

REVIEW

Perspectives toward Stochastic and Learned-by-Data Turbulence in Numerical Weather Prediction

METODIJA M. SHAPKALJEVSKI^a ^a *R and D Meteorology, Swedish Meteorological and Hydrological Institute, Norrköping, Sweden*

(Manuscript received 31 December 2022, in final form 16 November 2023, accepted 20 November 2023)

ABSTRACT: The increased social need for more precise and reliable weather forecasts, especially when focusing on extreme weather events, pushes forward research and development in meteorology toward novel numerical weather prediction (NWP) systems that can provide simulations that resolve atmospheric processes on hectometric scales on demand. Such high-resolution NWP systems require a more detailed representation of the nonresolved processes, i.e., usage of scale-aware schemes for convection and three-dimensional turbulence (and radiation), which would additionally increase the computation needs. Therefore, developing and applying comprehensive, reliable, and computationally acceptable parameterizations in NWP systems is of urgent importance. All operationally used NWP systems are based on averaged Navier–Stokes equations, and thus require an approximation for the small-scale turbulent fluxes of momentum, energy, and matter in the system. The availability of high-fidelity data from turbulence experiments and direct numerical simulations has helped scientists in the past to construct and calibrate a range of turbulence closure approximations (from the relatively simple to more complex), some of which have been adopted and are in use in the current operational NWP systems. The significant development of learned-by-data (LBD) algorithms over the past decade (e.g., artificial intelligence) motivates engineers and researchers in fluid dynamics to explore alternatives for modeling turbulence by directly using turbulence data to quantify and reduce model uncertainties systematically. This review elaborates on the LBD approaches and their use in NWP currently, and also searches for novel data-informed turbulence models that can potentially be used and applied in NWP. Based on this literature analysis, the challenges and perspectives to do so are discussed.

KEYWORDS: Turbulence; Inverse methods; Numerical weather prediction/forecasting; Parameterization; Subgrid-scale processes; Machine learning

I soon understood that there was little hope of developing a pure, closed theory, and, because of an absence of such a theory, the investigation must be based on hypotheses obtained on processing experimental data. (Kolmogorov, quoted by A. Tsinober)

1. Introduction

This study reviews the current state and perspectives of using stochastic and learned-by-data (LBD) methods in representing unresolved (subgrid scale) processes in operational numerical weather prediction (NWP) systems with a focus on turbulent representation.

The representation (parameterization) of subgrid-scale processes in NWP systems and global circulation models (GCM) is and will remain, a relevant cause of uncertainties. This is mainly due to the numerical approximation of nonlinear and complex

systems of multiscale interactions between atmospheric dynamics, radiation, clouds (and their microphysics), the boundary layer convective and turbulent mixing, and Earth's surface in NWP systems. As a consequence, one of the most important features in NWP is the resolution of numerical discretization systems, which determines the range of atmospheric scales (and thus the class of processes) that can be resolved, and the ones that have to be parameterized (Fig. 1). Intuitively, this will also affect NWP uncertainty and thus weather forecast precision.

Indeed, increasing the resolution of NWP models helps to improve weather forecasts and climate predictions by resolving more “finer” scale processes, such as deep moist convection (Wilhelmson and Wicker 2001). For instance, current operational storm-resolving models (often called convection-permitting models) run on horizontal grids of approximately 1–4 km and are capable of resolving deep convective clouds and thus improving the representation of land–atmosphere coupling (Sun and Pritchard 2016), convective organization (Muller and Bony 2015), and cumulonimbus-type extreme

 Denotes content that is immediately available upon publication as open access.

Corresponding author: Metodija M. Shapkaljevski, metodija.shapkaljevski@smhi.se

Publisher's Note: This article was revised on 6 February 2024 to correct the formatting of two variables from Eq. (1) that appear in the running text immediately following the equation.

DOI: 10.1175/WAF-D-22-0228.1

© 2024 American Meteorological Society. This published article is licensed under the terms of a Creative Commons Attribution 4.0 International (CC BY 4.0) License



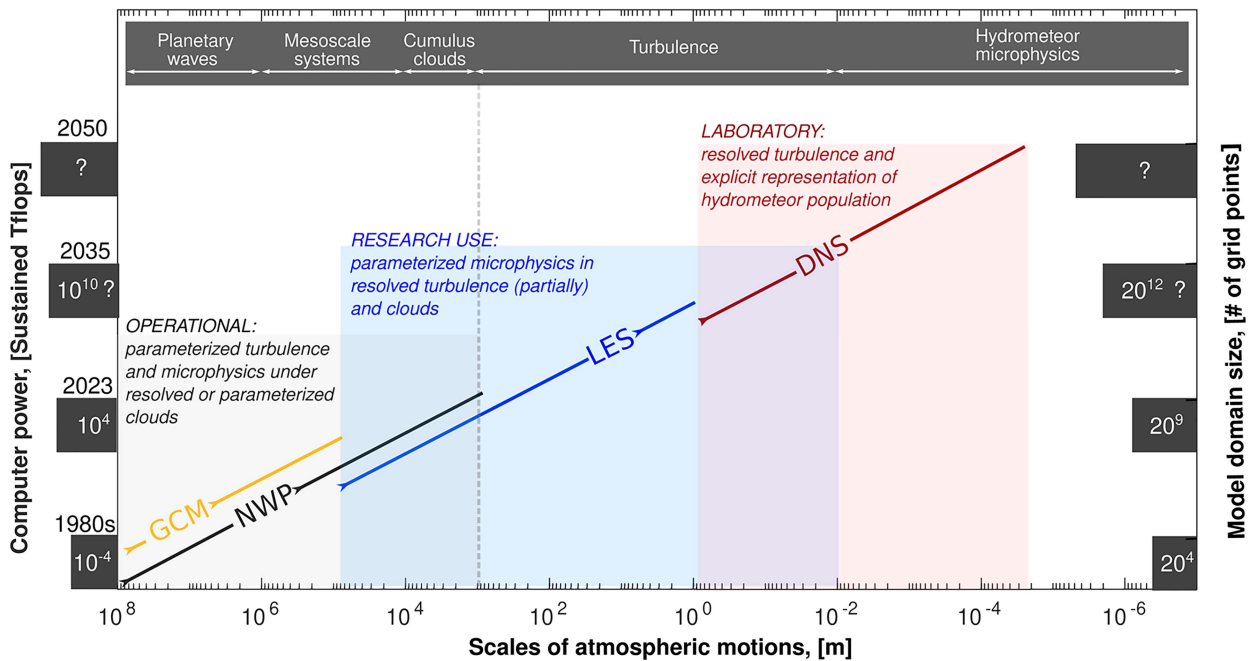


FIG. 1. Schematic of the proportionality between the resolved meteorological processes by typical mathematical systems [global circulation models (GCM), operational numerical weather prediction (NWP) systems, large-eddy simulations (LES), and direct numerical simulations (DNS)] and the computation power (infrastructure) needed to do the numerical computation of those systems. The scales of the atmospheric motions over which these meteorological processes and models operate are also shown (lower x axis and the colored boxes). The approximate past (1980s), current (2023), and projected (2035 and 2050) computer power [measured as sustained tera (10^{12}) floating-point operations per second (Tflops)] (left y axis), as well as the increasing model domain size (right y axis), mainly due to increasing model grid resolution (colored arrow lines) is based on the estimates by Schulthess et al. (2019), following a logarithmic scale.

convection and precipitation (Wilhelmson and Wicker 2001). By increasing the horizontal grid mesh of NWP systems up to hectometric scales ($O \sim 100$ m) on demand (e.g., such as in Destination Earth; https://ec.europa.eu/commission/presscorner/detail/en/IP_22_1977), some of the currently parameterized subgrid-scale processes related to shallow to deep convective transitions (e.g., cumuli- to stratocumuli-cloud organizations) are expected to be resolved, which additionally decrease the forecast uncertainty (Stevens et al. 2020). As a consequence, the shallow and deep convective parameterization schemes should be deactivated in the NWP model. However, in that regard the resolution limit is not trivial (Bryan et al. 2003). Expected improvements in weather forecasts are also related to activities in the area of generating and applying higher-resolution physiography data in NWP (e.g., from 300 m currently to 60-m horizontal resolution over Europe).

The grid size decrease in the operational NWP systems, consequently increasing the resolution of their grid mesh, unfortunately, comes with necessary large computation costs, and thus limits in the length and domain of the forecast. Running limited domain models requires information on the initial state of the system and the forcing at the boundaries of that domain. These are usually taken from the larger or global (host) models in a procedure that introduces an additional level of uncertainty. This technical aspect, together with some conceptual and more fundamental issues (Bryan et al. 2003), makes the high-resolution operational NWP simulations a highly

challenging task in the coming future. The forecast uncertainties, however, due to the parameterization of the small-scale turbulence and cloud-precipitation microphysics processes (due to the scales on which they operate) will remain present even when moving beyond hectometric-scale grid resolution (Fig. 1).

In the era of “Big Data,” scientific researchers in the field of meteorology and operational NWP system developers see the opportunity and the benefit of using LBD techniques to improve not only the short-range weather forecast (nowcasting) but the longer-range forecasts as well (Espeholt et al. 2022). LBD approaches [e.g., artificial intelligence (AI)] are already revolutionizing all components of the NWP systems, from observational analysis to postprocessing (Dewitte et al. 2021). Therefore, a logical question to be asked is, Can LBD models improve or fully replace the existing subgrid-scale parameterizations in NWP systems, thus contributing to decreasing the prediction uncertainty and improving the weather forecast?

Leaving the subject of cloud-precipitation microphysics out of further detailed discussion, this review investigates the offered alternatives to the classical *closure* approaches, which are currently used in NWP, to represent subgrid-scale turbulent processes. The offered literature study systematically elaborates on the current uncertainties and challenges in turbulence parameterization in operational NWP and seeks alternatives by using LBD approaches. Additionally, it discusses how these alternative methods advocate some of the drawbacks in the current parameterizations but also elaborates on the challenges of using these alternatives. The review

also summarizes the author's view on the future directions of turbulence representation in both research and operational NWP.

This review article offers the readership a concise synthesis of up-to-date knowledge in LBD frameworks that are or can be used to improve the computation of the Reynolds stress tensor. Thus, this review is not only useful to operational NWP developers but also for students and early-career researchers involved in NWP system development or meteorological science in general. The author finds it important to bring the advances in the LBD turbulence modeling in the field of theoretical and computational fluid dynamics to the NWP research audience and thus contribute to the NWP development. Thus, the overall goal of this review study is 1) to provide the readership with a synthesis of the current LBD modeling activities in NWP and 2) to discuss alternatives for turbulence representation in NWP by using stochastic and LBD models.

Initially in section 2, the general concept of LBD modeling is defined by providing a basic introduction of some common statistical definitions and terminology used in LBD modeling and throughout this review article (section 2a). This is then followed by a brief but informative overview of the current AI applications used in NWP in general (section 2b). Section 3 describes the uncertainties in modeling turbulence in NWP systems (section 3a) and provides an overview of modeling subgrid-scale turbulence by LBD techniques as a possible way forward toward decreasing these model uncertainties (section 3b). The requirements for the next-generation turbulence schemes in NWP are also discussed for stochastic and LBD methods (section 3c). The long-standing, and well-recognized problem of modeling the turbulent mixing in thermally stratified conditions is also a subject for discussion in section 3c. The summary of the discussion in this review study with some generalized conclusions about the perspectives toward using stochastic and LBD turbulence in NWP systems is presented in section 4.

