniques and test cases have been recently employed for *structural* modeling of the SGS stress. In the first study of its kind, Zanna and Bolton (2020) used relevance vector machine (RVM), a sparsity-promoting Bayesian linear regression technique, with a library of second-order velocity derivatives and their nonlinear combinations, to learn a closed-form closure model for the SGS momentum and buoyancy fluxes from *filtered* high-resolution simulations of ocean mesoscale turbulence. They found a closure that resembled the NGM, with close connections to earlier physics-based modeling work by Anstey and Zanna (2017). Although, the discovered closure performed well in *a priori* (offline) tests, it was unstable *a posteriori* (online), i.e., when it was coupled to a low-resolution ocean solver. Following the same general approach, more recently, Ross et al. (2023) proposed a novel equation-discovery approach combining linear regression and genetic programming (GP). This hybrid approach uses GP to discover the structure of the equation followed by linear regression to fine-tune the coefficients. In contrast to methods such as RVM, GP does not require an explicit library of features, instead, it uses a simple set of features and operations, and constructs expressions by successively applying operators and combining expressions. Similarly, in other disciplines, Reissmann et al. (2021) and Li et al. (2021) recently used gene expression programming (GEP) to discover SGS stress for

the Taylor-Green vortex and the 3D isotropic turbulence, respectively. They developed a nonlinear closure consisting of the local strain rate and rotation rate tensors, based on what is known as Pope tensors (Pope, 1975), which will be discussed later. Overall, these more recent studies found that GEP- and GP-based closures often outperform common baselines such as the Smagorinsky and the mixed models when turbulence statistics and flow structures are considered (Li et al., 2021; Reissmann et al., 2021; Ross et al., 2023). Note that there also have been a number of studies focused on equation-discovery for *functional* modeling, e.g., using techniques such as Ensemble Kalman inversion (Schneider, Stuart, & Wu, 2021; Schneider et al., 2020); see the Summary and Discussion.

In this study, we build on the work by Zanna and Bolton (2020) and use 2D-forced homogeneous isotropic turbulence (2D-FHIT) and Rayleigh-Bénard convection (RBC) to extend and expand their analysis in several directions:

1. We use RVM with an expansive high-order library to discover closures from DNS data for the SGS momentum flux tensor (2D-FHIT and RBC) and the SGS heat flux vector (RBC).
2. We conduct extensive robustness analysis of the discovered closures across a variety of flow configurations, filter types, and filter sizes, and examine the potential effects of numerical errors.
3. Further clarify the connections between the robustly discovered SGS momentum and heat flux closures, and the SGS closures obtained analytically from the truncated Taylor-series expansion of the filter's convolution integral, the NGM (Leonard, 1975).
4. Explain the physical reason for the unstable *a posteriori* LES with the discovered SGS closures, despite their high *a priori* accuracy in some metrics (such as CC).
5. Present a decomposition of the SGS tensor to the Leonard, cross, and Reynolds components, showing their relative importance and dependence on the filter type/size.
6. Based on these findings, we present a number of ideas for discovering stable and accurate SGS closures from the data in future work.

Note that while we focus on the use of RVM here, our findings and conclusions in (1)-(6) are applicable to any equation-discovery effort, and not just for SGS momentum and heat fluxes in geophysical turbulence, but for SGS modeling in any nonlinear dynamical system.

This paper is organized as follows. In Section 2, we provide an introduction to methodology, including the governing equations of test cases (2D-FHIT and RBC), filtering procedure for data and equations, RVM algorithm, and the employed library of the basis functions. Section 3 includes the discussion on the discovered closures, *a priori* and *a posteriori* tests, connection with the physics-based closures, and contribution of the Leonard, cross, and Reynolds components. Summary and Discussion are in Section 4.

2 Models, Methods, and Data

2.1 Filtering Procedure

In DNS, the velocity field, $\boldsymbol{u}(\boldsymbol{x},t)$, is resolved using high spatio-temporal resolutions down to all relevant scales. In LES, a low-pass filtering operation, denoted by $\overline{(.)}$, is performed on the equations and flow fields. The resulting filtered fields, for example, filtered velocity, $\overline{\boldsymbol{u}}\,(\boldsymbol{x},t)$, can be adequately resolved using relatively coarse spatio-temporal resolutions: the required grid spacing is proportional to the specified filter width, Δ, which is analogous to the size of the smallest eddies resolved in the LES (Pope, 2000; Sagaut, 2006). Using $\boldsymbol{u}(\boldsymbol{x},t)$ as an example, the general spatial filtering operation is defined by (Sagaut, 2006)

$$\overline{\boldsymbol{u}}(\boldsymbol{x},t) = G * \boldsymbol{u} \;= \int_{-\infty}^{\infty} G\,(\boldsymbol{r})\,\boldsymbol{u}(\boldsymbol{x}-\boldsymbol{r},t)d\boldsymbol{r}, \tag{1}$$

where $*$ is the convolution operator, and the integration is performed over the entire domain. The specified filter kernel, G, satisfies the normalization condition

$$\int_{-\infty}^{\infty} G(\boldsymbol{r})\, d\boldsymbol{r} = 1. \tag{2}$$

Subsequently, any flow field such as velocity can be decomposed into a filtered (resolved) part and SGS (residual) part:

$$\boldsymbol{u}(\boldsymbol{x}, t) = \overline{\boldsymbol{u}}(\boldsymbol{x}, t) + \boldsymbol{u}'(\boldsymbol{x}, t), \tag{3}$$

where $\boldsymbol{u}'$ is the SGS field. While this appears to be analogous to the Reynolds decomposition, an important distinction should be noted: the filtered residual field may not be strictly zero ($\overline{\boldsymbol{u}'} \neq 0$, thus $\overline{\overline{\boldsymbol{u}}} \neq \overline{\boldsymbol{u}}$), depending on the choice of the filter function (Sagaut, 2006). Further details about the filters used in this work (Gaussian, box, Gaussian + box, and sharp-spectral) are given in Appendix A.

2.2 Two-dimensional Forced Homogeneous Isotropic Turbulence (2D-FHIT)

We consider 2D-FHIT as the first test case. This canonical flow has been extensively used for testing novel physics-based and ML-based SGS closures for geophysical turbulence in the past decades (Boffetta & Ecke, 2012; Chandler & Kerswell, 2013; Guan et al., 2022a; Tabeling, 2002; Thuburn et al., 2014; Vallis, 2017; Verkley et al., 2019). The dimensionless continuity and momentum equations for 2D-FHIT in (x, y) spatial dimensions are:

$$\nabla \cdot \boldsymbol{u} = 0, \tag{4}$$

$$\frac{\partial \boldsymbol{u}}{\partial t} + (\boldsymbol{u} \cdot \nabla)\, \boldsymbol{u} = -\nabla p + \frac{1}{Re} \nabla^2 \boldsymbol{u} + \boldsymbol{\mathcal{F}} + \boldsymbol{\mathcal{R}}, \tag{5}$$

where $\boldsymbol{u} = (u, v)$ is the velocity, p is the pressure, $\boldsymbol{\mathcal{F}}$ represents a time-constant external forcing, $\boldsymbol{\mathcal{R}}$ is the Rayleigh drag, and Re is the Reynolds number. The domain is doubly periodic with length $L = 2\pi$.

The equations for LES are obtained by applying a homogeneous 2D filter (Eq. (1)) to Eqs. (4)-(5). The filtered continuity and momentum equations are:

$$\nabla \cdot \overline{\boldsymbol{u}} = 0, \tag{6}$$

$$\frac{\partial \overline{\boldsymbol{u}}}{\partial t} + (\overline{\boldsymbol{u}} \cdot \nabla)\, \overline{\boldsymbol{u}} = -\nabla \overline{p} + \frac{1}{Re} \nabla^2 \overline{\boldsymbol{u}} - \nabla \cdot \boldsymbol{\tau} + \overline{\boldsymbol{\mathcal{F}}} + \overline{\boldsymbol{\mathcal{R}}}, \tag{7}$$

where $\boldsymbol{\tau}$ is the SGS stress tensor:

$$\boldsymbol{\tau} = \begin{bmatrix} \tau_{xx} & \tau_{xy} \\ \tau_{yx} & \tau_{yy} \end{bmatrix} = \begin{bmatrix} \overline{u^2} - \overline{u}^2 & \overline{uv} - \overline{u}\,\overline{v} \\ \overline{uv} - \overline{u}\,\overline{v} & \overline{v^2} - \overline{v}^2 \end{bmatrix}. \tag{8}$$

A closure model is needed to represent τ_{xx}, τ_{xy} $(= \tau_{yx})$, and τ_{yy}, in terms of the resolved flow $(\overline{u}, \overline{v}, \overline{p})$. However, currently, this is not possible just using the first principles due to the presence of the $\overline{u^2}, \overline{uv}$, and $\overline{v^2}$ terms.

We study three cases of 2D-FHIT (Table 1), creating a variety of flows that differ in dominant length scales and energy/enstrophy cascade regimes. For DNS, as discussed in Appendix B, Eqs. (4)-(5) are numerically solved at high spatio-temporal resolutions using a Fourier-Fourier pseudo-spectral solver. For the LES, the same solver at lower spatio-temporal resolution is used (Appendix B).

2.3 Turbulent Rayleigh-Bénard Convection (RBC)

As our second test case, we use 2D turbulent RBC, a widely used canonical flow for buoyancy-driven turbulence (Chillà & Schumacher, 2012; Dabbagh et al., 2017; Hassanzadeh

Table 1. Physical and numerical parameters used in the 2D-FHIT cases. Cases with different flow regimes are produced by varying forcing wavenumber, (f_{k_x}, f_{k_y}), and Re. For each case, we use several filter types (Gaussian, box, Gaussian + box, and sharp-spectral filters) and filter sizes, $\Delta = \frac{2L}{N_{\mathrm{LES}}}$, where $N_{\mathrm{LES}} = \{32, 64, 128, 256\}$ for Cases K1 and K3 and $N_{\mathrm{LES}} = \{128, 256\}$ for Case K2. Here, N_{LES} and N_{DNS} are the number of points in each direction on the LES and DNS grids, respectively. $L = 2\pi$ is the length of the domain. Note that the lowest N_{LES} is chosen such that the LES resolution resolves at least 80% of the DNS kinetic energy (Pope, 2000). Filters are applied in both spatial dimensions for 2D-FHIT.

Cases	Re	(f_{k_x}, f_{k_y})	N_{DNS}
K1	20,000	$(4, 0)$	1024
K2	20,000	$(25, 25)$	1024
K3	100,000	$(4, 0)$	2048

et al., 2014; Kooloth et al., 2021; Lappa, 2009; Sondak et al., 2015), which in addition to the SGS (momentum) stress, requires closure modeling of the SGS heat flux (Pandey et al., 2022; Peng & Davidson, 2002; Wang et al., 2008). Under the Oberbeck-Boussinesq approximation, the dimensionless governing equations for the flow between horizontal walls at fixed temperatures (the bottom wall being warmer than the top) in (x, z) spatial dimensions are:

$$\nabla \cdot \boldsymbol{v} = 0, \tag{9}$$

$$\frac{\partial \boldsymbol{v}}{\partial t} + \boldsymbol{v} \cdot \nabla \boldsymbol{v} = -\nabla p + Pr \nabla^2 \boldsymbol{v} + Ra\, Pr\, \theta \hat{\boldsymbol{z}}, \tag{10}$$

$$\frac{\partial \theta}{\partial t} + \boldsymbol{v} \cdot \nabla \theta - w = \nabla^2 \theta, \tag{11}$$

where $\boldsymbol{v} = (u, w)$ is the velocity, θ is the temperature (T) departure from the conduction state, $\hat{\boldsymbol{z}}$ is the unit vector in the vertical direction, and Ra and Pr are the Rayleigh and Prandtl numbers, respectively. The domain is periodic in the horizontal direction with length $L = 6\pi$; no-slip boundary conditions are applied at the walls. We use three cases of turbulent RBC (Table 2) in which the Ra and Pr are varied.

To properly resolve the thin boundary layers in turbulent RBC, a pseudo-spectral solver with (non-uniform) Chebyshev collocation points in the vertical direction is used. However, filtering variables on a non-uniform grid can cause major errors in the diagnosed SGS terms, because the filters will not commute with spatial derivatives (Yalla et al., 2021). As a result, following the common practice for LES, we only filter the equations in the horizontal direction, where (uniform) Fourier collocation points are used. The LES equations obtained by applying a 1D filter along the horizontal direction, x, to Eqs. (9)-(11) are:

$$\nabla \cdot \overline{\boldsymbol{v}} = 0, \tag{12}$$

$$\frac{\partial \overline{\boldsymbol{v}}}{\partial t} + \overline{\boldsymbol{v}} \cdot \nabla \overline{\boldsymbol{v}} = -\nabla \overline{p} + Pr\, \nabla^2 \overline{\boldsymbol{v}} + Pr\, Ra\, \overline{\theta} \hat{\boldsymbol{z}} - \nabla \cdot \boldsymbol{\tau}, \tag{13}$$

$$\frac{\partial \overline{\theta}}{\partial t} + \overline{\boldsymbol{v}} \cdot \nabla \overline{\theta} - \overline{w} = \nabla^2 \overline{\theta} - \nabla \cdot \boldsymbol{J}, \tag{14}$$

where $\boldsymbol{\tau}$ is the SGS (momentum) stress tensor

$$\boldsymbol{\tau} = \begin{bmatrix} \tau_{xx} & \tau_{xz} \\ \tau_{zx} & \tau_{zz} \end{bmatrix} = \begin{bmatrix} \overline{u^2} - \overline{u}^2 & \overline{uw} - \overline{u}\,\overline{w} \\ \overline{uw} - \overline{u}\,\overline{w} & \overline{w^2} - \overline{w}^2 \end{bmatrix}, \tag{15}$$

and $\boldsymbol{J}$ is the SGS heat flux vector

$$\boldsymbol{J} = \begin{bmatrix} J_x \\ J_z \end{bmatrix} = \begin{bmatrix} \overline{u\theta} - \overline{u}\,\overline{\theta} \\ \overline{w\theta} - \overline{w}\,\overline{\theta} \end{bmatrix}. \tag{16}$$

Table 2. Physical and numerical parameters used in three cases of turbulent RBC. Cases with different flow regimes are produced by varying Ra and Pr. For each case, we use several filter types (Gaussian, box, Gaussian + box, and sharp-spectral cutoff filters) and filter size $\Delta = \Delta_x = \frac{2L}{N_{\text{LES}}}$, where $N_{\text{LES}} = \{128, 256\}$ for Case R1 and $N_{\text{LES}} = \{128, 256, 512\}$ for Cases R2 and R3. Here, N_{LES} is the number of points on the LES grid in the horizontal direction, x. N_x^{DNS} and N_z^{DNS} are the number of grid point on the DNS grid in the horizontal and vertical directions, respectively. $L = 6\pi$ is the length of the domain in the horizontal direction. Note that the lowest N_{LES} is chosen such that the LES resolution resolves at least 80% of the DNS kinetic energy (Pope, 2000). Filters are only applied along the horizontal direction.

Cases	Ra	Pr	$\left(N_x^{\text{DNS}}, N_z^{\text{DNS}}\right)$
R1	10^6	100	$(2048, 400)$
R2	40×10^6	7	$(2048, 400)$
R3	40×10^6	100	$(2048, 400)$

Here, in addition to $\boldsymbol{\tau}$, $\boldsymbol{J}$ needs a closure model too.

For DNS, as discussed in Appendix C, Eqs. (9)-(11) are numerically solved at high spatio-temporal resolutions using a Fourier-Chebyshev pseudo-spectral solver. For LES, the same solver with lower spatial resolution is used (Appendix C).

2.4 Filtered Direct Numerical Simulation (FDNS) Data

It should be highlighted that in this study with two canonical test cases, we consider DNS data as the "truth", and use filtered DNS (FDNS) data to discover the closures. However, in reality, performing DNS for many geophysical flows is computationally prohibitive. In such cases, high-resolution LES that adequately resolves the process of interest (e.g., ocean eddies, gravity waves, etc.) is often used as the truth to train the ML algorithms for SGS modeling (Yuval & O'Gorman, 2020; Zanna & Bolton, 2021; Shen et al., 2022; Sun et al., 2023).

Here, we compute FDNS variables on the LES grids, which are 4 to 64 times coarser than the DNS grid in both spatial dimensions for 2D-FHIT and one spatial dimension for RBC (see Tables 1-2). More specifically, we first apply the respective filter's transfer function (Tables A1 and A2) to the DNS data, and then coarse-grain the results onto the LES grid. Note that following some of the recent papers (Grooms et al., 2021; Guan et al., 2022a), we define "filtering" as an operation that removes the small scales but keeps the grid resolution (e.g., DNS), and "coarse-graining" as an operation that changes the grid size, e.g., from the DNS resolution to LES resolution. Note that $\boldsymbol{\tau}$ and $\boldsymbol{J}$ in Eqs. (7), (13), and (14) need to be on the LES grid.

