Neural-network parameterization of subgrid momentum transport in the atmosphere

Janni Yuval[1] and Paul A. O'Gorman[2]

[1]Massachusetts Institute of Technology
[2]MIT

November 21, 2022

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 parameterization of subgrid momentum transport that learns from coarse-grained data of a high-resolution atmospheric simulation in an idealized aquaplanet domain. We show that substantial subgrid momentum transport occurs due to convection and non-orographic gravity waves. The parameterization has a structure that ensures the conservation of momentum, and it has reasonable skill in predicting momentum fluxes associated with convection, although its skill is lower as compared to subgrid energy and moisture fluxes. The neural-network parameterization is implemented in the same atmospheric model at coarse resolution and leads to stable simulations. Overall, our results show that it is challenging to predict subgrid momentum fluxes and that machine-learning momentum parameterization gives promising results.

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, Massachusetts 02139, 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

Corresponding author: Janni Yuval, yaniyuval@gmail.com

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 parameterization of subgrid momentum transport that learns from coarse-grained data of a high-resolution atmospheric simulation in an idealized aquaplanet domain. We show that substantial subgrid momentum transport occurs due to convection and non-orographic gravity waves. The parameterization has a structure that ensures the conservation of momentum, and it has reasonable skill in predicting momentum fluxes associated with convection, although its skill is lower as compared to subgrid energy and moisture fluxes. The neural-network parameterization is implemented in the same atmospheric model at coarse resolution and leads to stable simulations. 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

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 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-networks parameterization 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 (Wu et al., 2007; Song et al., 2008; Woelfle et al., 2018). 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 parameterizations 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, machine learning has been been used extensively to emulate conventional parameterizations of convection (O'Gorman & Dwyer, 2018), radiation (Krasnopolsky et al., 2005; Belochitski & Krasnopolsky, 2021), 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-dimesional 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; Rasp, 2020; Brenowitz et al., 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), to learn fine-scale velocities at 100hPa 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 towards an observational analysis (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), and we present the structure the momentum parameterization (Section 4). Next, we investigate the skill of the parameterization in predicting subgrid momentum fluxes (Section 5), and we implement the parameterization in an atmospheric model at coarse resolution and study its effect on the climate (Section 6). Lastly, we give our conclusions (Section 7).

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 sea surface temperature (SST) following 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 12km for the high-resolution simulation (referred to as hi-res), while explicitly representing both convection and planetary scale circulations (Kuang et al., 2005; Pauluis & Garner, 2006; Garner et al., 2007; Boos et al., 2016; Fedorov et al., 2019). At vertical levels above 20km, 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 also ran coarse simulations with horizontal grid spacings of 48km and 96km, which correspond to coarsening the high-resolution grid by factors of 4 and 8, respectively. For each coarse grid spacing, we ran the following simulations:

- Simulations with no NN parameterizations (referred to as x4 and x8 for 48km and 96km grid spacing, respectively),
- Simulations with an NN parameterization only for subgrid effects on thermodynamic and moisture variables (referred to as x4-NN and 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.
- Simulations with an NN parameterization for thermodynamic and moisture variables, and additionally a separate NN parameterization for horizontal momentum variables (x4-NNMOM, x8-NNMOM).

We ran all 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 (x4 and 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 supplementary text S1 in YOH21 for more details).

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):

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

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

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 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

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

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:

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

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 dataset for training the NN parameterization is obtained from 3066 snapshots of 3-hourly model output taken from hi-res (see text S1 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 48 and 96km. 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 a neural network on this collocated grid for both inputs and outputs of the neural network (text S2 and Figure 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). 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 coarse graining factor of 8. Figure 1a,c shows the subgrid fluxes and tendencies for zonal momentum, and Figure 3a 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 are downgradient below 250hPa (Figure 1a). The associated tendencies due to subgrid vertical advection peak in the boundary layer (Figure 1c). 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 800hPa and 950hPa (Figure 1c). In the tropics, there are small negative tendencies in the middle troposphere and positive tendencies around 200-300hPa. 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 S3c), although Richter and Rasch (2008) showed that the Hadley cell strength is more sensitive to the subgrid zonal momentum tendency. Overall in the troposphere, the mean tendencies we calculated from hi-res have some similarities in pattern and magnitude with the tendencies obtained from a conventional convective momentum parameterization coupled to a GCM (Richter & Rasch, 2008). 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 also sharp positive and negative features near 50hPa that are likely related to the sponge layer (Figure 1c).

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 CMT parameterization we use is:

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

where α is an (empirical) regression coefficient, P is the surface precipitation, tilde represents a zonal- and time-mean, and the vertical levels ($z = 0.6$km and $z = 11.2$km) are chosen to approximately match the pressure levels chosen by Yang et al. (2013), which were 925hPa for the lower level and 200hPa for the upper level. We find that there is a high correlation between the simplified CMT parameterization to the calculated subgrid tendencies, and that the slope (α) is within the range found by Yang et al. (2013) (Figure 1e; 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 intervals with coherent structure over the depth of the troposphere (Figure 2a). The subgrid momentum flux peaks simultaneously with (convective) precipitation events (Figure 2a,b), and during these precipitation events the subgrid advective fluxes of energy (Figure 2d) and moisture (not shown) also peak over the depth of the troposphere, implying that these are convective events. Subgrid momentum fluxes occur also at times when no convection occurs, especially in the extratropical regions. For example, Figure S4 shows large momentum fluxes in non-convecting regions. During these time periods when no convection occurs but subgrid momentum fluxes are present, there is no substantial subgrid energy transport (Figure S4d) where the energy is the liquid/ice static energy which is conserved 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 and the source of the large variability of subgrid momentum fluxes in midlatitudes (Figure S5) is non-orographic gravity waves.