2. Learned-by-data modeling

a. Concept and definitions

As mentioned above, it is useful to introduce the reader to the basic concept of LBD models and provide some general description of the procedure (a minimum workflow) and terminology used when constructing LBD models. The author decided to use the term “learned-by-data” instead of commonly used “data-driven” to describe classes of statistical methods that use large amounts of data and computer algorithms to derive a model. Instead, data-driven refers to broader classes of methods used in statistical science in general including, for instance, simple linear fits or interpolations. This is not too different than the LBD method discussed in this study, but rather the latter can be seen as an extension to more complex statistical relations and using computer power to generalize nonlinear relations from data. Thus, a general form of all LBD models can be written as follows (Duraismy et al. 2019):

$$\tilde{\mathcal{M}} \equiv \mathcal{M}[\mathbf{w}; \mathcal{P}(\mathbf{w}); \mathbf{c}; \boldsymbol{\theta}; \boldsymbol{\delta}; \boldsymbol{\epsilon}_{\theta}], \quad (1)$$

where $\mathcal{M}$ represents a function (a model) of an array of independent variables $\mathbf{w}$ linked with a set of algebraic or differential operators $\mathcal{P}$ and parameters $\mathbf{c}$; $\boldsymbol{\theta}$ and $\boldsymbol{\epsilon}_{\theta}$ are the data and

their uncertainty, respectively; $\boldsymbol{\delta}$ represents the *discrepancy* of the model $\mathcal{M}$ to represent the *true* state $\boldsymbol{\theta}$ and is usually defined in terms of $\boldsymbol{\theta}$ and a set of features $\boldsymbol{\eta}$ given from prior knowledge, constraints, or directly from data. The target of the data modeling is usually predicting quantities of interest $\mathbf{q}(\tilde{\mathcal{M}})$ and proper model coefficients $\mathbf{c}$. Thus, the idea of LBD models is to use data to construct *calibrated models* $\tilde{\mathcal{M}} \neq \mathcal{M}$.

1) CALIBRATED MODELS

By assuming that the model coefficients (or parameters) $\mathbf{c}$ are the main source of model uncertainty, the general calibration model can be formulated as follows:

$$\tilde{\mathcal{M}} = \mathcal{M}[\mathbf{w}; \mathcal{P}(\mathbf{w}); \tilde{\mathbf{c}}_{\mathbf{q}}], \quad (2)$$

and it can consist of an experimental configuration that is similar to the prediction target. For instance, the measured/training data can be the same as $\mathbf{q}$ but in different conditions or configurations. The prediction accuracy is given by the difference in $\mathbf{q}$ when $\tilde{\mathbf{c}}_{\mathbf{q}}$ or $\mathbf{c}$ is used. Since no information for the uncertainties in the measurement/training data is included in this model, the calibration is *naïve*, and thus it is difficult to estimate the accuracy of the predictions.

More rigorous calibration models are based on statistical (Bayesian) *inference* and include the experimental (or measurement) uncertainty $\boldsymbol{\epsilon}_{\theta}$, as well as the potential discrepancy $\boldsymbol{\delta}$ between the model prediction and the training data:

$$\tilde{\mathcal{M}} = \mathcal{M}[\mathbf{w}; \mathcal{P}(\mathbf{w}); \tilde{\mathbf{c}}_{\theta}] + \boldsymbol{\delta} + \boldsymbol{\epsilon}_{\theta}. \quad (3)$$

The statistical inference problem usually requires intensive computation and is typically solved by Markov chain Monte Carlo (MCMC) techniques. The calibrated model $\tilde{\mathcal{M}}$ is usually deterministic (only the posterior probability mode is used), and the resulting model $\mathbf{q}(\tilde{\mathcal{M}})$ is a random quantity.

Further advances in the calibrated models include explicit representation of $\boldsymbol{\delta}$ in terms of $\boldsymbol{\theta}$ and previously known (or derived from data) features or constraints $\boldsymbol{\eta}$:

$$\tilde{\mathcal{M}} = \mathcal{M}[\mathbf{w}; \mathcal{P}(\mathbf{w}); \mathbf{c}(\boldsymbol{\theta}); \boldsymbol{\delta}(\boldsymbol{\theta}, \boldsymbol{\eta}); \boldsymbol{\epsilon}_{\theta}], \quad (4)$$

allowing for a more descriptive representation of the quantities than the ones characterized by the data.

The ability of the calibrated model to represent the data is beneficial for improved prediction in terms of relevant stochastic information on the prediction uncertainty. Therefore, the main goal is to develop methods to include the calibration within the prediction system.

2) IDENTIFYING MODEL PARAMETERS

Parameters $\mathbf{c}$ of a model $\mathcal{M}(\mathbf{c})$ for given data $\boldsymbol{\theta}$ with uncertainty $\boldsymbol{\epsilon}_{\theta}$ can be identified using methods for solving inverse problems (e.g., statistical inversion). This is usually done by minimizing the difference between $\boldsymbol{\theta}$ and the model output $\mathbf{o}[\mathcal{M}(\mathbf{c})]$. The *posterior* probability of $\mathbf{c}$ given $\boldsymbol{\theta}$ is then proportional to the probability of the model before applying any data, a *prior* (IP[$\mathbf{c}$]), and the probability of the model being consistent with the data, *likelihood* (IP [$\boldsymbol{\theta}|\mathbf{c}$]):

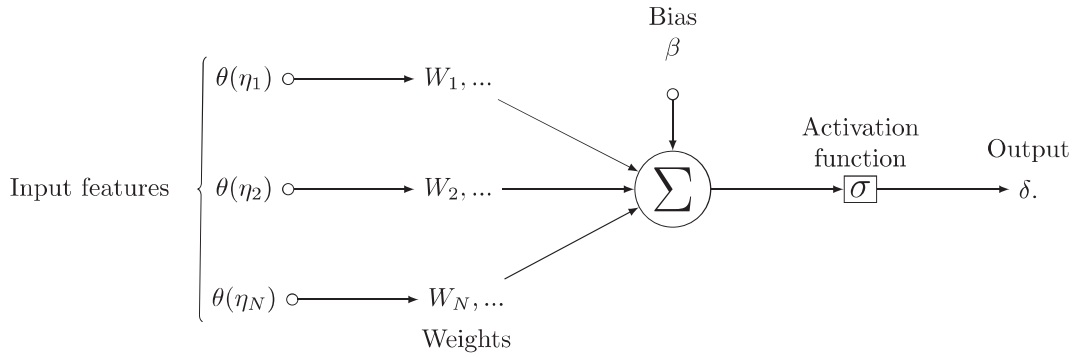


FIG. 2. Diagram of a conceptual deep neural network architecture that finds the discrepancy δ from given input data θ with characteristic features η ; $\mathbf{W}$ and β are the weight matrix and the bias vector, respectively [Eq. (6)], while the activation function σ [Eq. (7)] transforms the summed weighted input from a node into an output value to be fed to the next hidden layer or as output.

$$\text{IP}[\mathbf{c}|\theta] \propto \text{IP}[\theta|\mathbf{c}] \times \text{IP}[\mathbf{c}]. \quad (5)$$

The most commonly used approaches for solving inverse problems are the *maximum a posteriori* (MAP) and the *least squares* (LS) (Aster et al. 2018).

3) FOCUSING ON MODEL DISCREPANCY

The availability of large sets of data pushes forward the focus on finding functional forms of the discrepancy, δ in Eq. (3) instead of $\mathcal{M}$. A simple linear functional for δ , given dataset θ and vector or array of features η can be formed as follows:

$$\delta(\eta) = \mathbf{W}\eta + \beta, \quad (6)$$

where $\mathbf{W}$ and β are the weight matrix and the bias vector, respectively, which are obtained by solving optimization problems (LS for instance). This procedure is called training or *learning* of a model, and it is the basic building block in *neural networks*—a part of the *machine learning* (ML) techniques, which belongs to a more general field of AI. To represent nonlinear dependencies more complex functional forms for $\delta(\eta)$ have been developed by using intermediate layers l in Eq. (6):

$$\delta(\eta) = \mathbf{W}^{(l)} \sigma[\mathbf{W}^{(l-1)} \eta + \beta^{(l-1)}] + \beta^{(l)}, \quad (7)$$

where $l = 1, 2, 3, \dots, N$ and σ is an activation function (e.g., hyperbolic tangent) specified by the user. Equation (7) represents a composite form of a *deep neural network*, which can efficiently represent complex functional forms of the discrepancy between the model and the data. The computational procedure to solve Eq. (7) is commonly known as *deep learning* (DL). An example of a simple diagram of a deep neural network work-flow architecture that solves for the discrepancy δ in Eq. (7) given input data θ with features η can be constructed as given in Fig. 2.

b. Artificial intelligence in NWP

This section briefly summarizes and updates the review by Dewitte et al. (2021) by classifying the current AI applications and research in all aspects of the NWP, from short-range

(nowcasting) predictions to longer climate projections, and from observational analysis to postprocessing.

1) FROM NOWCASTING TO DECADEAL PROJECTIONS

For short-range forecasts of up to a few hours (nowcasting), and when focusing on specific meteorological quantities, the extrapolation of satellite and radar observations can provide more accurate prediction relative to the standard forecast by NWP (Haiden et al. 2011). Many of the recent short-range NWP systems, however, already use ML to enhance their capabilities (e.g., Flora et al. 2021). For precipitation, the use of DL techniques in that respect not only further improves the nowcasting (Shi et al. 2015), but also decreases the degree of computation (e.g., Kaae S nderby et al. 2020; Prudden et al. 2020; Ravuri et al. 2021). A recent study showed that a weather prediction system based on DL, which was *supervised* by an atmospheric simulation (MetNet-2), outperforms the standard physics-based nonprobabilistic NWP, as well as probabilistic ensemble NWP models, even when applied for a 12-h forecast (Esp holt et al. 2022). It was shown that the subseasonal-to-seasonal forecast can also benefit from using ML-based models (e.g., He et al. 2020; Weyn et al. 2020; Weyn-Vanhentenryck 2020; Hwang et al. 2019), especially when effects of processes such as Arctic amplification on midlatitude to subarctic weather are not represented in current NWP systems and climate models (Cohen et al. 2019). For longer (decadal) predictions, a convolutional neural network (CNN) model was “trained” first on historical simulations and subsequently on observations (reanalysis) to successfully predict the El Ni o–Southern Oscillation (ENSO) events up to 1.5-yr lead time, and thus outperform the state-of-the-art SINTEX-F dynamical forecast system (e.g., Ham et al. 2019). Additionally, DL-based models (CNN type) have shown a potential to predict extreme weather events (Chattopadhyay et al. 2020a,b), and provide warnings by clustering extreme weather patterns in climate data (Chattopadhyay et al. 2020a).

