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 S3
2. Figures S1 to S11
3. Table S1-S2

Here we describe the coarse-graining methods (text S1), the hemispherically asymmetric high-resolution simulation (text S2), and the train and test datasets and training protocol for the data obtained from the hemispherically asymmetric configuration (text S3). We also add several figures and tables to support our findings (Figs. S1-S11 and Tables S1-S2).

Text S1. 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 all 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):

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

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):

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

and quantities that are found on the horizontal w (horizontal) 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):

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

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

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

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 S4).

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:

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

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:

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

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

Text S2. High-resolution simulation in hemispherically asymmetric configuration

To help explore different configurations, we ran an additional high-resolution simulation in an aquaplanet configuration. In contrast to the high-resolution simulation introduced in the main text, this high-resolution simulation is not hemispherically symmetric. Instead the radiation is set to perpetual solstice and the SST is prescribed to have an off-equatorial maximum, which leads to major changes in the general circulation including the winds and convection. The SST is prescribed as:

$$SST = T_{\text{min}} + (T_{\text{max}} - T_{\text{min}}) \left(1 - 0.5\text{sin}^2[\frac{3}{2}(\phi - \phi_{\text{max}})] - 0.5\text{sin}^4[\frac{3}{2}(\phi - \phi_{\text{max}})]\right), \tag{8}$$

for $|\phi - \phi_{\text{max}}| < 60^\circ$ and $SST = T_{\text{min}}$ otherwise, where ϕ is the latitude which is calculated on the equatorial beta plane as dist $\times$ $360/(4 \times 10^7)$ where dist is the distance from the equator in meters, $\phi_{\text{max}} = 9^\circ$, $T_{\text{min}} = 0$°C and $T_{\text{max}} = 29$°C. This SST configuration was previously used by Fedorov, Muir, Boos, and Studholme (2019) to study tropical cyclones in an aquaplanet configuration where a hypohydrostatic rescaling factor was also used. The domain of the simulation has a width of $17,280$km and a meridional extent of $17,280$km. In all other aspects this simulation is set up in the same way as the one described in the main text.

To test the NN parameterizations, we ran simulations with a horizontal grid spacing of 96km (which correspond to coarsening the high-resolution grid by factor of 8):

• A simulation with no NN parameterizations (referred to as x8-asym).

• A simulation with an NN parameterization only for subgrid effects on thermodynamic and moisture variables (referred to as x8-asym-NN).

• A simulation with an NN parameterization for thermodynamic and moisture variables, and additionally a separate NN parameterization for horizontal momentum variables (x8-asym-NNMOM).

We ran all simulations for 300 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 200 days of each simulation. We ran these simulation for a shorter period compared to the hemispherically symmetric simulations since these simulations are run on a wider domain. The initial conditions for simulations with NN parameterizations are taken from the last time step of the simulation with no NN parameterization.

Text S3. Training and neural networks in hemispherically asymmetric configuration

Slight modifications in the inputs were needed when running the hemispherically asymmetic cofiguration (Text S2). For the NN that predicts radiation (NN1 in (Yuval et al., 2021)) we use slightly different inputs. Instead of the distance from equator, we use both SST and the incoming solar radiation. The reason that we make this change is that in the asymmetric configuration the SST and incoming solar radiation are not a function of the distance from equator. Furthermore, in the NN that predicts surface moisture and energy fluxes (NN2 in (Yuval et al., 2021)) we use the SST instead of the distance from equator.

The training data for the NN momentum parameterization in the asymmetric configuration is obtained from 200 days of 3-hourly snapshots of model output taken from the hi-res simulation (overall 1600 snapshots). This data was split into train and test datasets, where the first 160 days (1280 snapshots) were used for training, and the last 40 days (320 snapshots) were used as a test dataset. We note that since we used the exact same training protocol described in Text S2, no retuning of hyperparameters was done and the test data was only used to verify that the offline results are similar to the results for the hemispherically symmetric case. For each 3-hourly snapshot that was used during training, we reduced the training data set size by randomly subsampling 72 (out of 144) atmospheric columns at each latitude for each snapshot. This subsampling of the training data enables uploading all training data into the RAM during training. This results in training datasets size of $13,271,040$ for x8-asym. In all other aspects, the training procedures are identical to what is described in the main text.

