CAMBRIDGE
UNIVERSITY PRESS

POSITION PAPER

Machine learning for stochastic parametrization

Hannah M. Christensen[1] , Salah Kouhen[1] , Greta Miller[1] and Raghul Parthipan[2,3]

[1]Department of Physics, University of Oxford, Oxford, UK
[2]Department of Computer Science, University of Cambridge, Cambridge, UK
[3]British Antarctic Survey, Cambridge, UK
Corresponding author: Hannah M. Christensen; Email: hannah.christensen@physics.ox.ac.uk

Abstract

Atmospheric models used for weather and climate prediction are traditionally formulated in a deterministic manner. In other words, given a particular state of the resolved scale variables, the most likely forcing from the subgrid scale processes is estimated and used to predict the evolution of the large-scale flow. However, the lack of scale separation in the atmosphere means that this approach is a large source of error in forecasts. Over recent years, an alternative paradigm has developed: the use of stochastic techniques to characterize uncertainty in small-scale processes. These techniques are now widely used across weather, subseasonal, seasonal, and climate timescales. In parallel, recent years have also seen significant progress in replacing parametrization schemes using machine learning (ML). This has the potential to both speed up and improve our numerical models. However, the focus to date has largely been on deterministic approaches. In this position paper, we bring together these two key developments and discuss the potential for data-driven approaches for stochastic parametrization. We highlight early studies in this area and draw attention to the novel challenges that remain.

Impact Statement

Weather and climate predictions are relied on by users from industry, charities, governments, and the general public. The largest source of uncertainty in these predictions arises from approximations made when building the computer model used to make them. In particular, the representation of small-scale processes such as clouds and thunderstorms is a large source of uncertainty because of their complexity and their unpredictability. Machine learning (ML) approaches, trained to mimic high-quality datasets, present an unparalleled opportunity to improve the representation of these small-scale processes in models. However, it is important to account for the unpredictability of these processes while doing so. In this article, we demonstrate the untapped potential of such probabilistic ML approaches for improving weather and climate prediction.

1. Introduction

Weather and climate models exhibit long-standing biases in mean state, modes of variability, and the representation of extremes. These biases hinder progress across the World Climate Research Programme grand challenges. Understanding and reducing these biases is a key focus for the research community.

At the heart of weather and climate models are the physical equations of motion, which describe the atmosphere and ocean systems. To predict the evolution of the climate system, these equations are discretized in space and time. The resolution ranges from order 100 km and 30–60 minutes in a typical

climate model, through 10 km and 5–10 minutes in global numerical weather prediction (NWP) models, to one km and a few tens of seconds for state-of-the-art convection-permitting runs. The impact of unresolved scales of motion on the resolved scales is represented in models through *parametrisation* schemes (Christensen and Zanna, 2022). Many of the biases in weather and climate models stem from the assumptions and approximations made during this parametrization process (Hyder et al., 2018). Furthermore, despite their approximate nature, conventional parametrization schemes account for twice as much compute time as the dynamical core in a typical atmospheric model (Wedi et al., 2013).

Replacing existing parametrization schemes with statistical models trained using ML has great potential to improve weather and climate models both in terms of reduced computational cost and increased accuracy. For example, ML has been used to emulate existing parametrization schemes, including radiation (Chevallier et al., 1998; Krasnopolsky et al., 2005; Ukkonen et al., 2020) and convection (O'Gorman and Dwyer, 2018), realizing speed-ups of up to 80 times in the case of radiation. By emulating more complex (and therefore expensive) versions of a parametrization scheme, the accuracy of the weather or climate model can also be improved compared to the control simulation at minimal computational cost (Chantry et al., 2021; Gettelman et al., 2021). The accuracy of the climate model can be further improved by training ML models on high-fidelity data sets. For example, several authors have coarse grained high-resolution models which explicitly resolve deep convection to provide training data for ML models, leading to improved parametrizations (Gentine et al., 2018; Brenowitz et al., 2020). Recent research has focused on challenges, including online stability (Brenowitz and Bretherton, 2018; Brenowitz et al., 2020; Yuval and O'Gorman, 2020; Yuval et al., 2021), conservation properties (Beucler et al., 2021), and the ability of ML emulators to generalize and perform well in climate change scenarios (Beucler et al., 2021).

The schemes referenced above all implicitly assume that the grid-scale variables of state fully determine the parametrized impact of the subgrid scales (Palmer, 2019). This assumption is flawed. An alternative approach is *stochastic parametrisation* (Leutbecher et al., 2017), where the subgrid tendency is drawn from a probability distribution conditioned on the resolved scale state. Stochastic parametrization schemes are widely used in the weather and subseasonal-to-seasonal forecasting communities, where they have been shown to improve the reliability of forecasts (Leutbecher et al., 2017). Furthermore, the inclusion of stochastic parametrizations from the weather forecasting community into climate models has led to dramatic improvements in long-standing biases (Berner et al., 2017; Christensen et al., 2017). While rigorous theoretical ideas exist to demonstrate the need for stochasticity in fluid dynamical models (Gottwald et al., 2017), the approaches currently used tend to be pragmatic and ad hoc in their formulation (Berner et al., 2017). There is therefore great potential for data-driven approaches in this area, in which the uncertainty characteristics of subgrid scale processes are derived from observational or high-resolution model data sets (Christensen, 2020). In this article, we discuss the potential for ML to transform stochastic parametrization. In Section 2, we provide the physical motivation for stochastic parametrizations and give an overview of their benefits and limitations. In Section 3, we discuss the potential for ML and highlight some preliminary studies in this space. We conclude in Section 4 by issuing an invitation to the ML parametrization community to move toward such a probabilistic framework.

