JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2023MS003606

Special Section:
Machine learning application to Earth system modeling

Neural-Network Parameterization of Subgrid Momentum Transport in the Atmosphere

Janni Yuval[1] and **Paul A. O'Gorman**[1]

[1]Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA

Key Points:

- Subgrid momentum transport is calculated by coarse graining output of a three-dimensional high-resolution simulation of the atmosphere
- A neural-network parameterization has skill in predicting momentum transport during convective events
- The neural-network parameterization is implemented in the atmospheric model at coarse resolution and leads to stable simulations

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

J. Yuval,
yaniyuval@gmail.com

Citation:

Yuval, J., & O'Gorman, P. A. (2023). Neural-network parameterization of subgrid momentum transport in the atmosphere. *Journal of Advances in Modeling Earth Systems*, *15*, e2023MS003606. https://doi.org/10.1029/2023MS003606

Received 14 JAN 2023
Accepted 14 MAR 2023

Abstract Attempts to use machine learning to develop atmospheric parameterizations have mainly focused on subgrid effects on temperature and moisture, but subgrid momentum transport is also important in simulations of the atmospheric circulation. Here, we use neural networks to develop a subgrid momentum transport parameterization that learns from coarse-grained output of a high-resolution atmospheric simulation in an idealized aquaplanet domain. We show that substantial subgrid momentum transport occurs due to convection. The neural-network parameterization has skill in predicting momentum fluxes associated with convection, although its skill for subgrid momentum fluxes is lower compared to subgrid energy and moisture fluxes. The parameterization conserves momentum, and when implemented in the same atmospheric model at coarse resolution it leads to stable simulations and tends to reduce wind biases, although it over-corrects for one configuration tested. Overall, our results show that it is challenging to predict subgrid momentum fluxes and that machine-learning momentum parameterization gives promising results.

Plain Language Summary Variations in density in the atmosphere can lead to vertical motions that grow in amplitude (convection) or oscillate (gravity waves). Both convection and gravity waves have an important effect on the large-scale circulation of the atmosphere, but due to computational resource limitations these processes cannot be fully resolved in current climate models. To represent the effects of these processes on the simulated wind, climate models use simplified representations (known as parameterizations) which introduce inaccuracies in simulations. Here we use an atmospheric simulation over an ocean surface to develop a neural-network parameterization that predicts the effects of convection and gravity waves on the horizontal wind variables. We show that the neural-network parameterization has some skill in predicting the effects of convection on the wind, but little skill in predicting the effect of gravity waves except for a mean drag on the wind in the stratosphere. We implement this parameterization in an atmospheric model at coarse resolution and demonstrate that it corrects for biases in the mean wind although sometimes it overcorrects. Overall, our results show that neural-network parameterizations have the potential to improve the representation of the effects of subgrid processes on the wind in climate-model simulations.

1. Introduction

It is now well established that the vertical fluxes of horizontal momentum induced by convection and non-orographic gravity waves have an important effect on the general circulation of the atmosphere. Observations and reanalysis data imply that convective momentum transport (CMT) plays an important role in particular convective events (LeMone, 1983) and in regional and time averages (Carr & Bretherton, 2001; Lin et al., 2008). Model simulations imply that CMT influences large scale circulation patterns, such as the Hadley circulation, tropical precipitation, as well as equatorial surface wind stress and sea surface temperatures (SSTs) (Song et al., 2008; Woelfle et al., 2018; Wu et al., 2007). Furthermore, nonorographic gravity waves, which are generated by processes such as convection and geostrophic adjustment (Fritts & Nastrom, 1992; Lane et al., 2004), play an important role in setting the large scale circulation of the middle atmosphere (Alexander & Rosenlof, 2003; Orr et al., 2010) and contribute to the driving of the quasi-biennial oscillation (Dunkerton, 1997) and stratospheric semi-annual oscillation (Ray et al., 1998).

Due to limited resolution, climate models typically do not resolve the vertical fluxes of horizontal momentum induced by convection and small-scale non-orographic gravity waves. But designing simplified physical models to parameterize momentum fluxes is inherently challenging. For example, the role of unresolved pressure gradients across clouds is uncertain (Romps, 2012) and the sign of the subgrid convective momentum flux depends on the nature of convective organization (LeMone, 1983). One alternative to convectional parameter-

izations is to use machine learning (ML) parameterizations. ML parameterizations could potentially be more accurate, and they could learn a unified representation of CMT and non-orographic gravity waves that better represents the relationship between CMT and non-orographic gravity waves induced by convection (Lane & Moncrieff, 2010).

In recent years, ML has been used extensively to emulate conventional parameterizations of convection (O'Gorman & Dwyer, 2018), radiation (Belochitski & Krasnopolsky, 2021; Krasnopolsky et al., 2005), microphysics (Gettelman et al., 2020; Seifert & Rasp, 2020), and super-parameterizations (Rasp et al., 2018). These ML parameterization emulators have the potential to be almost as accurate as the parameterization they emulate at a fraction of the computational cost. ML has been also used to develop new parameterizations from output of three-dimensional high-resolution simulation to estimate the effect of subgrid processes on moisture and energy variables (Brenowitz & Bretherton, 2019; Yuval & O'Gorman, 2020; Yuval et al., 2021). These new parameterizations have the potential to substantially outperform existing parameterizations. One issue with ML parameterizations is that they may lead to instability when implemented in a coarse-resolution model (Brenowitz & Bretherton, 2018, 2019; Brenowitz et al., 2020; Rasp, 2020) but ensuring conservation of energy and accurate calculation of subgrid terms may help with this issue (Yuval et al., 2021).

Recently, there have been first attempts to use ML to emulate a conventional parameterization of gravity wave drag (Chantry et al., 2021; Espinosa et al., 2022), to learn fine-scale velocities at 100 hPa related to orographic gravity waves in a local region in Japan (Matsuoka et al., 2020), and to predict nudging wind tendencies learned from a hindcast simulation nudged toward an observational analysis (Bretherton et al., 2022; Watt-Meyer et al., 2021). ML has not yet been used to learn a new momentum transport parameterization from a high-resolution atmospheric model. In this study we use coarse-grained output of a high-resolution idealized model of a quasi-global atmosphere to develop neural network (NN) parameterizations for subgrid momentum transport.

We first describe the high-resolution simulation we use and how we coarse grain the output data from this simulation to obtain the training data for the NN momentum parameterization (Section 2). We then discuss the spatial and temporal structure of subgrid momentum transport and how it relates to convection (Section 3). We present the structure of the NN momentum parameterization and its training protocol and implementation (Section 4), and a conventional convection scheme we compare to for offline results (Section 5). Next, we investigate the skill of the NN parameterization in predicting subgrid momentum fluxes (Section 6), and we implement the parameterization in an atmospheric model at coarse resolution and study its effect on the climate (Section 7), including in different model configurations (Section 8). Finally, we give our conclusions (Section 9).

2. Methods

2.1. Simulations

Simulations were run using the System for Atmospheric Modeling (SAM) version 6.3 (Khairoutdinov & Randall, 2003). We use an aquaplanet configuration with specified SSTs following by default the qobs distribution (Neale & Hoskins, 2000) which is zonally and hemispherically symmetric. There are 48 vertical levels, and we use a quasi-global equatorial beta plane domain which has zonal width of 6,912 km, meridional extent of 17,280 km. To reduce computational expense, we use hypohydrostatic rescaling of the equations of motion with a scaling factor of 4. The hypohydrostatic rescaling increases the horizontal length scale of convection, which allows us to use a relatively coarse horizontal grid of 12 km for the high-resolution simulation (referred to as hi-res), while explicitly representing both convection and planetary scale circulations (Boos et al., 2016; Fedorov et al., 2019; Garner et al., 2007; Kuang et al., 2005; Pauluis & Garner, 2006). At vertical levels above 20 km, a sponge layer which dampens the horizontal wind components toward the horizontal mean is active at all latitudes, and consequently the stratospheric circulation is not realistic, and we mostly focus on the tropospheric circulation in this study. The hi-res simulation is the same simulation that was used for training in Yuval and O'Gorman (2020), hereinafter referred to as YOG20, and further details of the model configuration are given in YOG20.

We evaluate both the offline performance (i.e., the skill of NN-MOM when it is not coupled to SAM) and online performance (i.e., the skill when coupled to SAM). To test the online performance, we ran the following coarse

simulations with horizontal grid spacings of 96 km (which correspond to coarsening the high-resolution grid by factor of 8):

- A simulation with no NN parameterizations (referred to as x8),
- A simulation with an NN parameterization only for subgrid effects on thermodynamic and moisture variables (referred to as x8-NN); further details on the NN parameterization for the thermodynamic and moisture variables can be found in Yuval et al. (2021), hereinafter referred to as YOH21.
- A simulation with an NN parameterization for thermodynamic and moisture variables, and additionally a separate NN parameterization for horizontal momentum variables (x8-NNMOM).