2) FROM OBSERVATIONS TO POSTPROCESSING

The operational weather forecasts and climate monitoring are heavily based on meteorological observations. A typical workflow

in processing the meteorological observation in NWP systems integrates information retrieval, quality control, bias correction, and data assimilation (DA). The development and the usage of complex and more accurate DA techniques (e.g., 4D-Var; Rabier 2005) enormously improve the weather forecast, but they require long-term (several years) of dedicated human expertise and large computation resources. The concept of this type of DA technique, where a cost function minimization is used to improve the forecast, is very similar to the concept of DL where a cost function minimization is used to “teach” or “learn” the model. It has been shown, however, that the difference and the benefit of using DL techniques to combine information retrieval and data fusion is not only in their precision but also in their robustness in terms of construction and applicability (e.g., Moraux et al. 2019; Boukabara et al. 2020); for instance, the Multi-Instrument Inversion and Data Assimilation Preprocessing System (MIIDAPS-AI) (Boukabara et al. 2020) is an AI-based system, which integrates and analyses images and soundings from several satellites. Based on Dewitte et al. (2021), the MIIDAPS-AI was built in less than a year by two people and showed to be much faster (by a factor of 100) than a similar non-AI system.

Deep learning based on convolutional neural network (U-Net type; Ronneberger et al. 2015) techniques have also been successfully applied (Chen et al. 2020) to improve the classical postprocessing model output routine of the European Centre for Medium-Range Weather Forecasts (ECMWF)–Integrated Forecasting System (IFS) (Glahn and Lowry 1972), resulting in decreased temperature bias for 1° on average for a 10-day forecast. Similar DL architecture has been used to improve the postprocessed cloud cover in the French global weather system ARPEGE (Dupuy et al. 2021). Recently, attempts have been made to apply DL to postprocessing ensemble weather forecasts (e.g., Grönquist et al. 2021).

As previously mentioned in the introduction, improving the representations of the nonresolved processes in the NWP systems is a crucial task for decreased model bias and thus improved weather forecast and climate projection. In that respect, Schneider et al. (2017) provided a “learning” framework of using observations and high-resolution simulation to improve convection and cloud parameterization. O’Gorman and Dwyer (2018) showed how the commonly used in NWP systems moist convection schemes can be used to inform ML model to improve climate projections and extreme events. ML approaches have also been used together with an output from a near-global cloud-resolving model in developing a unified physics parameterization (Brenowitz and Bretherton 2018), or in representing the ill-known parts of a numerical atmospheric model (Brajard et al. 2019). Recently, several attempts have been made to use purely LBD cloud and precipitation parameterization (Brenowitz and Bretherton 2018; Gentine et al. 2018; Rasp et al. 2018). These ML-based methods showed a potential to replace the traditional (physics based) cloud and convection precipitation schemes in global models but also raised a criticism (Karpatne et al. 2017), mainly related to the lack of interpretability and the physical sense of the results, as well as the poor performance for conditions outside the training data. Given the lack of standardized reference training data (e.g., there is no benchmark cloud and precipitation data), a more generalized applicability of the purely

LBD approaches remains highly uncertain in the near future. Morrison et al. (2020a) instead proposed a hybrid approach in which physical and statistical models are combined with observation in a Bayesian observationally constrained statistical-physical scheme (BOSS). BOSS uses Bayesian inference (see section 2a) to constrain and improve existing physics-based process rate in the cloud-precipitation microphysics parameterization, as guided by observations from bulk measurements of cloud and precipitation. However, further development and emulation of physics-guided and observationally constrained (including ML techniques) cloud-precipitation schemes is a highly desirable task, together with support for new laboratory and observational campaigns Morrison et al. (2020b).

In conclusion, although several studies aim at providing a weather forecast by fully replacing the atmospheric dynamic and physics-based NWP systems with AI-based approaches (e.g., Kim et al. 2020; Zheng et al. 2020), best weather forecast results are achieved when using hybrid physics-guided and LBD NWP (e.g., Li et al. 2021; Wang et al. 2019; Espenholt et al. 2022).

3. Modeling turbulence

a. Background and uncertainties

From a mathematical point of view, turbulence can be described as a property of the numerical solution of the averaged Navier–Stokes (NS) equations and is a direct consequence of the averaging procedure. In NWP systems, where averaged NS are commonly used, turbulence represents the transfer of momentum, energy, and matter within the discretized volume of air due to small-scale fluctuations in vector and scalar fields. The numerical approximation of the averaged NS equations, however, always leads to uncertainties in their solution, and thus in predictions in general. This is due to the assumptions introduced in several stages in the numerical approximation procedure (Duraiamy et al. 2019). Because of their importance, a brief summary of the uncertainties caused by these assumptions is provided in what follows.

The first level of uncertainty (UL1) is introduced already in the spatiotemporal averaging procedure, indicated by angle brackets, of the governing equations, $\langle \mathcal{N}S(\cdot) \rangle$, resulting in an undetermined system, $\langle \mathcal{N}S(\cdot) \rangle \neq \mathcal{N}S(\langle \cdot \rangle)$, which requires modeling assumptions in order for the system to be closed: the well-known turbulence *closure* problem. This practically means that each of the infinite instantaneous vector (and scalar) fields are within the average field overall, but they can have dynamically different pathways—information that is missed in the averaging process. By definition, this type of inadequacy is irrecoverable.

The second level of uncertainty (UL2) is related to the selection, or the design, of the functional form of the closure model that is invoked to represent “the unknown” in the averaging process:

$$\langle \mathcal{N}S(\cdot) \rangle = \mathcal{N}S(\langle \cdot \rangle) + \mathcal{M}(), \quad (8)$$

where $\mathcal{N}S(\langle \cdot \rangle)$ represents the NWP grid-scale (and larger) processes that are resolved while $\mathcal{M}()$ represents the processes that are smaller than the NWP grid [subgrid-scale (SGS)]

processes; e.g., turbulence]. For instance, the SGS turbulent processes in high-resolution NWP systems are mainly defined by the Reynolds stress tensor $\mathcal{M}() = \nabla \tau$. Consequently, there are uncertainties related to the assumptions made for the functional form of the Reynolds stress tensor. Thus, there are several models for τ , ranging from more simple and computationally inexpensive (e.g., K -theory, or linear eddy viscosity models) but less accurate to more complex and computationally more intense but more accurate [e.g., large-eddy simulations (LES)]

The third level of uncertainty (UL3) is introduced in the functional within the model for a given set of selected averaged variables $\mathbf{w}$. The algebraic or differential form of these models typically represents some physical processes (e.g., convection, waves, boundary dynamics). For instance, models based on the turbulent kinetic energy (TKE) equation are the most commonly used in NWP systems currently.

The fourth level of uncertainty (UL4) is related to parameters, coefficients, or numbers $\mathbf{c}$ used in the model. In the above-mentioned TKE-equation, coefficients such as C_μ , Richardson critical number, Prandtl ($\mathbf{Pr}$), and Schmidt ($\mathbf{Sc}$) numbers, are examples of these types of uncertainties and are usually derived theoretically (asymptotic theory) or empirically.

The prediction of the NWP system for a quantity of interest $\mathbf{q}$ is schematically given as

$$\mathbf{q} = \mathbf{q}\{\mathcal{N}S(\langle \rangle); \mathcal{M}[\mathbf{w}; \mathcal{P}(\mathbf{w}); \mathbf{c}]\}, \quad (9)$$

and represents a balance between the sinks and sources of the atmospheric dynamics and thermodynamics due to the large-scale transport processes $[\mathcal{N}S(\langle \rangle)]$, which are expected to be resolved, and the smaller-scale (SGS) transport processes $[\mathcal{M}()]$, which are parameterized. It is then obvious that by increasing the grid resolution of the NWP system, smaller scales of atmospheric motion will be resolved, and only the smallest should be parameterized. But this comes with at least two important consequences. The first one is related to the required limitation of the domain size due to the increased computational needs for high-resolution simulations (Fig. 1). This leads to uncertainties in initial and boundary forcing in limited-area NWP systems, which are usually taken by global models. It has been shown that the uncertainties in high-resolution limited-area NWP systems due to the initial and boundary forcing can be decreased if a chain of models with similar constructions (numerical discretization and parameterization schemes) for $\mathcal{N}S(\langle \rangle)$ and $\mathcal{M}()$ are used (Schemann and Ebell 2020). The second consequence is more conceptual and is related to the uncertainties due to the link between the resolved and parameterized processes. For instance, if the currently constructed $\mathcal{M}()$ includes models of deep (and shallow) convection, it is not a trivial question how $\mathcal{M}()$ should look like when convection is resolved by $\mathcal{N}S(\langle \rangle)$ in high-resolution models (Bryan et al. 2003). Concerning turbulence and convection, a possible way forward is the use of scale-adaptive (or scale-aware) turbulence schemes and closures (e.g., Larson et al. 2012; Bogenschütz and Krueger 2013; Belochitski et al. 2016; Zanna et al. 2017; Olson et al. 2019a) in which the transfer of momentum, energy, and matter between $\mathcal{M}()$ and $\mathcal{N}S(\langle \rangle)$ is gradual following some previously established “resolution dependent” function. A more general review, however, of the recent

development in purely turbulence closure modeling has been provided by Durbin (2018).

b. Predicting turbulence by using LBD techniques

In the era of largely increased high-frequency (temporal and spatial) data availability, mainly from Direct Numerical Simulations (DNS) (Coupland 1993; Li et al. 2008), but also from measurement experiments (Comte-Bellot and Corrsin 1966), LBD approaches enable for a more comprehensive fusion of data and models resulting in attempts for improved turbulence closure (Tracey et al. 2015; Weatheritt and Sandberg 2017; Ling et al. 2016; Wang et al. 2017; Wu et al. 2018; Beck et al. 2018). Since the dominant cause of discrepancies in modeling turbulence in averaged NS is the tensor stresses (Oliver and Moser 2009), this section reviews the knowledge of using stochastic and LBD approaches to improve the prediction of the Reynolds stress in Reynolds averaged NS (RANS). The main idea in all presented studies here is to derive a model $\tilde{\mathcal{M}} = \mathcal{M}(\boldsymbol{\theta})$, which explicitly includes the predefined (or LBD) discrepancy function δ .

1) CALIBRATING COEFFICIENTS IN THE REYNOLDS STRESS

Several studies have focused on calibrating RANS model coefficients by integrating rigorous statistical (Bayesian) inference and a variety of turbulence data (e.g., DNS) while assuming simple discrepancy function in the process of prediction (Oliver and Moser 2011; Cheung et al. 2011; Edeling et al. 2014a,b; Lefantzi et al. 2015; Ray et al. 2016, 2018). Data from different flow types have been used in the above studies; for instance, Oliver and Moser (2011) and Cheung et al. (2011) have used DNS from plane channel flow, Edeling et al. (2014a,b) have used data from a number of wall-bounded flows, while Lefantzi et al. (2015) and Ray et al. (2016, 2018) have focused on jet-in-cross-flow cases. These studies have shown the added value of using data and statistics to calibrate parameters in the turbulence model and thus improve RANS, but they did not focus on directly predicting the Reynolds stress for more general and complex classes of flows.