2. Stochastic parametrization

2.1. Deterministic versus stochastic closure

The conceptual framework utilized by the ML parametrizations typically mirrors that of the schemes they replace. The Navier-Stokes and thermodynamic equations, which describe the atmosphere (and ocean), are discretized onto a spatial grid. The resultant grid-scale variables of state (e.g. temperature, horizontal winds, and humidity in the atmospheric case) for a particular location in space and time provide inputs to the parametrization schemes. The tendencies in these variables over one time step are computed by each scheme. The schemes are deterministic, local in horizontal space and time, but generally nonlocal in the

vertical, and thus can be thought of as acting within a subgrid column [for a review, see (Christensen and Zanna, 2022)].

Deterministic parametrization schemes typically assume that the grid box is large enough to contain many examples of the unresolved process while simultaneously being "small enough to cover only a fraction of a large-scale disturbance" (Arakawa and Schubert, 1974). The parametrization scheme is then tasked with computing the mean impact of a large ensemble of small-scale phenomena, all experiencing the same background state, onto the resolved scales: a well-posed problem. However, the Navier-Stokes equations are scale invariant (Lovejoy and Schertzer, 2013), leading to the emergence of power-law behavior in fluid flows, including in the atmosphere (Nastrom and Gage, 1985) and oceans (Storer et al., 2022). In other words, fluid motions are observed at a continuum of scales, and the spectral gap between resolved and unresolved scales required by deterministic parametrization schemes does not exist. There will always be variability in the real flow on scales similar to the truncation scale, such that it is not possible to determine the impact of the subgrid processes back on the grid-scale with certainty: the grid-scale variables cannot fully constrain the subgrid motions. Deterministic parametrizations will always be a source of error in predictions.

Evidence supporting these theoretical considerations can be provided by coarse-graining studies. For example, Christensen (2020) takes a km-scale simulation that accurately captures small-scale variability as a reference. This simulation is coarse-grained to the resolution of a typical NWP model and used to initialize a low-resolution forecast model which uses deterministic parametrization schemes. The true evolution of the coarse-grained high-resolution model is compared to that predicted by the forecast model. For a given deterministic forecast tendency, the high-resolution simulation shows a *distribution* of tendencies, as shown in Figure 1. While the mean of this distribution is captured by the parametrized model, there is substantial variability about the mean, which is not captured.

Stochastic parametrizations provide an alternative paradigm to deterministic schemes. Instead of representing the mean subgrid tendency, a stochastic approach represents one possible realization of the subgrid scale process (Arnold et al., 2013). This means a stochastic scheme can be constructed to capture the variability observed in high-resolution datasets but missing from deterministic schemes (Buizza et al., 1999), as shown in Figure 1.

To design a stochastic parametrization, a probability distribution must be specified, which represents the distribution of possible subgrid scale processes, which can be conditional on the observed scale state.
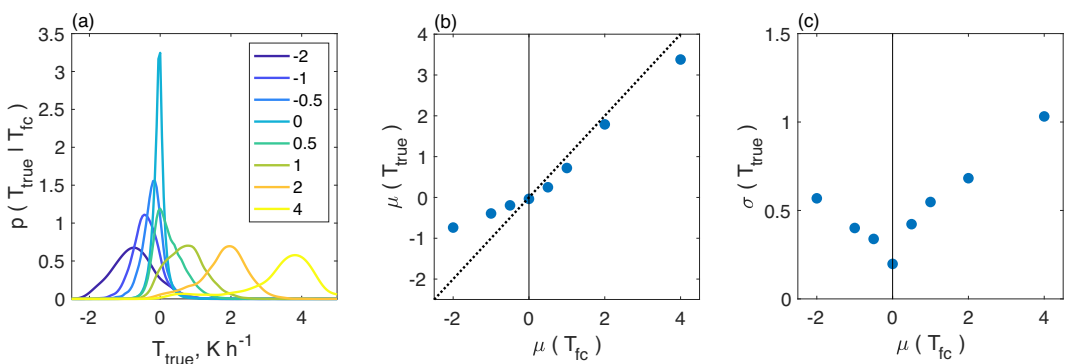
Figure 1. *Coarse-graining studies provide evidence for stochastic parametrizations. (a), the pdf of 'true' subgrid temperature tendencies derived from a high-resolution simulation is conditioned on the tendency predicted by a deterministic forecast model (T_{fc}: colors in legend). (b) Mean 'true' tendency conditioned on T_{fc}. For this forecast model, positive temperature tendencies are well calibrated, while negative temperature tendencies are biased cold. (c) Standard deviation of "true" tendency conditioned on T_{fc}. For this forecast model, the uncertainty in the "true" tendency increases with the magnitude of the low-resolution forecast tendency. Figure adapted from (Christensen, 2020).*

A random number is then generated and used to draw from the modeled distribution. Importantly, spatio-temporal correlations can be included in the draw. The resultant parametrization is no longer a subgrid scheme, but a *sur-grid* scheme, being nonlocal in space and time. This is better able to capture the scale invariance of the Navier-Stokes equations (Palmer, 2019). Evidence for the optimal spatio-temporal correlation scales can be provided by coarse-graining studies (Christensen, 2020).

We stress that it is impossible to know what the true subgrid tendency would have been. A stochastic scheme acknowledges this. Different calls to the stochastic scheme will produce different tendencies, for example, at different points in time in the model integration or in different realizations of an ensemble forecast. It is then possible to assess how the uncertainty in the parametrized subgrid tendency interacts with the rest of the model physics and ultimately leads to differences in the forecast. For this reason, we commonly refer to stochastic parametrizations as representing *model uncertainty.*

2.2. Benefits of stochastic parametrizations for weather and climate prediction

In Section 2.1, we argued that deterministic parametrization inevitably leads to errors in the predicted subgrid tendencies and that stochastic approaches allow us to capture the true variability in the impact of the subgrid scales. It is important to highlight that this is not simply a theoretical nicety but instead has large implications for the skill of the forecast model. This is because these small-scale errors in the forecast will not remain confined to the smallest scales. Instead, the chaotic nature of the atmosphere means errors will exponentially grow in time and cascade upscale in space (Lorenz, 1969), causing model simulations to diverge substantially from the true atmosphere.