We ran the simulations for 600 days, where the first 100 days in each simulation are considered as spin up, and the results for the time-mean fields are calculated from the last 500 days of each simulation. The initial conditions for simulations with NN parameterizations are taken from the last time step of the simulations with no NN parameterization (x8). Following YOH21, when running simulations with NN parameterizations, we do not use precipitating water as a prognostic variable, and we use for the conserved thermodynamic variable in the model a liquid ice static energy (H_L) that excludes precipitating water (see Text S1 in YOH21 for more details). To test the performance of our simulations in different configurations (Section 8), we also ran simulations with a horizontal grid spacing of 48 km which correspond to coarsening the high-resolution grid by factor of 4, and we trained and tested on a different high-resolution configuration that is hemispherically asymmetric (this simulation is referred to as hires-asym).

In this study, we do not run coarse-resolution simulation with conventional convection and boundary-layer parameterizations because SAM is not equipped with such parameterizations, and because the large scale organization of the tropics would be highly dependent on the specific choice of parameterizations (Möbis & Stevens, 2012). However, we use a conventional momentum parameterization based on Emanuel (1991), Emanuel and Živković-Rothman (1999), and Gregory et al. (1997) to compare the offline results of the NN parameterization to that of a conventional parameterization (Section 5).

2.2. Coarse-Graining and Calculation of Subgrid Terms

The zonal and meridional momentum equations used in SAM can be written as (Khairoutdinov & Randall, 2003):

$$\frac{\partial u}{\partial t} = -\frac{1}{\rho_0}\frac{\partial}{\partial x_i}(\rho_0 u_i u + F_{ui}) - \frac{\partial}{\partial x}\left(\frac{p'}{\rho_0}\right) + f v \tag{1}$$

$$\frac{\partial v}{\partial t} = -\frac{1}{\rho_0}\frac{\partial}{\partial x_i}(\rho_0 u_i v + F_{vi}) - \frac{\partial}{\partial y}\left(\frac{p'}{\rho_0}\right) - f u, \tag{2}$$

where u_i are the wind components ($u_i = (u, v, w)$, and u, v, w are the zonal, meridional, and vertical wind, respectively); p' is the pressure perturbation; $\rho_0(z)$ is the reference density profile and z is the vertical coordinate; F_{ui} and F_{vi} are the diffusive zonal and meridional momentum fluxes in the i direction, respectively, and f is the Coriolis parameter. The NN momentum parameterization we develop aims to account for unresolved vertical advective fluxes of horizontal momentum, to correct the surface fluxes of horizontal momentum to account for subgrid variability, and also to correct the calculation of the vertical diffusion of horizontal momentum.

We define the subgrid flux for a given variable as the difference between the coarse-grained high-resolution flux and the flux calculated at coarse resolution based on the coarse-grained prognostic variables. For example, the subgrid flux of the zonal momentum due to vertical advection is calculated as

$$(u)_{\text{adv}}^{\text{subg-flux}} = \rho_0\left(\overline{wu} - \overline{w}\ \overline{u}\right), \tag{3}$$

where overbars denote coarse-grained variables, the superscript "subg-flux" refers to subgrid fluxes, and subscript "adv" refers to vertical advection. The subgrid wind advective tendency is calculated as:

$$(u)_{\text{adv}}^{\text{subg-tend}} = -\frac{1}{\rho_0}\frac{\partial}{\partial z}(u)_{\text{adv}}^{\text{subg-flux}}, \tag{4}$$

where the superscript "subg-tend" refers to subgrid wind tendencies. For each high-resolution snapshot, we use the wind variables to calculate the subgrid fluxes associated with vertical advection of horizontal momentum

and the subgrid surface flux for the horizontal momentum variables. We do not include a parameterization for the subgrid effects of the pressure gradient and Coriolis forces, and we do not parametrize subgrid momentum transport in the horizontal. Note that subgrid pressure gradient forces are considered in plume-based CMT parameterizations (Gregory et al., 1997), but in that case it is the pressure gradient across the subgrid cloud that is considered as a step in estimating the subgrid vertical momentum flux, rather than the pressure gradient across the coarse grid box whose subgrid component is expected to be small. Subgrid surface fluxes generally act to amplify the drag on the surface wind because they account for the subgrid variability in the wind and the bulk law used depends nonlinearly on the wind.

The data set for training the NN parameterization is obtained from 3,066 snapshots of 3-hourly model output taken from hi-res (see Section 4.2 for details about training data and procedures). For each snapshot from hi-res, we coarse grain the prognostic variables (u, v, w, H_L and q_T, where q_T is the non-precipitating water mixing ratio), the vertical advective momentum fluxes, surface momentum fluxes and the vertical turbulent diffusivity used for the horizontal momentum variables.

Coarse-graining is generally performed by spatial averaging to horizontal grid spacings of 96 km. Since SAM uses a staggered C-grid (Arakawa & Lamb, 1977), coarse graining is a non-trivial task. Generally, subgrid terms may have two different types of contributions: (a) contributions that are directly related to the degradation in horizontal resolution and (b) contributions related to the staggered C-grid choice which means u, v, and w are evaluated at different locations in both the hi-res and coarse-resolution grids. For example, in Equation 3, $\overline{wu}$ is at a different horizontal location from $\overline{w}$ when coarse-graining to a C-grid. From a physical perspective, we are mostly interested in the first type of contribution since it represents the "missing" subgrid physics. Therefore, when presenting offline results not involving running the coarse model, we take into account only the contributions due to the degradation in horizontal resolution by coarse-graining all quantities to a collocated grid, calculating the subgrid terms using a modified vertical advection scheme (which does not assume a C-grid in the horizontal, but does assume a staggered grid in the vertical) and training an NN on this collocated grid for both inputs and outputs of the NN (Text S1 and Figure S1 in Supporting Information S1). However, calculating and predicting all subgrid terms on a collocated grid means that the inputs and outputs will have to be interpolated when implementing the parameterization in SAM. This is undesirable because interpolation of inputs and outputs changes the distribution of inputs and outputs compared to the training data. Therefore, the NN parameterization that is implemented in SAM is trained using coarse-grained variables that are kept on a C-grid, and the subgrid terms include both contributions due to the "missing" physics and the staggering of the C-grid (Figure S2 in Supporting Information S1). The choice of how to deal with the staggered grid affects the standard deviation of the subgrid terms (Figures S3 and S4 in Supporting Information S1), and future research could further investigate how best to coarse grain when simulations are run on a staggered grid.

3. Subgrid Momentum Transport

We first investigate the structure of the time- and zonal-mean subgrid momentum fluxes, their associated momentum tendencies and the potential sources for these fluxes. We focus on the subgrid terms calculated using a coarse graining factor of 8. Figures 1a and 1d shows the subgrid fluxes and tendencies for zonal momentum, and Figure 5a shows the climatological zonal-mean zonal wind for comparison. The zonal- and time-mean zonal momentum fluxes show broad coherent structure with a generally upward flux in the tropics and downward flux in midlatitudes, such that the fluxes are generally downgradient in the mean (although not necessarily instantaneously) below 250 hPa (Figure 1a). The associated tendencies due to subgrid vertical advection peak in the boundary layer (Figure 1d). The zonal component of the subgrid tendency tends to enhance the winds very close to the surface (for both extratropical westerlies and tropical easterlies) and to weaken winds above the surface in the lower troposphere at the levels between 800 and 950 hPa (Figure 1d). In the tropics, there are small negative tendencies in the middle troposphere and positive tendencies around 200–300 hPa. In the stratosphere, there are mostly negative zonal tendencies which are likely related to the drag effect of gravity waves on the mean flow, but there are also sharp positive and negative features near 50 hPa that are likely related to the sponge layer (Figure 1d). The meridional component of the subgrid tendency tends to decelerate the equatorward flow of the Hadley circulation near the surface and also slightly decelerate the flow at the upper branch of the Hadley circulation (Figure S5d in Supporting Information S1), although Richter and Rasch (2008) showed that the Hadley cell strength is more sensitive to the subgrid zonal momentum tendency.