2) PREDICTING THE REYNOLDS STRESS

It was demonstrated that the combination of machine learning and statistical inference can be used to derive LBD closures for turbulence modeling (Duraismy et al. 2015; Singh et al. 2017a,c; Zhu and Dinh 2020). While the previous studies focused on specific flow geometry and conditions (in both prior, and when building the posterior probability distribution), Tracey et al. (2013) developed a more generalized approach by using ML to reconstruct discrepancies in the anisotropy tensor of the Reynolds stress. Further development provided a strategy to predict the discrepancy not only in the anisotropy but also in the magnitude and the orientation of the Reynolds stress tensor (Wang et al. 2017). Wang et al. (2017) used the ML random forest method to build a functional mapping of the mean flow input features to the output. This led to the development of the Physics-Informed Machine Learning (PIML) framework (Wang et al. 2017) that is able to learn the functional form of Reynolds stress discrepancy in RANS simulations based on available data

from several flow geometries, as well as that “the learned discrepancy function can be used to improve Reynolds stresses in different flows where data are not available.” Recently, the latter work has been extended toward a comprehensive PIML framework for predictive turbulence modeling, including learning the Reynolds stress discrepancy function, predicting Reynolds stresses, and propagating to mean flow fields (Wu et al. 2018). The latter is an extremely important feature of the PIML turbulence model since it takes into account the nonlinear feedback between the main flow and the small-scale turbulence in the training process [Eq. (8)].

Weatheritt and Sandberg (2016, 2017) have developed a novel approach to predict the Reynolds stress by combining the traditional turbulence models and LBD methodology. Their algorithm is based on symbolic regression and gene expression programming and the end product is an algebraic turbulence model. As such, this model can be directly used in RANS, as the traditional physics-based turbulence models.

3) MODELING TURBULENCE CLOSURES BEYOND RANS

The combination of advanced turbulence modeling (e.g., DNS), statistical inference, and uncertainty quantification, as well as advances in ML (e.g., deep neural networks) algorithms, opened a new path in the research of LBD turbulence models (Beck et al. 2018). Increasing attention in LBD modeling of subgrid-scale stress in LES using artificial neural networks and DNS already have shown preliminary success in this field (Vollant et al. 2014; Gamahara and Hattori 2017; Vollant et al. 2017; Beck et al. 2018).

c. Next-generation turbulence schemes in NWP

One of the main goals of this paper is to inform readers how to use LBD methods to accelerate developments in NWP. Thus, a clear understanding of the requirements for the next-generation turbulence schemes is required.

1) SCALE-AWARE TURBULENCE

As mentioned before the focus on hectometric scales in NWP is becoming more in-demand, but it is naive to think that any contemporary turbulence scheme can be designed for any particular scale and have a bright future. With the increased complexity of atmospheric composition (aerosols and chemistry) also in demand, as well as the use of variable-grid-resolution dynamical cores (e.g., Skamarock et al. 2012), relatively coarse-resolution applications will be in demand for the foreseeable future. The Model Prediction Across Scales (MPAS; <https://mpas-dev.github.io/>) is an example of one such code that is now being considered for operational NWP in the United States. This is especially becoming truer as the climate and weather communities become more closely intertwined (Belušić et al. 2020). In that respect, scale-aware turbulence schemes are the only schemes that have a future (Honnert et al. 2011; Larson et al. 2012; Bogenschutz and Krueger 2013; Belochitski et al. 2016; Zanna et al. 2017; Olson et al. 2019a, among others). Any specific problems with LBD schemes associated with scale-awareness is yet to be resolved (Christensen and Zanna 2022). An example of such grid-dependent issues was admitted by Lagerquist et al. (2021)

for their radiation simulator and was discussed by Rasp and Thuerey (2021) as a lingering challenge for achieving good results at both climate and NWP scales. Moreover, the representation of three-dimensional turbulence is required for all truly scale-aware schemes. The importance of horizontal diffusion can be crude at coarse resolutions, but at higher resolutions, the horizontal stresses can be just as important as the vertical stresses (Goger et al. 2018). This is especially the case in complex terrain (Olson et al. 2019b). The recently developed LBD multiscale modeling frameworks for subgrid parameterizations in climate models (Yuval and O’Gorman 2020; Otness et al. 2023) could be a step farther in the right direction. Yuval and O’Gorman (2020) used the random forest to learn a subgrid parameterization from the coarse-grained output of a three-dimensional high-resolution idealized atmospheric model and showed insights into parameterization performance across length scales. By training NNs, Otness et al. (2023) showed proof of concept experiments that illustrate the potential advantages of decomposing the subgrid forcing problem into one across scales with a potential to represent multiscale aspects of the prediction problem explicitly. Whether these approaches can be applied to NWP systems will ultimately remain to be seen.

Related to scale awareness, the representation of nonlocal transport is extremely important for any contemporary turbulence scheme (e.g., Ādurán et al. 2018). Currently, the majority of physics-based and LBD turbulence parameterizations are based on a single-column approach. Thus, they only use information from single atmospheric columns, which is not ideal, because atmospheric processes, such as organized convective systems, can often cross irregularly through multiple grid boxes and involve nonvertical circulations. Therefore, some representation of nonlocal mixing should be considered for grid spacing down to 500 m (Angevine et al. 2020). On the side of the LBD modeling, it has been demonstrated that training neural networks by using nonlocal inputs improved offline prediction of a range of subgrid processes for atmospheric conditions associated with midlatitude fronts and convective instability Wang et al. (2022). These results imply that using nonlocal variables (horizontal and vertical wind divergences) and vertical velocity as inputs could improve the performance of ML parameterizations. However, the performance of this LBD approach remains a challenge in online simulations.

2) MOIST PHYSICS IN TURBULENCE SCHEMES

The strong connection between boundary layer turbulence and clouds in shallow-cumulus and stratocumulus regimes requires the inclusion of moist physics in any contemporary turbulence scheme. For physics-based schemes the trend in NWP development has been toward either the eddy diffusivity-mass flux (e.g., Suselj et al. 2019) or higher-order closure frameworks (e.g., Ādurán et al. 2018) for a more complete representation of moist-turbulent processes. Any LBD efforts must include extensive training in moist-turbulent regimes if they are to become competitive with the physics-based schemes. In that respect, a two-energy turbulence scheme (Ādurán et al. 2018) has been coupled to a simplified stochastic model (Bogenschutz and Krueger 2013) to achieve a more universal representation of the cloudy regimes, and thus

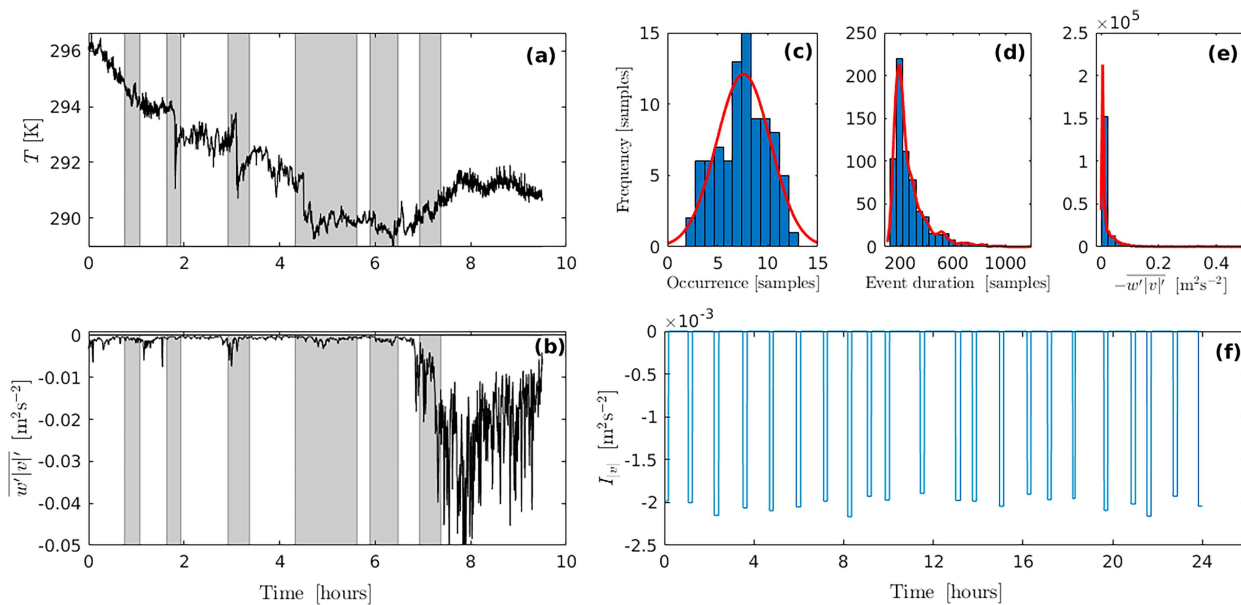


FIG. 3. Procedure of building a simple LBD turbulent intermittency model, showing application of the turbulent event detection (TED) method (Kang et al. 2014; Kang 2015) on eddy-covariance (a) temperature and (b) momentum data from Lindenberg observatory at 90-m height; estimating of distributions of selected intermittent characteristics over 8 h (night) for (c) frequency of occurrence of intermittent events, (d) length of intermittent events, and (e) magnitude of intermittent events [from measured signals (histograms), with fitted probability density functions (PDFs) (red solid lines)]; and (f) the build of turbulent intermittent forcing as a function of the intermittent characteristics.

improve the model performance in better separating between shallow convection and stratocumulus cases (Dúran et al. 2022).

3) INTERMITTENT TURBULENCE—A SPECIFIC ISSUE

According to Kolmogorov's turbulence theory (Kolmogorov 1941), turbulent mixing is governed by eddies with characteristic turbulent scales: velocity and length (diameter). These turbulent scales in the atmosphere span from very large eddies (meso-scale) that produce energy to very small eddies (microscales) that dissipate the energy. The interaction among the eddies at various scales passes energy from the larger eddies gradually to the smaller ones. This process is known as the turbulent energy cascade. This principle allows for a relatively simple representation and quantification of the turbulent mixing and is widely applied in atmospheric science. This Kolmogorov's principle, however, has some limitations, since it was developed under certain assumptions. Two of these conditions are the stationarity and homogeneity of the flow. Latest studies (Mahrt 2014), however, suggested that in strongly stratified atmospheric boundary layers (ABL), which decouple from the large-scale forcing (due to thermal stratification), the energy can be transferred from larger eddies to smaller ones directly instead of gradually, causing sporadic and intermittent energetic disorder locally. This process in the ABL causes intermittent mixing locally, or *intermittent turbulence*. In such cases, the classical turbulent energy cascade principle is insufficient in quantifying the turbulent mixing and needs modification. One of the difficulties in representing the intermittent turbulence in the ABL by deterministic approaches is its highly random distribution in time and space.

Indeed, NWP systems have difficulties in representing the intermittent character of the turbulent transport, and thus often underestimate the turbulent intensity in the ABL and near the surface (Belušić and Güttler 2010). This is especially problematic when representing the stably stratified flow (Sandu et al. 2013), where the intermittent turbulent transport becomes an important process to maintain the low-turbulent intensity flow, thus preventing unrealistic laminar flow (Acevedo and Fitzjarrald 2003). Consequently, important questions to be answered are the following (Mahrt 2014): Is intermittent turbulence a result of external forcing by motions with mesogamma and sub-meso-gamma scales? Can this scale interaction be estimated by observations? Last, can these intermittent effects be represented by LBD and stochastic models?