Since errors in our forecasts are inevitable, instead of a single "best guess" prediction, operational centers typically make a set (or "ensemble") of equally likely weather forecasts (Bauer et al., 2015). The goal is to produce a *reliable* forecast in which the observed evolution of the atmosphere is indistinguishable from individual ensemble members (Wilks, 2006). To generate the ensemble, the initial conditions of the members are perturbed to represent uncertainties in estimates of the current state of the atmosphere. However, if only initial condition uncertainty is accounted for, the resultant ensemble is overconfident, such that the observations routinely fall outside of the ensemble forecast. A reliable forecast must also account for model uncertainty (Buizza et al., 1999). Stochastic parametrizations have transformed the reliability of initialized forecasts (Buizza et al., 1999; Weisheimer et al., n.d.; Berner et al., 2017; Ollinaho et al., 2017), and so are widely used across the weather and subseasonal-to-seasonal forecasting communities. New developments in stochastic parametrizations can further improve the reliability of these forecasts (Christensen et al., 2017)

Climate prediction is a fundamentally different problem from weather forecasting (Lorenz, 1975). It seeks to predict the response of the Earth system to an external forcing (greenhouse gas emissions). The goal is to predict the change in the statistics of the weather over the coming decades, as opposed to a specific trajectory. It has been shown that stochastic parametrizations from the weather forecasting community can substantially improve climate models (Berner et al., 2017; Christensen et al., 2017; Strommen et al., 2019). Including stochasticity in models can improve long-standing biases in the mean state, such as the distribution of precipitation (Strommen et al., 2019), as well as biases in climate variability, such as the El Niño–Southern Oscillation (Christensen et al., 2017; Yang et al., 2019). Some evidence indicates that stochastic parametrizations can mimic the impacts of increasing the resolution of models (Dawson and Palmer, 2015; Vidale et al., 2021), likely by improving the representation of small-scale variability.

2.3. Current stochastic approaches

Stochastic parametrizations often couple with existing deterministic parametrizations. They can therefore be thought of as representing random errors in the deterministic scheme. One approach, the "Stochastically Perturbed Parametrisation tendencies" (SPPT) scheme, multiplies the sum of the output of the deterministic parametrizations by a spatio-temporally correlated random number with a mean of one

(Buizza et al., 1999; Sanchez et al., 2016; Leutbecher et al., 2017). An alternative approach takes parameters from within the parametrizations schemes and varies these stochastically to represent uncertainty in their values [Stochastically Perturbed Parameters (SPP): (Christensen et al., 2015; Ollinaho et al., 2017). Both these approaches are holistic (treating all the parametrized processes) but make key assumptions about the nature of model error. Other approaches are more physically motivated. For example, the Plant-Craig scheme (Plant and Craig, 2008) represents convective mass flux as following a Poisson distribution, motivated using ideas from statistical mechanics (Craig and Cohen, 2006). The scheme predicts the convective mass flux at cloud base, which can be used as the closure assumption in an existing deterministic deep convection scheme. These ideas were subsequently extended for shallow convection by Sakradzija et al. (2016).

A common conclusion across stochastic schemes is the need to include spatio-temporal correlations into the stochasticity in order for the scheme to have a significant impact on model skill. The need for correlations from a practical point of view complements the physical justification put forward in Section 2.1. Correlations are typically implemented using a first-order auto-regressive process in time and using a spectral pattern in space (Christensen et al., 2015; Johnson et al., 2019; Palmer et al., 2009). The correlation scale parameters can be tuned to maximize forecast skill.

3. ML for stochastic parametrization

3.1. Why could ML be useful?

ML solutions are a natural fit for stochastic parametrization. There is a long history of data-driven approaches proposed by the stochastic parametrization research community. For example, Christensen et al. (2015) use statistics derived from a data-assimilation approach to constrain the joint distribution of four uncertain parameters in a stochastic perturbed parameter scheme. Alternatively, the Markov Chain–Monte Carlo approaches of Crommelin and Vanden-Eijnden (2008) and Dorrestijn et al. (2013) begin by clustering high–fidelity data to produce a discrete number of realistic tendency profiles, before computing the conditional transition statistics between the states defined by the cluster centroids. On the other hand, *operationally* used stochastic parametrizations (e.g. SPPT, SPP) are pragmatic, with only limited evidence for the structure that they assume. There is therefore substantial room for improvement. We note that the cost of existing stochastic parametrizations, such as SPPT, are generally low, so parametrization improvement, not computational speed-up, is the principal goal here.

3.2. Predicting the PDF

Common to deterministic ML parametrizations, stochastic ML schemes must obey physical constraints, such as dependency on resolved-scale variables, correlations between subgrid tendencies, and conservation properties. In addition, there are two further challenges unique to ML for stochastic parametrization. The first is the need to learn the distribution, which represents uncertainty in the process of interest.

One approach is to couple a stochastic ML approach to an existing deterministic parametrization. A sub-component of the existing scheme can be identified, and the uncertainty in that component is represented using an ML framework. For example, Miller et al. (submitted) developed a probabilistic ML approach to replace a deterministic convection trigger function (see also (Ukkonen and Mäkelä, 2019)). A random forest was trained on large-scale atmospheric variables, such as temperature and humidity, selected from the variational analysis over the Southern Great Plains (USA) observational site (Tang et al., 2019). Since multiple subgrid atmospheric states are possible for single large-scale state, the random forest is used to predict the probability of convection occurrence thereby capturing uncertainty in the convective trigger. Figure 2 shows the reliability curve for the random forest estimates of convection occurrence. For this dataset, the random forest produced *reliable* convection estimates (i.e. the predicted probabilities match the conditional observed frequency of convection) with minimal hyperparameter adjustment (Miller et al., submitted). This is potentially because of the random forest's feature
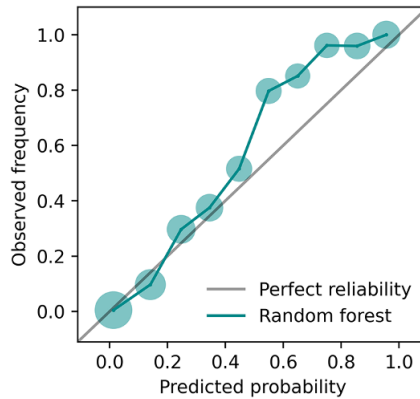
Figure 2. *Reliability curve for convection occurrence estimated by the random forest (green line), which is close to perfect reliability (grey line). The random forest was developed for use as a stochastic convection trigger function. The circle sizes are proportional to the log of the number of samples per bin; there are many more non-convection events (91%) than convection events (9%). Figure adapted from Miller et al. (submitted).*