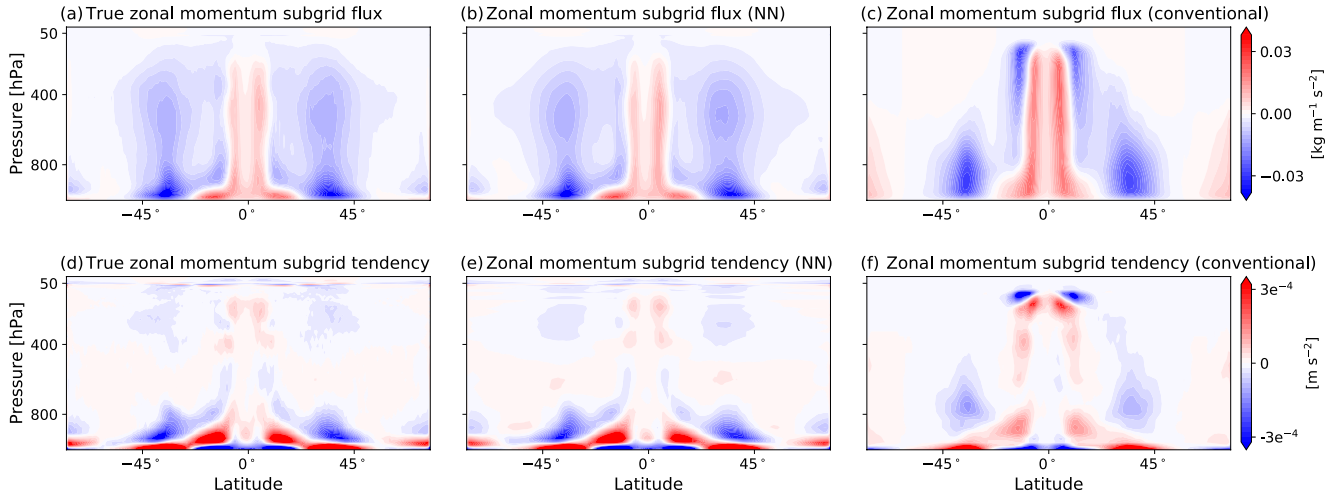

Figure 1. The time- and zonal-mean subgrid zonal momentum flux (a–c) and zonal wind tendency (d–f) from vertical advection as calculated from hires (a, d), and as estimated by (b, e) NN-MOM and (c, f) the conventional convection scheme (Section 5). Colors are saturated in panels (d–f) to highlight fluxes and tendencies outside of the boundary layer. The sponge layer is active above 50 hPa. All quantities are calculated from 501 snapshots of the x8 test data set.

We next make an indirect connection to observations by comparing the zonal momentum tendency we calculated using hi-res to a simplified CMT parameterization based on a plume model that was found to be consistent with unresolved zonal momentum transport in the upper troposphere above oceans in reanalysis data (Lin et al., 2008; Yang et al., 2013). The simplified CMT parameterization we use is:

$$(\tilde{u}(z = 11.2\text{km}))^{\text{subg-tend}}_{\text{adv}} \approx \alpha \tilde{P}(\tilde{u}(z = 0.6\text{km}) - \tilde{u}(z = 11.2\text{km})), \quad (5)$$

where α is an (empirical) regression coefficient, P is the surface precipitation, tilde represents a zonal- and time-mean, and the vertical levels ($z = 0.6$ and $z = 11.2$ km) are chosen to approximately match the pressure levels chosen by Yang et al. (2013), which were 925 for the lower level and 200 hPa for the upper level. We find that there is a high correlation between the simplified CMT parameterization and the calculated subgrid tendencies, and that the slope (α) is within the range found by Yang et al. (2013) (Figure 2a; we note that the data suggests that there should be a non zero intercept). The high correlation between the CMT approximation and the subgrid tendencies we calculate, and between the CMT approximation and reanalysis data suggests that the subgrid fluxes we calculate are realistic at least in some aspects.

Next, we investigate which processes give rise to subgrid vertical advective momentum transport. The subgrid momentum fluxes have large variability in time, and in the tropics they tend to peak sporadically for short time

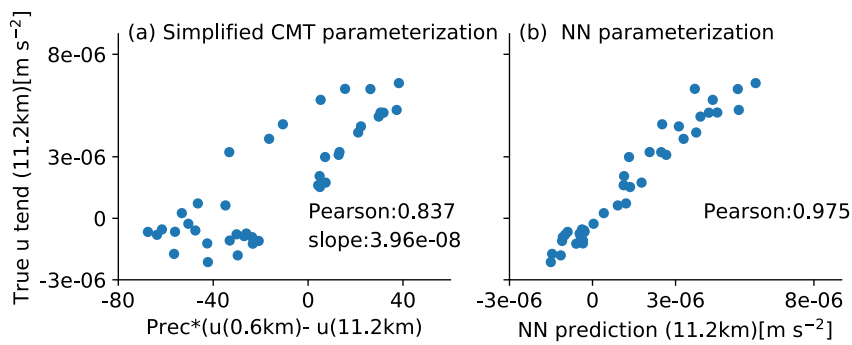

Figure 2. Scatterplots of the zonal- and time-mean of the subgrid zonal momentum tendency at an altitude of 11.2 km for individual tropical latitudes (defined as latitudes within 17.5° of the equator) versus (a) a simplified convective momentum parameterization (Equation 5) and (b) the NN-MOM predictions. Following Yang et al. (2013) only latitudes with mean precipitation rate greater than 2 mm day^{-1} are included. The Pearson correlation coefficients (r) are given in panels (a and b) and the slope (α) is given in panel (a) in units of day mm^{-1} s^{-1}. All quantities are calculated from 501 snapshots of the x8 test data set.

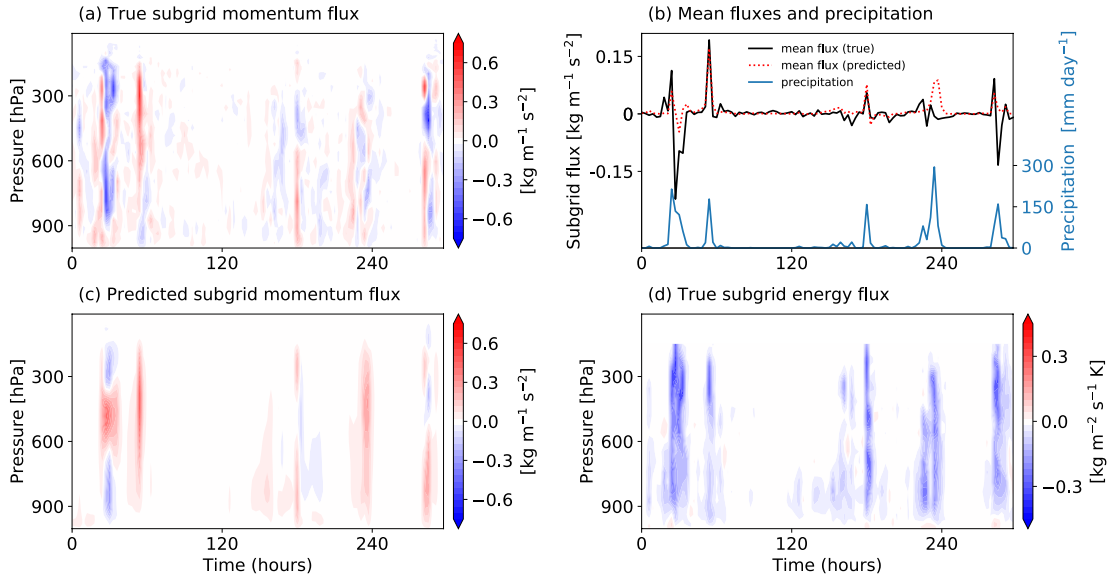

Figure 3. Time series of subgrid fluxes due to vertical advection for a tropical column at latitude 5.6° and a coarse-graining factor of 8: (a) true zonal momentum flux, (b) vertical-mean true (black) and neural network (NN)-MOM predicted (dotted red) zonal momentum flux, (c) NN-MOM predicted zonal momentum flux, (d) true subgrid energy (H_L) flux rescaled by the specific heat capacity. Panel (b) also shows the surface precipitation (blue) as a function of time. Time zero is taken to be the beginning of the presented time series which occurs in the statistical-equilibrium phase of the hi-res simulation.

intervals with coherent structure over the depth of the troposphere (Figure 3a). The subgrid momentum flux peaks simultaneously with precipitation events (Figures 3a and 3b), and during these precipitation events the subgrid advective fluxes of energy (Figure 3d) and moisture (not shown) also peak over the depth of the troposphere, implying that these peaks are deep convective events. Subgrid momentum fluxes occur also at times when no convection occurs, especially in the extratropical regions (e.g., Figure S6 in Supporting Information S1). When no convection occurs but subgrid momentum fluxes are present, there is no substantial subgrid energy transport (e.g., at 300 hPa in Figure S6d in Supporting Information S1). Note that the energy is the liquid/ice static energy which is the conserved thermodynamic variable in SAM. Linear gravity waves induce perturbations in u, v, and w that are in phase (or 180° out of phase) with each other but are 90° out of phase with a conserved thermodynamic variable (Andrews, 2010). Consequently, linear gravity waves can induce subgrid vertical momentum fluxes but do not induce substantial subgrid vertical fluxes of conserved thermodynamic variables. Therefore, we hypothesize that the source of these momentum fluxes that occur without convection and the source of the large variability of subgrid momentum fluxes in midlatitudes (Figure S3 in Supporting Information S1) is non-orographic gravity waves.