Recent studies on scale separation (Vercauteren and Klein 2015), by using LBD clustering techniques, demonstrated that the transfer of energy in increased atmospheric thermal stability conditions (Vercauteren et al. 2016, 2019) is due to larger sub-meso-motion without energy cascade. The turbulence in these increased thermal stability conditions is not in statistical equilibrium and the flow is inhomogeneous and nonstationary (Vercauteren et al. 2019). This is exactly why the classical turbulence theory seems not to work in conditions with increased thermal stability; it relies on homogeneity and stationarity.

Shapkaliyevski and Vecauteren (2019) proposed and presented a simple, and easy-to-apply in atmospheric models, LBD strategy to account for the intermittent turbulence (Fig. 3). The main assumption of this LBD framework is that the fluxes calculated based on the high-resolution (temporal) measurements [eddy

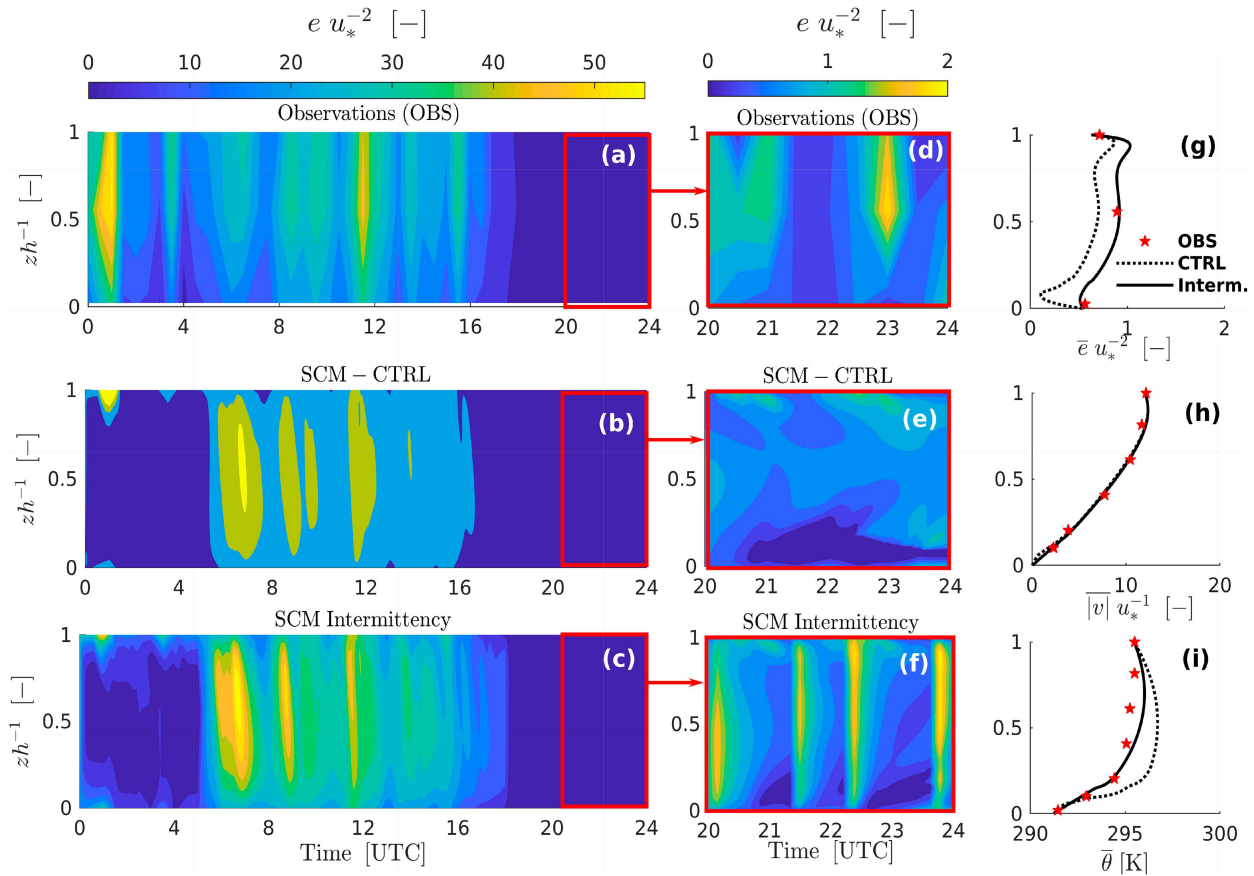


FIG. 4. (a)–(f) Contour plots of the instantaneous observed and modeled (with and without intermittent forcing) diurnal cycle of the TKE over Lindenberg observatory; also shown is the consequent effect of the intermittent forcing on the (g) prognostic TKE, (h) wind speed, and (i) temperature statistics during stable stratification [averages over the red-outlined rectangle in (a)–(c)].

covariance (EC)] carry the information of the small-scale turbulence and larger-scale (non-turbulent) motions ($\overline{w'\phi'_{\text{measured}}}$). Thus, one can separate (cluster) these two as resolved and parameterized ($R + P$) by the model ($\overline{w'\phi'_{R+P}}$), and “missed” by the model ($\overline{w'\phi'_{\text{residual}}}$). The “residual” part is expected to represent the intermittent turbulence I that should be accounted for in the atmospheric models ($\overline{w'\phi'_I} = \overline{w'\phi'_{\text{residual}}}$):

$$\overline{w'\phi'_I} = \overline{w'\phi'_{\text{measured}}} - \overline{w'\phi'_{R+P}}. \quad (10)$$

A turbulence event detection method (TED) (Kang et al. 2014; Kang 2015) has been applied to EC time series of temperature and momentum flux to separate and cluster between turbulent and nonturbulent (intermittent) motions (Figs. 3a,b). From a relatively large number of time series probability density functions (PDF), the characteristics of the intermittent signal (frequency of occurrence of the events, length of the events, and their magnitude) have been derived (histograms in Figs. 3c–e). Then, by fitting the PDFs (read solid lines in Figs. 3c–e) an *approximant* of the intermittent turbulence signal has been constructed (Fig. 3f).

To illustrate the simplicity of implementation in numerical schemes and to verify its effects on atmospheric boundary layer state, this framework has been applied in a single-column model (SCM) with one-and-a-half equation (e – l) turbulence closure to

account for the intermittent turbulence in the source term in the TKE (e) prognostic equation. The developed simple LBD framework to account for intermittent turbulence effects in large-scale atmospheric models showed potential to cover the underestimated turbulent kinetic energy in the thermally stratified ABL Fig. 4.

The latest advances in this field toward developing a more generalized LBD model for the intermittent turbulence in stably stratified atmospheric boundary layers have been conducted by Boyko (2022).

4. Challenges and perspectives

The present study reviews the current state of LBD application in numerical weather prediction (NWP) and seeks alternatives to the classical parameterizations of nonresolved processes in NWP with a focus on state-of-the-art developments in LBD turbulence modeling.

In summary, four general directions in the development of LBD turbulence models can be categorized: 1) models that use data to develop algebraic Reynolds stress based on symbolic regression (Weatheritt and Sandberg 2016, 2017); 2) models that use data and machine learning techniques to predict discrepancies in the source term in the turbulent

transport equations (Singh et al. 2017c,b), 3) models that use machine learning to directly predict Reynolds stress (Ling et al. 2016; Wang et al. 2017; Wu et al. 2018, 2017), and 4) models that use machine learning and direct numerical simulations to predict the subgrid-scale stress in LES (Vollant et al. 2017; Beck et al. 2018). The question of course is, Which of these LBD turbulence model developments are suitable for application in NWP?

The release of NWP systems operationally is followed by rigorous and systematic procedures in their research and development phase. These procedures can be summarized in terms of *applicability* and *interpretability*. For instance, the applicability of a new model is expected to (among other requirements): (i) improve the current model performance (or at least not to show worst performances), (ii) demonstrate robust capabilities in the complex range of atmospheric flows and conditions, (iii) show stable numerical implementation within the complex NWP system, (iv) to have acceptable computation requirements (balance between the forecast improvement overall due to its implementation, and computation power to do so). These are reasonable general criteria for most schemes, but additional specific requirements may be applied for certain schemes as well. The interpretability is more important in the research and development phase of the NWP system rather than in its operational use and refers to the ability of the model to be “interpreted” or physically understood (i.e., what the model does? What it improves and why?). In this section, the applicability and interpretability of the reviewed LBD turbulence models are discussed, as well as the challenges and perspectives toward their implementation in NWP systems.

a. Applicability

To the best of the author’s knowledge, the majority of the LBD turbulence models presented here are in their infancy and show satisfactory precision and predictive capabilities only for flows close to the training flows. There is a challenge when applying these models in real atmospheric flows in which various and complex flow configurations and conditions could appear at once (e.g., boundary layer flow over a rough surface (forest, cities, mountains) affected by nocturnal low-level jet and low-level stratocumulus clouds). When using training data with specific flow characteristics, it was demonstrated, however, that the model was able to predict an unknown flow with similar characteristics (Wang et al. 2017; Wu et al. 2018). Moreover, the robustness and the predictive capabilities of these turbulence models can be improved by using more diverse training data. As the name suggests, the key ingredient of the LBD models is the availability, and the quality of the data, based on which the model is expected to be developed. In that respect, the large amount of data from different NWP systems that is output daily can serve both, as training data of complex flows, as well as for validation of the model’s robustness; especially data from a chain of simulations with nested grids with increasing resolutions [from global (GCM) to beyond hectometric (LES), (e.g., Schemann and Ebell 2020)] can be of specific interest, since the cascade of energy from the large (global) circulation eddies to the smaller turbulent eddies (10–100 m in length) is systematically accounted

for. Using these data to inform and predict more computationally acceptable subgrid-scale stress models of the LES type [the fourth development direction, (Vollant et al. 2017; Beck et al. 2018)] could potentially increase the applicability of those models in the high-resolution NWP.

Another different, yet practically useful view is to consider the LBD approach as an augmentation of traditional turbulence models by adding flow-specific corrections into them. In that case, the LBD model can automatically turn on/off itself for unknown flow characteristics and reverts back to the traditional turbulent model instead. It was demonstrated that this could be achieved by providing a priori assessment of the prediction confidence (Wu et al. 2017) through automatically measuring the “distance” between different flows in a statistically rigorous and systematic way. Adding strategic constraints can also help to limit unphysical solutions by enforcing conservation laws (Beucler et al. 2021).