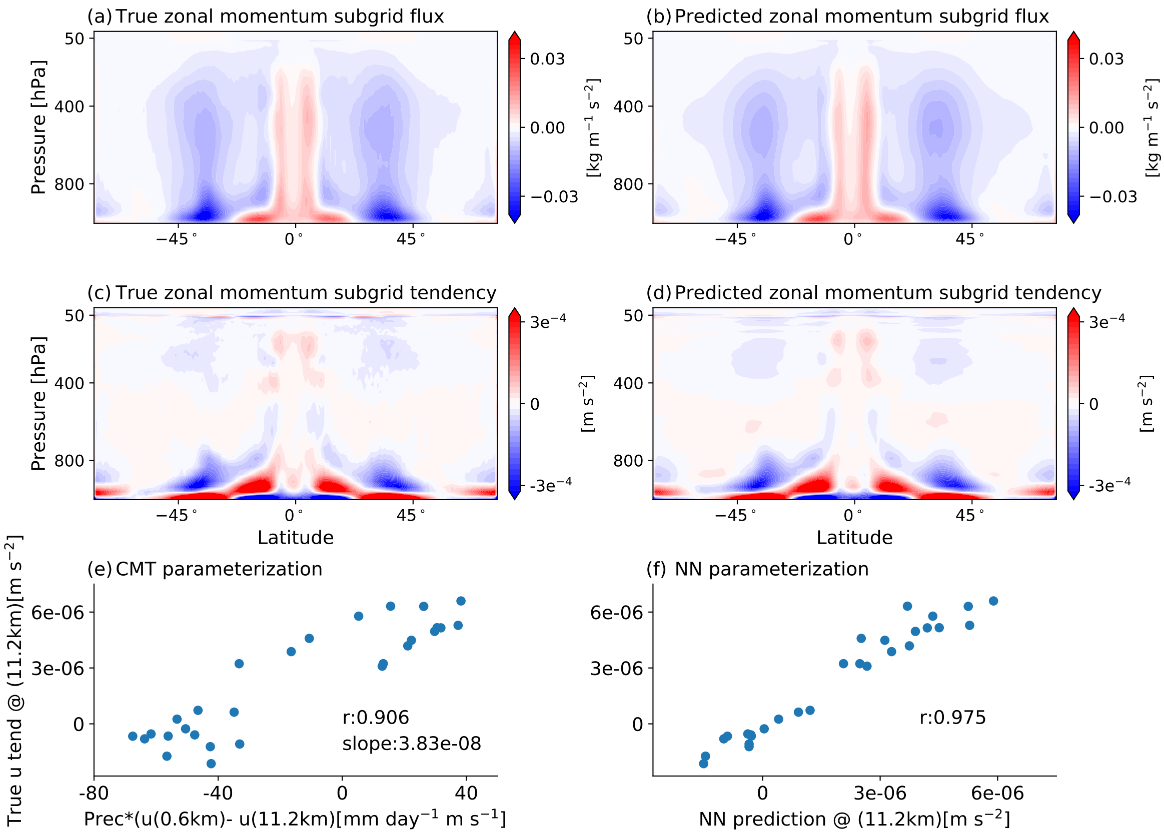

Figure 1. The time- and zonal-mean (a) true (calculated from hi-res) and (b) predicted (NN-MOM predictions) zonal momentum fluxes due to subgrid vertical advection and the associated time- and zonal-mean (c) true and (d) predicted zonal wind tendencies. Colors are saturated in panels a-d to highlight fluxes and tendencies outside of the boundary layer. The sponge layer is active above 50hPa. Panels (e,f) show scatterplots of the zonal- and time-mean of the subgrid zonal momentum tendency for individual tropical latitudes (defined as latitudes within 17.5° of the equator) vs. (e) a simplified convective momentum parameterization (equation 5) and (f) 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 (e), (f) and the slope (α) is given in (e) in units of day mm^{-1} s^{-1} . All quantities are calculated from 501 snapshots of the x8 test data set.

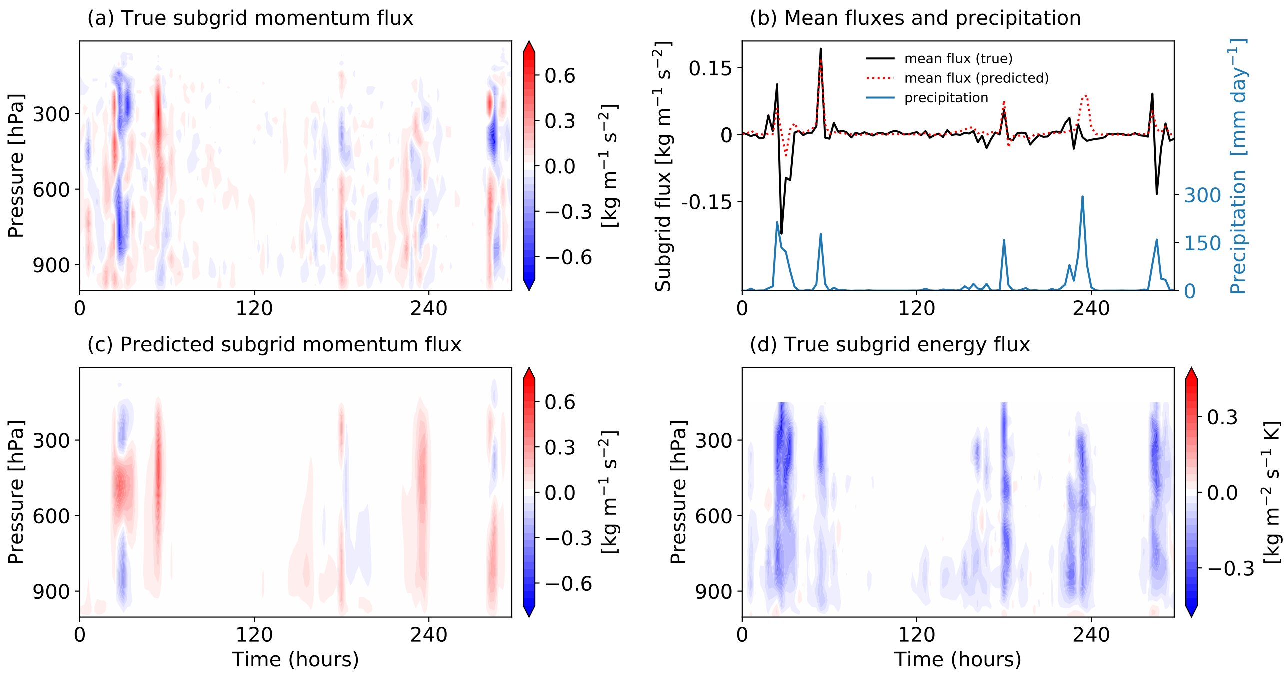

Figure 2. Time series of subgrid fluxes due to vertical advection for a tropical column at latitude 5.6° and a coarse-graining factor of 8: (a) true zonal momentum flux, (b) vertical-mean true (black) and NN-MOM predicted (dotted red) zonal momentum flux, (c) NN-MOM predicted zonal momentum flux, (d) true subgrid energy flux rescaled by the specific heat capacity. Panel (c) 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.

4 Neural network parameterization 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 ($\overline{D}_{\mathrm{mom}}$). Hence, the outputs for NN-MOM are

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

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 5.7km (lowest 15 model levels) because the magnitude of the diffusivity reduces with height. Above 5.7km the 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 ($\overline{D}_{\mathrm{mom}}$) 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.

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 13.9km ($\approx$ 134hPa). Using an NN momentum parameterization that includes q_{T}, 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.

5 Offline performance

We now investigate the offline performance of NN-MOM (i.e., the skill of NN-MOM when it is not coupled to SAM). NN-MOM captures accurately the zonal- and time-mean momentum transport (Figures 1b,d and S3b,d). Furthermore, in the tropical upper troposphere, NN-MOM is more accurate than the simplified conventional CMT parameterization in predicting the zonal- and time-mean vertical advective subgrid zonal momentum transport (Figure 1f).

The offline performance of NN-MOM at the instantaneous time scale is lower for predicting subgrid momentum fluxes due to vertical advection than for the subgrid surface fluxes and the diffusivity (Figure S6), especially at midlatitudes where the subgrid fluxes 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. In the tropics where convection occurs most frequently, NN-MOM has reasonable skill in predicting subgrid fluxes due to vertical advection, and this is particularly evident during convective events (Figure 2c). However, the skill of NN-MOM in all regions is substantially lower compared to the skill of an NN trained to predict moisture and energy fluxes due to subgrid vertical advection (Figure S7e,f). Predicting CMT might be more challenging than predicting moisture and energy convective transport because CMT can be both negative or positive (Figure 2a), 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 2d). We find that when a neural network is trained to predict the absolute value of subgrid momentum fluxes, 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 (Figure S7c,d).

6 Online performance

To investigate the effect of NN-MOM on the circulation, we compare simulations with an NN parameterization for thermodynamic and moisture variables but no momentum parameterization (x4-NN, x8-NN) to simulations with NN parameterizations both for thermodynamic, moisture and horizontal momentum variables (x4-NNMOM, x8-NNMOM). We note that NN-MOM is an additional neural network on top of 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 5.7km. This diffusivity is used at every time step in the momentum diffusion scheme for levels below 5.7km. We also compare to simulations with no NN parameterization (x4, 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 S1. Consistent with the results for x8-NN in YOH21, we find that use of the NN parameterization for thermodynamic and moisture variables in x4-NN and x8-NN brings the climatology much closer to hi-res than simulations without any NN parameterization. From here on, we will focus on the effect of the momentum parameterization in x4-NNMOM and x8-NNMOM which is smaller in magnitude compared to the effect of the energy and moisture parameterization but nonetheless important.