4. Neural Network Parameterization

4.1. Structure

The NN parameterization of momentum, referred to as NN-MOM, predicts the vertical profiles of the subgrid vertical advective fluxes of the horizontal momentum variables, surface wind subgrid fluxes, and coarse-grained vertical turbulent diffusivity used for momentum variables ($\overline{D}_{\text{mom}}$). Hence, the outputs for NN-MOM are

$$Y_{\text{NN-MOM}} = \left((u)^{\text{subg-flux}}_{\text{adv}}, (v)^{\text{subg-flux}}_{\text{adv}}, (u)^{\text{subg-flux}}_{\text{surf}}, (v)^{\text{subg-flux}}_{\text{surf}}, \overline{D}_{\text{mom}} \right), \tag{6}$$

where subscript "surf" refers to surface flux. Vertical advective subgrid fluxes are predicted at the 47 "half" model levels above the surface since at the surface and at the top half level the fluxes are always zero. The turbulent diffusivity is predicted only below 5.7 km (lowest 15 model levels) because the magnitude of the diffusivity reduces with height. Above 5.7 km the standard turbulent diffusivity calculated at coarse resolution from SAM is used. Overall, NN-MOM has $47 \times 2 + 1 \times 2 + 15 = 111$ outputs.

Predicting subgrid momentum fluxes due to subgrid vertical advection guarantee that the zonal and meridional momentum are conserved in each atmospheric column. Furthermore, predicting the coarse-grained vertical

turbulent diffusivity for momentum ($\overline{D}_{\text{mom}}$) ensures that turbulent momentum transport is downgradient and that diffusive processes do not introduce momentum sources or sinks. Predicting fluxes and diffusivities instead of tendencies is similar to the approach presented in YOH21 for subgrid effects on the thermodynamic and moisture variables. We predict the vertical diffusivity for the momentum variables such that our approach for the momentum variables is consistent with the approach we use for the thermodynamic and moisture variables, but we verified that the online results we present in Section 7 are not sensitive to whether or not we include the diffusivity in the parameterization.

The inputs for NN-MOM are the resolved vertical profiles of u, v, q_{T}, H_{L} and the distance to equator, $|y|$, which is a proxy for the Coriolis parameter. As in YOH21, we do not use q_{T} and H_{L} as inputs for levels above 13.9 km ($\approx$134 hPa). Using an NN momentum parameterization that includes q_{T} and H_{L} as inputs at all vertical levels in SAM leads to instability when the NN is implemented in SAM. This instability is possibly related to the small values of q_{T} in the stratosphere (leading to a normalization by very small number) or to an instability found also in previous studies that developed NN parameterization for thermodynamic and moisture variables (Brenowitz & Bretherton, 2019; Brenowitz et al., 2020, YOG20). Overall, NN-MOM has $48 \times 2 + 30 \times 2 + 1 = 157$ inputs.

4.2. Training and Implementation

The NN parameterization for energy and moisture variables is composed out of two different networks (Yuval et al., 2021). One NN (NN1) predicts the effect of subgrid vertical advection, sedimentation, microphysics, and radiation on the moisture and energy variables, and a second NN (NN2) predicts the turbulent diffusivity and moisture and energy correction for surface fluxes. NN-MOM is a separate NN in addition to NN1 and NN2. The training procedure for NN-MOM is overall similar to the training procedure done in YOH21 for training NN1 and NN2, but for completeness we repeat the description below including some small differences in the test data set, number of epochs and learning rate.

The training data for NN-MOM is obtained from 383.25 days of 3-hourly snapshots of model output taken from the hi-res simulation (overall 3,066 snapshots). This data was split into train and test datasets, where the first 320.625 days (2,565 snapshots) were used for training, and the last 62.625 days (501 snapshots) were used as a test data set. For each 3-hourly snapshot that was used during training, we reduced the training data set size by randomly subsampling atmospheric columns at each latitude for each snapshot. When using a coarse-graining factor of x4 we randomly sub-sampled 15 (out of 144) atmospheric columns at each latitude, and when using a coarse-graining factor of x8 we randomly sub-sampled 30 (out of 72) atmospheric columns at each latitude. This subsampling of the training data enables uploading all training data into the RAM during training. This results in training datasets size of 13,856,040, where a sample is defined as an individual atmospheric column for a given horizontal location and time step. To get better statistics for the calculation of the offline results, we did not subsample the test data when using a coarse-graining factor of x8, and the test data size is 6,492,960 samples. When using a coarse-graining factor of x4, the test set was randomly sub-sampled (using 15 out of 144 longitudes at each latitude), which allowed us to easily upload the whole test set to RAM, resulting in a test set of 2,705,400 samples. This x4 test set was used to verify we do not overfit, and we do not present results or plots that rely on this test set.

The NN training is performed in Python using PyTorch (Paszke et al., 2017). The weights and biases are optimized by the Adam optimizer (Kingma & Ba, 2014) combined with a cyclic learning rate (Smith, 2017). We use 1,024 samples in each batch and train over 8,000 batches before completing a full cycle in the learning rate. We use 10 epochs, where the first epoch is trained with a minimal learning rate of 0.0002 and a maximal learning rate of 0.002, and in the next five epochs both the minimal learning rate and maximal learning rate are reduced by 10% at each epoch. The last four epochs are trained after reducing both the minimum and maximum learning rates by a factor of 10 (giving a maximal learning rate of 0.000118). The results presented in this work are for NNs with 128 nodes at each hidden layer and rectified linear unit activations (ReLu) except in the output layer where no activation function was used. NNs have five densely connected layers.

Prior to training, each input (feature) of the NN momentum parameterization and the outputs were standardized by removing the mean and rescaling to unit variance. To standardize the subgrid momentum fluxes (and diffusivity) other than the surface fluxes, we calculated the mean and variance for standardization across 47 (15) vertical levels.

The NNs are stored as netcdf files, and then implemented in SAM using a Fortran module. Running SAM with the NN momentum parameterization requires 24% more CPU time than running the model without the NN momentum parameterization (i.e., only with NNs for thermodynamic and moisture variables). We note that we have not made efforts to reduce the computational cost of the parameterization at this point of development. Compiling matrix multiplication operation in Fortran can lead to very different computational cost with different compilers, and therefore, the computational cost of running NNs could be substantially different when using different compilers.

5. Conventional Momentum Parameterization

We compare the offline performance of the NN parameterization to that of a conventional convective parameterization. The conventional parameterization scheme used was the Emanuel convection scheme (Emanuel, 1991; Emanuel & Živković-Rothman, 1999) version 4.3c, in which the CMT parameterization is based on the scheme presented in Gregory et al. (1997). The method of calculating the cloud base mass flux (CBMF) in the convection scheme is based on a time dependent equation that requires memory (Emanuel, 1991). However, since we only outputted high-resolution snapshots every 3 hr we were not able to use a time dependent equation to calculate the parameterization's closure for CBMF. Therefore, instead of using the convection scheme to estimate the CBMF, we directly calculate the CBMF from hi-res for every snapshot, and use it as an input to the momentum parameterization. We expect this to substantially improve the performance of the momentum parameterization as the parameterization does not need to rely on the closure estimation of CBMF. The (coarse-grained) CBMF was calculated as:

$$\overline{CBMF}(i,j,z=1161\text{m}) = \frac{1}{N^2}\sum_{l=N(i-1)+1}^{l=Ni}\sum_{m=N(j-1)+1}^{m=Nj}\rho_{\text{w}}(z=1161\text{m})w'(l,m,z=1161\text{m}), \tag{7}$$

where ρ_{w} is the density on half levels, and w' is the deviation of the vertical velocity from the coarse-grained vertical velocity $\left(w' = w - \overline{w}\right)$ which is set to zero in grid points where the absolute value of w is smaller than 0.1 m s^{-1} or where the sum of cloud water and cloud ice mixing ratios is 0 (i.e., no cloud). We choose a threshold value of 0.1 m s^{-1} because it led to a good agreement between the instantaneous precipitation in SAM and the precipitation predicted by the conventional scheme. We define the cloud base level at a height of z = 1,161 m since it gave the best performance for the momentum parameterization (see below how the performance of the parameterization was calculated). We also tried using the first level above the lifting condensation level in each column as the definition for cloud base, but it lead to a deterioration in the results, and therefore we define the cloud base at a constant height.