bootstrapping and ensemble averaging methods. In addition, the random forest assigned a relatively even distribution of predicted probabilities greater than zero (Figure 2), indicating good *resolution*. Having trained a skillful and reliable model, explainable ML methods can be used to learn about the predictability of unresolved processes such as convection (Miller et al., submitted). In general, both random forests and neural networks have been found to produce reliable probabilities without postprocessing calibration, in comparison to other ML methods (Niculescu-Mizil and Caruana, 2005).

Instead of coupling to an existing parametrization, an alternative approach is to replace a conventional parametrization with an ML approach, which generates samples from the subgrid distribution, conditioned on the grid-scale variables. For example, Guillaumin and Zanna (2021) model subgrid momentum forcing using a parametric distribution, and train a neural network to learn the state-dependent parameters of this distribution (Guillaumin and Zanna, 2021). The distribution is then sampled from during model integration (Cheng et al., n.d.). Alternatively, Gagne et al. (2020) use a generative adversarial network (GAN) to generate samples conditioned on the resolved scale variables. When tested in online mode, the GAN was found to outperform baseline stochastic models and could produce reliable weather forecasts in a simple system. Nadiga et al. (2022) extend this work and demonstrate that a GAN trained using atmospheric reanalysis data can generate realistic profiles in offline tests. However, it is not known whether a GAN truly learns the target distribution (Arora et al., 2018). Behrens et al. (2024) train a Variational Encoder-Decoder (VED) network to represent unresolved moist-physics tendencies derived from a superparametrised run with the community atmosphere model: variability in the output profiles is generated by perturbing in the latent space of the encoder–decoder, giving improvements over a simple monte-carlo dropout approach.

Learning to predict the subgrid distribution is not sufficient for the implementation of a stochastic ML scheme. The second half of the problem concerns how to draw from the predicted distribution. Implementations of ML stochastic parametrizations to date have not generally addressed this half of the problem. For example, Cheng et al. (n.d.) choose to use noise uncorrelated in space and time to implement the mesoscale eddy parametrization of Guillaumin and Zanna (2021) in the Modular Ocean Model. Behrens et al. (Behrens et al., 2024) also use uncorrelated noise to sample from the VED network, which could explain the muted impact observed when using the scheme. In contrast, Gagne et al. (Gagne et al., 2020) feed correlated noise into their GAN to draw from the generator, though the characteristics of this noise, including standard deviation and correlation statistics, were tuned according to forecast skill and not learnt from the data.

3.3. Spatio-temporal correlations

The second key challenge is developing an ML approach to capture the correlations in subgrid tendencies across neighboring columns in space and time while still being practical to implement in a climate model. Any skillful ML approach must address this challenge.

Some early work has addressed temporal correlations. For example, Shin and Baik (Shin and Baik, 2022) modeled entrainment into convective clouds as a stochastic differential equation with parameters predicted by NNs. The resultant solution is analogous to the first-order autoregressive models widely used in stochastic parametrizations but for the case of continuous time. Alternatively, Parthipan et al. (Parthipan et al., 2023) use recurrent neural networks (RNNs) in a probabilistic framework to learn the temporal characteristics of subgrid tendencies in a toy atmospheric model. An RNN is a natural data-driven extension to simple autoregressive models. It can learn nonlinear temporal associations and also learn how many past states to use when predicting future tendencies. Parthipan et al. (2023) found that using an RNN to model the temporal dependencies resulted in improved performance over a first-order auto-regressive baseline (Gagne et al., 2020). This indicates that more complex correlation structures may improve forecasts in atmospheric models.

Spatial correlations are arguably more difficult to address. This is because of the need to implement any scheme in a model where subgrid columns are typically treated independently at integration stage, and the model is parallelized accordingly. However, inspiration can be taken from the approaches used in existing stochastic parametrization schemes. For example, the Neural GCM NWP emulator (Kochkov et al., 2024) uses spatio-temporally correlated fields much like those used in operational models to perturb the learnt physics module. The spatio-temporal scales were optimized during training, though most changed very little from their initialized values. Alternatively, Bengtsson et al. (Bengtsson et al., 2011; Bengtsson et al., 2013) have used a "game of life" cellular automaton (CA: Figure 3. a) to generate correlated fields for use in convection parametrization. The CA operates on a grid finer than that used by the model parametrizations, such that each CA grid cell represents one convective element. The CA self-organises in space and persists in time, introducing spatio-temporal correlations which can be coupled to an existing deterministic convection scheme (Bengtsson et al., 2013). However, the spatial correlations generated through this approach are extremely difficult to tune by hand (L. Bengtsson, pers. comm., 2018). One possible ML approach is to use a genetic algorithm (GA), which breeds and mutates successful rule sets, to navigate the large rule space. By using a fitness function based on the fractal dimension of the state after a certain number of model steps [e.g. following Christensen and Driver, 2021], a GA can discover cellular automata that produce a distribution of cells that more closely match those observed in clouds, as shown in Figure 3. Further work is needed to fully explore this possibility.


Figure 3. a. The classic cellular automata, the game of life, after 70 rule iterations on random initial conditions. b. A set of rules discovered through the use of a genetic algorithm after 70 iterations from a random initial condition. c. An example of fitness convergence for a genetic algorithm scheme.

An alternate approach to address spatial correlations incorporates information from neighboring grid cells, as is utilized in the stochastic convection scheme of Plant and Craig (Plant and Craig, 2008) or in the deterministic convolutional neural network proposed by Wang et al. (Wang et al., 2022). Here, accuracy gain from nonlocal inputs may be found to outweigh the increase in computational cost.