The filtering and coarse-graining are performed following Sagaut (2006) and Guan et al. (2022a). Briefly, using the velocity $\boldsymbol{u}\left(\boldsymbol{x}_{\text{DNS}}, t\right)$ as an example, and denoting the DNS grid and wavenumber as $\boldsymbol{x}_{\text{DNS}}$ and $\boldsymbol{k}_{\text{DNS}}$, we first transform the DNS velocity into the spectral space $\hat{\boldsymbol{u}}\left(\boldsymbol{k}_{\text{DNS}}, t\right)$, where $\hat{(\cdot)}$ means Fourier transformed. This is followed by applying the filter in the spectral space:

$$\overline{\hat{\boldsymbol{u}}}\left(\boldsymbol{k}_{\text{DNS}}, t\right) = \hat{G}\left(\boldsymbol{k}_{\text{DNS}}\right) \odot \hat{\boldsymbol{u}}\left(\boldsymbol{k}_{\text{DNS}}, t\right). \tag{17}$$

Here, $\hat{G}\left(\boldsymbol{k}_{\text{DNS}}\right)$ can be any of the transfer functions listed in Tables A1 and A2, and $\odot$ is the Hadamard (element-wise) multiplication. After the filtering operation, coarse-graining is performed to transform the filtered variable from the DNS to the LES grid. In this study,

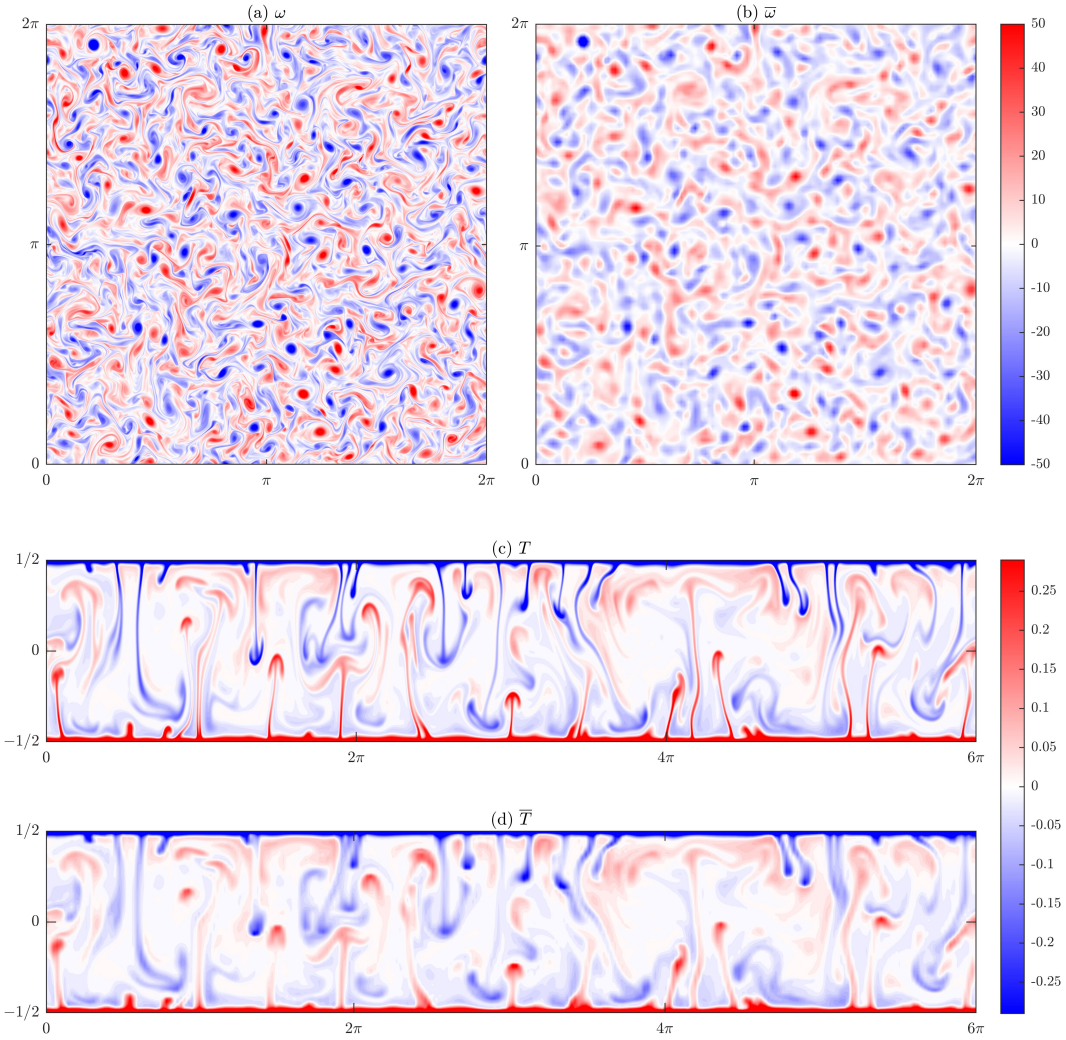

Figure 1. Snapshots of the (a) DNS vorticity field ω ($N_{\mathrm{DNS}} = 1024$) and the (b) FDNS vorticity field $\overline{\omega}$ ($N_{\mathrm{LES}} = 128$) for Case K2 (see Table 1). The (c) DNS temperature field T ($N_{\mathrm{DNS}} = 2048$), and the (d) FDNS temperature field $\overline{T}$ ($N_{\mathrm{LES}} = 256$) for Case R3 (see Table 2). The Gaussian filter is applied in both cases.

we perform coarse-graining in spectral space with cutoff $k_c = \pi/\Delta$, which for example in 2D, yields

$$\overline{\widehat{\boldsymbol{u}}}(\boldsymbol{k}_{\mathrm{LES}}, t) = \overline{\widehat{\boldsymbol{u}}}\left(|k_{\mathrm{DNS},x}| < k_c, |k_{\mathrm{DNS},y}| < k_c, t\right). \tag{18}$$

Hereafter, for brevity, we use the term "filtered" (still denoted by $\bar{\cdot}$) to mean "filtered" and then "coarse-grained".

Figure 1 shows the effects of filtering on the vorticity and temperature fields for 2D-FHIT and RBC, illustrating that the small-scale structures of ω and T are removed due to filtering and the fields are smoothed out.

2.5 The Equation-discovery Method

In this study, we employ the RVM (Tipping, 2001) to discover closed-form closures for each element of the $\boldsymbol{\tau}$ tensor and $\boldsymbol{J}$ vector from the FDNS data. RVM is a sparsity-

promoting Bayesian (linear) regression technique that has shown promise in applications involving dynamical systems (S. Zhang & Lin, 2018; Zanna & Bolton, 2020; Mojgani et al., 2022b). RVM relies on a pre-specified library of basis functions $\boldsymbol{\Phi}$; each column of this matrix is a basis, e.g., a linear or nonlinear combination of relevant variables such as velocity and temperature and/or their derivatives. The library should be expressive enough so that $\boldsymbol{s}$, a vectorized snapshot of a element of any $\boldsymbol{\tau}$ or $\boldsymbol{J}$, could be completely represented as

$$\boldsymbol{s}^{\mathrm{RVM}} = \boldsymbol{\Phi}\boldsymbol{c}. \tag{19}$$

The vector of regression weights, $\boldsymbol{c}$, is computed by minimizing the mean-squared error (MSE)

$$\mathrm{MSE} = \|\mathcal{S}^{\mathrm{RVM}} - \mathcal{S}^{\mathrm{FDNS}}\|_2^2, \tag{20}$$

where vector $\mathcal{S}$ consists of n samples of $\boldsymbol{s}$ stacked together. RVM assumes Gaussian prior distributions for each weight, and the width of the Gaussian posterior provides a measure of the weight's uncertainty. Sparsity is enforced via an iterative process: basis functions whose weights' uncertainties exceed a pre-specified hyperparameter (threshold), α, are removed (pruned), and Eq. (20) is minimized again. The iterations stop when all the remaining basis functions have uncertainties smaller than α. Larger α results in lower MSE but more terms in the discovered model (see below).

A critical step in using RVM (and most equation-discovery methods) is the choice of the library. Here, we have chosen the following libraries. For momentum stress, we use

$$\left[\frac{\partial^{(q_1+q_2)}A}{\partial x^{q_1}\partial y^{q_2}}\right]^{p_1}\left[\frac{\partial^{(q_4+q_5)}B}{\partial x^{q_4}\partial y^{q_5}}\right]^{p_2} \quad \text{or} \quad \left[\frac{\partial^{(q_1+q_2)}C}{\partial x^{q_1}\partial z^{q_2}}\right]^{p_1}\left[\frac{\partial^{(q_4+q_5)}D}{\partial x^{q_4}\partial z^{q_5}}\right]^{p_2}; \tag{21}$$

where $A, B = \overline{u}$ or $\overline{v}$ (2D-FHIT) and $C, D = \overline{u}$ or $\overline{w}$ (RBC). Note that experiments with including $\overline{\theta}$ in D yield the same results. For heat flux, we use

$$\left[\frac{\partial^{(q_1+q_2)}A}{\partial x^{q_1}\partial z^{q_2}}\right]^{p_1}\left[\frac{\partial^{(q_4+q_5)}\overline{\theta}}{\partial x^{q_4}\partial z^{q_5}}\right]^{p_2}, \tag{22}$$

where $A = \overline{u}, \overline{w}$, or $\overline{\theta}$ (RBC). These libraries are expansive, with integers $0 \leq q \leq 8$ and $0 \leq p \leq 2$, though the total derivative order is limited to 8th (there are a total of 546 and 614 terms in the libraries used for momentum and heat fluxes, respectively). The form of these libraries is motivated by the Galilean-invariant property of the SGS terms, and by past studies. For example, these libraries include Pope's tensors (Pope, 1975), which have been used in physics-based (Anstey & Zanna, 2017; Gatski & Speziale, 1993; Jongen & Gatski, 1998; Lund & Novikov, 1993) and equation-discovery (Li et al., 2021; Reissmann et al., 2021; Ross et al., 2023) approaches in the past (and include the structure of the Smagorinsky model; see below). Our library also includes the basis functions used by Zanna and Bolton (2020).

Note that all calculations for the libraries (and any computation in this work) is performed using the same spectral methods used for DNS and LES.

We have found it useful for interpretability of the outcome and improving the robustness of the algorithm to remove redundant terms using the continuity equation (e.g., using $\partial\overline{v}/\partial y = -\partial\overline{u}/\partial x, \partial^2\overline{v}/\partial y\partial x = -\partial^2\overline{u}/\partial x^2$, etc.). Also, we have found it essential to normalize each basis in $\boldsymbol{\Phi}$ to have a zero mean and a unit variance, because the amplitude of higher-order derivatives can be much larger than that of the lower-order ones.

Like any method, equation discovery using RVM has a number of strengths and weaknesses:

1. It is data efficient (Zanna & Bolton, 2020; Mojgani et al., 2022b). For example, here, we report the results with $n = 100$ FDNS samples, but even with $n = 1$, the results remain pratically the same.

2. It is more robust, in terms of convergence, compared to similar sparsity-promoting techniques (S. Zhang & Lin, 2018; Zanna & Bolton, 2020).
3. A pre-specified library is needed and it is assumed that the true answer (e.g., the SGS stress) can be represented as a linear combination of the chosen basis functions.
4. The pre-specified hyper-parameter α determines how parsimonious the discovered model is. Decreasing α leads to a smaller (likely, more interpretable) model at the expense of increasing the MSE. Here, we follow the model-selection literature (Mangan et al., 2017; Mojgani et al., 2022b) and objectively choose α using the L-curve, as shown later.
5. The answer can depend on the choice of the loss function. The RVM's MSE loss (Eq. (20)) is strictly following the principle of structural modeling, matching the flux between the FDNS and discovered model.

Note that the above strengths (1)-(2) are highly desirable while these weaknesses (3)-(5) are common among many equation-discovery methods, although techniques such as GP and GEP can address (3) and (5), for example using an evolving library. We will further discuss (3)-(5) in Section 4.

3 Results

In this section, we present and discuss the discovered closures, and analyze them *a priori* (offline) and *a posteriori* (online, coupled with LES). We then uncover the connections between the discovered closure and the NGM. For all results presented here, we use $n = 100$ FDNS samples from a training set and 20 FDNS samples from an independent testing set.

3.1 The Discovered Closures for SGS Momentum and Heat Fluxes

For each of the six cases in Tables 1-2, we separately discover closures for three elements of the SGS stress tensor, i.e., $\tau_{xx}, \tau_{xy} = \tau_{yx}$, and τ_{yy} for 2D-FHIT, and $\tau_{xx}, \tau_{xz} = \tau_{zx}$, and τ_{zz} for RBC. Additionally, we discover two elements of the SGS heat flux vector, i.e., J_x and J_z for RBC. We discover individual closures for 4 filter types: Gaussian, box, sharp-spectral, and Gaussian + box. The first three are common filter types, while the last one is motivated by a few recent studies (Zanna & Bolton, 2020; Guillaumin & Zanna, 2021). We also examine several filter sizes, Δ (see Tables 1-2), and the effect of varying α, which as mentioned earlier, is a key hyper-parameter in RVM.

We analyze the *a priori* performance of the discovered closures using the most commonly used metric: the average of CCs for testing samples (Sagaut, 2006; Maulik et al., 2019; Guan et al., 2023). For each element of $\boldsymbol{\tau}$ or $\boldsymbol{J}$, denoted below by τ for convenience, the CC for each testing sample is calculated between 2D patterns of τ from FDNS and τ predicted by the RVM-discovered closure for the corresponding filtered flow variables (e.g., $\overline{u}$, $\overline{v}$ etc.):

$$\mathrm{CC} = \frac{\langle \left(\tau^{\mathrm{RVM}} - \langle \tau^{\mathrm{RVM}} \rangle\right) \left(\tau^{\mathrm{FDNS}} - \langle \tau^{\mathrm{FDNS}} \rangle\right) \rangle}{\sqrt{\langle \left(\tau^{\mathrm{RVM}} - \langle \tau^{\mathrm{RVM}} \rangle\right)^2 \rangle} \sqrt{\langle \left(\tau^{\mathrm{FDNS}} - \langle \tau^{\mathrm{FDNS}} \rangle\right)^2 \rangle}}, \tag{23}$$

where $\langle \cdot \rangle$ is domain averaging. The same equation is also used for computing CC values of 2D patterns of inter-scale energy or enstrophy transfer, P (defined later).

As a representative example of the findings, Fig. 2(a)-(b) shows the averaged CC for τ_{yy} (K1-K3) and J_x (R1-R3) as α is increased. Figure 2(c)-(d) presents the number of terms in the discovered closures. With small α, the discovery is unsuccessful (CC=0; zero term). However, as α is further increased, for all cases, CC abruptly jumps to above $0.8-0.9$ with 1-2 discovered terms, and then gradually converges to 1 but with exponentially growing number of terms in the discovered closure. The CC-α relationship forms an "L-curve". The elbow of this curve indicates the α that balances accuracy and model size, and is extensively used

in the model-selection and equation-discovery literature to objectively choose α (Lawson & Hanson, 1995; Calvetti et al., 2000; Mangan et al., 2017; Goyal & Benner, 2022; Mojgani et al., 2022b). Examining all cases with other filter sizes and filter types reveals the same behavior as shown in Fig. 2, with the exception of the sharp-spectral filter. For this filter, the discovery is unsuccessful, leading to low CC and non-robust results; we will explain the reason of this failure later in this section.

We use the L-curve to determine the *optimal* α. In 2D-FHIT, there are two kinks in the curve around the elbow, corresponding to the discovery of closures with 1 and 2 terms, respectively (Fig. 2(a), (c)). Given the robust and asymptotic behavior in α after the second kink, we use the corresponding α to identify the discovered closure (see the black circles). We find that consistently, across Cases K1-K3, filter types, and filter sizes, this closure is of the form

$$\boldsymbol{\tau} = \begin{bmatrix} \tau_{xx} & \tau_{xy} \\ \tau_{yx} & \tau_{yy} \end{bmatrix} = \Delta^2 \begin{bmatrix} a_{xx}\left(\frac{\partial \overline{u}}{\partial x}\right)^2 + b_{xx}\left(\frac{\partial \overline{u}}{\partial y}\right)^2 & a_{xy}\frac{\partial \overline{u}}{\partial x}\frac{\partial \overline{v}}{\partial x} + b_{xy}\frac{\partial \overline{u}}{\partial y}\frac{\partial \overline{v}}{\partial y} \\ a_{xy}\frac{\partial \overline{u}}{\partial x}\frac{\partial \overline{v}}{\partial x} + b_{xy}\frac{\partial \overline{u}}{\partial y}\frac{\partial \overline{v}}{\partial y} & a_{yy}\left(\frac{\partial \overline{v}}{\partial x}\right)^2 + b_{yy}\left(\frac{\partial \overline{v}}{\partial y}\right)^2 \end{bmatrix}, \tag{24}$$

where a_{xx}, b_{xx}, a_{xy}, b_{xy}, a_{yy}, and b_{yy} are the discovered coefficients (Δ^2 is factored out to further highlight the independence of these coefficients from the filter size). Table 3 shows that these 6 coefficients are the same, and the same for Cases K1-K3, although they can depend on the filter type. This table also shows the average CC values of the discovered closure, which are around 0.99, demonstrating the accurate prediction of each element of the stress tensor and the excellent *a priori* (offline) performance of the discovered closure for a broad range of LES resolutions.

Following the same approach, we discover basically the same closure for $\boldsymbol{\tau}$ in RBC

$$\boldsymbol{\tau} = \begin{bmatrix} \tau_{xx} & \tau_{xz} \\ \tau_{zx} & \tau_{zz} \end{bmatrix} = \Delta^2 \begin{bmatrix} d_{xx}\left(\frac{\partial \overline{u}}{\partial x}\right)^2 & d_{xz}\frac{\partial \overline{u}}{\partial x}\frac{\partial \overline{w}}{\partial x} \\ d_{xz}\frac{\partial \overline{u}}{\partial x}\frac{\partial \overline{w}}{\partial x} & d_{zz}\left(\frac{\partial \overline{w}}{\partial x}\right)^2 \end{bmatrix}, \tag{25}$$

where, as before, d_{xx}, d_{xz}, and d_{zz} are the coefficients with Δ^2 factored out. Note that Eq. (25) is the same as Eq. (24), except that here, there is one term rather than two in each element of the tensor, which is a result of filtering (in RBC) performed only in the horizontal, x, direction. As before, Table 4 shows that these d coefficients are the same, and the same for Cases R1-R3, though varying with filter type. Like before, the discovered closure has fairly high CC values.

Again, following the same approach, we determine the optimal α for discovering the closure of $\boldsymbol{J}$. In Fig. 2(b), Case R1 has a clear elbow while Cases R2-R3 have two kinks around the elbow. Examining all cases and the number of discovered terms (Fig. 2(d)), we find that the single-term closures discovered at the first kink (circled) provide consistent and robust results. This closure is

$$\boldsymbol{J} = \begin{bmatrix} J_x \\ J_z \end{bmatrix} = \Delta^2 \begin{bmatrix} d_x \frac{\partial \overline{u}}{\partial x}\frac{\partial \overline{\theta}}{\partial x} \\ d_z \frac{\partial \overline{w}}{\partial x}\frac{\partial \overline{\theta}}{\partial x} \end{bmatrix}, \tag{26}$$

where d_x and d_z are the discovered coefficients with Δ^2 factored out. Table 5 shows that these d coefficients are the same, and the same for Cases R1-R3, but varying with filter type. As before, the discovered closure has a good *a priori* performance.

To summarize the findings, Eqs. (24)-(26) and Tables 3-5 show that

1. Closures of the same form are robustly discovered for $\boldsymbol{\tau}$ in two vastly different systems, 2D-FHIT and RBC. Even the closure for $\boldsymbol{J}$ overall has the same form, consisting of the products of the first-order derivatives of the variables involved in the nonlinearity of the SGS term.
2. Not just the form, but even the coefficients of the terms in the closures, are consistently the same as parameters such as Re, forcing wavenumber, Ra, or Pr are changed in Cases K1-K3 and R1-R3, leading to different dynamics. The coefficients are independent of the *fluid* and even the *flow* properties.
3. The form of the closures is independent of the filter type unless the sharp-spectral filter is used. The coefficients, once normalized by Δ^2, are independent of filter size, but depend on filter type.
4. The discovered closures have outstanding *a priori* performance, often with CC> 0.95 and even as high as 0.99. It should be noted that the CCs reported in these tables are averaged over a broad range of N_{LES}. The values of CC are higher for larger N_{LES}, i.e., smaller Δ.