Inclusion of NN-MOM leads to noticeable changes in the mean horizontal winds in x8-NNMOM compared to x8-NN (Figure 3c,f). The zonal wind weakens across the stratosphere (Figure 3c), such that biasses in the stratosphere of x8-NN are reduced (Figure 3b). Such a weakening of the stratospheric circulation is expected when introducing a subgrid parameterization for momentum due to gravity wave drag, although we note that the stratospheric zonal wind in the idealized simulation is not realistic. Furthermore, the subtropical wind is more easterly (Figure 3c), and the meridional wind near the surface and at the surface weakens (Figure 3f, Figure S8) again reducing biases in x8-NN (Figure 3e). Such an improvement in the simulation of surface winds could be important for coupled ocean-atmosphere simulations. Overall x8-NNMOM is much closer to the climatology of hi-res compared to x8-NN (Table S1). The pattern of changes in x4-NNMOM relative to x4-NN are similar to the changes in x8-NNMOM relative to x8-NN and oppose the biases relative to hi-res in many regions, but the magnitude of the change is bigger (Figure 3c,f,i,l) for reasons that remain unclear. As a result, the climatology of x4-NNMOM degrades compared to x4-NN (Table S1), due to an overshoot in the effect of NN-MOM on the circulation (e.g., in x4-NNMOM the meridional surface winds have a bias with opposite sign compare to x4-NN). 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).

7 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-

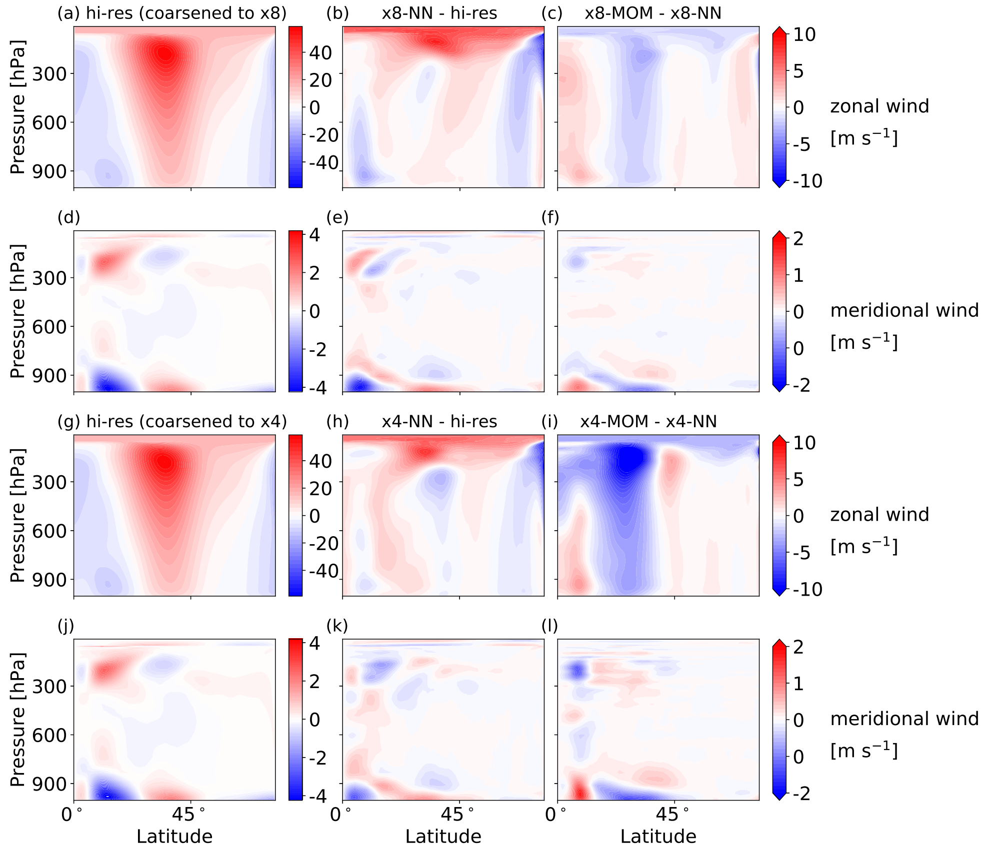

Figure 3. The zonal- and time-mean zonal and meridional wind as a function of pressure and latitude for simulations with coarser grids by factors of 8 (two upper rows) and 4 (two lower rows) compared to hi-res. First column shows the zonal- and time-mean zonal wind (a,g) and meridional wind (d,j) for hi-res (coarsened to x8 and x4, respectively). The second column (panels b, e, h, k) shows the difference between the coarse resolution simulations with NN parameterization for thermodynamic and moisture variables and hi-res, and the third column (panels c, f, i, l) show the difference between simulations with NN parameterization for thermodynamic, moisture and momentum variables and simulations with NN parameterization only for thermodynamic and moisture variables. The results were averaged over both hemispheres to obtain better statistics.

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 convective momentum transport. 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, but it has little 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 nontrivial task of determining the sign of convective momentum transport. 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. Simulations with the NN momentum parameterization run stably and without climate drift. We found that the momentum parameterization corrects some of the biases relative to hi-res in the near surface zonal and meridional wind and the zonal wind in the stratosphere. However, while in one coarse resolution there is an overall improvement in the simulation of the mean meridional, zonal and vertical winds as well as precipitation, in the other simulation (which has a finer grid spacing that is relatively close to that of hi-res) the inclusion of the NN momentum parameterization creates new biasses due to an overshoot in its effect on the circulation. 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. 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.

Acknowledgments
We thank Bill Boos for providing the output from the 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 (doi:10.5065/D6RX99HX) provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation. JY acknowledges support from the EAPS Houghton-Lorenz postdoctoral fellowship. POG acknowledges support from NSF awards AGS-1749986 and OAC 1835618.

Data availability

Associated code, processed data from online simulations, trained neural network parameterizations and (a link to) the output of the high-resolution simulation are available at zenodo.org (https://doi.org/10.5281/zenodo.5083483).

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: Atmospheres*, *108*(D19).

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*.

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.

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.

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*, 2727-2744.

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.

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: Atmospheres*, *102*(D22), 26053–26076.

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.

Fritts, D. C., & Nastrom, G. D. (1992). Sources of mesoscale variability of gravity waves. part ii: Frontal, convective, and jet stream excitation. *Journal of Atmospheric Sciences*, *49*(2), 111–127.

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.

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*, e2020MS002268.

Gregory, D., Kershaw, R., & Inness, P. (1997). Parametrization of momentum transport by convection. ii: Tests in single-column and general circulation models. *Q. J. R. Meteorol. Soc.*, *123*(541), 1153–1183.

Kershaw, R., & Gregory, D. (1997). Parametrization of momentum transport by convection. i: Theory and cloud modelling results. *Q. J. R. Meteorol. Soc.*, *123*(541), 1133–1151.

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.

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.

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). Retrieved from https://doi.org/10.1029/2004GL021024 doi: 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.

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.

LeMone, M. A. (1983). Momentum transport by a line of cumulonimbus. *Journal of Atmospheric Sciences*, *40*(7), 1815–1834.

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.

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.

Moncrieff, M. W. (1992). Organized convective systems: Archetypal dynamical models, mass and momentum flux theory, and parametrization. *Q. J. R. Meteorol. Soc.*, *118*(507), 819–850.

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

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.

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.

Pauluis, O., & Garner, S. (2006). Sensitivity of radiative–convective equilibrium simulations to horizontal resolution. *Journal of the atmospheric sciences*, *63*(7), 1910–1923.

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.

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*, 9684–9689.

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: Atmospheres*, *103*(D2), 1759–1774.

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.

Romps, D. M. (2012). On the equivalence of two schemes for convective momentum transport. *Journal of the Atmospheric Sciences*, *69*(12), 3491–3500.

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.

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.

Watt-Meyer, O., Brenowitz, N. D., Clark, S. K., Henn, B., Kwa, A., McGibbon, J. J., … Bretherton, C. S. (2021). Correcting weather and climate models by machine learning nudged historical simulations.

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.

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.

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.

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.

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.