To obtain a fair comparison between NN-MOM, which was tuned to fit to the subgrid momentum flux obtained from hi-res, and the conventional convective momentum scheme, we also tune the conventional scheme to fit as best as possible to the time- and zonal-mean subgrid momentum tendencies. We found it best to tune the coefficient C_u (Gregory et al., 1997, Equation 6) which controls the strength of CMT. C_u was modified between 0.2 and 0.9 in intervals of 0.1, and the optimal value that was found was $C_u = 0.6$. To save computation time for the optimization, only 30 snapshots from the test set were used during the optimization process (overall $(180-2)\times 72\times 30 = 384,480$ atmospheric columns, where we did not use the columns that are found at the poleward boundaries of our domain). The optimal value was chosen as those for which the time- and zonal mean zonal momentum convective tendencies were closest (in terms of R^2) to the subgrid tendencies obtained from hi-res. We note that the pressure levels that were used as inputs to the Emanuel scheme were the reference pressure plus the perturbation pressure (namely, the exact pressure levels) since using the reference pressure levels without the pressure perturbation led to slightly worse performance.

6. Offline Performance

We now investigate the offline performance of NN-MOM and compare it to a conventional CMT scheme (Section 5) and to the simplified CMT scheme (Equation 5). NN-MOM captures accurately the zonal- and time-mean momentum transport (Figures 1b and 1e, Figures S5b and S5e in Supporting Information S1). In contrast, the mean tendencies and fluxes calculated from hi-res have only some similarities in pattern and magnitude with the fluxes/tendencies obtained from the conventional convective momentum parameterization, and there are

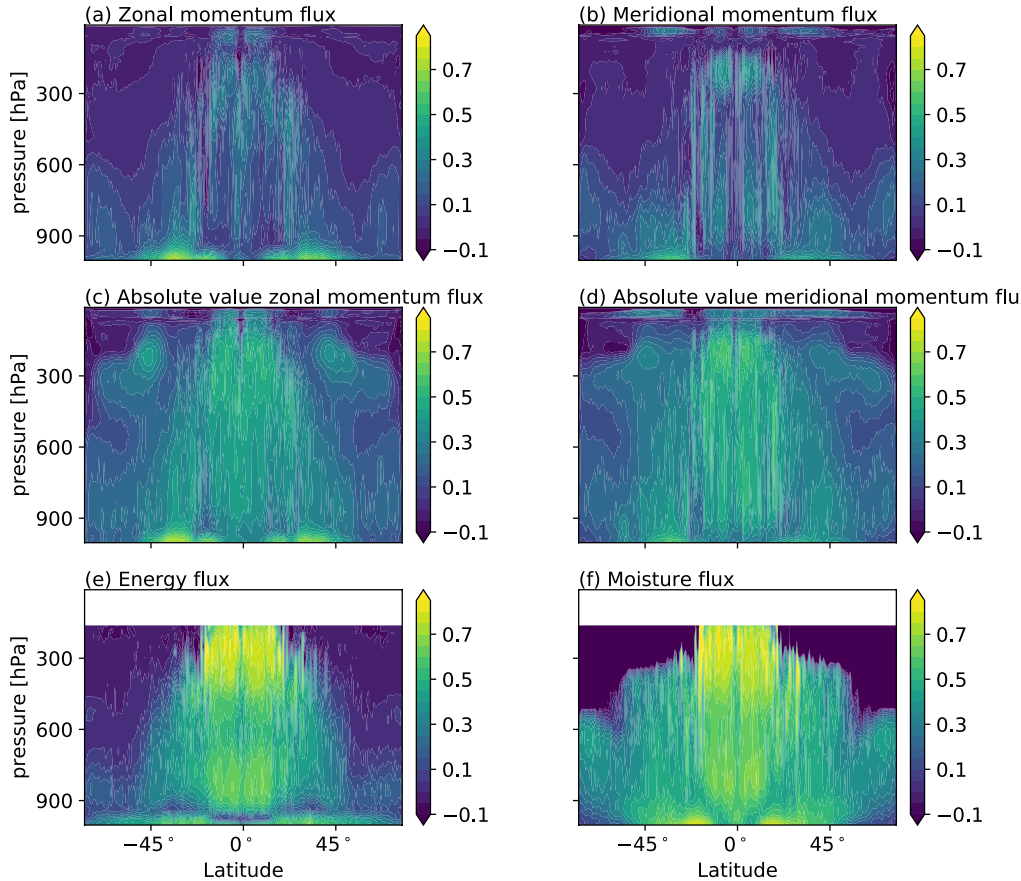

Figure 4. The coefficient of determination (R^2) for offline performance of at x8 of NN-MOM that predicts the subgrid fluxes of (a) zonal and (b) meridional momentum, of an NN that was trained to predict the absolute value of subgrid fluxes of (c) zonal and (d) meridional momentum, and of the NN that predicts the subgrid fluxes of (e) energy (H_L) and (f) non-precipitating water mixing ratio. The NN parameterization that predicts the subgrid energy and moisture fluxes is described in detail in YOH21.

noticeable differences. Overall, the zonal- and time-mean momentum transport calculated from the conventional scheme is substantially less accurate than NN-MOM (Figures 1c and 1f and Figures S5c and S5f in Supporting Information S1). We note that the conventional scheme is not expected to capture subgrid fluxes that originate from gravity waves or are related to the sponge layer that is included in our simulations and therefore the comparison between NN-MOM and the conventional scheme should be confined to the troposphere. We also find that in the tropical upper troposphere, NN-MOM is more accurate than the simplified conventional CMT parameterization with one tunable parameter (Equation 5) in predicting the zonal- and time-mean vertical advective subgrid zonal momentum tendency in the tropical upper troposphere (Figure 2b).

Considering next the instantaneous time scale, we find that the NN has modest skill in predicting subgrid fluxes due to vertical advection in parts of the troposphere particularly in the tropics and near the surface (Figures 4a and 4b). This skill is particularly evident during convective events (Figure 3c). NN-mom is substantially more accurate than the conventional scheme which has widespread negative R^2 values at the instantaneous time scale (Figure S7 in Supporting Information S1). A limitation of our comparison with the conventional parameterization is the use of a 12 km grid in hi-res even if this is mitigated by the use of hypohydrostatic rescaling, and thus future work should make such a comparison using a higher resolution simulation to calculate subgrid fluxes. The offline performance of NN-MOM at the instantaneous time scale is best for the subgrid surface fluxes and the diffusivity (Figure S8 in Supporting Information S1), especially at midlatitudes where the subgrid fluxes due to advection might be associated to a greater extent with non-orographic gravity waves rather than convection. Accurately predicting fluxes due to non-orographic gravity waves in a column parameterization will be especially difficult when gravity waves are propagating horizontally in addition to vertically. The skill of NN-MOM in all regions is

Table 1
Online Performance as Measured by the Root Mean Square Error of the Time- and Zonal-Mean Zonal Wind, Meridional Wind, and Eddy Kinetic Energy

	RMSE relative to hi-res		
	x8	x8-NN	x8-NNMOM
Zonal wind (m s^{-1})	4.51	3.21	2.71
Meridional wind (m s^{-1})	0.63	0.25	0.17
EKE (m^2 s^{-2})	42.89	35.12	34.70
	Correlation (bias in NN vs. correction of NNMOM)		
Zonal wind (r)		−0.56	
Meridional wind (r)		−0.77	
EKE (r)		−0.24	

Note. The RMSE is calculated relative to hi-res for the coarse-resolution simulations with no ML parameterization (x8), the simulations with the NN parameterization only for thermodynamic and moisture variables (x8-NN) and for the simulations with NN parameterization for thermodynamic, moisture, and horizontal momentum variables (x8-NNMOM). The eddy kinetic energy is defined with respect to the zonal and time mean. The correlation coefficient between the bias in x8-NN (x8-NN minus x8) and the correction from including momentum (x8-NNMOM minus x8-NN) is also given, where a high negative value indicates the correction has the correct spatial pattern and sign.

substantially lower compared to the skill of an NN trained to predict moisture and energy fluxes due to subgrid vertical advection (compare Figures 4a and 4b to Figures 4e and 4f). Predicting CMT might be more challenging than predicting moisture and energy convective transport because CMT can be both negative or positive (Figure 3a), depending on the spatial organization of clouds (Moncrieff, 1992). In contrast, energy and moisture are always transported in the same direction during convecting events (Figure 3d). We find that when an NN is trained to predict the absolute value of subgrid momentum fluxes (Figures 4c and 4d), it has substantially better skill compared to NN-MOM, implying that it is difficult for the network to learn the sign of subgrid momentum transport.