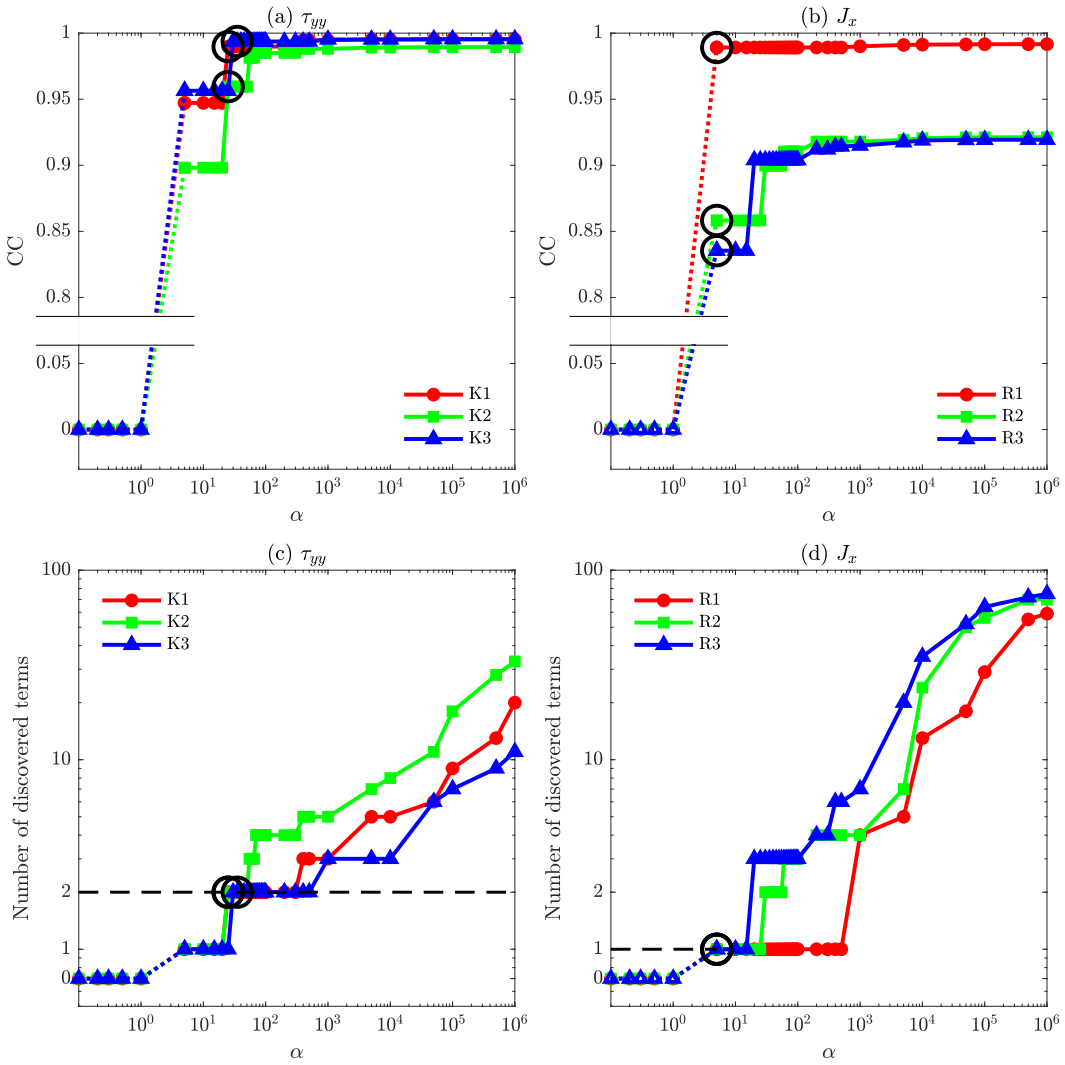

Figure 2. Representative examples of the the effects of increasing the sparsity-level hyper-parameter, α, on the CC and number of terms in the discovered closure. (a), (c): τ_{yy} (2D-FHIT) and (b), (d): J_x (RBC). A Gaussian filter with $N_{\text{LES}} = 128$ (K1-K3) and $N_{\text{LES}} = 256$ (R1-R3) is used, but the same behavior is observed with any other N_{LES} and filter type (except for the sharp-spectral, see the text). In general, for small α (< 1), no closure is discovered (CC=0, zero term). With increasing α, the CC converges to ~ 1 (a more accurate *a priori* closure) but at the expense of a larger closure with many more terms (note the logarithmic scale of the y axes in panels (c)-(d)). However, the CC-α relationship forms an "L-curve", whose elbow indicates the α that balances accuracy and model size (see the text).

Table 3. Coefficients in Eq. (24), the robustly discovered closure for $\boldsymbol{\tau}$ for 2D-FHIT (note that Δ^2 is included in the coefficients). For Cases K1-K3 and different filter types, the mean and standard deviation of the discovered coefficients over different N_{LES} are reported (see Tables 1). The average CC of the closure for each element of $\boldsymbol{\tau}$ is shown in parentheses. The last column shows the analytically derived coefficients for the NGM (see Section 3.2).

Case	Filter	τ_{xx}		τ_{xy}		τ_{yy}		Mean	NGM
		$\left(\frac{\partial \overline{u}}{\partial x}\right)^2$	$\left(\frac{\partial \overline{u}}{\partial y}\right)^2$	$\frac{\partial \overline{u}}{\partial x}\frac{\partial \overline{v}}{\partial x}$	$\frac{\partial \overline{u}}{\partial y}\frac{\partial \overline{v}}{\partial y}$	$\left(\frac{\partial \overline{v}}{\partial x}\right)^2$	$\left(\frac{\partial \overline{v}}{\partial y}\right)^2$		
K1	Gaussian	$\frac{\Delta^2}{11.53 \pm 0.36}$ (0.99)	$\frac{\Delta^2}{11.83 \pm 0.21}$	$\frac{\Delta^2}{11.8 \pm 0.15}$ (0.99)	$\frac{\Delta^2}{11.72 \pm 0.18}$	$\frac{\Delta^2}{11.84 \pm 0.16}$ (0.99)	$\frac{\Delta^2}{11.63 \pm 0.34}$	$\frac{\Delta^2}{11.72 \pm 1.39}$ (0.99)	$\frac{\Delta^2}{12}$
	Box	$\frac{\Delta^2}{11.3 \pm 0.43}$ (0.99))	$\frac{\Delta^2}{11.42 \pm 0.54}$	$\frac{\Delta^2}{11.43 \pm 0.43}$ (0.99)	$\frac{\Delta^2}{11.35 \pm 0.43}$	$\frac{\Delta^2}{11.4 \pm 0.48}$ (1.00)	$\frac{\Delta^2}{11.38 \pm 0.43}$	$\frac{\Delta^2}{11.38 \pm 2.73}$ (0.99)	$\frac{\Delta^2}{12}$
	Gaussian + box	$\frac{\Delta^2}{5.63 \pm 0.28}$ (0.99)	$\frac{\Delta^2}{5.76 \pm 0.26}$	$\frac{\Delta^2}{5.79 \pm 0.17}$ (0.99)	$\frac{\Delta^2}{5.74 \pm 0.18}$	$\frac{\Delta^2}{5.79 \pm 0.19}$ (0.99)	$\frac{\Delta^2}{5.67 \pm 0.28}$	$\frac{\Delta^2}{5.73 \pm 1.35}$ (0.99)	$\frac{\Delta^2}{6}$
K2	Gaussian	$\frac{\Delta^2}{10.87 \pm 0.44}$ (0.99)	$\frac{\Delta^2}{11.33 \pm 0.38}$	$\frac{\Delta^2}{11.58 \pm 0.15}$ (0.98)	$\frac{\Delta^2}{11.58 \pm 0.15}$	$\frac{\Delta^2}{10.33 \pm 0.38}$ (0.99)	$\frac{\Delta^2}{10.88 \pm 0.44}$	$\frac{\Delta^2}{11.26 \pm 1.94}$ (0.99)	$\frac{\Delta^2}{12}$
	Box	$\frac{\Delta^2}{11.09 \pm 0.47}$ (0.99)	$\frac{\Delta^2}{11.11 \pm 0.39}$	$\frac{\Delta^2}{10.71 \pm 0.46}$ (0.99)	$\frac{\Delta^2}{10.72 \pm 0.46}$	$\frac{\Delta^2}{11.11 \pm 0.43}$ (0.99)	$\frac{\Delta^2}{11.10 \pm 0.41}$	$\frac{\Delta^2}{10.91 \pm 2.51}$ (0.99)	$\frac{\Delta^2}{12}$
	Gaussian + box	$\frac{\Delta^2}{5.43 \pm 0.31}$ (0.99)	$\frac{\Delta^2}{5.61 \pm 0.28}$	$\frac{\Delta^2}{5.56 \pm 0.15}$ (0.97)	$\frac{\Delta^2}{5.56 \pm 0.15}$	$\frac{\Delta^2}{5.61 \pm 0.27}$ (0.99)	$\frac{\Delta^2}{5.44 \pm 0.22}$	$\frac{\Delta^2}{5.51 \pm 0.19}$ (0.98)	$\frac{\Delta^2}{6}$
K3	Gaussian	$\frac{\Delta^2}{11.5 \pm 0.37}$ (0.99)	$\frac{\Delta^2}{11.81 \pm 0.22}$	$\frac{\Delta^2}{11.78 \pm 0.14}$ (0.99)	$\frac{\Delta^2}{11.70 \pm 0.18}$	$\frac{\Delta^2}{11.85 \pm 0.14}$ (0.99)	$\frac{\Delta^2}{11.66 \pm 0.30}$	$\frac{\Delta^2}{11.72 \pm 1.35}$ (0.99)	$\frac{\Delta^2}{12}$
	Box	$\frac{\Delta^2}{11.61 \pm 0.43}$ (0.99)	$\frac{\Delta^2}{11.75 \pm 0.53}$	$\frac{\Delta^2}{11.78 \pm 0.41}$ (0.94)	$\frac{\Delta^2}{11.73 \pm 0.43}$	$\frac{\Delta^2}{11.79 \pm 0.46}$ (0.99)	$\frac{\Delta^2}{11.70 \pm 0.41}$	$\frac{\Delta^2}{11.73 \pm 2.66}$ (0.99)	$\frac{\Delta^2}{12}$
	Gaussian + box	$\frac{\Delta^2}{5.61 \pm 0.28}$ (0.99)	$\frac{\Delta^2}{5.75 \pm 0.25}$	$\frac{\Delta^2}{5.78 \pm 0.16}$ (0.99)	$\frac{\Delta^2}{5.73 \pm 0.18}$	$\frac{\Delta^2}{5.79 \pm 0.19}$ (0.99)	$\frac{\Delta^2}{5.70 \pm 0.24}$	$\frac{\Delta^2}{5.73 \pm 1.30}$ (0.99)	$\frac{\Delta^2}{6}$

manuscript submitted to *Journal of Advances in Modeling Earth Systems (JAMES)*

3.2 The Nonlinear Gradient Model (NGM): Taylor-series Expansion of the SGS Term

A closer examination of Eq. (24) reveals that this closure is indeed the NGM (this includes both the form and the coefficients, within the uncertainty range). This connection was already pointed out by Zanna and Bolton (2020), although the implications and findings such as 1-4 mentioned in the previous subsection were not further discussed in their short letter.

First, let's briefly review the NGM (Leonard, 1975; Clark et al., 1979; Sagaut, 2006). As a simple illustration of the idea behind this model, Appendix D presents the derivation of the NGM using a 1D arbitrary field, $a(x)$. Taylor-series expansion of $a(x - r_x)$ around $a(x)$ (Eq. (D2)) simplifies the convolution integral of the filtering operation (Eq. (D1)) such that $\overline{a}(x)$ can be written in terms of $a(x)$ and its derivatives, with coefficients that depend only on the moments of the the filter's kernel, G (Eq. (D4)). Using $\overline{u}^2$ and $\overline{u^2}$ as $a(x)$, we eventually arrive at an analytically derived closure for τ_{xx} with error $\mathcal{O}\left(\Delta^4\right)$ (Eq. (D12)). In 2D with filtering applied in both directions (like our 2D-FHIT), the NGM is (Sagaut, 2006)

$$\boldsymbol{\tau}_{\mathrm{2D}}^{\mathrm{NGM}} = \begin{bmatrix} \tau_{xx} & \tau_{xy} \\ \tau_{yx} & \tau_{yy} \end{bmatrix} = c_\tau \Delta^2 \begin{bmatrix} \left(\dfrac{\partial \overline{u}}{\partial x}\right)^2 + \left(\dfrac{\partial \overline{u}}{\partial y}\right)^2 & \dfrac{\partial \overline{u}}{\partial x}\dfrac{\partial \overline{v}}{\partial x} + \dfrac{\partial \overline{u}}{\partial y}\dfrac{\partial \overline{v}}{\partial y} \\ \dfrac{\partial \overline{u}}{\partial x}\dfrac{\partial \overline{v}}{\partial x} + \dfrac{\partial \overline{u}}{\partial y}\dfrac{\partial \overline{v}}{\partial y} & \left(\dfrac{\partial \overline{v}}{\partial x}\right)^2 + \left(\dfrac{\partial \overline{v}}{\partial y}\right)^2 \end{bmatrix} + \mathcal{O}\left(\Delta^4\right), \tag{27}$$

where c_τ depends on the filter's kernel. Similarly, for the 2D RBC with filtering only in the x direction, the NGM is

$$\boldsymbol{\tau}_{\mathrm{1D}}^{\mathrm{NGM}} = \begin{bmatrix} \tau_{xx} & \tau_{xz} \\ \tau_{zx} & \tau_{zz} \end{bmatrix} = d_\tau \Delta^2 \begin{bmatrix} \left(\dfrac{\partial \overline{u}}{\partial x}\right)^2 & \dfrac{\partial \overline{u}}{\partial x}\dfrac{\partial \overline{w}}{\partial x} \\ \dfrac{\partial \overline{u}}{\partial x}\dfrac{\partial \overline{w}}{\partial x} & \left(\dfrac{\partial \overline{w}}{\partial x}\right)^2 \end{bmatrix} + \mathcal{O}\left(\Delta^4\right). \tag{28}$$

As emphasized in Appendix D, there is nothing specific to momentum flux or even turbulence (or even physical systems) in the derivation of NGM. In fact, for the *filtered quadratic nonlinearity of any two arbitrary variables*, one arrives at the same expression with $\mathcal{O}\left(\Delta^4\right)$ accuracy. For example, following this derivation, for the SGS heat flux, we obtain

$$\boldsymbol{J}^{\mathrm{NGM}} = \begin{bmatrix} J_x \\ J_z \end{bmatrix} = d_J \Delta^2 \begin{bmatrix} \dfrac{\partial \overline{u}}{\partial x}\dfrac{\partial \overline{\theta}}{\partial x} \\ \dfrac{\partial \overline{w}}{\partial x}\dfrac{\partial \overline{\theta}}{\partial x} \end{bmatrix} + \mathcal{O}\left(\Delta^4\right), \tag{29}$$

where like c_τ and d_τ, d_J only depends on the filter's kernel.

Computing c_τ, d_τ, d_J for each of the filter types used in this study, we confirm that the discovered closures for the SGS stress are basically the NGM (Eqs. (27)-(28)), and in the case of the SGS heat flux, an NGM-like (Eq. (29)) closure (see Tables 3–5).

Based on the above analyses, we can now explain the findings (1)-(4) in Section 3.1. Closures of the same structure are robustly discovered for both SGS momentum and heat fluxes in two vastly different turbulent flows (and independent of parameters such as Re, Ra, Pr, and forcing) because the first term in the Taylor-series expansion dominates the SGS flux. As a result, in equation-discovery using common loss functions such as MSE and evaluation metrics such as CC, which aim at closely matching τ or J, NGM or NGM-like closures are discovered (if the library is expansive enough to include all the relevant terms). We emphasize that this would be the case with discovering the representation of the filtered nonlinearity of any two arbitrary variables. As already observed, the coefficients of the discovered closure become even closer to those of NGM as Δ is decreased (thus reducing potential contributions from the truncated $\mathcal{O}\left(\Delta^4\right)$ terms).

Table 4. Coefficients in Eq. (25), the robustly discovered closure for $\boldsymbol{\tau}$ for RBC (note that Δ^2 is included in the coefficients). For Cases R1-R3 and different filter types, the mean and standard deviation of the discovered coefficients over different N_{LES} are reported (see Table 2). The average CC of the closure for each element of $\boldsymbol{\tau}$ is shown in parentheses. The last column shows the analytically derived coefficients for the NGM (see Section 3.2).

Case	Filter	τ_{xx}	τ_{xy}	τ_{yy}	Mean	NGM
		$\left(\frac{\partial \overline{u}}{\partial x}\right)^2$	$\frac{\partial \overline{u}}{\partial x}\frac{\partial \overline{w}}{\partial x}$	$\left(\frac{\partial \overline{w}}{\partial x}\right)^2$		
R1	Gaussian	$\frac{\Delta^2}{10.89 \pm 0.39}$ (0.98)	$\frac{\Delta^2}{11.01 \pm 0.41}$ (0.97)	$\frac{\Delta^2}{10.55 \pm 0.75}$ (0.92)	$\frac{\Delta^2}{10.98 \pm 0.49}$ (0.95)	$\frac{\Delta^2}{12}$
	Box	$\frac{\Delta^2}{10.45 \pm 0.94}$ (0.98)	$\frac{\Delta^2}{10.21 \pm 0.93}$ (0.94)	$\frac{\Delta^2}{10.32 \pm 0.87}$ (0.91)	$\frac{\Delta^2}{10.35 \pm 0.97}$ (0.93)	$\frac{\Delta^2}{12}$
	Gaussian + box	$\frac{\Delta^2}{5.35 \pm 0.56}$ (0.93)	$\frac{\Delta^2}{5.37 \pm 0.35}$ (0.94)	$\frac{\Delta^2}{5.21 \pm 0.48}$ (0.89)	$\frac{\Delta^2}{5.29 \pm 0.66}$ (0.91)	$\frac{\Delta^2}{6}$
R2	Gaussian	$\frac{\Delta^2}{11.35 \pm 0.41}$ (0.98)	$\frac{\Delta^2}{11.82 \pm 0.36}$ (0.88)	$\frac{\Delta^2}{9.7 \pm 0.54}$ (0.81)	$\frac{\Delta^2}{10.62 \pm 0.79}$ (0.89)	$\frac{\Delta^2}{12}$
	Box	$\frac{\Delta^2}{10.52 \pm 0.65}$ (0.97)	$\frac{\Delta^2}{9.38 \pm 0.5}$ (0.90)	$\frac{\Delta^2}{9.11 \pm 0.59}$ (0.86)	$\frac{\Delta^2}{10.01 \pm 0.44}$ (0.91)	$\frac{\Delta^2}{12}$
	Gaussian + box	$\frac{\Delta^2}{5.48 \pm 0.24}$ (0.98)	$\frac{\Delta^2}{5.33 \pm 0.12}$ (0.92)	$\frac{\Delta^2}{5.00 \pm 0.23}$ (0.93)	$\frac{\Delta^2}{5.27 \pm 0.28}$ (0.94)	$\frac{\Delta^2}{6}$
R3	Gaussian	$\frac{\Delta^2}{11.22 \pm 0.16}$ (0.94)	$\frac{\Delta^2}{11.34 \pm 0.41}$ (0.93)	$\frac{\Delta^2}{10.51 \pm 1.03}$ (0.91)	$\frac{\Delta^2}{11.02 \pm 0.79}$ (0.93)	$\frac{\Delta^2}{12}$
	Box	$\frac{\Delta^2}{10.17 \pm 0.32}$ (0.93)	$\frac{\Delta^2}{9.94 \pm 0.64}$ (0.93)	$\frac{\Delta^2}{9.44 \pm 1.32}$ (0.92)	$\frac{\Delta^2}{9.85 \pm 0.95}$ (0.92)	$\frac{\Delta^2}{12}$
	Gaussian + box	$\frac{\Delta^2}{5.46 \pm 0.10}$ (0.93)	$\frac{\Delta^2}{5.55 \pm 0.12}$ (0.90)	$\frac{\Delta^2}{4.87 \pm 0.66}$ (0.88)	$\frac{\Delta^2}{5.3 \pm 0.54}$ (0.90)	$\frac{\Delta^2}{6}$