3.4. Structure

The question of finding the optimal structure for an ML model is common to all domains. Whilst tremendous success has come in the language domain from transformer-based models (Vaswani et al., 2017), the same tools have not similarly transformed the modeling of continuous processes. In fact, in the physical domain, there is an ongoing search for the most appropriate architecture. For example, the data-driven NWP model GraphCast (Lam et al., 2022) uses a graph neural network as the backbone, while FourCastNet (Pathak et al., 2022) and Pangu-Weather (Bi et al., 2023) each develop separate inductive biases to use on top of a transformer backbone.

This problem is also encountered when building stochastic ML parametrizations. For example, developing a custom ML architecture was key when developing a probabilistic ML parametrization for cloud cover for the European Centre for Medium-range Weather Forecasts single-column model (SCM) (Parthipan, 2024). It was found that a simple feed-forward neural network struggled to create and remove clouds appropriately. This is because transitions between cloud-containing and cloudless states are relatively rare in the training data. One solution to this problem was to use a mixed continuous-discrete model, which separated the task into i) predicting the probability of a cloud-containing state and ii) given a cloud-containing state, predicting the amount of cloud. Such models are often used for situations when the probability distribution has a discrete spike at the origin combined with a continuous positive distribution (Weld and Leemis, 2017). Figure 4 shows how this improvement in ML model structure results in substantial improvements in predicted cloud cover. However, this was a bespoke solution for the chosen problem. It is not clear whether a general architectural leap could bring enormous benefits to the field of physical modeling, as the transformer did for language.
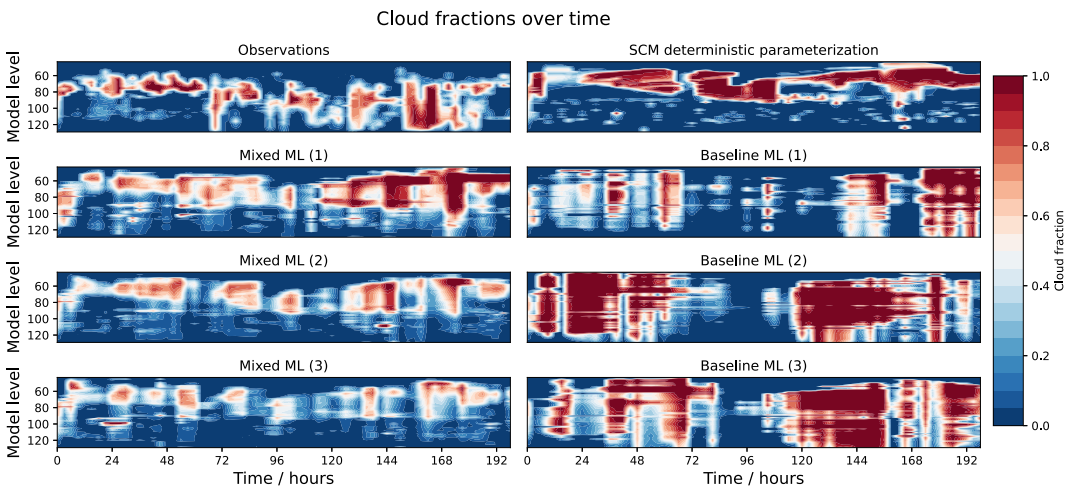
Figure 4. *Cloud fractions as a function of height (model levels) for forecasts of 200 hours. Observed cloud fraction is compared to that from the operational deterministic parametrization, and to two stochastic ML models. The Baseline ML model is a simple feed-forward neural network, whilst the Mixed ML model separates the task of modeling into a binary categorization and continuous prediction problem. These are probabilistic models, and three sampled trajectories are shown for both. The mixed model is better able to create and remove cloud. Adapted from Parthipan (2024).*

4. Conclusions

Sub-grid parametrizations are a large source of error in weather and climate forecasts. ML is transforming this field, leading to reduced errors and large speed-ups in model predictions. However, a key widely used assumption in ML parametrization is that of determinism. We have argued that this assumption is flawed and that substantial progress could be made by moving to a probabilistic framework. We discuss the first attempts to develop such stochastic ML parametrizations and highlight remaining challenges including learning sur-grid correlation structures and suitable architectures. Fortunately, ample training data exists to learn such probabilistic parametrizations. The high-resolution datasets, which are coarse-grained and used to train deterministic ML parametrizations are suitable for this task [e.g. Brenowitz et al., 2020; Yuval and O'Gorman, 2020; Beucler et al., 2021]. In addition, a new multi-model training dataset is in production as part of the Model Uncertainty Model Intercomparison Project (MUMIP: https://mumip. web.ox.ac.uk), which will be ideal for ML stochastic parametrizations. Exploring this dataset will be the focus of future work.

References

Arakawa A and Schubert WH (1974) Interaction of a Cumulus Cloud ensemble with the large-scale environment, part I. *Journal of the Atmospheric Sciences 31*, 674–701.

Arnold HM, **Moroz IM and Palmer TN** (2013) Stochastic parameterizations and model uncertainty in the Lorenz'96 system. *Philosophical Transactions of the Royal Society A. 371* (1991) 1–22.

Arora S, **Risteski A and Zhang Y** (2018) Do GANs learn the distribution? some theory and empirics. In *International Conference on Learning Representations.* https://openreview.net/forum?id=BJehNfW0-.

Bauer P, **Thorpe A and Brunet G** (2015) The quiet revolution of numerical weather prediction. *Nature 525*(7567), 47–55, sep http://doi.org/10.1038/nature14956. http://www.nature.com/doifinder/10.1038/nature14956.