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

Supporting Information for "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, Massachusetts 02139, USA

Contents of this file

1. Text S1 to S2
2. Figures S1 to S9
3. Table S1

Here we describe the train and test datasets, training protocol, architecture of the neural networks and the implementation of the neural network (NN) parameterization in SAM (text S1), and the coarse-graining methods we use (text S2). We also add several figures to support our findings and a table showing the online performance of simulations we run (Figs. S1-S9 and Table S1).

Text S1. 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. The NN parameterization for horizontal momentum variables (NN-MOM) is a separate NN in addition to NN1 and NN2, and its inputs and outputs are described in details in the manuscript. 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 dataset, number of epochs and learning rate. The training data for the NN momentum parameterization is obtained from 383.25 days of 3-hourly snapshots of model output taken from the hi-res simulation (overall 3066 snapshots). This data was split into train and test datasets, where the first 320.625 days (2565 snapshots) were used for training, and the last 62.625 days (501 snapshots) were used as a test dataset. 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 to 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 implemented 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 1024 samples in each batch and train over 8000 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 NNs are stored as netcdf files, and then implemented in SAM using a Fortran module. 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.

We do not present results for the NN momentum parameterization at a coarser grid spacing of 192km (corresponding to coarse graining factor of 16) since we find that the exact structure and width of the ITCZ at this grid spacing is sensitive to the structure of the NN parameterization of moisture and energy, which makes it difficult to choose a baseline simulation to compare against. This sensitivity is possibly related to results shown in a previous study that found that in an aquaplanet configuration with hemispherically symmetric SST the ITCZ structure is very sensitive to the exact parameterization (Möbis & Stevens, 2012).

Text S2. Coarse-graining on a collocated grid and on a staggered C-grid

As explained in the manuscript, we use two different coarse graining protocols: (a) coarse graining on a collocated grid and (b) coarse graining on a staggered C-grid. Coarse graining on the collocated grid is used for the offline results presented in the paper, while coarse graining on the staggered grid is needed for learning the parameterization that is actually used online in SAM which uses a staggered C-grid.

In order to get coarse-grained variables on a collocated grid, each variable is coarse-grained slightly differently, depending on which grid it is found on in hi-res (Figure S1). Specifically, quantities that are found on the horizontal u grid (i.e., u, vertical fluxes of zonal momentum, zonal momentum surface fluxes) are coarse grained using (red circles in Figure S1):

$$\overline{A}(i,j,k) = \frac{1}{N^2}\Big[\sum_{l=N(i-1)+2}^{l=Ni}\ \sum_{m=N(j-1)+1}^{m=Nj} A(l,m,k) + \frac{1}{2}\sum_{m=N(j-1)+1}^{m=Nj} A(N(i-1)+1,m,k) + \frac{1}{2}\sum_{m=N(j-1)+1}^{m=Nj} A(Ni+1,m,k)\Big], \tag{1}$$

where A is the high-resolution variable, $\overline{A}$ is the coarse-grained variable, N is the coarse graining factor, k is the index of the vertical level, and i, j (l, m) are the discrete indices of the longitudinal and latitudinal coordinates at coarse resolution (high resolution). Similarly quantities that are found on the horizontal v grid (i.e., v, vertical fluxes of meridional momentum, meridional momentum surface fluxes) are coarse grained using (red triangles in Figure S1):

$$\overline{A}(i,j,k) = \frac{1}{N^2}\Big[\sum_{l=N(i-1)+1}^{l=Ni}\ \sum_{m=N(j-1)+2}^{m=Nj} A(l,m,k) + \frac{1}{2}\sum_{l=N(i-1)+1}^{l=Ni} A(l,N(j-1)+1,k) + \frac{1}{2}\sum_{l=N(i-1)+1}^{l=Ni} A(l,Nj+1,k)\Big], \tag{2}$$

and quantities that are found on the horizontal w grid (e.g., w, H_{L}, q_{T}) are coarse grained using (red stars in Figure S1):

$$\overline{A}(i,j,k) = \frac{1}{N^2}\sum_{l=N(i-1)+1}^{l=Ni}\ \sum_{m=N(j-1)+1}^{m=Nj} A(l,m,k). \tag{3}$$

Likewise, in order to get coarse-grained variables on a staggered C-grid, each variable is coarse grained slightly differently, depending on which grid it is found on in hi-res (Figure S2). To obtain coarse-grained variables on a C-grid, the quantities that are found on the horizontal u grid are coarse-grained using (red circles in Figure S2):

$$\begin{aligned}\overline{A}(i,j,k) \;=\; & \frac{1}{N^2}\Big[\sum_{l=N(i-1.5)+2}^{l=N(i-0.5)}\;\sum_{m=N(j-1)+1}^{m=Nj} A(l,m,k)\\ & +\frac{1}{2}\sum_{m=N(j-1)+1}^{m=Nj} A(N(i-1.5)+1,m,k)\\ & +\frac{1}{2}\sum_{m=N(j-1)+1}^{m=Nj} A(N(i-0.5)+1,m,k)\Big],\end{aligned} \tag{4}$$

the quantities that are found on the horizontal v grid are coarse-grained using (red triangles in Figure S2):

$$\begin{aligned}\overline{A}(i,j,k) \;=\; & \frac{1}{N^2}\Big[\sum_{l=N(i-1)+1}^{l=Ni}\;\sum_{m=N(j-1.5)+2}^{m=N(j-0.5)} A(l,m,k)\\ & +\frac{1}{2}\sum_{l=N(i-1)+1}^{l=Ni} A(l,N(j-1.5)+1,k)\\ & +\frac{1}{2}\sum_{l=N(i-1)+1}^{l=Ni} A(l,N(j-0.5)+1,k)\Big],\end{aligned} \tag{5}$$

and the quantities that are found on the horizontal w grid are coarse-grained using equation 3 (red stars in Figure S2). Note that N is assumed to be even.

The two different coarse-graining protocols we use lead to a similar mean subgrid fluxes (not shown), but calculating the subgrid fluxes using a C-grid leads to a much larger variability in the fluxes because of the spatial staggering of the coarse-grained vertical momentum fluxes and the resolved (coarse-grained) w, and this effect is particularly large in the extratropical jet regions (Figure S9).

We note that different choices of coarse graining protocols require modifications of some of the numerical schemes when calculating subgrid terms. For example, when using variables on a staggered C-grid, both resolved flux and the flux in hi-res due to vertical advection of zonal momentum are calculated as:

$$(u)_{\mathrm{adv}}^{\mathrm{subg-flux}}(i,j,k) = \frac{\rho_w(k)}{4}(w(i,j,k)+w(i-1,j,k))(u(i,j,k)+u(i,j,k-1)). \tag{6}$$

where ρ_w is the reference density profiles defined on the w grid, which are staggered vertically, and i, j, k are the discrete indices of the longitudinal, latitudinal and vertical coordinates, respectively. In contrast, when using a (horizontal) collocated grid, the resolved flux due to vertical advection of zonal momentum is calculated as:

$$(u)_{\mathrm{adv}}^{\mathrm{subg-flux}}(i,j,k) = \frac{\rho_w(k)}{2}w(i,j,k)(u(i,j,k)+u(i,j,k-1)), \tag{7}$$

while the hi-res flux is calculated using equation 6.

References

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

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.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., ... Lerer, A. (2017). Automatic differentiation in pytorch. In *Neural information processing sys-*

tems workshop, 2017.

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

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.