Table 5. Coefficients in Eq. (26), the robustly discovered closure for $\boldsymbol{J}$ for RBC (note that Δ^2 is included in the coefficients). For Cases R1-R3 and different filter types, the mean and standard deviation of the discovered coefficients over different N_{LES} are reported (see Table 2). The average CC of the closure for each element of $\boldsymbol{J}$ is shown in parentheses. The last column shows the analytically derived coefficients for the NGM (see Section 3.2)

Case	Filter	J_x	J_z	Mean	NGM
		$\overline{\frac{\partial \overline{u}}{\partial x}\frac{\partial \overline{\theta}}{\partial x}}$	$\overline{\frac{\partial \overline{w}}{\partial x}\frac{\partial \overline{\theta}}{\partial x}}$		
R1	Gaussian	$\frac{\Delta^2}{10.54 \pm 0.66}$ (0.93)	$\frac{\Delta^2}{10.3 \pm 0.87}$ (0.90)	$\frac{\Delta^2}{10.88 \pm 1.3}$ (0.92)	$\frac{\Delta^2}{12}$
	Box	$\frac{\Delta^2}{9.11 \pm 0.86}$ (0.93)	$\frac{\Delta^2}{9.00 \pm 0.65}$ (0.92)	$\frac{\Delta^2}{9.05 \pm 0.80}$ (0.93)	$\frac{\Delta^2}{12}$
	Gaussian + box	$\frac{\Delta^2}{5.32 \pm 0.3}$ (0.96)	$\frac{\Delta^2}{5.31 \pm 0.5}$ (0.90)	$\frac{\Delta^2}{5.31 \pm 0.45}$ (0.93)	$\frac{\Delta^2}{6}$
R2	Gaussian	$\frac{\Delta^2}{11.27 \pm 0.2}$ (0.89)	$\frac{\Delta^2}{10.9 \pm 0.4}$ (0.85)	$\frac{\Delta^2}{11.12 \pm 0.37}$ (0.87)	$\frac{\Delta^2}{12}$
	Box	$\frac{\Delta^2}{9.7 \pm 0.11}$ (0.90)	$\frac{\Delta^2}{9.3 \pm 0.23}$ (0.84)	$\frac{\Delta^2}{9.5 \pm 0.67}$ (0.87)	$\frac{\Delta^2}{12}$
	Gaussian + box	$\frac{\Delta^2}{5.55 \pm 0.08}$ (0.91)	$\frac{\Delta^2}{5.1 \pm 0.22}$ (0.85)	$\frac{\Delta^2}{5.32 \pm 0.78}$ (0.88)	$\frac{\Delta^2}{6}$
R3	Gaussian	$\frac{\Delta^2}{9.75 \pm 0.47}$ (0.84)	$\frac{\Delta^2}{9.21 \pm 0.34}$ (0.83)	$\frac{\Delta^2}{9.46 \pm 0.97}$ (0.83)	$\frac{\Delta^2}{12}$
	Box	$\frac{\Delta^2}{9.87 \pm 0.23}$ (0.80)	$\frac{\Delta^2}{9.5 \pm 0.22}$ (0.81)	$\frac{\Delta^2}{9.68 \pm 0.57}$ (0.81)	$\frac{\Delta^2}{12}$
	Gaussian + box	$\frac{\Delta^2}{4.78 \pm 0.12}$ (0.83)	$\frac{\Delta^2}{4.52 \pm 0.34}$ (0.80)	$\frac{\Delta^2}{4.65 \pm 0.77}$ (0.81)	$\frac{\Delta^2}{6}$

The connection to the analytical derivation also explains why the coefficients in the discovered models are *independent* of the fluid or even the flow properties (Ra, Re, Pr) and only depend on the filter size (Δ) and filter type. For the Gaussian and box filters we obtain $c_\tau = d_\tau = d_J = 1/12$: this is because the parameters of the Gaussian filter are chosen such that Gaussian and box filters' kernels have the same second moment (Pope, 2000). For Gaussian + box filter, the coefficients are 1/6 because the kernel of this filter is convolution of the Gaussian and box filter kernels. For the sharp-spectral filter, the moments are indefinite, this is why there is no NGM discovery with this filter (and we will discuss later why the discovery fails altogether). Note that coarse-graining done here via cutoff in the spectral space does not change c_τ, d_τ and d_τ; however, if coarse-graining is done by other techniques such as box averaging, then the coefficients might change (note that the NGM coefficient for Gaussian + box filter is half of the coefficient of either filter; see Tables 3-5).

In short, one can explain the effects of different filter kernels and coarse-graining strategies on the discovered closures following the analytically derivable NGM (see Appendix D and Sagaut (2006)).

An important implication of the above findings and discussions is that the discovered closure may not be unique and can depend on the filtering and coarse-graining procedure: it depends on the filter type (and up to a factor, on the filter size). This is not a problem of equation-discovery; in fact, the SGS fluxes diagnosed from FDNS are not unique and depend on the filtering and coarse-graining procedure (this is further shown in Fig. 4 and discussed at the end of this section). This has implications not just for equation-discovery, but more broadly, for the ongoing efforts on learning SGS parameterizations for various processes from high-fidelity data using ML. See Sun et al. (2023) for extensive discussions about this issue focused on the data-driven SGS modeling of atmospheric gravity waves.

The next key question is about the accuracy and stability of LES of the 2D-FHIT and RBC with the NGM closures, $\boldsymbol{\tau}^{\mathrm{NGM}}$ and $\boldsymbol{J}^{\mathrm{NGM}}$. However, before discussing the *a posteriori* (online) performance of NGM closures, we address one more issue, and that is about any potential influence from numerical calculations in our equation discovery.

3.2.1 Effects of numerical discretization

The appearance of gradients of velocity (or temperature) in Eqs. (27)-(29) might suggest to some that the discovered equations represent the truncated terms of finite difference/volume discretization schemes (the methods used in Zanna and Bolton (2020)). The discussions in their paper and the comprehensive analyses here should leave no ambiguity that Eqs. (27)-(29) represent the physics of the SGS fluxes, rather than any numerical error. Still, we wish to discuss a few more points here, as numerical errors from finite difference/volume discretizations or from aliasing (in spectral calculations) can certainly contaminate equation discovery.

All numerical calculations in this study are performed using Fourier and Chebyshev spectral methods. Moreover, we have repeated our calculations of the SGS fluxes and of the basis functions in the library after de-aliasing based on the 2/3 rule (Orszag, 1971). Furthermore, we have performed discovery on fluxes that are only filtered but not coarse-grained (thus they remain on the high-resolution DNS grid). The outcomes of all these experiments are Eqs. (27)-(29), demonstrating that the discovered closures do not contain any contributions from numerical errors.

3.3 *A posteriori* (Online) Tests and Inter-scale Energy/Enstrophy Transfer

For all 6 cases and all tested N_{LES}, the LES runs with NGM closures are unstable: high-wave number features appear and the simulations eventually blow up (not shown). This is consistent with the findings of Zanna and Bolton (2020), who only found stable LES once the SGS momentum fluxes predicted by the discovered closure were attenuated. More

generally, this is also consistent with extensive studies in the 1990s (though mainly focused on 3D turbulence), which found that LES with the NGM closure is unstable (Liu et al., 1994; Leonard, 1997; Vreman et al., 1997; Borue & Orszag, 1998; Meneveau & Katz, 2000; Pope, 2000; S. Chen et al., 2003, 2006). The exact reason(s) for the instabilities remain unclear but these studies found that in general, in 3D turbulence, NGM has insufficient dissipation and/or too much backscattering; see, e.g., the discussions in Leonard (1997, 2016) and Sagaut (2006). As a result, later studies focused more on eddy-viscosity closures, or on NGM with backscattering removed or combined with eddy-viscosity, the so-called mixed models (Winckelmans et al., 1998; Cottet, 1996; Balarac et al., 2013). Such versions of NGM have been used in some geophysical flows, e.g., for atmospheric boundary layer (Lu & Porté-Agel, 2010, 2014; Khani & Porté-Agel, 2017; Khani & Waite, 2020; Khani & Porté-Agel, 2022) and oceanography (Khani & Dawson, 2023).

In 2D turbulence with filtering done in both directions, such as our 2D-FHIT cases, the NGM has a clear major shortcoming: it cannot capture any energy transfer between the subgrid and resolved scales, despite capturing the enstrophy transfer well (S. Chen et al., 2003, 2006; Nadiga, 2008). To further explore this issue, first note that the rate of kinetic energy transfer between the resolved and subgrid scales, P_τ, is (Pope, 2000)

$$P_\tau = -\tau^r_{ij}\overline{S}_{ij}, \tag{30}$$

where summation over repeated indices is implied. $\overline{S}$ and τ^r are the 2D filtered rate of strain tensor and the anisotropic part of the SGS stress tensor (see Appendix E for details). In 2D turbulence with filtering done in both directions, using $\boldsymbol{\tau}^{\mathrm{NGM}}$ in the above equation shows that $P_\tau^{\mathrm{NGM}}(x, y, t)$ is identically zero (see Appendix E). This is demonstrated numerically in Table 6, which also shows that NGM captures both forward transfer and backscatter of SGS *enstrophy* fairly well (CC>0.95). Therefore, despite the high CC of $\boldsymbol{\tau}^{\mathrm{NGM}}$ with $\boldsymbol{\tau}^{\mathrm{FDNS}}$, and even a fairly accurate inter-scale enstrophy transfer, NGM cannot capture any inter-scale *energy* transfer, indicating a major failure from a functional modeling perspective (note that in this context, "inter-scale" means between the *resolved* and *subgrid* scales). A physical/mathematical interpretation of this failure is that while NGM reproduces the structure of $\boldsymbol{\tau}$ remarkably well, it does not at all capture the correlations between the $\boldsymbol{\tau}$ and $\overline{\mathbf{S}}$ tensors, e.g., the angles between their principle directions (Leonard, 2016).

This inability to represent any inter-scale energy transfer is likely the reason for the instabilities of LES with NGM closure in Cases K1-K3 (and generally, in 2D turbulence). But how about in RBC? In Cases R1-R3, filtering is conducted only in the horizontal direction, and as a result, P_τ^{NGM} is not identically zero. In fact, in these cases, the forward transfer and backscatter of both kinetic energy and enstrophy are captured fairly well by NGM, with CC often above 0.95 (Table 7). However, a deeper examination shows that the backscatter (anti-diffusion) of inter-scale SGS potential energy, measured as P_J (see Appendix E), may not be captured well, specially at low N_{LES} (Table 7). Poor representation of backscattering can certainly lead to instabilities, as for example shown by Guan et al. (2022a) for 2D turbulence.

To further explore other potential shortcomings of NGM, we have also examined the spectra of elements of $\boldsymbol{\tau}^{\mathrm{NGM}}$ and $\boldsymbol{J}^{\mathrm{NGM}}$ in comparison to those from FDNS (Figure 3). This analysis shows that the spectra of SGS momentum and heat fluxes are captured well across wavenumbers, even at high wavenumbers, indicating that NGM performs well in this *a priori* (offline) metric.

To summarize the above analyses: we find all LES with NGM closures for 2D-FHIT and RBC cases to become unstable even at high LES resolutions. Understanding the reason(s) for this poor *a posteriori* (online) performance is essential to make further progress. Examining a few functional and structural metrics beyond CC of SGS fluxes (e.g., inter-scale energy/enstrophy transfers, spectra) point to only one major shortcoming that is relevant to 2D-FHIT (and any 2D turbulent flow): NGM cannot capture any inter-scale kinetic energy transfer, which is likely the reason for the instabilities. This is not an issue in RBC, for

Table 6. The average correlation coefficient (CC) between inter-scale energy transfer (P_τ) or enstrophy transfer (P_Z) patterns of the SGS momentum stresses from FDNS and from NGM closure for Cases K1-K3 and different N_{LES}. The CC of P_τ for both forward transfer (> 0) and backscatter (< 0) of SGS energy is "undefined" since $P_\tau^{NGM} = 0$ everywhere for 2D-FHIT (in general, $P_\tau^{FDNS} \neq 0$). On the contrary, the forward transfer and backscatter of SGS enstrophy are captured well by the NGM. The Gaussian filter is used in FDNS.

Cases	$N_{\text{LES}} = 32$	$N_{\text{LES}} = 64$	$N_{\text{LES}} = 128$	$N_{\text{LES}} = 256$
CC for P_τ ($P_\tau > 0, P_\tau < 0$)				
K1 K2 K3	undefined (undefined, undefined)			
CC for P_Z ($P_Z > 0, P_Z < 0$)				
K1	0.98 (0.98, 0.97)	0.98 (0.98, 0.97)	0.98 (0.98, 0.97)	0.98 (0.98, 0.96)
K2	-	-	0.98 (0.98, 0.97)	0.99 (0.99, 0.98)
K3	0.98 (0.98, 0.97)	0.98 (0.98, 0.96)	0.97 (0.97, 0.95)	0.96 (0.97, 0.93)

which we only identify one shortcoming, and that is the poor representation of backscatter (anti-diffusion) of potential energy, specially at low LES resolution. These findings indicate that the poor *a posteriori* (online) performance of NGM might have different causes in different flows and requires more extensive investigations.

Before discussing ideas for addressing these challenges in future work, we present more analyses in two areas: a closer examination of the physics included in the library (Subsection 3.4) and the decomposition of the SGS fluxes and the sensitivity of the diagnosed fluxes to the filter type/size (Subsection 3.5).

3.4 A Physics-guided Library: Pope Tensors

In Section 3.1, we consider an expansive library of basis functions combining the low- and high-order derivatives and polynomials of velocity and temperature. Under certain assumptions, smaller but physics-informed libraries can be devised. For example, Boussinesq (1877) hypothesized that for a nearly homogeneous, incompressible, high-Re flow, the anisotropic SGS stress $\boldsymbol{\tau}^{\text{r}}$ (Eq. (E2)) is only a function of the filtered rates of strain $\overline{\mathbf{S}}$ (Eq. (E1)) and rotation $\overline{\boldsymbol{\Omega}}$ (Eq. (32)) tensors:

$$\boldsymbol{\tau}^{\text{r}} = \boldsymbol{\tau}^{\text{r}}\left(\overline{\mathbf{S}}, \overline{\boldsymbol{\Omega}}\right), \tag{31}$$

$$\overline{\boldsymbol{\Omega}} = \frac{1}{2}\begin{bmatrix} 0 & \dfrac{\partial \overline{u}}{\partial y} - \dfrac{\partial \overline{v}}{\partial x} \\ \dfrac{\partial \overline{v}}{\partial x} - \dfrac{\partial \overline{u}}{\partial y} & 0 \end{bmatrix}. \tag{32}$$

Note that in Eqs. (7) and (13) and in general, only $\nabla \cdot \boldsymbol{\tau}^{\text{r}}$ has to be parameterized as the rest of $\nabla \cdot \boldsymbol{\tau}$ can be absorbed into $\nabla \overline{p}$ (Sagaut, 2006). Owing to Cayley-Hamilton theorem (Gantmakher, 2000), $\boldsymbol{\tau}^{\text{r}}$ can be represented as a linear combination of a finite number of tensors, the so-called Pope tensors (Pope, 1975). In 2D, there are only 3 Pope tensors $\mathbf{Z}$, thus

$$\boldsymbol{\tau}^{\text{r}} = \sum_{n=0}^{2} \zeta^{(n)}\left(I_1, I_2\right) \mathbf{Z}^{(n)}. \tag{33}$$

Table 7. The average correlation coefficient (CC) between inter-scale kinetic energy transfer (P_τ) or enstrophy transfer (P_Z) or potential energy transfer (P_J) patterns of the SGS fluxes from FDNS and from NGM closure for Cases R1-R3 and different $N_{\rm LES}$. Note that for RBC, filtering is conducted in only one direction (x), therefore, P_τ is not identically zero. Here, the forward transfer and backscatter of SGS kinetic energy and enstrophy are overall captured well, specially as $N_{\rm LES}$ increases. However, the backscatter of SGS potential energy is not well captured, specially at low LES resolutions. The Gaussian filter is used in FDNS. See Appendix E for the definition of P_J.

Cases	$N_{\rm LES}=128$	$N_{\rm LES}=256$	$N_{\rm LES}=512$
CC for P_τ ($P_\tau>0, P_\tau<0$)			
R1	0.94 (0.96, 0.85)	0.99 (0.99, 0.98)	–
R2	0.97 (0.81, 0.98)	0.98 (0.91, 0.98)	0.99 (0.97, 1.00)
R3	0.79 (0.81, 0.74)	0.88 (0.92, 0.81)	0.96 (0.97, 0.93)
CC for P_Z ($P_Z>0, P_Z<0$)			
R1	1.00 (1.00, 0.99)	1.00 (1.00, 1.00)	–
R2	0.99 (0.96, 1.00)	0.99 (0.98, 1.00)	1.00 (1.00, 1.00)
R3	0.96 (0.95, 0.96)	0.99 (0.99, 0.98)	1.00 (1.00, 0.99)
CC for P_J ($P_J>0, P_J<0$)			
R1	0.89 (0.89, 0.15)	0.97(0.97, 0.46)	–
R2	0.76 (0.75, 0.65)	0.91 (0.91, 0.63)	0.98 (0.98, 0.76)
R3	0.77 (0.75, 0.40)	0.87 (0.86, 0.39)	0.94 (0.94, 0.44)

The three Pope's tensors are $\mathbf{Z}^{(0)}=\mathbf{I}$, $\mathbf{Z}^{(1)}=\overline{\mathbf{S}}$, and

$$\mathbf{Z}^{(2)}=\overline{\mathbf{S}}\,\overline{\mathbf{\Omega}}-\overline{\mathbf{\Omega}}\,\overline{\mathbf{S}}=-\frac{1}{2}\begin{bmatrix}\left(\frac{\partial\overline{u}}{\partial y}\right)^2-\left(\frac{\partial\overline{v}}{\partial x}\right)^2 & 2\left(\frac{\partial\overline{u}}{\partial x}\frac{\partial\overline{v}}{\partial x}+\frac{\partial\overline{u}}{\partial y}\frac{\partial\overline{v}}{\partial y}\right)\\ 2\left(\frac{\partial\overline{u}}{\partial x}\frac{\partial\overline{v}}{\partial x}+\frac{\partial\overline{u}}{\partial y}\frac{\partial\overline{v}}{\partial y}\right) & -\left(\frac{\partial\overline{u}}{\partial y}\right)^2+\left(\frac{\partial\overline{v}}{\partial x}\right)^2\end{bmatrix}, \tag{34}$$

which is related to the anisotropic part of the NGM stress. In fact, $\boldsymbol{\tau}^{\rm NGM-r}=-\Delta^2\mathbf{Z}^{(2)}/12$ (see Eq. (E3)). Note that this is also the physics-based closure derived in Anstey and Zanna (2017). Coefficients $\zeta^{(n)}$ are functions of invariants $I_1=\mathrm{tr}(\overline{\mathbf{S}}^2)$ and $I_2=\mathrm{tr}(\overline{\mathbf{\Omega}}^2)$. The standard Smagorinsky model is $\zeta^{(1)}(I_1)\mathbf{Z}^1$.