Behrens G, **Beucler T**, **Iglesias-Suarez F**, **Yu S**, **Gentine P**, **Pritchard M**, **Schwabe M and Eyring V** (2024) Improving atmospheric processes in earth system models with deep learning ensembles and stochastic parameterizations. *arXiv preprint: 2402.03079.*

Bengtsson L, **Körnich H**, **Källén E and Svensson G** (2011) Large-scale dynamical response to subgrid-scale organization provided by cellular automata. *Journal of the Atmospheric Sciences 68*(12), 3132–3144.

Bengtsson L, **Steinheimer M**, **Bechtold P and Geleyn J-F** (2013) A stochastic parametrization for deep convection using cellular automata. *Quarterly Journal of the Royal Meteorological Society 139*(675), 1533–1543.

Berner J, **Achatz U**, **Batte L**, **De La Cámara A**, **Christensen H**, **Colangeli M**, **Coleman DR**, **Crommelin D**, **Dolaptchiev S**, **Franzke CLE** et al. Stochastic parameterization: Towards a new view of weather and climate models. *Bulletin of the American Meteorological Society 98*, 565–588, 2017.

Beucler T, **Gentine P**, **Yuval J**, **Gupta A**, **Peng L**, **Lin J**, **Yu S**, **Rasp S**, **Ahmed F**, **O'Gorman PA**, **Neelin JD**, **Lutsko NJ and Pritchard M** (2021) Climate-invariant machine learning. *arXiv preprint arXiv:2112.08440*.

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

Bi K, **Xie L**, **Zhang H**, **Chen X**, **Gu X and Tian Q** (2023) Accurate medium-range global weather forecasting with 3d neural networks. *Nature 619*(7970), 533–538.

Brenowitz ND, **Beucler T**, **Pritchard M and Bretherton CS** (2020) Interpreting and stabilizing machine-learning parametrizations of convection. *Journal of the Atmospheric Sciences 77*, 4357–4375,. http://doi.org/10.1175/JAS-D-20-0082.1.

Brenowitz ND and Bretherton CS (2018) Prognostic validation of a neural network unified physics parameterization. *Geophysical Research Letters*, *45* (1228) 6289–6298.

Buizza R, **Miller M and Palmer TN** (1999) Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Quarterly Journal of the Royal Meteorological Society 125*(560), 2887–2908. http://doi.org/10.1256/smsqj.56005.

Chantry M, **Hatfield S**, **Dueben P**, **Polichtchouk I and Palmer T** (2021) Machine learning emulation of gravity wave drag in numerical weather forecasting. *Journal of Advances in Modeling Earth Systems 13*(7), e2021MS002477. https://doi.org/10.1029/2021MS002477. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021MS002477. e2021MS002477 2021MS002477.

Cheng Z, **Perezhogin P**, **Gultekin C**, **Adcroft A**, **Fernandez-Granda C and Zanna L** (2023) Implementation and evaluation of a machine learned mesoscale eddy parameterization into a numerical ocean circulation model. *arXiv preprint: 2303.00962*.

Chevallier F, **Cheruy F**, **Scott NA and Chedin A** (1998) A neural network approach for a fast and accurate computation of a longwave radiative budget. *Journal of Applied Meteorology 37*(11), 1385–1397, http://doi.org/10.1175/1520-0450(1998)037<1385:annafa>2.0.co;2.

Christensen HM (2020) Constraining stochastic parametrisation schemes using high-resolution simulations. *Quarterly Journal of the Royal Meteorological Society 146*(727), 938–962.

Christensen HM, **Berner J**, **Coleman DRB and Palmer TN** (2017) Stochastic parameterization and El Niño-southern oscillation. *Journal of Climate 30*(1), 17–38, http://doi.org/10.1175/JCLI-D-16-0122.1.

Christensen HM and Driver OGA (2021) The fractal nature of clouds in global storm-resolving models. *Geophysical Research Letters 48*(23), n/a–n/a, http://doi.org/10.1029/2021GL095746.

Christensen HM, **Lock S-J**, **Moroz IM and Palmer TN** (2017) Introducing independent patterns into the stochastically perturbed parametrization tendencies (SPPT) scheme. *Quarterly Journal of the Royal Meteorological Society 143*(706), 2168–2181

Christensen HM, **Moroz IM and Palmer TN** (2015) Simulating weather regimes: Impact of stochastic and perturbed parameter schemes in a simple atmospheric model. *Climate Dynamics 44*(7), 2195–2214.

Christensen HM, **Moroz IM and Palmer TN** (2015) Stochastic and perturbed parameter representations of model uncertainty in convection parameterization. *Journal of the Atmospheric Sciences 72*(6), 2525–2544.

Christensen HM and Zanna L (2022) Parametrization in weather and climate models. *Oxford Research Encyclopedias, Climate Science 12*(12), e2020MS002226 http://doi.org/10.1029/2020MS002226. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020MS002226.e2020MS002226/10.1029/2020MS002226.

Craig GC and Cohen BG (2006) Fluctuations in an equilibrium convective ensemble. Part I: Theoretical formulation. *Journal of the Atmospheric Sciences 63*(8), 1996–2004, http://doi.org/10.1175/JAS3709.1.

Crommelin D and Vanden-Eijnden E (2008) Subgrid-scale parametrisation with conditional markov chains. *Journal of the Atmospheric Sciences 65*(8), 2661–2675.

Dawson A and Palmer TN (2015) Simulating weather regimes: Impact of model resolution and stochastic parametrisation. *Climate Dynamics 44*(7), 2177–2193.

Dorrestijn J, **Crommelin DT**, **Siebesma AP and Jonker HJJ** (2013) Stochastic parameterization of shallow cumulus convection estimated from high-resolution model data. *Theoretical and Computational Fluid Dynamics 27*(1–2), 133–148.

Gagne DJ, **Christensen HM**, **Subramanian AC and Monahan AH** (2020) Machine learning for stochastic parameterization: Generative adversarial networks in the Lorenz'96 model. *Journal of Advances in Modeling Earth Systems 21*(3), https://doi.org/10.1029/2019MS001896.