Most of the comprehensive LBD models require very high computing resources in the learning phase. However, they are much more computationally efficient in their actual prediction run when compared with, for instance, using LES runs (Beck et al. 2018). Therefore, it can be expected that the computation requirements for integrating and using these models in NWP systems would be generally satisfied.

b. Interpretability

Increased challenges of using purely LBD models in NWP are expected because of the lack of interpretability due to the “black box” nature of machine learning. It seems therefore that the group of LBD models that use the symbolic regression approach [the first group (Weatheritt and Sandberg 2016, 2017)] is more suitable than the other groups (groups 2–4), since they aim at finding Reynolds stress models in analytic forms, and thus being more interpretable. This is true when constructing a simple analytical expression of the Reynolds stress based on one specific class of flow. However, for a given training data with the complex nature of the atmospheric flow, the resulting analytical expression with a larger number of terms may still be too difficult to interpret. On the other hand, using the purely LBD models in combination with physically based turbulence closures, for instance, “learning” parameters within the model, would help in addressing both (interpretability and applicability) challenges. The interpretability of machine learning models (groups 2–4) can be improved by applying regularization methodology (James et al. 2013), and thus constructing simple architecture that can be more easily interpreted.

c. Key limitations and future directions

Based on the analysis in this literature survey, and to help bridge the disconnect between the research (engineering) community and the NWP community in developing LBD turbulence models, the following concluding notes and future directions can be revealed:

- Stochastic and LBD methods are already included in many parts of the weather forecast. While applications like post-processing seem like impressive successes for LBD applications, general forecasting still seems far from competitive for

all forecast ranges, all resolutions, and all variables. Yet, the stochastic and LBD modeling approaches are useful tools for improving parameterizations in numerical weather prediction systems. Thus, their most plausible usage in NWP systems in the near future is expected when used together with the physics-based turbulence models.

- Increasing the flow complexity in the “learning” process of the LBD turbulence models by including a system of NWP outputs with different grid resolutions could potentially increase the capability and applicability of these models for simulating the full range of atmospheric turbulent flows. However, while the turbulence modeling benefits from the high-fidelity LES and DNS turbulence data, there are no standardized benchmark data that take into account the full range of atmospheric scales and phenomena. Therefore, actions for producing standardized benchmark training data are required before emulating physics-based turbulence schemes with stochastic and LBD models.

These directions could help to build on the message that training for all regimes and relevant variables is crucial for general NWP applications. Building the AI architecture toward comprehensive full-range NWP emulation, however, will remain an open question for the foreseeable future.

... the new computer model is still not as good at predicting rain as some clever mathematical formulas that were developed 20 years ago. Maybe we first have to teach the computer model more about calculus before it can learn to predict rain. (Seifert and Rasp 2020)

Acknowledgments. The author acknowledges A Consortium for Convection-Scale Modeling Research and Development (the ACCORD consortium) for supporting the research that helped to create this review study. Moreover, the author expresses gratitude to the three anonymous reviewers for their valuable comments and suggestions during the review process; special acknowledgments go to Reviewer 2 for the suggestions and contributions in designing section 3c. The author declares that there are no conflicts of interest.

Data availability statement. Any data of public interest used in this review can be freely accessed upon request using the author’s contact details.

REFERENCES

- Acevedo, O. C., and D. R. Fitzjarrald, 2003: In the core of the night-effects of intermittent mixing on a horizontally heterogeneous surface. *Bound.-Layer Meteor.*, **106**, 1–33, <https://doi.org/10.1023/A:1020824109575>.
- Angevine, W. M., J. Olson, J. J. Gristey, I. Glenn, G. Feingold, and D. D. Turner, 2020: Scale awareness, resolved circulations, and practical limits in the MYNN-EDMF boundary layer and shallow cumulus scheme. *Mon. Wea. Rev.*, **148**, 4629–4639, <https://doi.org/10.1175/MWR-D-20-0066.1>.
- Aster, R. C., B. Borchers, and C. H. Thurber, 2018: *Parameter Estimation and Inverse Problems*. 3rd ed. Elsevier, 404 pp.
- Beck, A. D., D. G. Flad, and C. Munz, 2018: Deep neural networks for data-driven turbulence models. arXiv, 1806.04482v3, <https://doi.org/10.48550/arXiv.1806.04482>.
- Belochitski, A., S. K. Krueger, S. Moorthi, P. Bogenschutz, and R. Pincus, 2016: Applying an economical scale-aware pdf-based turbulence closure model in NOAA NCEP GCMS. *2016 Fall Meeting*, San Francisco, CA, Amer. Geophys. Union, Abstract A33D-0264.
- Belušić, D., and I. Güttler, 2010: Can mesoscale models reproduce meandering motions? *Quart. J. Roy. Meteor. Soc.*, **136**, 553–565, <https://doi.org/10.1002/qj.606>.
- , and Coauthors, 2020: HCLIM38: A flexible regional climate model applicable for different climate zones from coarse to convection-permitting scales. *Geosci. Model Dev.*, **13**, 1311–1333, <https://doi.org/10.5194/gmd-13-1311-2020>.
- Beucler, T., M. Pritchard, S. Rasp, J. Ott, P. Baldi, and P. Gentile, 2021: Enforcing analytic constraints in neural networks emulating physical systems. *Phys. Rev. Lett.*, **126**, 098302, <https://doi.org/10.1103/PhysRevLett.126.098302>.
- Bogenschutz, P. A., and S. K. Krueger, 2013: A simplified pdf parameterization of subgrid-scale clouds and turbulence for cloud-resolving models. *J. Adv. Model. Earth Syst.*, **5**, 195–211, <https://doi.org/10.1002/jame.20018>.
- Boukabara, S., E. Maddy, N. Shahroudi, R. N. Hoffman, T. Connor, S. Upton, and J. E. Ten Hoeve III, 2020: Artificial intelligence (AI) techniques to enhance satellite data use for nowcasting and NWP/data assimilation. *19th Conf. on Artificial Intelligence for Environmental Science*, Boston, MA, Amer. Meteor. Soc., 2B.3, <https://ams.confex.com/ams/2020Annual/webprogram/Paper364227.html>.
- Boyko, V., 2022: Data-driven modeling of intermittent turbulence in the stably stratified atmospheric boundary layer. Ph.D. thesis, Freien Universität Berlin, Berlin, Germany, 225 pp., <https://doi.org/10.17169/refubium-35137>.
- Brajard, J., A. Charantonis, and J. Sirven, 2019: Representing ill-known parts of a numerical model using a machine learning approach. arXiv, 1903.07358v1, <https://doi.org/10.48550/arXiv.1903.07358>.
- Brenowitz, N. D., and C. S. Bretherton, 2018: Prognostic validation of a neural network unified physics parameterization. *Geophys. Res. Lett.*, **45**, 6289–6298, <https://doi.org/10.1029/2018GL078510>.
- Bryan, G. H., J. C. Wyngaard, and J. M. Fritsch, 2003: Resolution requirements for the simulation of deep moist convection. *Mon. Wea. Rev.*, **131**, 2394–2416, [https://doi.org/10.1175/1520-0493\(2003\)131<2394:RRFTSO>2.0.CO;2](https://doi.org/10.1175/1520-0493(2003)131<2394:RRFTSO>2.0.CO;2).
- Chattopadhyay, A., P. Hassanzadeh, and S. Pasha, 2020a: Predicting clustered weather patterns: A test case for applications of convolutional neural networks to spatio-temporal climate data. *Sci. Rep.*, **10**, 1317, <https://doi.org/10.1038/s41598-020-57897-9>.
- , E. Nabizadeh, and P. Hassanzadeh, 2020b: Analog forecasting of extreme-causing weather patterns using deep learning. *J. Adv. Model. Earth Syst.*, **12**, e2019MS001958, <https://doi.org/10.1029/2019MS001958>.
- Chen, K., P. Wang, X. Yang, N. Zhang, and D. Wang, 2020: A model output deep learning method for grid temperature forecasts in Tianjin area. *Appl. Sci.*, **10**, 5808, <https://doi.org/10.3390/app10175808>.
- Cheung, S. H., T. A. Oliver, E. E. Prudencio, S. Prudhomme, and R. D. Moser, 2011: Bayesian uncertainty analysis with applications to turbulence modeling. *Reliab. Eng. Syst. Saf.*, **96**, 1137–1149, <https://doi.org/10.1016/j.ress.2010.09.013>.
- Christensen, H., and L. Zanna, 2022: Parametrization in weather and climate models. *Oxford Research Encyclopedia of Climate Science*, Oxford University Press, <https://doi.org/10.1093/acrefore/9780190228620.013.826>.