Our expansive library, described in Eqs. (21)-(22), includes the individual terms to discover $\mathbf{Z}^{(n)}$ ($n=1,2,3$); however, we have always found the NGM stress, $\boldsymbol{\tau}^{\rm NGM}$. To see whether the results will change with a discovery only done on the anisotropic SGS stress, $\boldsymbol{\tau}^{\rm r}$, and with a smaller library that only has the terms relevant to the Pope tensors, we have conducted more experiments with 3 libraries for Cases K1-K3. The first library only includes the 3 Pope tensors $\{\mathbf{Z}^{(1)},\mathbf{Z}^{(2)},\mathbf{Z}^{(3)}\}$, the second library only includes the 6 non-zero elements of these tensors, and the third library only includes the 8 terms that compromise these 6:

$$\left\{1,\frac{\partial\overline{u}}{\partial x},\frac{\partial\overline{u}}{\partial y},\frac{\partial\overline{v}}{\partial x},\left(\frac{\partial\overline{u}}{\partial y}\right)^2,\left(\frac{\partial\overline{v}}{\partial x}\right)^2,\frac{\partial\overline{u}}{\partial x}\frac{\partial\overline{v}}{\partial x},\frac{\partial\overline{u}}{\partial y}\frac{\partial\overline{v}}{\partial y}\right\}. \tag{35}$$

The RVM with any of these libraries robustly discovers $\boldsymbol{\tau}^{\rm NGM-r}=-\Delta^2\mathbf{Z}^{(2)}/12$, without $\mathbf{Z}^{(1)}$ (or $\mathbf{Z}^{(0)}$) showing up (thus, no Smagorinsky/eddy viscosity-like term). Needless to say, LES with this closure in unstable.

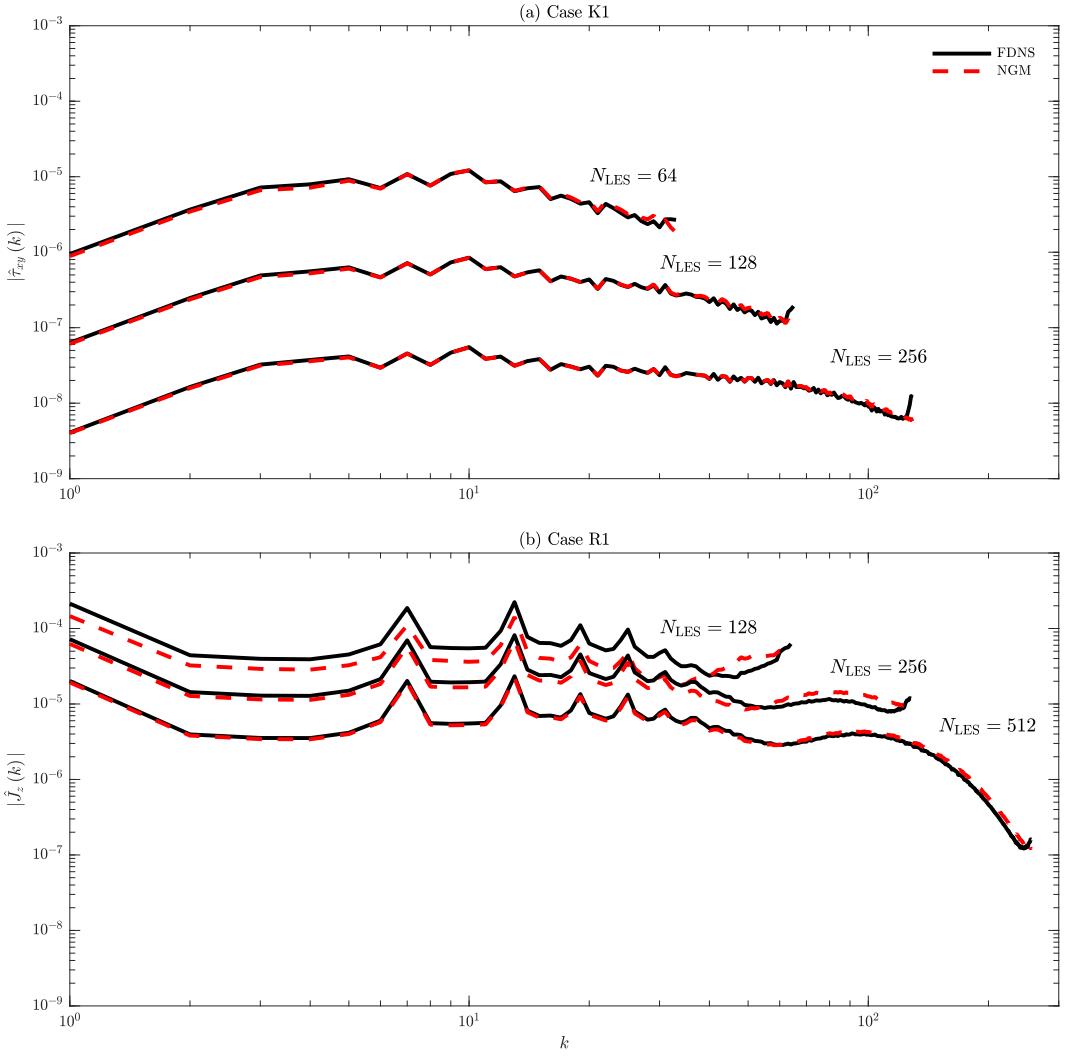

Figure 3. Examples of the spectra of SGS fluxes predicted using NGM compared to those diagnosed using FDNS (the truth). (a) τ_{xy} from Case K1 and (b) J_z from Case R1 for 3 different N_{LES}. A Gaussian filter is used for FDNS, but the same behavior is observed for box and Gaussian+box filters. Here, $|\hat{\cdot}|$ is the modulus of Fourier coefficients.

The above analyses show the prevalence of NGM: it emerges whether the full or just the anisotropic part of the SGS stress tensor is discovered, and whether an expansive or a small physics-guided library is used.

3.5 Decomposition of SGS Fluxes: Leonard, Cross, and Reynolds Stresses

As discussed earlier, whether a closure could be successfully discovered from FDNS data or if the NGM could be derived depend on the choice of the filter. The latter was explained based on the dependence of the derived closure on the moments' of the filter kernel. Furthermore, the coefficients of the discovered closure and the analytically derived coefficients of the NGM depend on the choice of the filters (Tables 3-5). Here, we further demonstrate the sensitivity of the diagnosed FDNS SGS flux (which is treated as truth in offline/supervised learning data-driven modeling approaches) to the choice of the filter, and then decompose the flux into its three components to gain further insight.

The top row of Fig. 4 show examples of SGS τ in 2D-FHIT diagnosed from the FDNS data using different filter types. It is clear that the diagnosed fluxes are not unique and particularly different between Gaussian/box filters and sharp-spectral filter (similar differences can be seen in SGS momentum and heat fluxes in RBC). This sensitivity, which has important implications for data-driven SGS modeling efforts (Sun et al., 2023), has been known for a long time in the LES community (Leonard, 1975; Sagaut, 2006). The Gaussian and box filters extract fairly similar features, even of the almost the same amplitude (which is due to their matched kernels' second moments). The Gaussian+box filter captures similar features but with a factor of ~ 2 difference in amplitude (related to the factor of 2 difference in NGM coefficients). However, the sharp-spectral filter captures very different features that have much smaller length scales and amplitudes. This is the reason that RVM fails to discover any closure from these FDNS data. We also point out that in Guan et al. (2022a, 2023), deep convolutional neural networks (CNNs) could not be successfully trained on FDNS data obtained using sharp-spectral *cutoff* filters, while high CC and stable/accurate LES runs in different systems were achieved using CNNs trained on FDNS data obtained through the Gaussian filter. Note that Ross et al. (2023) successfully trained CNNs (and performed equation-discovery) using a "smoothed" sharp-spectral filter that had exponential decay at high wavenumbers (rather than a cutoff). These findings further show the importance of how the "true" SGS fluxes are diagnosed for offline/supervised learning.

To see the reason for this difference, we decompose the SGS tensor using $\boldsymbol{u} = \overline{\boldsymbol{u}} + \boldsymbol{u}'$. Leonard (1975) introduced a decomposition of the SGS tensor into three components. However, since two of these components were not Galilean-invariant (Speziale, 1985), a Galilean-invariant decomposition was later proposed by Germano (1986):

$$\boldsymbol{\tau} = \mathbf{L} + \mathbf{C} + \mathbf{R}. \tag{36}$$

Here, $\mathbf{L}$, $\mathbf{C}$, and $\mathbf{R}$ are the Leonard, cross, and Reynolds stresses, which in 2D-FHIT are

$$\mathbf{L} = \begin{bmatrix} L_{xx} & L_{xy} \\ L_{yx} & L_{yy} \end{bmatrix} = \begin{bmatrix} \overline{\overline{u}^{\,2}} - \overline{\overline{u}}^{\,2} & \overline{\overline{u}\,\overline{v}} - \overline{\overline{u}}\,\overline{\overline{v}} \\ \overline{\overline{u}\,\overline{v}} - \overline{\overline{u}}\,\overline{\overline{v}} & \overline{\overline{v}^{\,2}} - \overline{\overline{v}}^{\,2} \end{bmatrix}, \tag{37}$$

$$\mathbf{C} = \begin{bmatrix} C_{xx} & C_{xy} \\ C_{yx} & C_{yy} \end{bmatrix} = \begin{bmatrix} 2\left(\overline{\overline{u}u'} - \overline{\overline{u}}\,\overline{u'}\right) & \overline{\overline{u}v'} + \overline{u'}\,\overline{v} - \overline{\overline{u}}\,\overline{v'} - \overline{u'}\,\overline{\overline{v}} \\ \overline{\overline{u}v'} + \overline{u'}\,\overline{v} - \overline{\overline{u}}\,\overline{v'} - \overline{u'}\,\overline{\overline{v}} & 2\left(\overline{\overline{v}v'} - \overline{\overline{v}}\,\overline{v'}\right) \end{bmatrix}, \tag{38}$$

$$\mathbf{R} = \begin{bmatrix} R_{xx} & R_{xy} \\ R_{yx} & R_{yy} \end{bmatrix} = \begin{bmatrix} \overline{u'^{\,2}} - \overline{u'}^{\,2} & \overline{u'v'} - \overline{u'}\,\overline{v'} \\ \overline{u'v'} - \overline{u'}\,\overline{v'} & \overline{v'^{\,2}} - \overline{v'}^{\,2} \end{bmatrix}. \tag{39}$$

$\boldsymbol{\tau}$ and $\boldsymbol{J}$ of RBC can be decomposed in the same fashion. The most familiar component, the Reynolds stress, represents interactions in the unresolved scales that project onto the resolved scale. The cross stress represents the direct interactions between the unresolved and resolved scales that project onto the resolved scale. The Leonard stress includes interactions between the resolved scales not captured by the low-resolution LES grid. See Leonard (1975) and McDonough (2007) for more discussions.

The relative importance of these three components in $\boldsymbol{\tau}$ and $\boldsymbol{J}$ depends on the filter type and size (and even the flow characteristics). Rows 2-4 of Fig. 4 show examples of the Leonard, cross, and Reynolds stress components of τ_{xy}. For Gaussian, box, and Gaussian+box filters, the Leonard stress dominates, followed by cross and then Reynolds stress. However, for sharp-spectral, only the Reynolds stress has coherent structures that more or less resemble the Reynolds stress from Gaussian/box filters. The strong dependence on filter type comes from the fact that for Gaussian and box filters, $\overline{\boldsymbol{u}'} \neq 0$ and $\overline{\overline{\boldsymbol{u}}} \neq \overline{\boldsymbol{u}}$, leading to non-zero Leonard and cross stresses. However, for the sharp-spetcral filters, $\overline{\boldsymbol{u}'} = 0$ and $\overline{\overline{\boldsymbol{u}}} = \overline{\boldsymbol{u}}$, resulting in Reynolds stress as the only non-zero component, except for potential contributions from numerical errors.

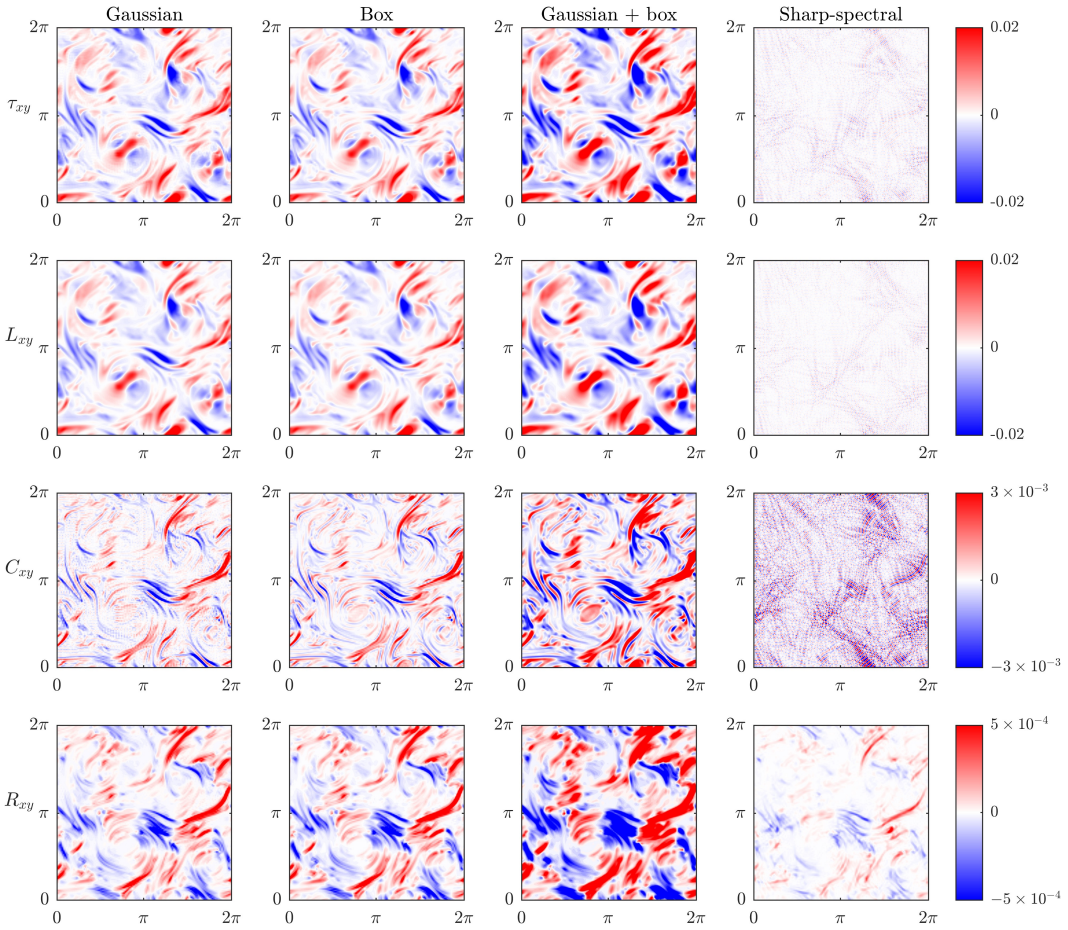

Figure 4. The first row shows examples of snapshots of the SGS stress, τ_{xy}, for Case K1, diagnosed from FNDS data using different filters and $N_{\text{LES}} = 128$ (see Table 1). Rows 2-4 show the three components of this τ_{xy}: the Leonard stress, L_{xy}, cross stress, C_{xy}, and Reynolds stress, R_{xy}. Note the substantially different ranges of the colorbars.

As for the dependency on filter size, as Δ increases, the relative importance of Reynolds stress increases: See Fig. 5 for examples from Cases K3 (τ_{xx}) and R3 (J_z). Finally, note that the relative importance of these three components might depend on the flow itself. For example, in 3 km-resolution regional simulations of the tropics, Sun et al. (2023) found that the vertical (horizontal) flux of the SGS zonal momentum is dominated by the Reynolds (Leonard) stress, which was attributed to the substantial differences of the filtered vertical wind and the filtered zonal or meridional winds.

The above analyses further explain the strong dependency of the diagnosed "true" SGS flux and the discovered closures on the filter type and size. These analyses also show that depending on the filter type/size, the Reynolds stress may not be the only component of the SGS flux that needs to be parameterized. In fact, the Leonard and cross stresses might be even larger and have to be included in the calculation of the total SGS flux and in the closure. Needless to say, these sensitivities have major implications for the "true" SGS flux that is fed into the RVM or any equation-discovery algorithm (and more broadly, any ML algorithm).

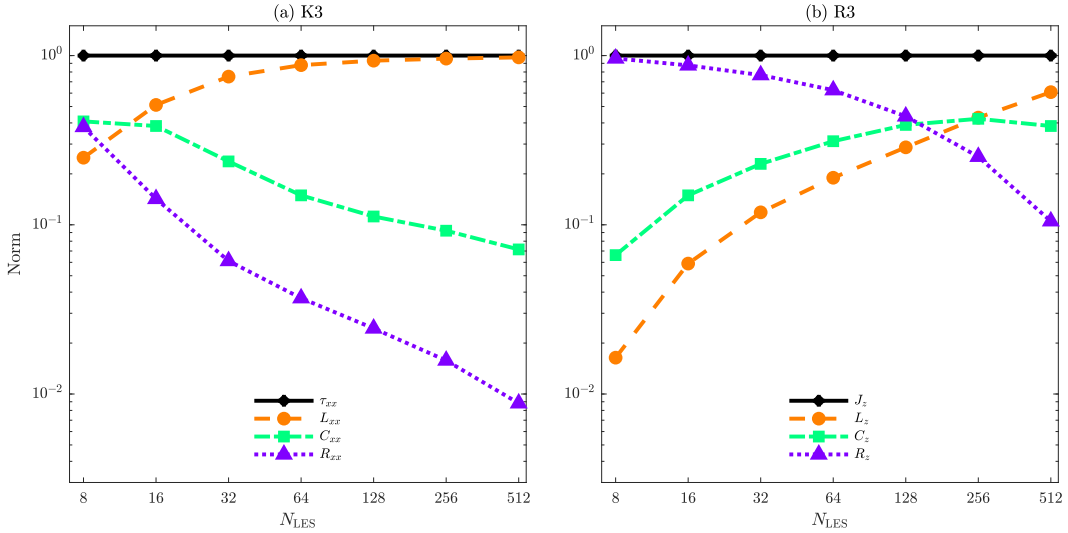

Figure 5. The L_2-norm of the SGS components versus N_{LES}. (a) τ_{xx} from Case K3. (b) J_z from Case R3. The contribution of SGS components is dependent on filter size: as N_{LES} decreases, i.e., Δ increases, the relative importance of Reynolds stress (Leonard stress) increases (decreases). Norm of all the SGS components are normalized by the respective SGS flux's norm. A Gaussian filter is used, but the same behavior is observed for the box and Gauassian+box filters.

4 Summary and Discussion

In this work, we have used relevance vector machine (RVM) to discover subgrid-scale (SGS) closures from filtered direct numerical simulation (DNS) data for both the SGS momentum flux tensor (in 2D forced homogeneous isotropic turbulence, 2D-FHIT, and Rayleigh-Bénard convection, RBC) and the SGS heat flux vector (in RBC). The expansive library includes derivatives of velocity (and temperature) up to 8th order (calculated using spectral methods) and their quadratic combinations. We have conducted extensive robustness analysis of the discovered closures across a variety of flow configurations (changing Re, Ra, Pr, and the forcing wavenumber), filter types (Gaussian, box, Gaussian + box, and sharp-spectral cutoff), and filter sizes.