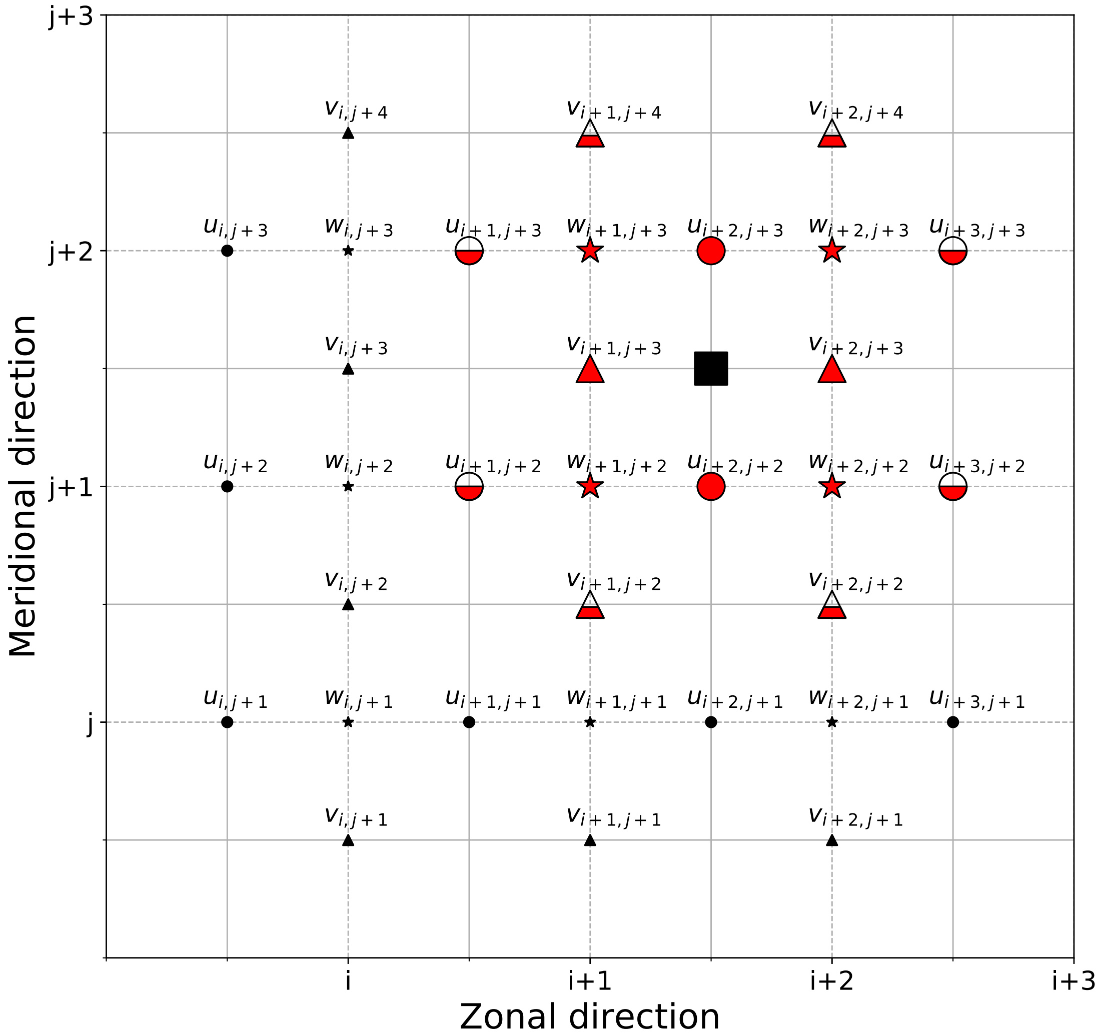

Figure S1. Illustration of the C-grid used in SAM and the coarse graining procedure done to achieve a **collocated grid** for the coarse-grained variables using a coarse-graining factor of 2 for simplicity. Circles, triangles and stars represent the high-resolution grids of u, v, and w, respectively. Red symbols show the grid points that are averaged for the coarse graining procedure that gives a point on the collocated coarse grid represented by the square. Symbols that are half-filled are weighted with a factor of 0.5 when coarse graining (see equations 1 and 2). Note that the square shows the collocated coarse-grained grid for u, v, and w.

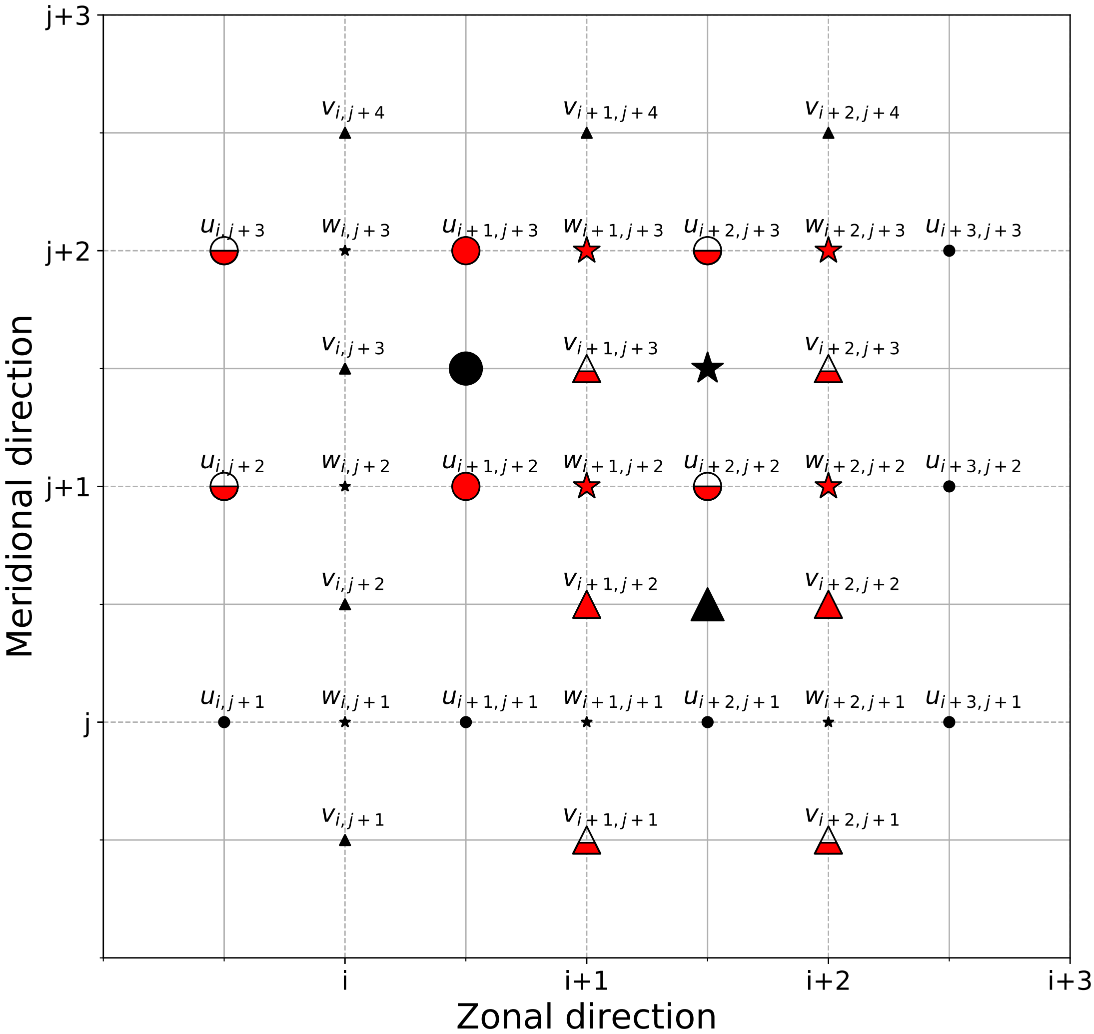

Figure S2. Illustration of the C-grid used in SAM and the coarse graining procedure done to achieve a **C-grid** for the coarse-grained variables using a coarse-graining factor of 2 for simplicity. Red and smaller black circles, triangles and stars represent the high-resolution grids of u, v, and w, respectively. Red symbols show grid points that are averaged for the coarse graining procedure that gives a coarse grid represented by the larger black circle, triangle and star (representing grid points in the coarse C-grid of u, v, and w, respectively). Symbols that are half-filled are weighted with a factor of 0.5 when coarse graining (see equations 4 and 5).