7. Online Performance

To investigate the effect of NN-MOM on the circulation, we compare a simulation with an NN parameterization for thermodynamic and moisture variables but no momentum parameterization (x8-NN) to a simulation with NN parameterizations for thermodynamic, moisture and horizontal momentum variables (x8-NNMOM). We note that NN-MOM is a separate NN in addition to the neural networks that predict moisture and energy related quantities. NN-MOM predicts at every time step the subgrid horizontal momentum fluxes, and from these fluxes we diagnose the subgrid tendencies which are added to the resolved tendencies. NN-MOM also predicts at every time step the coarse-grained vertical diffusivity for the horizontal momentum variables at levels below 5.7 km. This diffusivity is used at every time step in the momentum diffusion scheme for levels below 5.7 km. We also compare to a simulation with no NN parameterization (x8). All simulations we run are stable and do no exhibit climate drift for the 500 days after spinup. Results are summarized in terms of root mean square error (RMSE) for climatological variables compared to coarse-grained hi-res in Table 1.

In YOH21 we already showed that the use of the NN parameterization for thermodynamic and moisture variables in x8-NN brings the climatology of precipitation statistics much closer to hi-res than simulations without any NN parameterization. Here we focus on the effect of the momentum parameterization in x8-NNMOM on the zonal and time-mean horizontal winds and eddy kinetic energy. Inclusion of NN-MOM leads to noticeable changes in the mean horizontal winds in x8-NNMOM compared to x8-NN (Figures 5c and 5f), and it overall corrects biases in x8-NN as shown by the biases and corrections of opposite signs and similar patterns and magnitudes in Figure 5b versus Figures 5c and 5e versus Figure 5f, and by the reduced RMSE values when momentum parameterization is included (Table 1). The zonal wind weakens across the stratosphere (Figure 5c), such that biases in the stratosphere of x8-NN possibly due to subgrid gravity-wave breaking or interactions with the sponge layer are reduced (Figure 5b). Furthermore, the subtropical westerly jet is weakened (Figure 5c), and the meridional wind near the surface and at the surface weakens (Figures 5f and 6a) again reducing biases in x8-NN (Figure 5e). Such

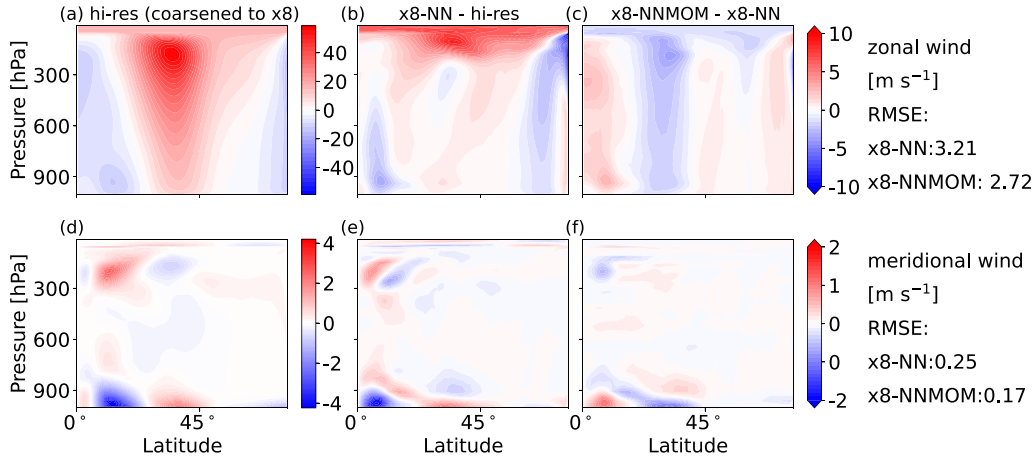

Figure 5. The zonal- and time-mean zonal and meridional wind as a function of pressure and latitude for simulations with coarser grids by factors of 8 compared to hi-res. First column shows the zonal- and time-mean zonal wind (a) and meridional wind (d) for hi-res (coarsened to x8). The second column (panels (b, e)) shows the difference between the coarse resolution simulation with neural network (NN) parameterization for thermodynamic and moisture variables and hi-res, and the third column (panels (c, f)) show the difference between a simulation with NN parameterization for thermodynamic, moisture, and momentum variables and a simulation with NN parameterization only for thermodynamic and moisture variables. The results were averaged over both hemispheres to obtain better statistics. The root mean square error of the time- and zonal-mean horizontal winds relative to hi-res is also given.

an improvement in the simulation of surface winds could be important for coupled ocean-atmosphere simulations. Overall x8-NNMOM is closer to the climatology of hi-res compared to x8-NN (Table 1).

8. Performance in Different Configurations

To explore the performance in different configurations, we trained parameterizations and ran simulations for (a) a finer coarse grid spacing of 48 km (x4, x4-NN, x4-NNMOM) and (b) based on a high-resolution simulation with a hemispherically asymmetric configuration (x8-asym, x8-NN-asym, x8-NNMOM-asym; described in Text S2 in Supporting Information S1). To train NNs for the asymmetric configuration, we use different inputs because radiation and SST are no longer just functions of distance to the equator, we repeat the coarse-graining processes described in Section 2.2 for hires-asym, and we train new NN parameterizations with an almost identical protocol (described in Text S3 in Supporting Information S1). Testing NN parameterizations in the hemispherically asymmetric configuration is especially interesting since we did not retune any NN hyperparameters. Note, we are not testing generalization between different training and testing climates which would require different inputs for good performance (Beucler et al., 2021; O'Gorman & Dwyer, 2018). We find that all simulations we tested for the asymmetric and x4 configurations lead to stable simulations. These configurations were not included in any previous publication, and therefore we show in Figure S9 in Supporting Information S1 that the representation of precipitation distribution and extreme events is substantially improved when including only the NN parameterization for thermodynamic and moisture variables, which is consistent with the results shown in YOH21.

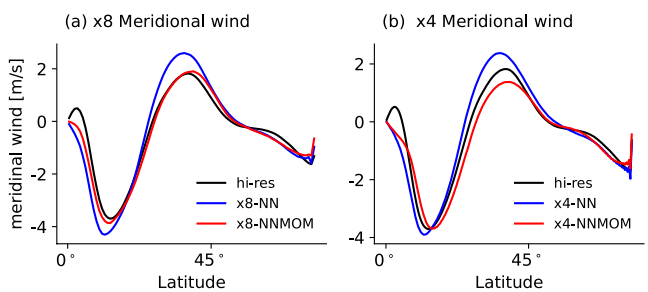

Figure 6. The zonal- and time-mean meridional surface wind as a function of latitude for simulations with a coarser grid by a factor of (a) 8 and (b) 4 compared to hi-res. Results are shown for simulations with a neural network (NN) parameterization for thermodynamic and moisture variables (blue), simulations with NN parameterizations for thermodynamic, moisture, and momentum variables (red) and hires (black; coarsened to x8 and x4, respectively).

The pattern of changes in x4-NNMOM relative to x4-NN is similar to the changes in x8-NNMOM relative to x8-NN and oppose the biases relative to hi-res in many regions (Table S1 in Supporting Information S1 shows a negative correlation between the bias in the mean wind variables and the effect of momentum parameterization on the simulation), but the magnitude of the change is too large (Figure S10 in Supporting Information S1) for reasons that remain unclear. As a result, the wind climatology of x4-NNMOM degrades compared to x4-NN (Table S1 in Supporting Information S1) due to an overshoot in the effect of NN-MOM on the circulation (e.g., Figure 6 shows that in x4-NNMOM the meridional surface winds have a bias with opposite sign compared to x4-NN).

Next, we focus on the effect of the momentum parameterization in the hemispherically asymmetric simulations. We find that the simulation without momentum parameterization has quite different wind biases compared to the hemispherically symmetric case (Figures S11b and S11e in Supporting Information S1), and that including the momentum parameterization leads to more accurate simulation of the mean wind (Table S2 and Figure S11 in Supporting Information S1).

9. Conclusions

In this study, we calculated subgrid momentum fluxes by coarse graining output from a three-dimensional high-resolution simulation, and we developed an NN momentum parameterization for vertical fluxes of horizontal momentum that was implemented in an atmospheric model at coarse resolution. To our knowledge this is the first machine-learning momentum parameterization that has learned from a high-resolution model of the atmosphere and implemented in the model at coarse resolution.