Based on these analyses, except for when the sharp-spectral filter is used (see below), we have robustly discovered the *same closure* for the SGS stress in 2D-FHIT and RBC. We have further shown that this closure model is in fact the NGM, which can be derived analytically from the first term of the Taylor-series expansion of the convolution integral. The discovered SGS heat flux in RBC is also consistent with the truncated Taylor-series expansion. We have demonstrated a few important points about these discovered closures:

1. They all have high CC (often $> 0.9 - 0.95$) with the true SGS terms obtained from filtered DNS data, i.e., excellent performance based on this commonly used *a priori* test metric. The same closure is discovered regardless of the system because the expansion's first term dominates the MSE loss function of RVM.
2. Despite this high CC, all *a posteriori* (online) tests result in unstable LES. This is consistent with the past findings about the NGM in the LES community (mainly for 3D turbulence) and in the climate community (for geophysical turbulence). Here, we argue that the inability of NGM to capture any inter-scale kinetic energy transfer in 2D-FHIT (or any 2D flow filtered in both directions) is likely the reason for the instability. For RBC, where filtering is done only in one direction, deeper investigations into the spectra of the SGS fluxes and inter-scale enstrophy and potential energy

transfer, pointed to another likely reason for the instability: poor representation of the backscatter of SGS potential energy. This suggests that the poor *a posteriori* (online) performance of NGM might have different reasons in different flows.

3. The exact form of the discovered closure depends on the filter type and the filter size, Δ. For filters with compact support (i.e., all filters used here except for sharp-spectral cutoff), the structure of the closures is the same, but coefficients are different (still, consistent with the Taylor-series expansion, as shown in the appendices). For the sharp-spectral cutoff filter, the equation discovery fails, again, consistent with the fact that the Taylor-series expansion cannot be conducted, a known issue in the LES literature (Sagaut, 2006). Again, note that with a smoothed sharp-spectral filter, Ross et al. (2023) successfully performed equation-discovery.

As a side note, while the terms of the discovered closures might look like truncation error of finite difference/volume discretization, in our work, all calculations (DNS solver, SGS terms, library) are done using spectral methods. This further shows, along with the Taylor-series expansion results, that the discovered closures are indeed representing the physics of the SGS terms, rather than any numerical error.

As an additional piece of analysis, we also present the decomposition of the SGS terms to the Leonard, cross, and Reynolds terms. We show that the Leonard and then cross terms often dominate the total SGS term, though the relative amplitude of these terms decreases as the filter size increases. However, this analysis shows that only computing the Reynolds momentum stress or heat flux can lead to discovering an inaccurate closure (and in general, in data-driven SGS modeling, in too-small SGS fluxes). That said, the relative importance of these 3 terms depends on the filter type and size, and likely, on the flow's spatial spectrum (Sun et al., 2023).

The analyses presented in this paper are aimed at highlighting the promises and challenges of the equation-discovery approach to SGS modeling. On one hand, it is promising that this approach robustly discovers closures that could be closely connected with those mathematically derived, and could be easily interpreted and analyzed in terms of turbulence physics. On the other hand,

a) The commonly used MSE loss function, or similar loss functions, will be always dominated by the leading term(s) of the Taylor-series expansion. Thus, sparsity-promoting equation-discovery techniques, at least with the common derivative/polynomial-based libraries, will always find the NGM (if all the relevant terms exists in the library). Note that this is true for the closure of any SGS process, as the Taylor-series expansion applies to any compact filter. *Thus, this point and many of the main points of this paper are relevant beyond just SGS modeling for turbulence, but also SGS modeling of other nonlinear, multi-scale processes in the Earth system.*

b) Given that our diagnoses show shortcomings of the NGM with functional modeling metrics (e.g., inter-scale energy transfer), one idea is to include such physics constrains in the loss function. For example, Guan et al. (2023) demonstrated that a loss functions that combines structural and functional modeling constraints can enhance the *a priori* and *a posteriori* performance of the data-driven closure model in the small-data regime. More functional-modeling physics constraints (as domain averaged or wavenumber-dependent quantities) can be included in the loss function, which can potentially close the gap between *a priori* (offline) and *a posteriori* (online) performance. While the loss function of some techniques such as RVM may not be flexible to change beyond MSE, other methods such as GP/GEP or symbolic regression provide such flexibility (Ross et al., 2023; Cranmer, 2023). Also, equation-discovery using neural network-based algorithms has gained popularity recently, as for example, their loss functions can be very flexible given the use of backpropagation for training (Z. Chen et al., 2021). That said, "spectral bias" (Xu et al., 2019), the fundamental challenge of neural networks in learning high wavenumbers, can become

an issue when equation-discovery is the goal; see Mojgani et al. (2022a) for an example and a solution in a quasi-geostrophic turbulence testcase.

c) The fault may not entirely (or at all) lie with the MSE loss function. Guan et al. (2022a) showed that a deep CNN with basically the same MSE loss function as the one used here (which only accounts for structural modeling) can learn a closure for 2D turbulence that has CC>0.95 and leads to stable and accurate LES (and accurate inter-scale transfers; see Guan et al. (2023)). But a major difference between the CNN and RVM is that the former does not use a pre-defined set of basis functions, but rather, *learns* them. Recent work by Subel et al. (2023) has shown that the CNN of Guan et al. (2022a) learned a set of low-pass, high-pass, and band-pass Gabor filters. As another major difference, the CNN's sparsity is not user-defined, but rather, comes from over-parameterization.

d) Related to (c), the discovered closures can depend on the choice of the library. This issue can be addressed by trying more expansive libraries (though this can lead to non-robust discoveries) or as mentioned earlier, by using methods such as GP or GEP, which allow the library to evolve (see Schmidt and Lipson (2009); Udrescu and Tegmark (2020); Ross et al. (2023)). Libraries inspired by the CNNs' basis functions or distilled from other deep neural networks could be explored as well (Subel et al., 2023; Cranmer et al., 2020). Furthermore, there are studies, e.g., based on the Mori-Zwanzig formalism, suggesting that memory has to be included in closures (Wouters & Lucarini, 2013; Parish & Duraisamy, 2017). Hence, basis functions that include temporal information (as already used in Ross et al. (2023)) should be further explored in future work.

e) Choosing the hyper-parameter(s) that determine the level of sparsity might require more thoughts too. While the L-curve criterion has shown success in many problems, the metrics for which the curve is calculated for should be further investigated. The common *a priori* metrics such as CC of SGS fluxes are completely incapable of identifying shortcomings from a functional modeling perspective, such as lack of inter-scale energy transfer or poor representation of backscattering, which can be diagnosed using additional metrics. Note that a high CC of SGS fluxes has been found as the *necessary* but not *sufficient* condition for a successful closure (Meneveau, 1994).

f) Aside from all of the above issues related to the discovery algorithm, what needs to be discovered (the "truth") should be further examined. The discovered closures can depend on the filter type/size and the methodology (e.g., calculating Reynolds stress or the full SGS stress), because what is diagnosed as the "true" SGS flux from DNS has such dependencies. This has important implications for any data-driven SGS modeling approach, including those using deep neural networks or any other statistical learning method (Fatkullin & Vanden-Eijnden, 2004; Zanna & Bolton, 2021; Guan et al., 2022a; Sun et al., 2023).

We point out that there are other approaches to equation-discovery of SGS closures that are more directly focused on functional modeling. One is based on learning a closure from the differences between the evolved *states* of a high-resolution and a low-resolution simulation (Lang et al., 2016; Mojgani et al., 2022a, 2022b). The other is to learn from the differences between the evolved long-term *statistics* of such simulations (Schneider et al., 2020; Schneider, Stuart, & Wu, 2021; Schneider et al., 2022). These approaches would partially or entirely resolve the issues (a), (b), and (f) mentioned above, although challenges (d) and (e) would remain. Furthermore, the *a priori* performance of such closures and challenges in interpretability arising from numerical errors accumulated during evolutions are left to be further investigated.

In summary, equation-discovery is a promising approach to developing interpretable, practical, stable, and accurate SGS closures for various complex processes. However, further work, particularly on physics-guided loss functions (that for example, contain both structural

Table A1. List of 1D filters and their kernel and transfer functions. All filters are implemented in the spectral space, i.e., by applying their transfer function on Fourier-transformed variables. Here, r and k are coordinates in the physical space and spectral space, respectively. $\odot$ is the Hadamard product and $\hat{(.)}$ is the Fourier transform.

Filter	Kernel function	Transfer function
General	$G(r)$	$\hat{G}(k) = \int_{-\infty}^{\infty} e^{i2\pi kr} G(r)\, dr$
Gaussian (G_{G})	$\left(\frac{6}{\pi\Delta^2}\right)^{\frac{1}{2}} \exp\left(-\frac{6r^2}{\Delta^2}\right)$	$\exp\left(-\frac{k^2\Delta^2}{24}\right)$
Box (G_{B})	$\begin{cases} \frac{1}{\Delta}, & \text{if } r \le \frac{\Delta}{2} \\ 0, & \text{otherwise} \end{cases}$	$\frac{\sin(\frac{1}{2}k\Delta)}{\frac{1}{2}k\Delta}$
Gaussian + box (G_{GB})	$G_{\mathrm{G}}(r) * G_{\mathrm{B}}(r)$	$\hat{G}_{\mathrm{G}}(k) \odot \hat{G}_{\mathrm{B}}(k)$
Sharp-spectral cutoff (G_{S})	$\frac{\sin\left(\frac{\pi r}{\Delta}\right)}{\pi r}$	$\begin{cases} 1, & \text{if } \left(k_c - \lvert k\rvert \ge 0\right), k_c = \frac{2\pi}{\Delta} \\ 0, & \text{otherwise} \end{cases}$

and functional modeling components), physics- and data-guided libraries, and better metrics are needed.

Appendix A Filtering Procedure

In this work, we explore the most commonly used filters in LES and climate modeling: the Gaussian filter, the box filter, the Gaussian + box filter, and the sharp-spectral filter (Sagaut, 2006; Grooms et al., 2021). The box filter (also known as the top-hat filter) is simply the average of a variable over a box of dimension Δ; for instance, in 1D space, $\overline{u}(x,t)$ is the average of $u(x_\circ, t)$ over $x - \Delta/2 < x_\circ < x + \Delta/2$. The Gaussian filter's kernel is $G(r) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{r-\mu}{\sigma}\right)^2\right)$, with zero mean, $\mu = 0$, and variance, $\sigma^2 = \Delta^2/12$. These values are chosen to match the second moments of the Gaussian and box filters following Leonard (1975). The kernel for Gaussian + box filter is the convolution of the Gaussian and box filter kernels, which is equivalent to using a Gaussian filter followed by a box filter. The sharp-spectral cutoff filter simply removes the wavenumbers beyond a cutoff, k_c. The 1D filters used in this work are listed in Table A1, and the 2D filters are listed in Table A2. Note that all of these 4 filters commute with the spatial and temporal derivative operators on uniform grids (Pope, 2000; Sagaut, 2006).

Appendix B The 2D-FHIT Numerical Solver

The numerical solver is the same as the one used in Guan et al. (2022a). Briefly, we solve Eqs. (4)-(5) in the vorticity-streamfunction, $\omega - \psi$, formulation, where

$$\omega = (\nabla \times \boldsymbol{u}) \cdot \hat{\boldsymbol{z}}. \tag{B1}$$

Table A2. List of 2D filters and their kernel and transfer functions. All filters are implemented in the spectral space, i.e., by applying their transfer function on Fourier-transformed variables. Here, $\boldsymbol{r}$ and $\boldsymbol{k}$ are coordinates in the physical space and spectral space, respectively, with $\boldsymbol{r} = (r_x, r_y)$, $|\boldsymbol{r}|^2 = r_x^2 + r_y^2$, $\boldsymbol{k} = (k_x, k_y)$, and $|\boldsymbol{k}|^2 = k_x^2 + k_y^2$. $\odot$ is the Hadamard product and $\hat{(.)}$ is the Fourier transform.

Filter	Kernel function	Transfer function
General	$G(\boldsymbol{r})$	$\hat{G}(\boldsymbol{k}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{i2\pi(k_x r_x + k_y r_y)} G(\boldsymbol{r})\, d\boldsymbol{r}$
Gaussian (G_{G})	$\frac{6}{\pi\Delta^2}\exp\left(-\frac{6\,\lvert\boldsymbol{r}\rvert^2}{\Delta^2}\right)$	$\exp\left(-\frac{\lvert\boldsymbol{k}\rvert^2\,\Delta^2}{24}\right)$
Box (G_{B})	$\begin{cases} \frac{1}{\Delta^2}, & \text{if } (r_x, r_y) \le \frac{\Delta}{2} \\ 0, & \text{otherwise} \end{cases}$	$\frac{\sin(\frac{1}{2}k_x\Delta)\sin\left(\frac{1}{2}k_y\Delta\right)}{\left(\frac{1}{2}k_x\Delta\right)\left(\frac{1}{2}k_y\Delta\right)}$
Gaussian + box (G_{GB})	$G_{\mathrm{G}}(\boldsymbol{r}) * G_{\mathrm{B}}(\boldsymbol{r})$	$\hat{G}_{\mathrm{G}}(\boldsymbol{k}) \odot \hat{G}_{\mathrm{B}}(\boldsymbol{k})$
Sharp-spectral (G_{S})	$\frac{\sin\left(\frac{\pi\lvert\boldsymbol{r}\rvert}{\Delta}\right)}{\pi\,\lvert\boldsymbol{r}\rvert}$	$\begin{cases} 1, & \text{if } \left(k_c - \lvert\boldsymbol{k}\rvert \ge 0\right), k_c = \frac{2\pi}{\Delta} \\ 0, & \text{otherwise} \end{cases}$

With this formulation, the governing equations are

$$\nabla^2\psi = -\omega, \tag{B2}$$

$$\frac{\partial\omega}{\partial t} + \mathcal{N}(\omega,\psi) = \frac{1}{Re}\nabla^2\omega - \chi\omega - f, \tag{B3}$$

where $\mathcal{N}(\omega,\psi)$ is

$$\mathcal{N}(\omega,\psi) = \frac{\partial\psi}{\partial y}\frac{\partial\omega}{\partial x} - \frac{\partial\psi}{\partial x}\frac{\partial\omega}{\partial y}. \tag{B4}$$

f is a deterministic forcing (Chandler & Kerswell, 2013; Kochkov et al., 2021):

$$f = f_{k_x}\cos(f_{k_x}x) + f_{k_y}\cos(f_{k_y}y), \tag{B5}$$

where f_{k_x} and f_{k_y} are the forcing wavenumbers and $\chi = 0.1$ represents the Rayleigh drag coefficient. Comparing Eq. (5) with Eq. (B3), it is evident that $\nabla\times\boldsymbol{\mathcal{R}} = -\chi\omega$ and $\nabla\times\boldsymbol{\mathcal{F}} = -f$.

In DNS, Eqs. (B2)-(B3) are solved in a doubly periodic domain using a Fourier-Fourier pseudo-spectral solver with second-order Adams-Bashforth and Crank Nicholson time-integration schemes for the advection and viscous terms, respectively (time step Δt_{DNS}). For LES, we use the same solver with lower spatio-temporal resolution: We use N_{LES} that is 8 to 64 times smaller than N_{DNS}, and $\Delta t_{\mathrm{LES}} = 10\Delta t_{\mathrm{DNS}}$.

Appendix C The RBC Numerical Solver

We solve Eqs. (9)-(11) using a Fourier-Chebyshev pseudo-spectral solver (Khodkar et al., 2019; Khodkar & Hassanzadeh, 2018). Briefly, using the $\omega-\psi$ formulation, the dimen-

sionless governing equations become

$$\nabla^2\psi = -\omega \tag{C1}$$

$$\frac{\partial\omega}{\partial t} + \mathcal{N}(\omega,\psi) = Pr\nabla^2\omega + Pr\,Ra\,\theta\hat{\boldsymbol{z}}, \tag{C2}$$

$$\frac{\partial\theta}{\partial t} + \mathcal{M}(\theta,\psi) + \frac{\partial\psi}{\partial x} = \nabla^2\theta, \tag{C3}$$

where $\mathcal{N}(\omega,\psi)$ and $M(\theta,\psi)$ are

$$\mathcal{N}(\omega,\psi) = \frac{\partial\psi}{\partial z}\frac{\partial\omega}{\partial x} - \frac{\partial\psi}{\partial x}\frac{\partial\omega}{\partial z}, \quad \mathcal{M}(\theta,\psi) = \frac{\partial\psi}{\partial z}\frac{\partial\theta}{\partial x} - \frac{\partial\psi}{\partial x}\frac{\partial\theta}{\partial z}. \tag{C4}$$

For DNS, we solve Eqs. (C1)-(C3) in domain $(6\pi, 1)$. Periodic boundary conditions are imposed in the horizontal direction and no-slip and fixed temperature boundary conditions are imposed on the horizontal walls. Second-order Adams-Bashforth and Crank Nicholson time integration schemes are used for the advection and viscous terms, respectively. Table 2 presents the N_{DNS} and N_{LES} for each case. For LES, we use the same solver but with lower resolution in the horizontal direction.