- Cohen, J., D. Coumou, J. Hwang, L. Mackey, P. Orenstein, S. Totz, and E. Tziperman, 2019: S2S reboot: An argument for greater inclusion of machine learning in subseasonal to seasonal forecasts. *Wiley Interdiscip. Rev.: Climate Change*, **10**, e00567, <https://doi.org/10.1002/wcc.567>.
- Comte-Bellot, G., and S. Corrsin, 1966: The use of a contraction to improve the isotropy of grid-generated turbulence. *J. Fluid Mech.*, **25**, 657–682, <https://doi.org/10.1017/S0022112066000338>.
- Coupland, J., 1993: ERCOFTAC classic database. European Research Community on Flow, Turbulence and Combustion, https://www.ercotac.org/products_and_services/classic_collection_database/.
- Dewitte, S., J. P. Cornelis, R. Müller, and A. Munteanu, 2021: Artificial intelligence revolutionises weather forecast, climate monitoring and decadal prediction. *Remote Sens.*, **13**, 3209, <https://doi.org/10.3390/rs13163209>.
- Dupuy, F., and Coauthors, 2021: ARPEGE cloud cover forecast postprocessing with convolutional neural network. *Wea. Forecasting*, **36**, 567–586, <https://doi.org/10.1175/WAF-D-20-0093.1>.
- Duraisamy, K., Z. J. Zhang, and A. P. Singh, 2015: New approaches in turbulence and transition modeling using data-driven techniques. *53rd AIAA Aerospace Sciences Meeting*, Kissimmee, FL, American Institute of Aeronautics and Astronautics, 14 pp., <https://deepblue.lib.umich.edu/handle/2027.42/140497>.
- , G. Iaccarino, and H. Xiao, 2019: Turbulence modeling in the age of data. *Annu. Rev. Fluid Mech.*, **51**, 357–377, <https://doi.org/10.1146/annurev-fluid-010518-040547>.
- Đurán, I. B., J.-F. Geleyn, F. Váňa, J. Schmidli, and R. Brožková, 2018: A turbulence scheme with two prognostic turbulence energies. *J. Atmos. Sci.*, **75**, 3381–3402, <https://doi.org/10.1175/JAS-D-18-0026.1>.
- , M. Sakradzija, and J. Schmidli, 2022: The two-energies turbulence scheme coupled to the assumed PDF method. *J. Adv. Model. Earth Syst.*, **14**, e2021MS002922, <https://doi.org/10.1029/2021MS002922>.
- Durbin, P. A., 2018: Some recent developments in turbulence closure modeling. *Annu. Rev. Fluid Mech.*, **50**, 77–103, <https://doi.org/10.1146/annurev-fluid-122316-045020>.
- Edeling, W. N., P. Cinnella, and R. P. Dwight, 2014a: Predictive rans simulations via Bayesian model-scenario averaging. *J. Comput. Phys.*, **275**, 65–91, <https://doi.org/10.1016/j.jcp.2014.06.052>.
- , —, and H. Bijl, 2014b: Bayesian estimates of parameter variability in the k - ϵ turbulence model. *J. Comput. Phys.*, **258**, 73–94, <https://doi.org/10.1016/j.jcp.2013.10.027>.
- Espeholt, L., and Coauthors, 2022: Deep learning for twelve hour precipitation forecasts. *Nat. Commun.*, **13**, 5145, <https://doi.org/10.1038/s41467-022-32483-x>.
- Flora, M. L., C. K. Potvin, P. S. Skinner, S. Handler, and A. McGovern, 2021: Using machine learning to generate storm-scale probabilistic guidance of severe weather hazards in the Warn-on-Forecast System. *Mon. Wea. Rev.*, **149**, 1535–1557, <https://doi.org/10.1175/MWR-D-20-0194.1>.
- Gamahara, M., and Y. Hattori, 2017: Searching for turbulence models by artificial neural network. *Phys. Rev. Fluids*, **2**, 054604, <https://doi.org/10.1103/PhysRevFluids.2.054604>.
- Gentine, P., M. Pritchard, S. Rasp, G. Reinaudi, and G. Yacalis, 2018: Could machine learning break the convection parameterization deadlock? *Geophys. Res. Lett.*, **45**, 5742–5751, <https://doi.org/10.1029/2018GL078202>.
- Glahn, H. R., and D. A. Lowry, 1972: The use of model output statistics (MOS) in objective weather forecasting. *J. Appl. Meteor.*, **11**, 1203–1211, [https://doi.org/10.1175/1520-0450\(1972\)011<1203:TUOMOS>2.0.CO;2](https://doi.org/10.1175/1520-0450(1972)011<1203:TUOMOS>2.0.CO;2).
- Goger, B., M. W. Rotach, A. Gohm, O. Fuhrer, I. Stiperski, and A. A. Holtslag, 2018: The impact of three-dimensional effects on the simulation of turbulence kinetic energy in a major alpine valley. *Bound.-Layer Meteor.*, **168**, 1–27, <https://doi.org/10.1007/s10546-018-0341-y>.
- Grönquist, P., C. Yao, T. Ben-Nun, N. Dryden, P. Dueben, S. Li, and T. Hoefler, 2021: Deep learning for post-processing ensemble weather forecasts. *Philos. Trans. Roy. Soc.*, **A379**, 20200092, <https://doi.org/10.1098/rsta.2020.0092>.
- Haider, T., A. Kann, C. Wittmann, G. Pistotnik, B. Bica, and C. Gruber, 2011: The integrated nowcasting through comprehensive analysis (INCA) system and its validation over the eastern alpine region. *Wea. Forecasting*, **26**, 166–183, <https://doi.org/10.1175/2010WAF2222451.1>.
- Ham, Y.-G., J.-H. Kim, and J.-J. Luo, 2019: Deep learning for multi-year ENSO forecasts. *Nature*, **573**, 568–572, <https://doi.org/10.1038/s41586-019-1559-7>.
- He, S., X. Li, T. DelSole, P. Ravikumar, and A. Banerjee, 2020: Sub-seasonal climate forecasting via machine learning: Challenges, analysis, and advances. arXiv, 2006.07972v2, <https://doi.org/10.48550/arXiv.2006.07972>.
- Honnert, R., V. Masson, and F. Couvreux, 2011: A diagnostic for evaluating the representation of turbulence in atmospheric models at the kilometeric scale. *J. Atmos. Sci.*, **68**, 3112–3131, <https://doi.org/10.1175/JAS-D-11-061.1>.
- Hwang, J., P. Orenstein, J. Cohen, K. Pfeiffer, and L. Mackey, 2019: Improving subseasonal forecasting in the western U.S. with machine learning. *Proc. 25th ACM SIGKDD Int. Conf. on Knowledge Discovery & Data Mining*, Anchorage, AK, Association for Computing Machinery, 2325–2335, <https://doi.org/10.1145/3292500.3330674>.
- James, G., D. Witten, T. Hastie, and R. Tibshirani, 2013: *An Introduction to Statistical Learning*. Vol. 112. Springer, 426 pp.
- Kaae Sønderby, C., and Coauthors, 2020: MetNet: A neural weather model for precipitation forecasting. arXiv, 2003.12140v2, <https://doi.org/10.48550/arXiv.2003.12140>.
- Kang, Y., 2015: Detection, classification and analysis of events in turbulence time series. *Bull. Aust. Math. Soc.*, **91**, 521–522, <https://doi.org/10.1017/S0004972715000106>.
- , D. Belušić, and K. Smith-Miles, 2014: Detecting and classifying events in noisy time series. *J. Atmos. Sci.*, **71**, 1090–1104, <https://doi.org/10.1175/JAS-D-13-0182.1>.
- Karpatne, A., and Coauthors, 2017: Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE Trans. Knowledge Data Eng.*, **29**, 2318–2331, <https://doi.org/10.1109/TKDE.2017.2720168>.
- Kim, K.-S., J.-B. Lee, M.-I. Roh, K.-M. Han, and G.-H. Lee, 2020: Prediction of ocean weather based on denoising autoencoder and convolutional LSTM. *J. Mar. Sci. Eng.*, **8**, 805, <https://doi.org/10.3390/jmse8100805>.
- Kolmogorov, A. N., 1941: The local structure of turbulence in incompressible viscous fluid for very large Reynolds' number. *Dokl. Akad. Nauk SSSR*, **30**, 301–305.
- Lagerquist, R., D. Turner, I. Ebert-Uphoff, J. Stewart, and V. Hagerty, 2021: Using deep learning to emulate and accelerate a radiative transfer model. *J. Atmos. Oceanic Technol.*, **38**, 1673–1696, <https://doi.org/10.1175/JTECH-D-21-0007.1>.
- Larson, V. E., D. P. Schanen, M. Wang, M. Ovchinnikov, and S. Ghan, 2012: PDF parameterization of boundary layer clouds in models with horizontal grid spacings from 2 to 16 km. *Mon. Wea. Rev.*, **140**, 285–306, <https://doi.org/10.1175/MWR-D-10-05059.1>.

- Lefantzi, S., J. Ray, S. Arunajatesan, and L. Dechant, 2015: Estimation of k - ϵ parameters using surrogate models and jet-in-crossflow data. Sandia National Laboratories, Sandia Tech. Rep. SAND2015-0707, 46 pp., <http://www.osti.gov/scitech/biblio/1170402>.
- Li, Y., and Coauthors, 2008: A public turbulence database cluster and applications to study Lagrangian evolution of velocity increments in turbulence. *J. Turbul.*, **9**, N31, <https://doi.org/10.1080/14685240802376389>.
- , J. Lang, L. Ji, J. Zhong, Z. Wang, Y. Guo, and S. He, 2021: Weather forecasting using ensemble of spatial-temporal attention network and multi-layer perceptron. *Asia-Pac. J. Atmos. Sci.*, **57**, 533–546, <https://doi.org/10.1007/s13143-020-00212-3>.
- Ling, J., A. Kurzawski, and J. Templeton, 2016: Reynolds averaged turbulence modelling using deep neural networks with embedded invariance. *J. Fluid Mech.*, **807**, 155–166, <https://doi.org/10.1017/jfm.2016.615>.
- Mahrt, L., 2014: Stably stratified atmospheric boundary layers. *Annu. Rev. Fluid Mech.*, **46**, 23–45, <https://doi.org/10.1146/annurev-fluid-010313-141354>.
- Moraux, A., S. Dewitte, B. Cornelis, and A. Munteanu, 2019: Deep learning for precipitation estimation from satellite and rain gauges measurements. *Remote Sens.*, **11**, 2463, <https://doi.org/10.3390/rs11212463>.
- Morrison, H., M. van Lier-Walqui, M. R. Kumjian, and O. P. Prat, 2020a: A Bayesian approach for statistical–physical bulk parameterization of rain microphysics. Part I: Scheme description. *J. Atmos. Sci.*, **77**, 1019–1041, <https://doi.org/10.1175/JAS-D-19-0070.1>.
- , and Coauthors, 2020b: Confronting the challenge of modeling cloud and precipitation microphysics. *J. Adv. Model. Earth Syst.*, **12**, e2019MS001689, <https://doi.org/10.1029/2019MS001689>.
- Muller, C., and S. Bony, 2015: What favors convective aggregation and why? *Geophys. Res. Lett.*, **42**, 5626–5634, <https://doi.org/10.1002/2015GL064260>.
- O’Gorman, P. A., and J. G. Dwyer, 2018: Using machine learning to parameterize moist convection: Potential for modeling of climate, climate change, and extreme events. *J. Adv. Model. Earth Syst.*, **10**, 2548–2563, <https://doi.org/10.1029/2018MS001351>.
- Oliver, T., and R. Moser, 2009: Uncertainty quantification for RANS turbulence model predictions. *62nd Annual Meeting of the APS Division of Fluid Dynamics*, Minneapolis, MN, American Physical Society, Abstract LC-004, <https://meetings.aps.org/Meeting/DFD09/Event/111733>.
- , and —, 2011: Bayesian uncertainty quantification applied to RANS turbulence models. *J. Phys.: Conf. Ser.*, **318**, 042032, <https://doi.org/10.1088/1742-6596/318/4/042032>.
- Olson, J. B., J. S. Kenyon, W. Angevine, J. M. Brown, M. Pagowski, and K. Sušelj, 2019a: A description of the MYNN-EDMF scheme and the coupling to other components in WRF-ARW. NOAA Tech. Memo OAR GSD-61, 42 pp., <https://repository.library.noaa.gov/view/noaa/19837>.
- , and Coauthors, 2019b: Improving wind energy forecasting through numerical weather prediction model development. *Bull. Amer. Meteor. Soc.*, **100**, 2201–2220, <https://doi.org/10.1175/BAMS-D-18-0040.1>.
- Otness, K., L. Zanna, and J. Bruna, 2023: Data-driven multiscale modeling of subgrid parameterizations in climate models. *arXiv*, 2303.17496v1, <https://doi.org/10.48550/arXiv.2303.17496>.
- Prudden, R., S. Adams, D. Kangin, N. Robinson, S. Ravuri, S. Mohamed, and A. Arribas, 2020: A review of radar-based nowcasting of precipitation and applicable machine learning techniques. *arXiv*, 2005.04988v1, <https://doi.org/10.48550/arXiv.2005.04988>.
- Rabier, F., 2005: Overview of global data assimilation developments in numerical weather-prediction centres. *Quart. J. Roy. Meteor. Soc.*, **131**, 3215–3233, <https://doi.org/10.1256/qj.05.129>.
- Rasp, S., and N. Thuerey, 2021: Data-driven medium-range weather prediction with a Resnet pretrained on climate simulations: A new model for WeatherBench. *J. Adv. Model. Earth Syst.*, **13**, e2020MS002405, <https://doi.org/10.1029/2020MS002405>.
- , M. S. Pritchard, and P. Gentile, 2018: Deep learning to represent subgrid processes in climate models. *Proc. Natl. Acad. Sci. USA*, **115**, 9684–9689, <https://doi.org/10.1073/pnas.1810286115>.
- Ravuri, S., and Coauthors, 2021: Skilful precipitation nowcasting using deep generative models of radar. *Nature*, **597**, 672, <https://doi.org/10.1038/s41586-021-03854-z>.
- Ray, J., S. Lefantzi, S. Arunajatesan, and L. Dechant, 2016: Bayesian parameter estimation of a k - ϵ model for accurate jet-in-crossflow simulations. *AIAA J.*, **54**, 2432–2448, <https://doi.org/10.2514/1.J054758>.
- , —, —, and —, 2018: Learning an eddy viscosity model using shrinkage and Bayesian calibration: A jet-in-crossflow case study. *ASME J. Risk Uncertainty*, **4B**, 011001, <https://doi.org/10.1115/1.4037557>.
- Ronneberger, O., P. Fischer, and T. Brox, 2015: U-Net: Convolutional networks for biomedical image segmentation. *18th International Conference on Medical Image Computing and Computer-Assisted Intervention*, N. Navab et al., Eds., Springer, 234–241, https://link.springer.com/chapter/10.1007/978-3-319-24574-4_28.
- Sandu, I., A. Beljaars, P. Bechtold, T. Mauritsen, and G. Balsamo, 2013: Why is it so difficult to represent stably stratified conditions in numerical weather prediction (NWP) models? *J. Adv. Model. Earth Syst.*, **5**, 117–133, <https://doi.org/10.1002/jame.20013>.
- Schemann, V., and K. Ebell, 2020: Simulation of mixed-phase clouds with the ICON large-eddy model in the complex arctic environment around Ny-Ålesund. *Atmos. Chem. Phys.*, **20**, 475–485, <https://doi.org/10.5194/acp-20-475-2020>.
- Schneider, T., S. Lan, A. Stuart, and J. Teixeira, 2017: Earth system modeling 2.0: A blueprint for models that learn from observations and targeted high-resolution simulations. *Geophys. Res. Lett.*, **44**, 12 396–12 417, <https://doi.org/10.1002/2017GL076101>.
- Schulthess, T. C., P. Bauer, N. Wedi, O. Fuhrer, T. Hoeffler, and C. Schär, 2019: Reflecting on the goal and baseline for exascale computing: A roadmap based on weather and climate simulations. *Comput. Sci. Eng.*, **21**, 30–41, <https://doi.org/10.1109/MCSE.2018.2888788>.
- Seifert, A., and S. Rasp, 2020: Potential and limitations of machine learning for modeling warm-rain cloud microphysical processes. *J. Adv. Model. Earth Syst.*, **12**, e2020MS002301, <https://doi.org/10.1029/2020MS002301>.
- Shapkalijevski, M., and N. Vecauteren, 2019: Introducing a simple stochastic framework to account for turbulent intermittency in atmospheric boundary-layer parameterizations. *Workshop Conservation Principles, Data and Uncertainty in Atmosphere-Ocean Modelling*, TRR 181 “Energy transfers in Atmosphere and Ocean” Center for Earth System Research and Sustainability, Universität Hamburg, Universität Potsdam (SFB 1294) and Freie Universität Berlin (SFB 1114), 33 pp., https://www.trr-energytransfers.de/fileadmin/user_upload/documents/events/WSP-2019/WSP-2019.pdf.