We first studied the character and climatology of the subgrid momentum fluxes based on the coarse-graining approach. Subgrid momentum transport in the tropics occurs primarily due to CMT. In the extratropics, subgrid momentum fluxes have large variability in the vicinity of the jet, possibly due to gravity waves excited by baroclinic instability. We showed that the zonal- and time-mean subgrid momentum tendencies in the tropical upper troposphere are broadly consistent with a simple approximation of CMT that was previously found to reproduce residuals in the resolved momentum budget in reanalysis in that region (Lin et al., 2008; Yang et al., 2013).

Next, we developed an NN momentum parameterization for vertical fluxes of horizontal momentum. The NN predicts fluxes instead of tendencies, which guarantees that the NN obeys momentum conservation in each atmospheric column. The NN has skill in predicting momentum fluxes during convecting events and the mean subgrid fluxes in all regions, and it performs better in offline tests than a conventional convective momentum parameterization, but it has little instantaneous skill in regions of large variability near the jets. We showed that it is more difficult to predict subgrid momentum transport compared to subgrid moisture or energy transport, and this is likely due to the difficulty in predicting momentum transport by gravity waves and the non-trivial task of determining the sign of CMT. Indeed, we found that an NN that is trained to predict the absolute value of subgrid momentum transport performs substantially better compared to an NN that is trained to predict subgrid momentum fluxes (including their sign). Future studies could further investigate how to design neural networks that have better performance for this task and specifically what inputs are needed for good accuracy in all regions.

Finally, we implemented the NN momentum parameterization in the atmospheric model at two different coarse resolutions and for two different model configurations. Simulations with the NN momentum parameterization run stably and without climate drift in all cases. The momentum parameterization reduces the wind biases relative to high-resolution simulations when running at 96 km grid spacing and for both the default hemispherically symmetric configuration and an alternate hemispherically asymmetric configuration. However, when the NN momentum parameterization is implemented at a finer grid spacing of 48 km, the NN momentum parameterization overshoots in its correction of the circulation. These results might imply that additional processes that were not parameterized might be important at smaller grid spacings and would partly compensate the effect of the vertical momentum flux, such as the horizontal subgrid momentum fluxes or pressure differences across the coarse grid cell (which does not need to be exactly equal to the coarse-grained pressure differences), or possibly the parameterization we developed is not sufficiently accurate at this resolution. Interestingly, an overshoot in the effect of an ML parameterization was also found when an ML momentum parameterization was implemented in an ocean model (Zanna & Bolton, 2020). We choose not to retune NN-mom based on online performance (e.g., by simply reducing the tendencies by a fixed factor) in order to focus attention on possible missing processes at x4 and for greater consistency in training and online testing. The staggering of momentum variables on the model grid poses challenging for learning a momentum parameterization and future work could investigate how best to deal with this issue which may improve online performance at all resolutions. In a related study, Wang et al. (2022) show that including non-local inputs improves offline performance for the NN momentum parameterization developed here, and future work should investigate its online performance. Overall, our results show that using high-resolution simulations to evaluate subgrid fluxes provides useful information for the design of parameterizations, and that NN parameterization for momentum is a promising alternative to existing parameterizations.

Data Availability Statement

Open Research Associated code, Software utilized to create the simulations, notebooks to generate figures, processed data from online simulations, trained neural network parameterizations and (a link to) the output of hi-res are available at zenodo.org (https://doi.org/10.5281/zenodo.7537443).

Acknowledgments
We thank Bill Boos and Alexey Fedorov for providing the output from the hemispherically symmetric high-resolution simulation. This research was made possible by Schmidt Futures, a philanthropic initiative founded by Eric and Wendy Schmidt, as part of its Virtual Earth System Research Institute (VESRI). We acknowledge high-performance computing support from Cheyenne (https://doi.org/10.5065/D6RX99HX) provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation. This work used Bridges2 at Pittsburgh Supercomputing Center through allocation EES220003 from Advanced Cyberinfrastructure Coordination Ecosystemm: Services & Support (ACCES) program, which is supported by National Science Foundation Grants #2138259, #2138286, #2138307, #2137603, and #2138296. POG acknowledges support from NSF awards AGS-1749986 and OAC 1835618.

References

Alexander, M. J., & Rosenlof, K. H. (2003). Gravity-wave forcing in the stratosphere: Observational constraints from the upper atmosphere research satellite and implications for parameterization in global models. *Journal of Geophysical Research*, *108*(D19), 4597. https://doi.org/10.1029/2003jd003373

Andrews, D. G. (2010). *An introduction to atmospheric physics*. Cambridge University Press.

Arakawa, A., & Lamb, V. R. (1977). Computational design of the basic dynamical processes of the UCLA general circulation model. *General circulation models of the atmosphere*, *17*(Supplement C), 173–265.

Belochitski, A., & Krasnopolsky, V. (2021). Robustness of neural network emulations of radiative transfer parameterizations in a state-of-the-art general circulation model. arXiv preprint arXiv:2103.07024.

Beucler, T., Pritchard, M., Yuval, J., Gupta, A., Peng, L., Rasp, S., et al. (2021). Climate-invariant machine learning. arXiv preprint arXiv:2112.08440.

Boos, W. R., Fedorov, A., & Muir, L. (2016). Convective self-aggregation and tropical cyclogenesis under the hypohydrostatic rescaling. *Journal of the Atmospheric Sciences*, *73*(2), 525–544. https://doi.org/10.1175/jas-d-15-0049.1

Brenowitz, N. D., Beucler, T., Pritchard, M., & Bretherton, C. S. (2020). Interpreting and stabilizing machine-learning parametrizations of convection. arXiv preprint arXiv:2003.06549.

Brenowitz, N. D., & Bretherton, C. S. (2018). Prognostic validation of a neural network unified physics parameterization. *Geophysical Research Letters*, *45*(12), 6289–6298. https://doi.org/10.1029/2018gl078510

Brenowitz, N. D., & Bretherton, C. S. (2019). Spatially extended tests of a neural network parametrization trained by coarse-graining. *Journal of Advances in Modeling Earth Systems*, *11*(8), 2727–2744. https://doi.org/10.1029/2019ms001711

Bretherton, C., Henn, B., Kwa, A., Brenowitz, N., McGibbon, J., Clark, S., et al. (2022). Correcting coarse-resolution weather and climate models by machine learning from global storm-resolving simulations. *Authorea Preprints*.

Carr, M. T., & Bretherton, C. S. (2001). Convective momentum transport over the tropical Pacific: Budget estimates. *Journal of the Atmospheric Sciences*, *58*(13), 1673–1693. https://doi.org/10.1175/1520-0469(2001)058<1673:cmtott>2.0.co;2

Chantry, M., Hatfield, S., Duben, P., Polichtchouk, I., & Palmer, T. (2021). Machine learning emulation of gravity wave drag in numerical weather forecasting. arXiv preprint arXiv:2101.08195.

Dunkerton, T. J. (1997). The role of gravity waves in the quasi-biennial oscillation. *Journal of Geophysical Research*, *102*(D22), 26053–26076. https://doi.org/10.1029/96jd02999

Emanuel, K. A. (1991). A scheme for representing cumulus convection in large-scale models. *Journal of the Atmospheric Sciences*, *48*(21), 2313–2329. https://doi.org/10.1175/1520-0469(1991)048<2313:asfrcc>2.0.co;2

Emanuel, K. A., & Živković-Rothman, M. (1999). Development and evaluation of a convection scheme for use in climate models. *Journal of the Atmospheric Sciences*, *56*(11), 1766–1782. https://doi.org/10.1175/1520-0469(1999)056<1766:daeoac>2.0.co;2

Espinosa, Z. I., Sheshadri, A., Cain, G. R., Gerber, E. P., & DallaSanta, K. J. (2022). Machine learning gravity wave parameterization generalizes to capture the QBO and response to increased CO_2. *Geophysical Research Letters*, *49*(8), e2022GL098174. https://doi.org/10.1029/2022gl098174

Fedorov, A. V., Muir, L., Boos, W. R., & Studholme, J. (2019). Tropical cyclogenesis in warm climates simulated by a cloud-system resolving model. *Climate Dynamics*, *52*(1–2), 107–127. https://doi.org/10.1007/s00382-018-4134-2

Fritts, D. C., & Nastrom, G. D. (1992). Sources of mesoscale variability of gravity waves. Part ii: Frontal, convective, and jet stream excitation. *Journal of the Atmospheric Sciences*, *49*(2), 111–127. https://doi.org/10.1175/1520-0469(1992)049<0111:somvog>2.0.co;2