Gentine P, **Pritchard M**, **Rasp S**, **Reinaudi G and Yacalis G** (2018) Could machine learning break the convection parameterization deadlock? *Geophysical Research Letters*. *45*(1116) 5742–5751. https://doi.org/10.1029/2018GL078202

Gettelman A, **Gagne DJ**, **Chen C-C**, **Christensen MW**, **Lebo ZJ**, **Morrison H and Gantos G** (2021) Machine learning the warm rain process. *Journal of Advances in Modeling Earth Systems 13*(2), e2020MS002268. doi: http://doi.org/10.1029/2020MS002268. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020MS002268. e2020MS002268 2020MS002268.

Gottwald GA, **Crommelin DT and Franzke CLE** (2017) Stochastic climate theory. In CLE Franzke and T.J O'Kane (eds), *Nonlinear and Stochastic Climate Dynamics*. Cambridge, UK and New York: Cambridge University Press, pp. 209–240, http://doi.org/10.1017/9781316339251.009.

Guillaumin AP and Zanna L (2021) Stochastic-deep learning parameterization of ocean momentum forcing. *Journal of Advances in Modeling the Earth System 13*(9), e2021MS002534.

Hyder P, **Edwards JM**, **Allan RP**, **Hewitt HT**, **Bracegirdle TJ**, **Gregory JM**, **Wood RA**, **Meijers AJS**, **Mulcahy J**, **Field P**, **Furtado K**, **Bodas-Salcedo A**, **Williams KD**, **Copsey D**, **Josey SA**, **Liu C**, **Roberts CD**, **Sanchez C**, **Ridley J**, **Thorpe L**, **Hardiman SC**, **Mayer M**, **Berry DI and Belcher SE** (2018) Critical Southern Ocean climate model biases traced to atmospheric model cloud errors. *Nature Communications 9*(1), http://doi.org/10.1038/s41467-018-05634-2.

Johnson SJ, **Stockdale TN**, **Ferranti L**, **Balmaseda MA**, **Molteni F**, **Magnusson L**, **Tietsche S**, **Decremer D**, **Weisheimer A**, **Balsamo G**, et al. (2019) SEAS5: the new ECMWF seasonal forecast system. *Geoscientific Model Development 12*(3), 1087–1117.

Kochkov D, **Yuval J**, **Langmore I**, **Norgaard P**, **Smith J**, **Mooers G**, **Klöwer M**, **Lottes J**, **Rasp S**, **Düben P**, **Hatfield S**, **Battaglia P**, **Sanchez-Gonzalez A**, **Willson M**, **Brenner MP and Hoyer S** (2024) Neural general circulation models for weather and climate. *arXiv preprint arXiv:2311.07222*.

Krasnopolsky VM, **Fox-Rabinovitz MS and Chalikov DV** (2005) New approach to calculation of atmospheric model physics: Accurate and fast neural network emulation of longwave radiation in a climate model. *Monthly Weather Review 133*(12), 3721–3723.

Lam R, **Sanchez-Gonzalez A**, **Willson M**, **Wirnsberger P**, **Fortunato M**, **Pritzel A**, **Ravuri S**, **Ewalds T**, **Alet F**, **Eaton-Rosen Z**, et al. (2022) Graphcast: learning skillful medium-range global weather forecasting. *arXiv preprint arXiv:2212.12794*.

Leutbecher M, **Lock SJ**, **Ollinaho P**, **Lang STK**, **Balsamo G**, **Bechtold P**, **Bonavita M**, **Christensen HM**, **Diamantakis M**, **Dutra E**, et al. (2017) Stochastic representations of model uncertainties at ECMWF: state of the art and future vision. *Quarterly Journal of the Royal Meteorological Society 143*(707), 2315–2339. http://doi.org/10.1002/qj.3094.

Lorenz EN (1969) The predictability of a flow which possesses many scales of motion. *Tellus 21*(3), 289–307.

Lorenz EN (1975) Climatic predictability. In *The Physical Basis of Climate and Climate Modeling*, vol. *16* WMO, Geneva: GARP Publication Series, pp. 132–136.

Lovejoy S and Schertzer D (2013) *The Weather and Climate: Emergent Laws and Multifractal Cascades*. Cambridge, UK: Cambridge University Press.

Miller GA, **Stier P and Christensen HM** (submitted) Characterizing uncertainty in deep convection triggering using explainable machine learning. *Journal of the Atmospheric Sciences*.

Nadiga BT, **Sun X and Nash C** (2022) Stochastic parameterization of column physics using generative adversarial networks. *Environmental Data Science 1*, e22, http://doi.org/10.1017/eds.2022.32.

Nastrom GD and Gage KS (1985) A climatology of atmospheric wavenumber spectra of wind and temperature observed by commercial aircraft. *Journal of the Atmospheric Sciences 42*(9), 950–960.

Niculescu-Mizil A and Caruana R (2005) Predicting good probabilities with supervised learning. In *Proc. 22nd International Conference on Machine Learning (ICML'05)*. https://www.cs.cornell.edu/alexn/papers/calibration.icml05.crc.rev3.pdf.

O'Gorman PA and Dwyer JG (2018) Using machine learning to parameterize moist convection: potential for modeling of climate, climate change and extreme events. *Journal of Advances in Modelling Earth Systems*. *10* (10) 2548–2563.

Ollinaho P, **Lock S-J**, **Leutbecher M**, **Bechtold P**, **Beljaars A**, **Bozzo A**, **Forbes RM**, **Haiden T**, **Hogan RJ and Sandu I** (2017) Towards process-level representation of model uncertainties: Stochastically perturbed parametrisations in the ECMWF ensemble. *Quarterly Journal of the Royal Meteorological Society 143*(702), 408–422.

Palmer TN (2019) Stochastic weather and climate models. *Nature Reviews Physics 1*(7) 463–471, http://doi.org/10.1038/s42254-019-0062-2.

Palmer TN, **Buizza R**, **Doblas-Reyes F**, **Jung T**, **Leutbecher M**, **Shutts GJ**, **Steinheimer M and Weisheimer A** (2009) Stochastic parametrization and model uncertainty. *Tech. Memo 598* October 8, 2009 ECMWF, Reading UK, 42.