Appendix D Taylor-series Expansion of the SGS Flux for a 1D Field

Let's focus on a spatially 1D field $a(x)$ (dependence on t is not explicitly written for brevity). The filtering operation's convolution integral (Eq. (1)) becomes

$$\overline{a}(x) = G * a = \int_{-\infty}^{\infty} G(r_x)\, a(x - r_x)\, dr_x, \tag{D1}$$

The Taylor-series expansion of $a(x - r_x)$ around $a(x)$ gives

$$a(x - r_x) = a(x) - \frac{1}{1!}\frac{\partial a(x)}{\partial x} r_x + \frac{1}{2!}\frac{\partial^2 a(x)}{\partial x^2} r_x^2 + \dots \tag{D2}$$

Substituting this into Eq. (D1) and using $a = a(x)$, $\overline{a} = \overline{a}(x)$ for brevity yields

$$\overline{a} = \int_{-\infty}^{\infty} G(r_x)\, a dr_x - \int_{-\infty}^{\infty} G(r_x)\frac{\partial a}{\partial x} r_x dr_x + \frac{1}{2!}\int_{-\infty}^{\infty} G(r_x)\frac{\partial^2 a}{\partial x^2} r_x^2 dr_x + \dots \tag{D3}$$

$$= a\int_{-\infty}^{\infty} G(r_x)\, dr_x - \frac{\partial a}{\partial x}\int_{-\infty}^{\infty} G(r_x)\, r_x dr_x + \frac{1}{2!}\frac{\partial^2 a}{\partial x^2}\int_{-\infty}^{\infty} G(r_x)\, r_x^2 dr_x + \dots \tag{D4}$$

The second line follows the first line considering that a and its derivatives do not depend on the variable of integration, r_x. In Eq. (D4), $\overline{a}$ depends on a and its derivatives, with coefficients that only depend on the filter type and size through moments of the kernel, G. For example, for a Gaussian filter (Table A1)

$$\int_{-\infty}^{\infty} G(r_x)\, dr_x = 1, \int_{-\infty}^{\infty} G(r_x)\, r_x dr_x = 0, \int_{-\infty}^{\infty} G(r_x)\, r_x^2 dr_x = \frac{\Delta^2}{12}. \tag{D5}$$

Note that all the odd moments are 0, resulting in $\mathcal{O}(\Delta^4)$ as the order of the truncated terms once the moments in Eq. (D5) are substituted in Eq. (D4):

$$\overline{a} = a + \frac{1}{2!}\frac{\Delta^2}{12}\frac{\partial^2 a}{\partial x^2} + \mathcal{O}(\Delta^4). \tag{D6}$$

To calculate a term like $\tau_{xx} = \overline{u^2} - \overline{u}^2$, we first use $a = u$ in Eq. (D6) and then square it to arrive at

$$\overline{u}^2 = u^2 + 2u\left(\frac{1}{2!}\frac{\Delta^2}{12}\frac{\partial^2 u}{\partial x^2}\right) + \mathcal{O}(\Delta^4). \tag{D7}$$

Next, we use $a = u^2$ in Eq. (D6) to obtain

$$\overline{u^2} = u^2 + \frac{1}{2!}\frac{\Delta^2}{12}\frac{\partial^2 u^2}{\partial x^2} + \mathcal{O}\left(\Delta^4\right), \tag{D8}$$

$$= u^2 + \frac{2}{2!}\frac{\Delta^2}{12}\left(\left(\frac{\partial u}{\partial x}\right)^2 + u\frac{\partial^2 u}{\partial x^2}\right) + \mathcal{O}\left(\Delta^4\right). \tag{D9}$$

Using Eq. (D7) and Eq. (D9) we find

$$\tau_{xx} = \overline{u^2} - \overline{u}^2 = \frac{\Delta^2}{12}\left(\frac{\partial u}{\partial x}\right)^2 + \mathcal{O}\left(\Delta^4\right). \tag{D10}$$

Note that this expression depends on u rather than $\overline{u}$, which is what we desire. Next, we use $a = \partial u/\partial x$ in Eq. (D6) to obtain

$$\frac{\partial \overline{u}}{\partial x} = \frac{\partial u}{\partial x} + \frac{1}{2!}\frac{\Delta^2}{12}\frac{\partial^3 u}{\partial x^3} + \mathcal{O}\left(\Delta^4\right). \tag{D11}$$

Using this expression in Eq. (D10) yields an analytically derived closure for τ_{xx} with error $\mathcal{O}\left(\Delta^4\right)$

$$\tau_{xx}^{\mathrm{NGM}} = \overline{u^2} - \overline{u}^2 = \frac{\Delta^2}{12}\left(\frac{\partial \overline{u}}{\partial x}\right)^2. \tag{D12}$$

This is the NGM (Leonard, 1975; Sagaut, 2006). Four issues should be emphasized here

i. This procedure can be followed for any filter type. However, the Taylor series is divergent for filters such as sharp-spectral, whose kernel's second-order moment is indefinite; for such filters, NGM does not exist (Meneveau & Katz, 2000; Sagaut, 2006).
ii. The same procedure can be followed to derive NGM for higher dimensions, e.g., τ_{xx}^{NGM}, τ_{xy}^{NGM}, and τ_{yy}^{NGM} in 2D; see Sagaut (2006).
iii. The coefficients in NGM depend on the filter' kernel and its moments (Eq. (D5)). For Gaussian and top-hat, the parameters of the kernels are chosen to match their first moment, leading to $\Delta^2/12$ coefficient for both. However the coefficients differ for higher-order terms (Sagaut, 2006).
iv. The procedure presented above is not specific to turbulence or even dynamical systems. The procedure and its outcome are valid for the filtered quadratic nonlinearity of any two variables, even random variables.

Appendix E Subgrid-scale Energy and Enstrophy Transfers

The filtered rate of train tensor $\overline{\mathbf{S}}$ and the anisotropic part of the SGS stress tensor $\boldsymbol{\tau}^{\mathrm{r}}$ are

$$\overline{\mathbf{S}} = \begin{bmatrix} \dfrac{\partial \overline{u}}{\partial x} & \dfrac{1}{2}\left(\dfrac{\partial \overline{u}}{\partial y} + \dfrac{\partial \overline{v}}{\partial x}\right) \\ \dfrac{1}{2}\left(\dfrac{\partial \overline{u}}{\partial y} + \dfrac{\partial \overline{v}}{\partial x}\right) & \dfrac{\partial \overline{v}}{\partial y} \end{bmatrix}, \tag{E1}$$

$$\boldsymbol{\tau}^{\mathrm{r}} = \boldsymbol{\tau} - \frac{1}{2}\mathrm{tr}\left(\boldsymbol{\tau}\right)\mathbf{I}, \tag{E2}$$

where $\mathbf{I}$ is the identity matrix. In 2D with filtering in both directions, the anisotropic part of the SGS stress tensor from the NGM is

$$\boldsymbol{\tau}^{\mathrm{NGM\text{-}r}} = \boldsymbol{\tau}^{\mathrm{NGM}} - \frac{1}{2}\mathrm{tr}\left(\boldsymbol{\tau}^{\mathrm{NGM}}\right)\mathbf{I}. \tag{E3}$$

$$\boldsymbol{\tau}_{\mathrm{2D}}^{\mathrm{NGM\text{-}r}} = \frac{\Delta^2}{12}\begin{bmatrix} \dfrac{1}{2}\left(\left(\dfrac{\partial \overline{u}}{\partial y}\right)^2 - \left(\dfrac{\partial \overline{v}}{\partial x}\right)^2\right) & \dfrac{\partial \overline{u}}{\partial x}\dfrac{\partial \overline{v}}{\partial x} + \dfrac{\partial \overline{u}}{\partial y}\dfrac{\partial \overline{v}}{\partial y} \\ \dfrac{\partial \overline{u}}{\partial x}\dfrac{\partial \overline{v}}{\partial x} + \dfrac{\partial \overline{u}}{\partial y}\dfrac{\partial \overline{v}}{\partial y} & -\dfrac{1}{2}\left(\left(\dfrac{\partial \overline{u}}{\partial y}\right)^2 - \left(\dfrac{\partial \overline{v}}{\partial x}\right)^2\right) \end{bmatrix}. \tag{E4}$$

Inserting this and $\overline{\mathbf{S}}$ (Eq. (E1)) into Eq. (30) shows zero point-wise inter-scale (kinetic) energy transfer in NGM: $P_\tau^{\text{NGM}}(x,y,t)=0$.

In buoyancy-driven turbulence such as RBC, the total inter-scale energy transfer rate P_E is the sum of the rate of transfer of kinetic energy (P_τ) due to SGS momentum fluxes and potential energy (P_J) due to SGS heat fluxes (Eidson, 1985; Peng & Davidson, 2002):

$$\begin{aligned} P_E &= P_\tau + P_J \\ &= -\tau_{ij}^r \overline{S}_{ij} - Ra\,Pr\,J_z. \end{aligned} \tag{E5}$$

Given the 1D filtering used in RBC, $\boldsymbol{\tau}_{\text{1D}}^{\text{NGM-r}}$ becomes

$$\boldsymbol{\tau}_{\text{1D}}^{\text{NGM-r}} = \frac{\Delta^2}{12}\begin{bmatrix} \frac{1}{2}\left(\left(\frac{\partial\overline{u}}{\partial x}\right)^2 - \left(\frac{\partial\overline{w}}{\partial x}\right)^2\right) & \frac{\partial\overline{u}}{\partial x}\frac{\partial\overline{w}}{\partial x} \\ \frac{\partial\overline{u}}{\partial x}\frac{\partial\overline{w}}{\partial x} & -\frac{1}{2}\left(\left(\frac{\partial\overline{u}}{\partial x}\right)^2 - \left(\frac{\partial\overline{w}}{\partial x}\right)^2\right) \end{bmatrix}, \tag{E6}$$

and P_τ^{NGM} is not strictly zero: The resulting production of subgrid-scale (SGS) energy transfer for NGM is

$$P_\tau^{\text{NGM}} = -\frac{\Delta^2}{12}\left(\frac{\partial^3\overline{u}}{\partial x^3} + \frac{\partial\overline{u}}{\partial x}\frac{\partial\overline{u}}{\partial z}\frac{\partial\overline{w}}{\partial x}\right). \tag{E7}$$

$$P_J^{\text{NGM}} = -Ra\,Pr\,\frac{\Delta^2}{12}\frac{\partial\overline{w}}{\partial x}\frac{\partial\overline{\theta}}{\partial x} \tag{E8}$$

Similarly, one can define the inter-scale enstrophy transfer for 2D-FHIT and RBC as (S. Chen et al., 2003)

$$P_Z = -\nabla\overline{\omega}\cdot\left(\overline{\mathbf{u}\omega} - \overline{\mathbf{u}}\,\overline{\omega}\right). \tag{E9}$$

Open Research

The codes and data used in this work can be found at `https://github.com/jakharkaran/EqsDiscovery_2D-FHIT_RBC` and `https://doi.org/10.5281/zenodo.7500647`, respectively.

Acknowledgments

We thank Ian Grooms, Sina Khani, Tony Leonard, and Charles Meneveau for insightful discussions. We are grateful to Moein Darman, Hamid Pahlavan, and Qiang Sun for helpful comments on the manuscript. This work was supported by an award from the ONR Young Investigator Program (N00014-20-1-2722), a grant from the NSF CSSI program (OAC-2005123), and by the generosity of Eric and Wendy Schmidt by recommendation of the Schmidt Futures program. Computational resources were provided by NSF XSEDE (allocation ATM170020) and NCAR's CISL (allocations URIC0004 and URIC0009).

References

Alexander, M., & Dunkerton, T. (1999). A spectral parameterization of mean-flow forcing due to breaking gravity waves. *Journal of the Atmospheric Sciences*, *56*(24), 4167–4182.

Anstey, J. A., & Zanna, L. (2017). A deformation-based parametrization of ocean mesoscale eddy Reynolds stresses. *Ocean Modelling*, *112*, 99–111.

Balaji, V. (2021). Climbing down Charney's ladder: machine learning and the post-Dennard era of computational climate science. *Philosophical Transactions of the Royal Society A*, *379*(2194), 20200085.

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.

Berner, J., Achatz, U., Batte, L., Bengtsson, L., De La Camara, A., Christensen, H. M., ... others (2017). Stochastic parameterization: Toward a new view of weather and climate models. *Bulletin of the American Meteorological Society*, *98*(3), 565–588.

Beucler, T., Pritchard, M., Rasp, S., Ott, J., Baldi, P., & Gentine, P. (2021). Enforcing analytic constraints in neural networks emulating physical systems. *Physical Review Letters*, *126*(9), 098302.

Boffetta, G., & Ecke, R. E. (2012). Two-dimensional turbulence. *Annual review of fluid mechanics*, *44*, 427–451.

Bolton, T., & Zanna, L. (2019). Applications of deep learning to ocean data inference and subgrid parameterization. *Journal of Advances in Modeling Earth Systems*, *11*(1), 376–399.

Borue, V., & Orszag, S. A. (1998). Local energy flux and subgrid-scale statistics in three-dimensional turbulence. *Journal of Fluid Mechanics*, *366*, 1–31.

Boussinesq, J. (1877). *Essai sur la théorie des eaux courantes [Essay on the theory of running waters]*. Impr. nationale.

Brenowitz, N. D., & Bretherton, C. S. (2018). Prognostic validation of a neural network unified physics parameterization. *Geophysical Research Letters*, *45*(12), 6289–6298.

Brunton, S. L., Noack, B. R., & Koumoutsakos, P. (2020). Machine learning for fluid mechanics. *Annual review of fluid mechanics*, *52*, 477–508.

Brunton, S. L., Proctor, J. L., & Kutz, J. N. (2016). Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the national academy of sciences*, *113*(15), 3932–3937.

Calvetti, D., Morigi, S., Reichel, L., & Sgallari, F. (2000). Tikhonov regularization and the L-curve for large discrete ill-posed problems. *Journal of Computational and Applied Mathematics*, *123*(1), 423-446.

Carati, D., Winckelmans, G. S., & Jeanmart, H. (2001). On the modelling of the subgrid-scale and filtered-scale stress tensors in large-eddy simulation. *Journal of Fluid Mechanics*, *441*, 119–138.

Cessi, P. (2008). An energy-constrained parameterization of eddy buoyancy flux. *Journal of physical oceanography*, *38*(8), 1807–1819.

Chai, X., & Mahesh, K. (2012). Dynamic-equation model for large-eddy simulation of compressible flows. *Journal of Fluid Mechanics*, *699*, 385–413.

Chandler, G. J., & Kerswell, R. R. (2013). Invariant recurrent solutions embedded in a turbulent two-dimensional Kolmogorov flow. *Journal of Fluid Mechanics*, *722*, 554–595.

Chattopadhyay, A., Subel, A., & Hassanzadeh, P. (2020). Data-driven super-parameterization using deep learning: Experimentation with multiscale Lorenz 96 systems and transfer learning. *Journal of Advances in Modeling Earth Systems*, *12*(11), e2020MS002084.

Chen, S., Ecke, R. E., Eyink, G. L., Rivera, M., Wan, M., & Xiao, Z. (2006). Physical mechanism of the two-dimensional inverse energy cascade. *Physical review letters*, *96*(8), 084502.

Chen, S., Ecke, R. E., Eyink, G. L., Wang, X., & Xiao, Z. (2003). Physical mechanism of the two-dimensional enstrophy cascade. *Physical review letters*, *91*(21), 214501.

Chen, Y., Luo, Y., Liu, Q., Xu, H., & Zhang, D. (2022). Symbolic genetic algorithm for discovering open-form partial differential equations (SGA-PDE). *Physical Review Research*, *4*(2), 023174.

Chen, Z., Liu, Y., & Sun, H. (2021). Physics-informed learning of governing equations from scarce data. *Nature communications*, *12*(1), 1–13.

Cheng, Y., Giometto, M. G., Kauffmann, P., Lin, L., Cao, C., Zupnick, C., ... others (2022). Deep learning for subgrid-scale turbulence modeling in large-eddy simulations of the convective atmospheric boundary layer. *Journal of Advances in Modeling Earth*

Systems, e2021MS002847.
Chillà, F., & Schumacher, J. (2012). New perspectives in turbulent Rayleigh-Bénard convection. *The European Physical Journal E*, *35*(7), 1–25.
Clare, M. C., Sonnewald, M., Lguensat, R., Deshayes, J., & Balaji, V. (2022). Explainable artificial intelligence for bayesian neural networks: Toward trustworthy predictions of ocean dynamics. *Journal of Advances in Modeling Earth Systems*, *14*(11), e2022MS003162.
Clark, R. A., Ferziger, J. H., & Reynolds, W. C. (1979). Evaluation of subgrid-scale models using an accurately simulated turbulent flow. *Journal of fluid mechanics*, *91*(1), 1–16.
Cottet, G.-H. (1996). Artificial viscosity models for vortex and particle methods. *Journal of Computational Physics*, *127*(2), 299–308.
Cranmer, M. (2023). Interpretable machine learning for science with PySR and SymbolicRegression. jl. *arXiv preprint arXiv:2305.01582*.
Cranmer, M., Sanchez Gonzalez, A., Battaglia, P., Xu, R., Cranmer, K., Spergel, D., & Ho, S. (2020). Discovering symbolic models from deep learning with inductive biases. *Advances in Neural Information Processing Systems*, *33*, 17429–17442.
Dabbagh, F., Trias, F., Gorobets, A., & Oliva, A. (2017). A priori study of subgrid-scale features in turbulent Rayleigh-Bénard convection. *Physics of Fluids*, *29*(10), 105103.
Duraisamy, K. (2021). Perspectives on machine learning-augmented Reynolds-averaged and large eddy simulation models of turbulence. *Physical Review Fluids*, *6*(5), 050504.
Eidson, T. M. (1985). Numerical simulation of the turbulent Rayleigh–Bénard problem using subgrid modelling. *Journal of Fluid Mechanics*, *158*, 245–268.
Fabre, Y., & Balarac, G. (2011). Development of a new dynamic procedure for the Clark model of the subgrid-scale scalar flux using the concept of optimal estimator. *Physics of Fluids*, *23*(11), 115103.
Fatkullin, I., & Vanden-Eijnden, E. (2004). A computational strategy for multiscale systems with applications to Lorenz 96 model. *Journal of Computational Physics*, *200*(2), 605–638.
Fox-Kemper, B., Adcroft, A., Böning, C. W., Chassignet, E. P., Curchitser, E., Danabasoglu, G., ... others (2019). Challenges and prospects in ocean circulation models. *Frontiers in Marine Science*, *6*, 65.
Frezat, H., Le Sommer, J., Fablet, R., Balarac, G., & Lguensat, R. (2022). A posteriori learning for quasi-geostrophic turbulence parametrization. *Journal of Advances in Modeling Earth Systems*, *14*(11), e2022MS003124.
Gallet, B., & Ferrari, R. (2020). The vortex gas scaling regime of baroclinic turbulence. *Proceedings of the National Academy of Sciences*, *117*(9), 4491–4497.
Gantmakher, F. R. (2000). *The theory of matrices* (Vol. 131). American Mathematical Soc.
Gatski, T. B., & Speziale, C. G. (1993). On explicit algebraic stress models for complex turbulent flows. *Journal of fluid Mechanics*, *254*, 59–78.
Gentine, P., Eyring, V., & Beucler, T. (2021). Deep learning for the parametrization of subgrid processes in climate models. *Deep Learning for the Earth Sciences: A Comprehensive Approach to Remote Sensing, Climate Science, and Geosciences*, 307–314.
Germano, M. (1986). A proposal for a redefinition of the turbulent stresses in the filtered Navier-Stokes equations. *The Physics of fluids*, *29*(7), 2323–2324.
Germano, M. (1992). Turbulence: The filtering approach. *Journal of Fluid Mechanics*, *238*, 325–336.
Ghosal, S., Lund, T. S., & Moin, P. (1993). A local dynamic model for large eddy simulation. *Annual Research Briefs, 1992*.
Goyal, P., & Benner, P. (2022). Discovery of nonlinear dynamical systems using a Runge-Kutta inspired dictionary-based sparse regression approach. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *478*(2262), 20210883.
Grooms, I., Lee, Y., & Majda, A. J. (2015). Numerical schemes for stochastic backscatter in the inverse cascade of quasigeostrophic turbulence. *Multiscale Modeling & Simulation*,

13(3), 1001–1021.

Grooms, I., Loose, N., Abernathey, R., Steinberg, J., Bachman, S. D., Marques, G., ... Yankovsky, E. (2021). Diffusion-based smoothers for spatial filtering of gridded geophysical data. *Journal of Advances in Modeling Earth Systems*, *13*(9), e2021MS002552.

Guan, Y., Chattopadhyay, A., Subel, A., & Hassanzadeh, P. (2022a). Stable a posteriori LES of 2D turbulence using convolutional neural networks: Backscattering analysis and generalization to higher Re via transfer learning. *Journal of Computational Physics*, *458*, 111090.

Guan, Y., Subel, A., Chattopadhyay, A., & Hassanzadeh, P. (2023). Learning physics-constrained subgrid-scale closures in the small-data regime for stable and accurate LES. *Physica D: Nonlinear Phenomena*, *443*, 133568.

Guillaumin, A. P., & Zanna, L. (2021). Stochastic-deep learning parameterization of ocean momentum forcing. *Journal of Advances in Modeling Earth Systems*, *13*(9), e2021MS002534.