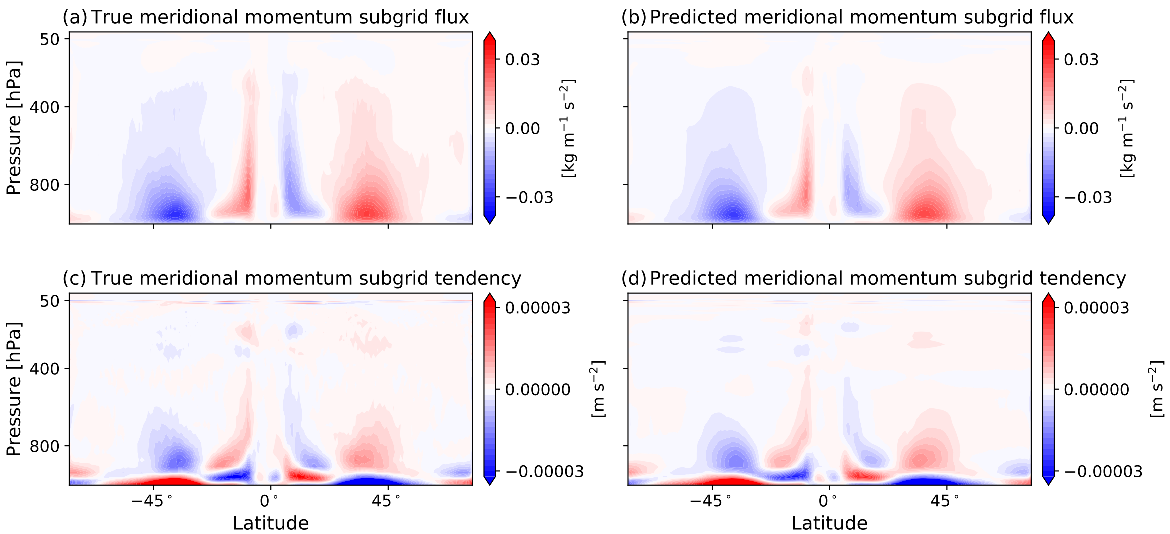

Figure S3. The time- and zonal-mean (a) true (calculated from hi-res) and (b) predicted (NN-MOM predictions) meridional momentum fluxes due to subgrid vertical advection and the associated time- and zonal-mean (c) true and (d) predicted meridional wind tendencies. Colors are saturated to highlight fluxes and tendencies outside of the boundary layer. The sponge layer is active above 50hPa. All quantities are calculated from 501 snapshots of the x8 test data set.

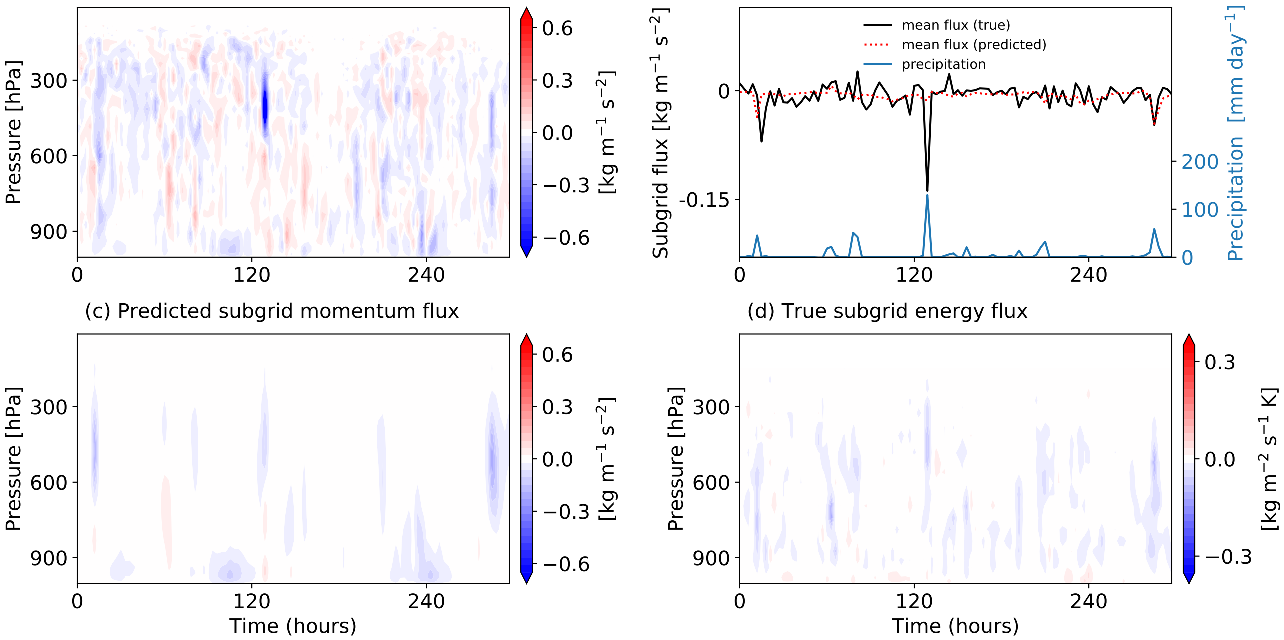

Figure S4. Time series of subgrid fluxes due to vertical advection for an extratropical column at latitude 35.3° and a coarse-graining factor of 8: (a) true zonal momentum flux, (b) vertical-mean true (black) and NN-MOM predicted (dotted red) zonal momentum flux, (c) NN-MOM predicted zonal momentum flux, (d) true subgrid energy flux rescaled by the specific heat capacity. Panel (c) 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. The scale of the colorbars is chosen to be the same to Figure 2 in the main paper.

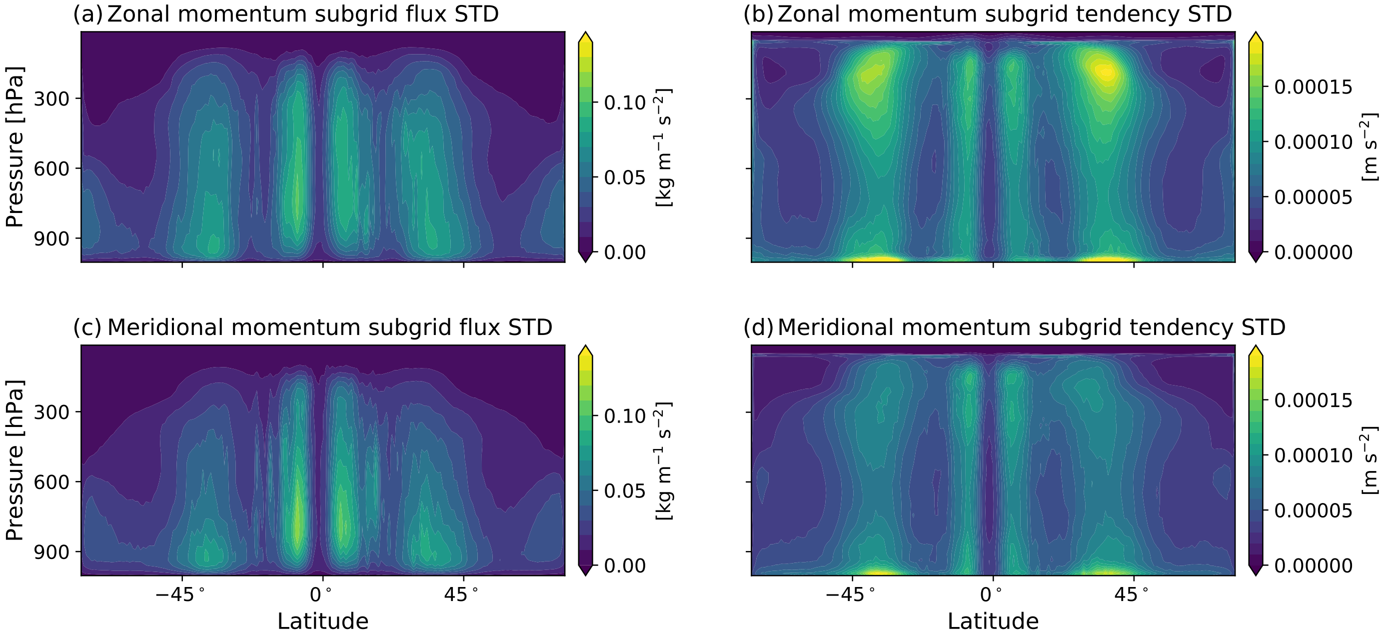

Figure S5. The standard deviation of subgrid terms due to vertical advection as a function of pressure and latitude when using coarse graining on a **collocated grid** for (a) zonal momentum flux, (b) zonal wind tendency, (c) meridional momentum flux, and (d) meridional wind tendency, as calculated from 501 snapshots of x8 test data set.