- Shi, X., Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, 2015: Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *Advances in Neural Information Processing Systems*, M. Jordan, Y. LeCun, and S. Solla, Eds., MIT, 802–810, <https://proceedings.neurips.cc/paper/2015/file/07563a3fe3bbe7e3ba84431ad9d055af-Paper.pdf>.
- Singh, A. P., K. Duraisamy, and Z. J. Zhang, 2017a: Augmentation of turbulence models using field inversion and machine learning. *55th AIAA Aerospace Sciences Meeting*, Grapevine, TX, American Institute of Aeronautics and Astronautics, 18 pp., <https://deepblue.lib.umich.edu/handle/2027.42/143032>.
- , R. Matai, A. Mishra, K. Duraisamy, and P. A. Durbin, 2017b: Data-driven augmentation of turbulence models for adverse pressure gradient flows. *23rd AIAA Computational Fluid Dynamics Conf.*, Denver, CO, American Institute of Aeronautics and Astronautics, 20 pp., <https://doi.org/10.2514/6.2017-3626>.
- , S. Medida, and K. Duraisamy, 2017c: Machine-learning-augmented predictive modeling of turbulent separated flows over airfoils. *AIAA J.*, **55**, 2215–2227, <https://doi.org/10.2514/1.J055595>.
- Skamarock, W. C., J. B. Klemp, M. G. Duda, L. D. Fowler, S.-H. Park, and T. D. Ringler, 2012: A multiscale nonhydrostatic atmospheric model using centroidal Voronoi tessellations and C-grid staggering. *Mon. Wea. Rev.*, **140**, 3090–3105, <https://doi.org/10.1175/MWR-D-11-00215.1>.
- Stevens, B., and Coauthors, 2020: The added value of large-eddy and storm-resolving models for simulating clouds and precipitation. *J. Meteor. Soc. Japan*, **98**, 395–435, <https://doi.org/10.2151/jmsj.2020-021>.
- Sun, J., and M. S. Pritchard, 2016: Effects of explicit convection on global land-atmosphere coupling in the superparameterized CAM. *J. Adv. Model. Earth Syst.*, **8**, 1248–1269, <https://doi.org/10.1002/2016MS000689>.
- Suselj, K., M. J. Kurowski, and J. Teixeira, 2019: On the factors controlling the development of shallow convection in eddy-diffusivity/mass-flux models. *J. Atmos. Sci.*, **76**, 433–456, <https://doi.org/10.1175/JAS-D-18-0121.1>.
- Tracey, B., K. Duraisamy, and J. Alonso, 2013: Application of supervised learning to quantify uncertainties in turbulence and combustion modeling. *51st AIAA Aerospace Sciences Meeting Including the New Horizons Forum and Aerospace Exposition*, Grapevine (Dallas/Ft. Worth Region), TX, American Institute of Aeronautics and Astronautics, 18 pp., <https://doi.org/10.2514/6.2013-259>.
- , —, and J. J. Alonso, 2015: A machine learning strategy to assist turbulence model development. *53rd AIAA Aerospace Sciences Meeting*, Kissimmee, FL, American Institute of Aeronautics and Astronautics, 22 pp., <https://doi.org/10.2514/6.2015-1287>.
- Vercauteren, N., and R. Klein, 2015: A clustering method to characterize intermittent bursts of turbulence and interaction with submesoscale motions in the stable boundary layer. *J. Atmos. Sci.*, **72**, 1504–1517, <https://doi.org/10.1175/JAS-D-14-0115.1>.
- , L. Mahrt, and R. Klein, 2016: Investigation of interactions between scales of motion in the stable boundary layer. *Quart. J. Roy. Meteor. Soc.*, **142**, 2424–2433, <https://doi.org/10.1002/qj.2835>.
- , V. Boyko, D. Faranda, and I. Stiperski, 2019: Scale interactions and anisotropy in stable boundary layers. *Quart. J. Roy. Meteor. Soc.*, **145**, 1799–1813, <https://doi.org/10.1002/qj.3524>.
- Vollant, A., G. Balarac, G. Geraci, and C. E. Corre, 2014: Optimal estimator and artificial neural network as efficient tools for the subgrid-scale scalar flux modeling. *Proc. Summer Program*, Stanford, CA, Stanford University, 10 pp., https://web.stanford.edu/group/ctr/Summer/SP14/09_Large-eddy_simulation/08_vollant.pdf.
- , —, and C. Corre, 2017: Subgrid-scale scalar flux modelling based on optimal estimation theory and machine-learning procedures. *J. Turbul.*, **18**, 854–878, <https://doi.org/10.1080/14685248.2017.1334907>.
- Wang, B., J. Lu, Z. Yan, H. Luo, T. Li, Y. Zheng, and G. Zhang, 2019: Deep uncertainty quantification: A machine learning approach for weather forecasting. *Proc. 25th ACM SIGKDD Int. Conf. on Knowledge Discovery & Data Mining*, Anchorage, AK, Association for Computing Machinery, 2087–2095, <https://doi.org/10.1145/3292500.3330704>.
- Wang, J.-X., J.-L. Wu, and H. Xiao, 2017: Physics-informed machine learning approach for reconstructing Reynolds stress modeling discrepancies based on DNS data. *Phys. Rev. Fluids*, **2**, 034603, [10.1103/PhysRevFluids.2.034603](https://doi.org/10.1103/PhysRevFluids.2.034603).
- Wang, P., J. Yuval, and P. A. O’Gorman, 2022: Non-local parameterization of atmospheric subgrid processes with neural networks. *J. Adv. Model. Earth Syst.*, **14**, e2022MS002984, <https://doi.org/10.1029/2022MS002984>.
- Weatheritt, J., and R. Sandberg, 2016: A novel evolutionary algorithm applied to algebraic modifications of the RANS stress-strain relationship. *J. Comput. Phys.*, **325**, 22–37, <https://doi.org/10.1016/j.jcp.2016.08.015>.
- , and —, 2017: The development of algebraic stress models using a novel evolutionary algorithm. *Int. J. Heat Fluid Flow*, **68**, 298–318, <https://doi.org/10.1016/j.ijheatfluidflow.2017.09.017>.
- Weyn, J. A., D. R. Durran, and R. Caruana, 2020: Improving data-driven global weather prediction using deep convolutional neural networks on a cubed sphere. *J. Adv. Model. Earth Syst.*, **12**, e2020MS002109, <https://doi.org/10.1029/2020MS002109>.
- Weyn-Vanhenenryck, J., 2020: *Sub-Seasonal Forecasting Using Large Ensembles of Data-Driven Global Weather Prediction Models*. University of Washington, 125 pp.
- Wilhelmson, R. B., and L. J. Wicker, 2001: Numerical modeling of severe local storms. *Severe Convective Storms, Meteor. Monogr.*, No. 28, Amer. Meteor. Soc., 123–166, <https://doi.org/10.1175/0065-9401-28.50.123>.
- Wu, J.-L., J.-X. Wang, H. Xiao, and J. Ling, 2017: A priori assessment of prediction confidence for data-driven turbulence modeling. *Flow Turbul. Combust.*, **99**, 25–46, <https://doi.org/10.1007/s10494-017-9807-0>.
- , H. Xiao, and E. Paterson, 2018: Physics-informed machine learning approach for augmenting turbulence models: A comprehensive framework. *Phys. Rev. Fluids*, **3**, 074602, <https://doi.org/10.1103/physrevfluids.3.074602>.
- Yuval, J., and P. A. O’Gorman, 2020: Stable machine-learning parameterization of subgrid processes for climate modeling at a range of resolutions. *Nat. Commun.*, **11**, 3295, <https://doi.org/10.1038/s41467-020-17142-3>.
- Zanna, L., P. P. Mana, J. Anstey, T. David, and T. Bolton, 2017: Scale-aware deterministic and stochastic parametrizations of eddy-mean flow interaction. *Ocean Modell.*, **111**, 66–80, <https://doi.org/10.1016/j.ocemod.2017.01.004>.
- Zheng, G., X. Li, R.-H. Zhang, and B. Liu, 2020: Purely satellite data-driven deep learning forecast of complicated tropical instability waves. *Sci. Adv.*, **6**, eaba1482, <https://doi.org/10.1126/sciadv.aba1482>.
- Zhu, Y., and N. Dinh, 2020: A data-driven approach for turbulence modeling. arXiv, 2005.00426v1, <https://doi.org/10.48550/arXiv.2005.00426>.