Hassanzadeh, P., Chini, G. P., & Doering, C. R. (2014). Wall to wall optimal transport. *Journal of fluid mechanics*, *751*, 627–662.

Herman, M. J., & Kuang, Z. (2013). Linear response functions of two convective parameterization schemes. *Journal of Advances in Modeling Earth Systems*, *5*(3), 510–541.

Hewitt, H. T., Roberts, M., Mathiot, P., Biastoch, A., Blockley, E., Chassignet, E. P., ... others (2020). Resolving and parameterising the ocean mesoscale in earth system models. *Current Climate Change Reports*, *6*(4), 137–152.

Jansen, M. F., & Held, I. M. (2014). Parameterizing subgrid-scale eddy effects using energetically consistent backscatter. *Ocean Modelling*, *80*, 36–48.

Jansen, M. F., Held, I. M., Adcroft, A., & Hallberg, R. (2015). Energy budget-based backscatter in an eddy permitting primitive equation model. *Ocean Modelling*, *94*, 15–26.

Jongen, T., & Gatski, T. (1998). General explicit algebraic stress relations and best approximation for three-dimensional flows. *International Journal of Engineering Science*, *36*(7-8), 739–763.

Juricke, S., Danilov, S., Koldunov, N., Oliver, M., Sein, D., Sidorenko, D., & Wang, Q. (2020). A kinematic kinetic energy backscatter parametrization: From implementation to global ocean simulations. *Journal of Advances in Modeling Earth Systems*, *12*(12).

Khani, S., & Dawson, C. N. (2023). A gradient based subgrid-scale parameterization for ocean mesoscale eddies. *Journal of Advances in Modeling Earth Systems*, e2022MS003356.

Khani, S., & Porté-Agel, F. (2017). A modulated-gradient parametrization for the large-eddy simulation of the atmospheric boundary layer using the weather research and forecasting model. *Boundary-Layer Meteorology*, *165*, 385–404.

Khani, S., & Porté-Agel, F. (2022). A gradient tensor–based subgrid-scale parameterization for large-eddy simulations of stratified shear layers using the weather research and forecasting model. *Monthly Weather Review*, *150*(9), 2279–2298.

Khani, S., & Waite, M. L. (2020). An anisotropic subgrid-scale parameterization for large-eddy simulations of stratified turbulence. *Monthly Weather Review*, *148*(10), 4299–4311.

Khodkar, M. A., & Hassanzadeh, P. (2018). Data-driven reduced modelling of turbulent Rayleigh–Bénard convection using DMD-enhanced fluctuation–dissipation theorem. *Journal of Fluid Mechanics*, *852*.

Khodkar, M. A., Hassanzadeh, P., Nabi, S., & Grover, P. (2019). Reduced-order modeling of fully turbulent buoyancy-driven flows using the Green's function method. *Physical Review Fluids*, *4*(1), 013801.

Kochkov, D., Smith, J. A., Alieva, A., Wang, Q., Brenner, M. P., & Hoyer, S. (2021). Machine learning–accelerated computational fluid dynamics. *Proceedings of the National Academy of Sciences*, *118*(21).

Kooloth, P., Sondak, D., & Smith, L. M. (2021). Coherent solutions and transition to

turbulence in two-dimensional Rayleigh-Bénard convection. *Physical Review Fluids*, *6*(1), 013501.

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

Lang, M., Jan Van Leeuwen, P., & Browne, P. (2016). A systematic method of parameterisation estimation using data assimilation. *Tellus A: Dynamic Meteorology and Oceanography*, *68*(1), 29012.

Lappa, M. (2009). *Thermal convection: Patterns, evolution and stability*. John Wiley & Sons.

Lawson, C. L., & Hanson, R. J. (1995). *Solving least squares problems*. Society for Industrial and Applied Mathematics.

Leonard, A. (1975). Energy cascade in large-eddy simulations of turbulent fluid flows. In *Advances in geophysics* (Vol. 18, pp. 237–248). Elsevier.

Leonard, A. (1997). Large-eddy simulation of chaotic convection and beyond. In *35th aerospace sciences meeting and exhibit* (p. 204).

Leonard, A. (2016). Large-eddy simulation of the Navier-Stokes equations: deconvolution, particle methods, and super-resolution. In T. Sengupta, S. Lele, K. Sreenivasan, & P. Davidson (Eds.), *IUTAM symposium on advances in computation, modeling and control of transitional and turbulent flows* (pp. 1–16). World Scientific.

Li, H., Zhao, Y., Wang, J., & Sandberg, R. D. (2021). Data-driven model development for large-eddy simulation of turbulence using gene-expression programing. *Physics of Fluids*, *33*(12), 125127.

Lilly, D. K. (1992). A proposed modification of the Germano subgrid-scale closure method. *Physics of Fluids A: Fluid Dynamics*, *4*(3), 633–635.

Liu, S., Meneveau, C., & Katz, J. (1994). On the properties of similarity subgrid-scale models as deduced from measurements in a turbulent jet. *Journal of Fluid Mechanics*, *275*, 83–119.

Lu, H., & Porté-Agel, F. (2010). A modulated gradient model for large-eddy simulation: application to a neutral atmospheric boundary layer. *Physics of Fluids*, *22*(1), 015109.

Lu, H., & Porté-Agel, F. (2014). On the development of a dynamic non-linear closure for large-eddy simulation of the atmospheric boundary layer. *Boundary-layer meteorology*, *151*(3), 429–451.

Lund, T. S., & Novikov, E. (1993). Parameterization of subgrid-scale stress by the velocity gradient tensor. *Annual Research Briefs, 1992*.

Mamalakis, A., Barnes, E. A., & Ebert-Uphoff, I. (2022). Investigating the fidelity of explainable artificial intelligence methods for applications of convolutional neural networks in geoscience. *Artificial Intelligence for the Earth Systems*, *1*(4), e220012.

Mangan, N. M., Kutz, J. N., Brunton, S. L., & Proctor, J. L. (2017). Model selection for dynamical systems via sparse regression and information criteria. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *473*(2204), 20170009.

Maulik, R., San, O., Rasheed, A., & Vedula, P. (2018). Data-driven deconvolution for large eddy simulations of Kraichnan turbulence. *Physics of Fluids*, *30*(12), 125109.

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

McDonough, J. M. (2007). *Introductory lectures on turbulence: Physics, mathematics and modeling*. Mechanical Engineering Textbook Gallery. Retrieved from `https://uknowledge.uky.edu/me_textbooks/2`

Meneveau, C. (1994). Statistics of turbulence subgrid-scale stresses: Necessary conditions and experimental tests. *Physics of Fluids*, *6*(2), 815–833.

Meneveau, C., & Katz, J. (2000). Scale-invariance and turbulence models for large-eddy simulation. *Annual Review of Fluid Mechanics*, *32*(1), 1–32.

Mojgani, R., Chattopadhyay, A., & Hassanzadeh, P. (2022b). Discovery of interpretable structural model errors by combining Bayesian sparse regression and data assimilation:

A chaotic Kuramoto-Sivashinsky test case. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *32*(6), 061105.

Mojgani, R., Chattopadhyay, A. K., & Hassanzadeh, P. (2022a). Discovering interpretable structural model errors in climate models. In *Neurips 2022 workshop on tackling climate change with machine learning*. Retrieved from `https://www.climatechange.ai/papers/neurips2022/8`

Moser, R. D., Haering, S. W., & Yalla, G. R. (2021). Statistical properties of subgrid-scale turbulence models. *Annual Review of Fluid Mechanics*, *53*, 255–286.

Nadiga, B. (2008). Orientation of eddy fluxes in geostrophic turbulence. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *366*(1875), 2489–2508.

Nadiga, B. (2010). Stochastic vs. deterministic backscatter of potential enstrophy in geostrophic turbulence. *Stochastic Physics and Climate Modeling, edited by T. Palmer and P. Williams (Cambridge University Press, Cambridge, England, 2009)*.

O'Kane, T. J., & Frederiksen, J. (2008). Statistical dynamical subgrid-scale parameterizations for geophysical flows. *Physica Scripta*, *2008*(T132), 014033.

Orszag, S. A. (1971). On the elimination of aliasing in finite-difference schemes by filtering high-wavenumber components. *Journal of Atmospheric Sciences*, *28*(6), 1074–1074.

Palmer, T. N. (2001). A nonlinear dynamical perspective on model error: A proposal for non-local stochastic-dynamic parametrization in weather and climate prediction models. *Quarterly Journal of the Royal Meteorological Society*, *127*(572), 279–304.

Pandey, S., Teutsch, P., Mäder, P., & Schumacher, J. (2022). Direct data-driven forecast of local turbulent heat flux in Rayleigh–Bénard convection. *Physics of Fluids*, *34*(4), 045106.

Parish, E. J., & Duraisamy, K. (2017). Non-Markovian closure models for large eddy simulations using the Mori-Zwanzig formalism. *Physical Review Fluids*, *2*(1), 014604.

Pawar, S., San, O., Rasheed, A., & Vedula, P. (2020). A priori analysis on deep learning of subgrid-scale parameterizations for Kraichnan turbulence. *Theoretical and Computational Fluid Dynamics*, *34*(4), 429–455.

Peng, S.-H., & Davidson, L. (2002). On a subgrid-scale heat flux model for large eddy simulation of turbulent thermal flow. *International Journal of Heat and Mass Transfer*, *45*(7), 1393–1405.

Pope, S. (1975). A more general effective-viscosity hypothesis. *Journal of Fluid Mechanics*, *72*(2), 331–340.

Pope, S. (2000). *Turbulent flows*. Cambridge university press.

Prakash, A., Jansen, K. E., & Evans, J. A. (2021). Optimal clipping of the gradient model for subgrid stress closure. In *Aiaa scitech 2021 forum* (p. 1665).

Rasp, S., Pritchard, M. S., & Gentine, P. (2018). Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences*, *115*(39), 9684–9689.

Reissmann, M., Hasslberger, J., Sandberg, R. D., & Klein, M. (2021). Application of gene expression programming to a-posteriori LES modeling of a Taylor Green vortex. *Journal of Computational Physics*, *424*, 109859.

Ross, A., Li, Z., Perezhogin, P., Fernandez-Granda, C., & Zanna, L. (2023). Benchmarking of machine learning ocean subgrid parameterizations in an idealized model. *Journal of Advances in Modeling Earth Systems*, *15*(1), e2022MS003258.

Rudy, S. H., Brunton, S. L., Proctor, J. L., & Kutz, J. N. (2017). Data-driven discovery of partial differential equations. *Science advances*, *3*(4), e1602614.

Sadourny, R., & Basdevant, C. (1985). Parameterization of subgrid scale barotropic and baroclinic eddies in quasi-geostrophic models: Anticipated potential vorticity method. *Journal of Atmospheric Sciences*, *42*(13), 1353–1363.

Sagaut, P. (2006). *Large eddy simulation for incompressible flows: an introduction*. Springer Science & Business Media.

Schaeffer, H. (2017). Learning partial differential equations via data discovery and sparse optimization. *Proceedings of the Royal Society A: Mathematical, Physical and Engi-*

neering Sciences, *473*(2197), 20160446.
Schmidt, M., & Lipson, H. (2009). Distilling free-form natural laws from experimental data. *science*, *324*(5923), 81–85.
Schneider, T., Jeevanjee, N., & Socolow, R. (2021). Accelerating progress in climate science. *Physics Today*, *74*(6), 44–51.
Schneider, T., Lan, S., Stuart, A., & Teixeira, J. (2017). Earth system modeling 2.0: A blueprint for models that learn from observations and targeted high-resolution simulations. *Geophysical Research Letters*, *44*(24), 12–396.
Schneider, T., Stuart, A. M., & Wu, J.-L. (2020). Imposing sparsity within ensemble Kalman inversion. *arXiv preprint arXiv:2007.06175*.
Schneider, T., Stuart, A. M., & Wu, J.-L. (2021). Learning stochastic closures using ensemble Kalman inversion. *Transactions of Mathematics and Its Applications*, *5*(1), tnab003.
Schneider, T., Stuart, A. M., & Wu, J.-L. (2022). Ensemble Kalman inversion for sparse learning of dynamical systems from time-averaged data. *Journal of Computational Physics*, 111559.
Schneider, T., Teixeira, J., Bretherton, C. S., Brient, F., Pressel, K. G., Schär, C., & Siebesma, A. P. (2017). Climate goals and computing the future of clouds. *Nature Climate Change*, *7*(1), 3–5.
Shen, Z., Sridhar, A., Tan, Z., Jaruga, A., & Schneider, T. (2022). A library of large-eddy simulations forced by global climate models. *Journal of Advances in Modeling Earth Systems*, *14*(3), e2021MS002631.
Shutts, G. (2005). A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, *131*(612), 3079–3102.
Smagorinsky, J. (1963). General circulation experiments with the primitive equations: I. the basic experiment. *Monthly weather review*, *91*(3), 99–164.
Sondak, D., Smith, L. M., & Waleffe, F. (2015). Optimal heat transport solutions for Rayleigh–Bénard convection. *Journal of Fluid Mechanics*, *784*, 565–595.
Speziale, C. G. (1985). Galilean invariance of subgrid-scale stress models in the large-eddy simulation of turbulence. *Journal of fluid mechanics*, *156*, 55–62.
Sridhar, A., Tissaoui, Y., Marras, S., Shen, Z., Kawczynski, C., Byrne, S., ... others (2022). Large-eddy simulations with climatemachine v0. 2.0: A new open-source code for atmospheric simulations on GPUs and CPUs. *Geoscientific Model Development*, *15*(15), 6259–6284.
Srinivasan, K., Chekroun, M. D., & McWilliams, J. C. (2023). Turbulence closure with small, local neural networks: Forced two-dimensional and \β-plane flows. *arXiv preprint arXiv:2304.05029*.
Stensrud, D. J. (2009). *Parameterization schemes: Keys to understanding numerical weather prediction models*. Cambridge University Press.
Subel, A., Chattopadhyay, A., Guan, Y., & Hassanzadeh, P. (2021). Data-driven subgrid-scale modeling of forced Burgers turbulence using deep learning with generalization to higher Reynolds numbers via transfer learning. *Physics of Fluids*, *33*(3), 031702.
Subel, A., Guan, Y., Chattopadhyay, A., & Hassanzadeh, P. (2023). Explaining the physics of transfer learning in data-driven turbulence modeling. *PNAS Nexus*, pgad015.
Sullivan, P. P., McWilliams, J. C., & Moeng, C.-H. (1994). A subgrid-scale model for large-eddy simulation of planetary boundary-layer flows. *Boundary-Layer Meteorology*, *71*(3), 247–276.
Sun, Y. Q., Hassanzadeh, P., Alexander, M. J., & Kruse, C. G. (2023). Quantifying 3D gravity wave drag in a library of tropical convection-permitting simulations for data-driven parameterizations. *Journal of Advances in Modeling Earth Systems (in press)*. Retrieved from `https://doi.org/10.22541/essoar.167126083.37839360/v1`
Tabeling, P. (2002). Two-dimensional turbulence: A physicist approach. *Physics reports*, *362*(1), 1–62.

Tan, Z., Kaul, C. M., Pressel, K. G., Cohen, Y., Schneider, T., & Teixeira, J. (2018). An extended eddy-diffusivity mass-flux scheme for unified representation of subgrid-scale turbulence and convection. *Journal of Advances in Modeling Earth Systems*, *10*(3), 770–800.

Thuburn, J., Kent, J., & Wood, N. (2014). Cascades, backscatter and conservation in numerical models of two-dimensional turbulence. *Quarterly Journal of the Royal Meteorological Society*, *140*(679), 626–638.

Tipping, M. E. (2001). Sparse Bayesian learning and the relevance vector machine. *Journal of machine learning research*, *1*(Jun), 211–244.

Udrescu, S.-M., & Tegmark, M. (2020). Ai Feynman: A physics-inspired method for symbolic regression. *Science Advances*, *6*(16), eaay2631.

Vallis, G. K. (2017). *Atmospheric and oceanic fluid dynamics*. Cambridge University Press.

Verkley, W. T., Severijns, C. A., & Zwaal, B. A. (2019). A maximum entropy approach to the interaction between small and large scales in two-dimensional turbulence. *Quarterly Journal of the Royal Meteorological Society*, *145*(722), 2221–2236.

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*(2), 025114.

Vreman, B., Geurts, B., & Kuerten, H. (1996). Large-eddy simulation of the temporal mixing layer using the Clark model. *Theoretical and Computational Fluid Dynamics*, *8*(4), 309–324.

Vreman, B., Geurts, B., & Kuerten, H. (1997). Large-eddy simulation of the turbulent mixing layer. *Journal of fluid mechanics*, *339*, 357–390.

Wang, B. C., Yee, E., Bergstrom, D. J., & Iida, O. (2008). New dynamic subgrid-scale heat flux models for large-eddy simulation of thermal convection based on the general gradient diffusion hypothesis. *Journal of Fluid Mechanics*, *604*, 125–163.

Winckelmans, G., Wray, A., & Vasilyev, O. (1998). Testing of a new mixed model for les: The leonard model supplemented by a dynamic smagorinsky term. In *Summer program* (pp. 367–388).

Wouters, J., & Lucarini, V. (2013). Multi-level dynamical systems: Connecting the Ruelle response theory and the Mori-Zwanzig approach. *Journal of Statistical Physics*, *151*(5), 850–860.

Xie, C., Wang, J., Li, H., Wan, M., & Chen, S. (2019). Artificial neural network mixed model for large eddy simulation of compressible isotropic turbulence. *Physics of Fluids*, *31*(8), 085112.

Xu, Z.-Q. J., Zhang, Y., Luo, T., Xiao, Y., & Ma, Z. (2019). Frequency principle: Fourier analysis sheds light on deep neural networks. *arXiv preprint arXiv:1901.06523*.

Yalla, G. R., Oliver, T. A., Haering, S. W., Engquist, B., & Moser, R. D. (2021). Effects of resolution inhomogeneity in large-eddy simulation. *Physical Review Fluids*, *6*(7), 074604.

Yuval, J., & O'Gorman, P. A. (2020). Stable machine-learning parameterization of subgrid processes for climate modeling at a range of resolutions. *Nature communications*, *11*(1), 1–10.

Zanna, L., & Bolton, T. (2020). Data-driven equation discovery of ocean mesoscale closures. *Geophysical Research Letters*, *47*(17), e2020GL088376.

Zanna, L., & Bolton, T. (2021). Deep learning of unresolved turbulent ocean processes in climate models. *Deep Learning for the Earth Sciences: A Comprehensive Approach to Remote Sensing, Climate Science, and Geosciences*, 298–306.

Zanna, L., Mana, P. P., Anstey, J., David, T., & Bolton, T. (2017). Scale-aware deterministic and stochastic parametrizations of eddy-mean flow interaction. *Ocean Modelling*, *111*, 66–80.

Zhang, S., & Lin, G. (2018). Robust data-driven discovery of governing physical laws with error bars. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *474*(2217), 20180305.

Zhang, X., Schneider, T., Shen, Z., Pressel, K. G., & Eisenman, I. (2022). Seasonal cycle of

idealized polar clouds: Large eddy simulations driven by a GCM. *Journal of Advances in Modeling Earth Systems*, *14*(1), e2021MS002671.

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