Parthipan R (2024) *ML Tools for Modelling Dynamical Systems: Existing Deficiencies and New Approaches*. PhD thesis, University of Cambridge.

Parthipan R, **Christensen HM**, **Hosking JS and Wischik DJ** (2023 Using probabilistic machine learning to better model temporal patterns in parameterizations: a case study with the lorenz 96 model. *Geoscientific Model Development 16*(15), 4501–4519.

Pathak J, **Subramanian S**, **Harrington P**, **Raja S**, **Chattopadhyay A**, **Mardani M**, **Kurth T**, **Hall D**, **Li Z**, **Azizzadenesheli K**, et al. Fourcastnet: A global data-driven high-resolution weather model using adaptive fourier neural operators. *arXiv preprint arXiv:2202.11214*, 2022.

Plant RS and Craig GC (2008) A stochastic parameterization for deep convection based on equilibrium statistics. *Journal of the Atmospheric Sciences 65*(1), 87–105. ISSN 00224928. http://doi.org/10.1175/2007JAS2263.1.

Sakradzija M, **Seifert A and Dipankar A** (2016) A stochastic scale-aware parameterization of shallow cumulus convection across the convective gray zone. *Journal of Advances in Modeling Earth Systems 8*(2), 786–812, http://doi.org/10.1002/2016MS000634. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016MS000634.

Sanchez C, **Williams KD and Collins M** (2016) Improved stochastic physics schemes for global weather and climate models. *Quarterly Journal of the Royal Meteorological Society 142*(694), 147–159.

Shin J and Baik J-J (2022) Parameterization of stochastically entraining convection using machine learning technique. *Journal of Advances in Modeling Earth Systems 14*(5), http://doi.org/10.1029/2021MS002817.

Storer BA, **Buzzicotti M**, **Khatri H**, **Griffies SM and Aluie H** (2022) Global energy spectrum of the general oceanic circulation. *Nature Communications 13*(1), http://doi.org/10.1038/s41467-022-33031-3.

Strommen K, **Christensen HM**, **Macleod D**, **Juricke S and Palmer TN** (2019) Progress towards a probabilistic Earth system model: Examining the impact of stochasticity in the atmosphere and land component of EC-Earth v3.2. *Geoscientific Model Development 12*(7), http://doi.org/10.5194/gmd-12-3099-2019.

Tang S, **Tao C**, **Xie S and Zhang M** (2019) Description of the arm large-scale forcing data from the constrained variational analysis (varanal) version 2 (August). https://www.arm.gov/publications/tech_reports/doe-sc-arm-tr-222.pdf.

Ukkonen P and Mäkelä A (2019) Evaluation of machine learning classifiers for predicting deep convection. *Journal of Advances in Modeling Earth Systems 11*(6), 1784–1802 http://doi.org/10.1029/2018MS001561. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018MS001561.

Ukkonen P, **Pincus R**, **Hogan RJ**, **Nielsen KP and Kaas E** (2020) Accelerating radiation computations for dynamical models with targeted machine learning and code optimization. *Journal of Advances in Modeling Earth Systems 12*(12), e2020MS002226, http://doi.org/10.1029/2020MS002226. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020MS002226. e2020MS002226.

Vaswani A, **Shazeer N**, **Parmar N**, **Uszkoreit J**, **Jones L**, **Gomez AN**, **Kaiser Ł and Polosukhin I** (2017) Attention is all you need. *Advances in Neural Information Processing Systems 30*. Curran Associates, Inc.

Vidale PL, **Hodges K**, **Vanniere B**, **Davini P**, **Roberts MJ**, **Strommen K**, **Weisheimer A**, **Plesca E and Corti S** (2021) Impact of stochastic physics and model resolution on the simulation of tropical cyclones in climate GCMs. *Journal of Climate 34*(11), 4315–4341, http://doi.org/10.1175/JCLI-D-20-0507.1. https://journals.ametsoc.org/view/journals/clim/34/11/JCLI-D-20-0507.1.xml.

Wang P, **Yuval J and O'Gorman PA** (2022) Non-local parameterization of atmospheric subgrid processes with neural networks. *Journal of Advances in Modeling Earth Systems 14*(10), e2022MS002984, http://doi.org/10.1029/2022MS002984.

Wedi NP, **Hamrud M and Mozdzynski G** (2013) The ECMWF model, progress and challenges. In *Seminar on Recent Developments in Numerical Methods for Atmosphere and Ocean Modelling*. ECMWF, pp. 1–14.

Weisheimer A, **Corti S**, **Palmer TN and Vitart F**. 2014 Addressing model error through atmospheric stochastic physical parametrizations: impact on the coupled ECMWF seasonal forecasting system. *Philosophical Transactions of the Royal Society A. 372*(2018), 1–21.

Weld C and Leemis L (2017) Modeling mixed type random variables. In *2017 Winter Simulation Conference (WSC)*. IEEE, pp. 1595–1606.

Wilks D. (2006) Statistical Methods in the Atmospheric Sciences, volume *91* of *International Geophysics Series*, 2nd Edn. Amsterdam, Netherlands: Elsevier.

Yang C, **Christensen HM**, **Corti S**, **von Hardenberg J and Davini P** (2019) The impact of stochastic physics on the El Niño Southern Oscillation in the EC-Earth coupled model. *Climate Dynamics 53*, 2843–2859, http://doi.org/10.1007/s00382-019-04660-0.

Yuval J and O'Gorman PA (2020) Stable machine-learning parameterization of subgrid processes for climate modeling at a range of resolutions. *Nature Communications 11*(3295), 1–10, ISSN 2041-1723. http://doi.org/10.1038/s41467-020-17142-3.

Yuval J, **O'Gorman PA and Hill CN** (2021) Use of neural networks for stable, accurate and physically consistent parameterization of subgrid atmospheric processes with good performance at reduced precision. *Geophysical Research Letters* 1–15. ISSN 0094-8276. http://doi.org/10.1029/2020gl091363.