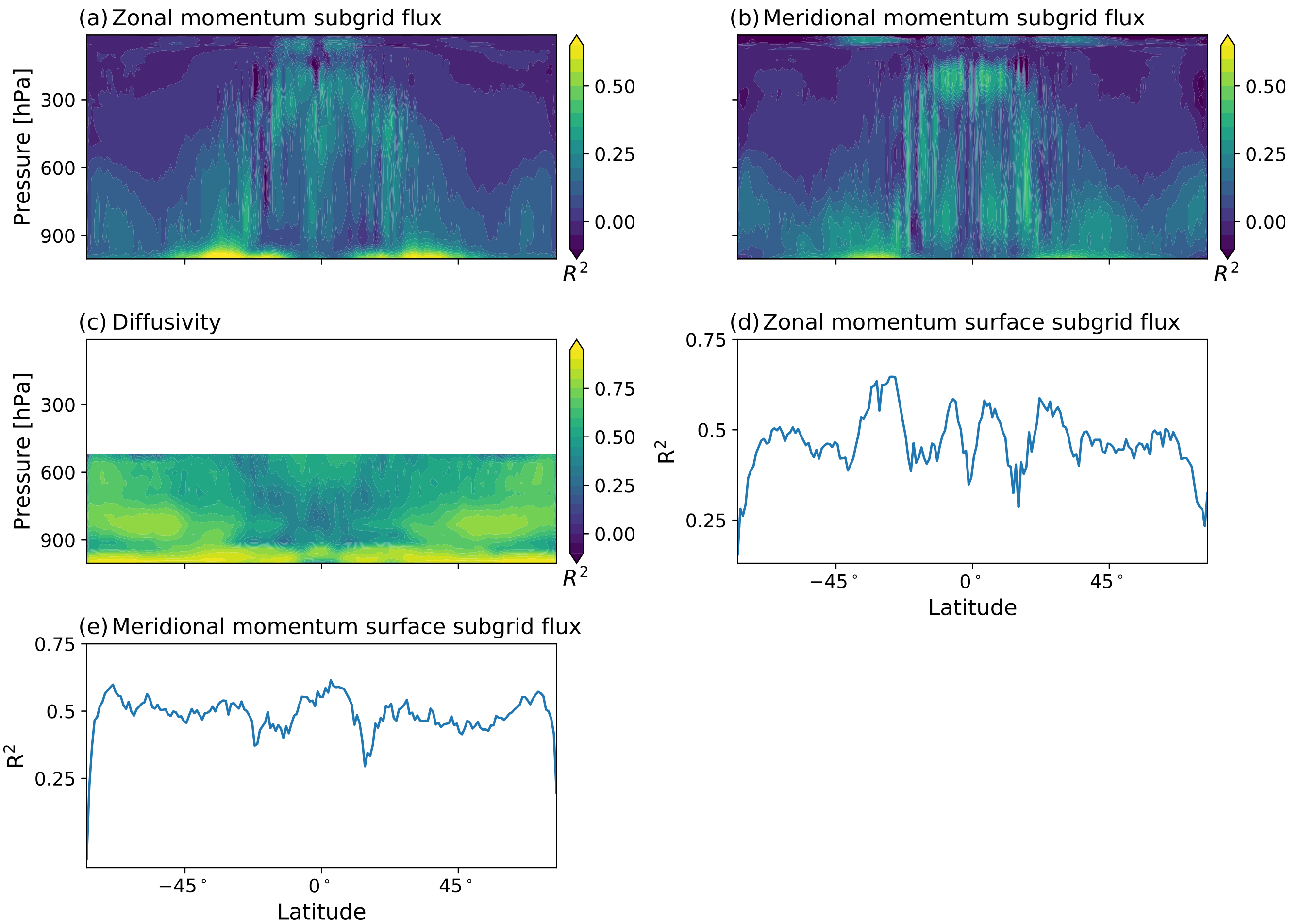

Figure S6. The coefficient of determination (R^2) for offline performance of NN-MOM for the: (a) subgrid zonal momentum fluxes due to vertical advecton (b) subgrid meridional momentum fluxes due to vertical advecton and (c) turbulent diffusivity as a function of pressure and latitude, and of (d) subgrid surface zonal momentum flux and (e) subgrid surface meridional momentum flux as a function of latitude. The coefficient of determination is calculated from 501 snapshots of x8 test data set.