Garner, S. T., Frierson, D. M. W., Held, I. M., Pauluis, O., & Vallis, G. K. (2007). Resolving convection in a global hypohydrostatic model. *Journal of the Atmospheric Sciences*, *64*(6), 2061–2075. https://doi.org/10.1175/jas3929.1

Gettelman, A., Gagne, D. J., Chen, C.-C., Christensen, M., Lebo, Z., Morrison, H., & Gantos, G. (2020). Machine learning the warm rain process. *Journal of Advances in Modeling Earth Systems*, *13*(2), e2020MS002268. https://doi.org/10.1029/2020MS002268

Gregory, D., Kershaw, R., & Inness, P. (1997). Parametrization of momentum transport by convection. II: Tests in single-column and general circulation models. *Quarterly Journal of the Royal Meteorological Society*, *123*(541), 1153–1183. https://doi.org/10.1002/qj.49712354103

Khairoutdinov, M. F., & Randall, D. A. (2003). Cloud resolving modeling of the ARM summer 1997 IOP: Model formulation, results, uncertainties, and sensitivities. *Journal of the Atmospheric Sciences*, *60*(4), 607–625. https://doi.org/10.1175/1520-0469(2003)060<0607:crmota>2.0.co;2

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Krasnopolsky, V. M., Fox-Rabinovitz, M. S., & Chalikov, D. V. (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*(5), 1370–1383. https://doi.org/10.1175/mwr2923.1

Kuang, Z., Blossey, P. N., & Bretherton, C. S. (2005). A new approach for 3D cloud-resolving simulations of large-scale atmospheric circulation. *Geophysical Research Letters*, *32*(2), L02809. https://doi.org/10.1029/2004GL021024

Lane, T. P., Doyle, J. D., Plougonven, R., Shapiro, M. A., & Sharman, R. D. (2004). Observations and numerical simulations of inertia–gravity waves and shearing instabilities in the vicinity of a jet stream. *Journal of the Atmospheric Sciences*, *61*(22), 2692–2706. https://doi.org/10.1175/jas3305.1

Lane, T. P., & Moncrieff, M. W. (2010). Characterization of momentum transport associated with organized moist convection and gravity waves. *Journal of the Atmospheric Sciences*, *67*(10), 3208–3225. https://doi.org/10.1175/2010jas3418.1

LeMone, M. A. (1983). Momentum transport by a line of cumulonimbus. *Journal of the Atmospheric Sciences*, *40*(7), 1815–1834. https://doi.org/10.1175/1520-0469(1983)040<1815:mtbalo>2.0.co;2

Lin, J.-L., Mapes, B. E., & Han, W. (2008). What are the sources of mechanical damping in Matsuno–Gill-type models? *Journal of Climate*, *21*(2), 165–179. https://doi.org/10.1175/2007jcli1546.1

Matsuoka, D., Watanabe, S., Sato, K., Kawazoe, S., Yu, W., & Easterbrook, S. (2020). Application of deep learning to estimate atmospheric gravity wave parameters in reanalysis data sets. *Geophysical Research Letters*, *47*(19), e2020GL089436. https://doi.org/10.1029/2020gl089436

Möbis, B., & Stevens, B. (2012). Factors controlling the position of the intertropical convergence zone on an aquaplanet. *Journal of Advances in Modeling Earth Systems*, *4*, M00A04. https://doi.org/10.1029/2012ms000199

Moncrieff, M. W. (1992). Organized convective systems: Archetypal dynamical models, mass and momentum flux theory, and parametrization. *Quarterly Journal of the Royal Meteorological Society*, *118*(507), 819–850. https://doi.org/10.1002/qj.49711850703

Neale, R. B., & Hoskins, B. J. (2000). A standard test for AGCMs including their physical parametrizations: I: The proposal. *Atmospheric Science Letters*, *1*(2), 101–107. https://doi.org/10.1006/asle.2000.0019

O'Gorman, P. A., & Dwyer, J. G. (2018). Using machine learning to parameterize moist convection: Potential for modeling of climate, climate change, and extreme events. *Journal of Advances in Modeling Earth Systems*, *10*(10), 2548–2563. https://doi.org/10.1029/2018ms001351

Orr, A., Bechtold, P., Scinocca, J., Ern, M., & Janiskova, M. (2010). Improved middle atmosphere climate and forecasts in the ECMWF model through a nonorographic gravity wave drag parameterization. *Journal of Climate*, *23*(22), 5905–5926. https://doi.org/10.1175/2010jcli3490.1

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., et al. (2017). Automatic differentiation in pytorch. In *Neural information processing systems workshop, 2017*.

Pauluis, O., & Garner, S. (2006). Sensitivity of radiative–convective equilibrium simulations to horizontal resolution. *Journal of the Atmospheric Sciences*, *63*(7), 1910–1923. https://doi.org/10.1175/jas3705.1

Rasp, S. (2020). Coupled online learning as a way to tackle instabilities and biases in neural network parameterizations: General algorithms and Lorenz 96 case study (v1. 0). *Geoscientific Model Development*, *13*(5), 2185–2196. https://doi.org/10.5194/gmd-13-2185-2020

Rasp, S., Pritchard, M. S., & Gentine, P. (2018). Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(39), 9684–9689. https://doi.org/10.1073/pnas.1810286115

Ray, E. A., Alexander, M. J., & Holton, J. R. (1998). An analysis of the structure and forcing of the equatorial semiannual oscillation in zonal wind. *Journal of Geophysical Research*, *103*(D2), 1759–1774. https://doi.org/10.1029/97jd02679

Richter, J. H., & Rasch, P. J. (2008). Effects of convective momentum transport on the atmospheric circulation in the community atmosphere model, version 3. *Journal of Climate*, *21*(7), 1487–1499. https://doi.org/10.1175/2007jcli1789.1

Romps, D. M. (2012). On the equivalence of two schemes for convective momentum transport. *Journal of the Atmospheric Sciences*, *69*(12), 3491–3500. https://doi.org/10.1175/jas-d-12-068.1

Seifert, A., & Rasp, S. (2020). Potential and limitations of machine learning for modeling warm-rain cloud microphysical processes. *Journal of Advances in Modeling Earth Systems*, *12*(12), e2020MS002301. https://doi.org/10.1029/2020ms002301

Smith, L. N. (2017). Cyclical learning rates for training neural networks. In *2017 IEEE winter conference on applications of computer vision* (pp. 464–472). WACV).

Song, X., Wu, X., Zhang, G. J., & Arritt, R. W. (2008). Understanding the effects of convective momentum transport on climate simulations: The role of convective heating. *Journal of Climate*, *21*(19), 5034–5047. https://doi.org/10.1175/2008jcli2187.1

Wang, P., Yuval, J., & O'Gorman, P. A. (2022). Non-local parameterization of atmospheric subgrid processes with neural networks. *Journal of Advances in Modeling Earth Systems*, *14*(10), e2022MS002984. https://doi.org/10.1029/2022ms002984

Watt-Meyer, O., Brenowitz, N. D., Clark, S. K., Henn, B., Kwa, A., McGibbon, J., et al. (2021). Correcting weather and climate models by machine learning nudged historical simulations. *Geophysical Research Letters*, *48*(15), e2021GL092555. https://doi.org/10.1029/2021GL092555

Woelfle, M., Yu, S., Bretherton, C., & Pritchard, M. (2018). Sensitivity of coupled tropical Pacific model biases to convective parameterization in CESM1. *Journal of Advances in Modeling Earth Systems*, *10*(1), 126–144. https://doi.org/10.1002/2017ms001176

Wu, X., Deng, L., Song, X., & Zhang, G. J. (2007). Coupling of convective momentum transport with convective heating in global climate simulations. *Journal of the Atmospheric Sciences*, *64*(4), 1334–1349. https://doi.org/10.1175/jas3894.1

Yang, W., Seager, R., & Cane, M. A. (2013). Zonal momentum balance in the tropical atmospheric circulation during the global monsoon mature months. *Journal of the Atmospheric Sciences*, *70*(2), 583–599. https://doi.org/10.1175/jas-d-12-0140.1

Yuval, J., & O'Gorman, P. A. (2020). Stable machine-learning parameterization of subgrid processes for climate modeling at a range of resolutions. *Nature Communications*, *11*(1), 1–10. https://doi.org/10.1038/s41467-020-17142-3

Yuval, J., O'Gorman, P. A., & Hill, C. N. (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*, *48*(6), e2020GL091363. https://doi.org/10.1029/2020gl091363

Zanna, L., & Bolton, T. (2020). Data-driven equation discovery of ocean mesoscale closures. *Geophysical Research Letters*, *47*(17), e2020GL088376. https://doi.org/10.1029/2020gl088376