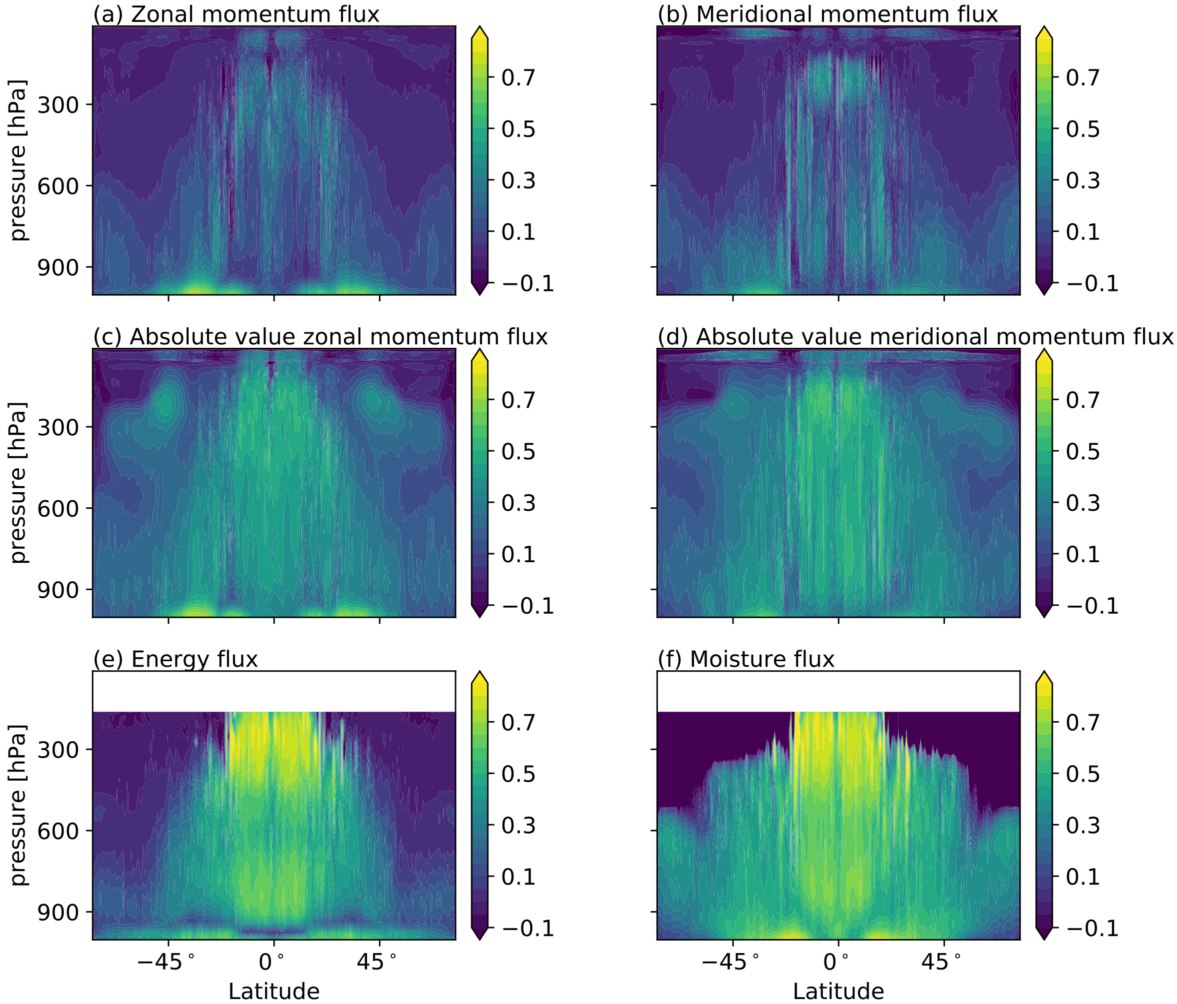

Figure S7. Comparing offline performance (defined as the coefficient of determination) at x8 of NN-MOM that predicts the subgrid fluxes of (a) zonal and (b) meridional momentum, of an NN that was trained to predict the absolute value of subgrid fluxes (NN-MOM-ABS) of (c) zonal and (d) meridional momentum, 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.

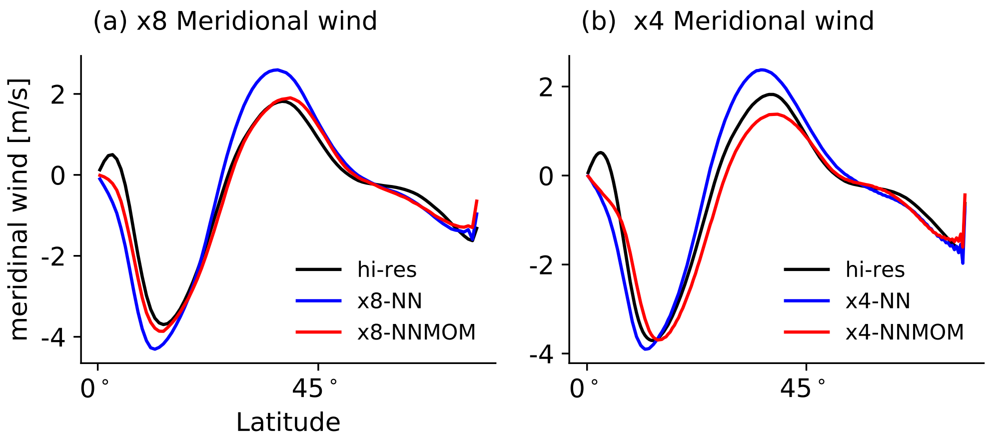

Figure S8. The zonal- and time-mean meridional surface wind as a function of latitude for simulations with coarser grids by factors of (a) 8 and (b) 4 compared to hi-res. Results are shown for simulations with NN parameterization for thermodynamic and moisture variables (blue), simulations with NN parameterization for thermodynamic, moisture and momentum variables (red) and hires (black; coarsened to x8 and x4, respectively).

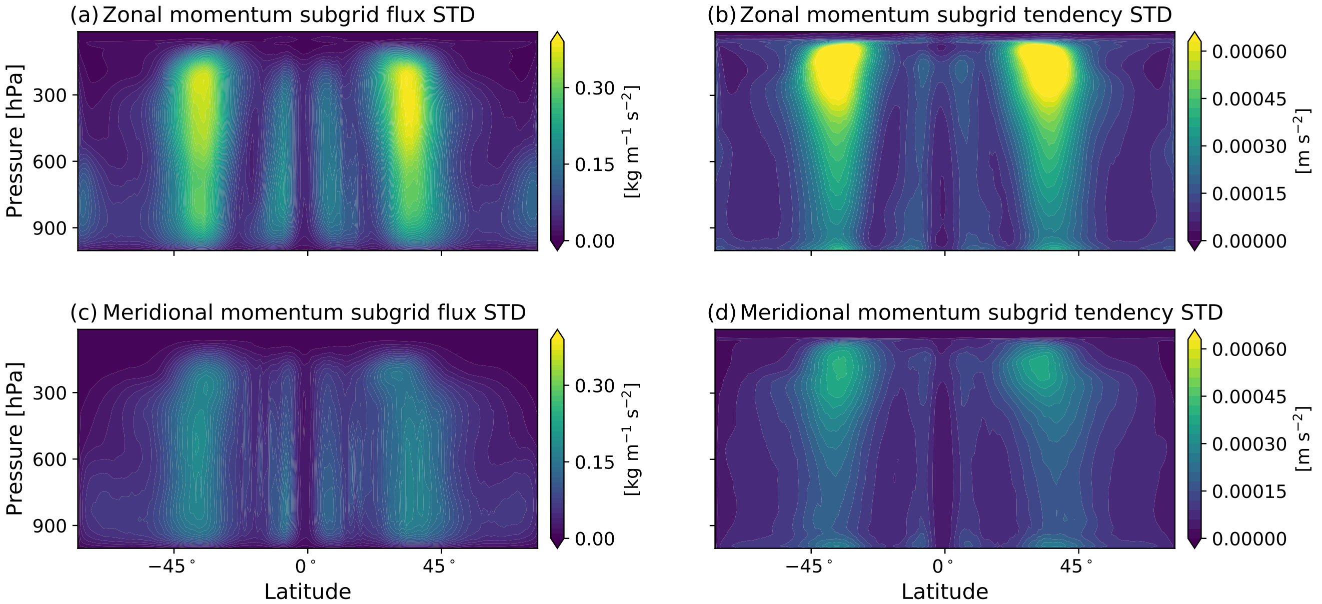

Figure S9. The standard deviation of subgrid terms due to vertical advection as a function of pressure and latitude when using coarse graining on a **staggered C-grid** for (a) zonal momentum flux, (b) zonal wind tendency, (c) meridional momentum flux, and (d) meridional wind tendency, as calculated from 501 snapshots of x8 test data set.

	RMSE relative to hi-res					
	x8	x8-NN	x8-NNMOM	x4	x4-NN	x4-NNMOM
Zonal wind [m s^{-1}]	4.51	3.21	2.71	4.48	2.70	2.85
Meridional wind [m s^{-1}]	0.63	0.25	0.17	0.66	0.21	0.23
Vertical wind [cm s^{-1}]	0.383	0.079	0.057	0.402	0.067	0.123
EKE [m^2 s^{-2}]	42.89	34.91	34.70	41.70	28.22	34.23
Precipitation [mm day^{-1}]	4.69	0.71	0.44	4.80	0.39	1.13
Streamfunction [kg m^{-1} s^{-1} $\times 10^2$]	9.69	2.17	1.36	9.87	2.06	3.20

Table S1. Online performance as measured by the root mean square error (RMSE) of the time- and zonal-mean zonal wind, meridional wind, vertical wind, eddy kinetic energy, precipitation, and the mass streamfunction. The RMSE is calculated relative to hi-res for the coarse-resolution simulations with no ML parameterization (x4, x8), the simulations with the NN parameterization only for thermodynamic and moisture variables (x4-NN, x8-NN) and for the simulations with NN parameterization for thermodynamic, moisture and horizontal momentum variables (x4-NNMOM, x8-NNMOM). The mass streamfunction is defined here as $1/g \int_p^{p_s} [v] dp$, where g is the gravitational acceleration, p_s is the pressure surface and $[v]$ is the zonal- and time-mean meridional velocity. The eddy kinetic energy is defined with respect to the zonal and time mean.