References

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.

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.

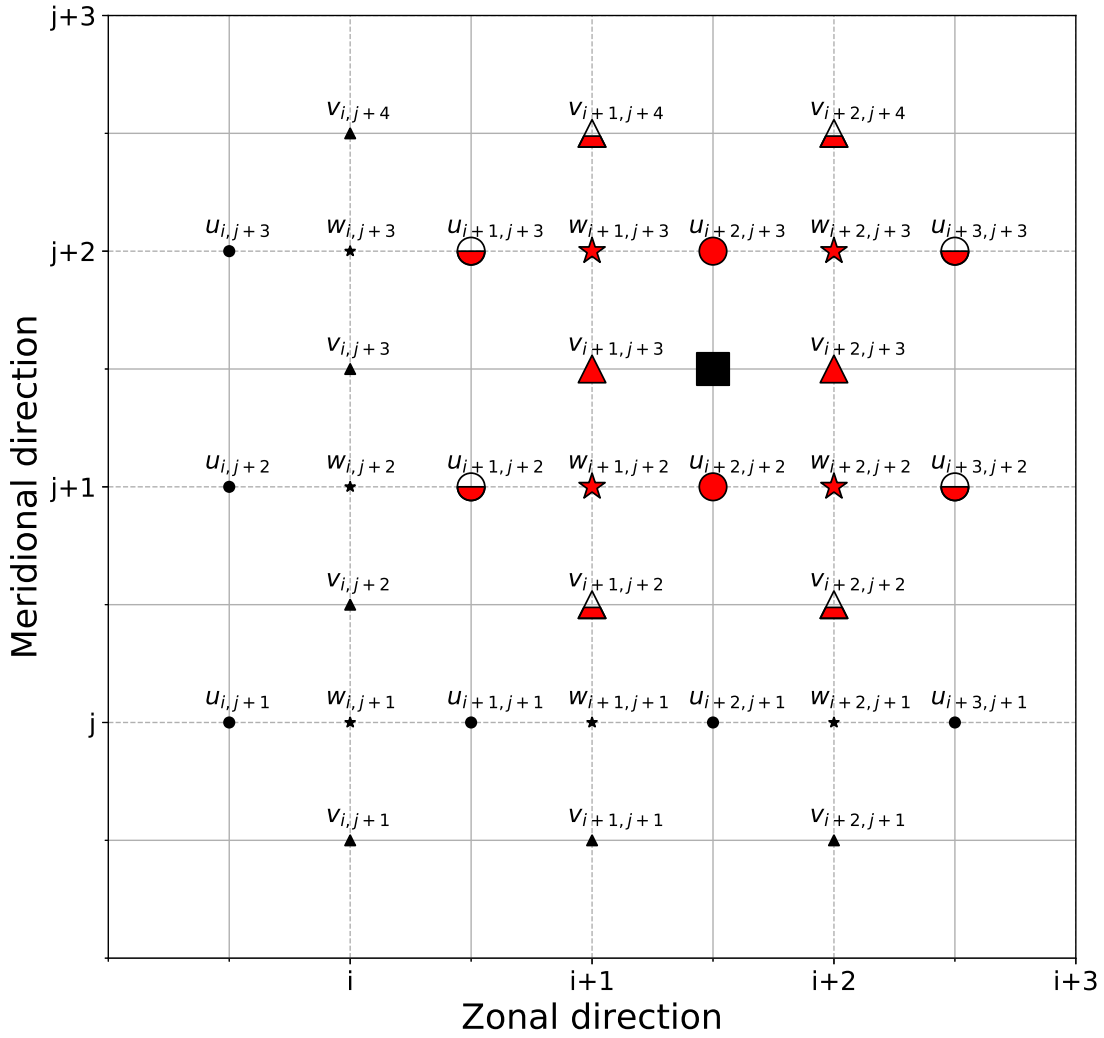

Figure S1. Illustration of the C-grid used in SAM and the coarse graining procedure done to achieve a **collocated grid** for the coarse-grained variables using a coarse-graining factor of 2 for simplicity. Circles, triangles and stars represent the high-resolution grids of u, v, and w, respectively. 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.

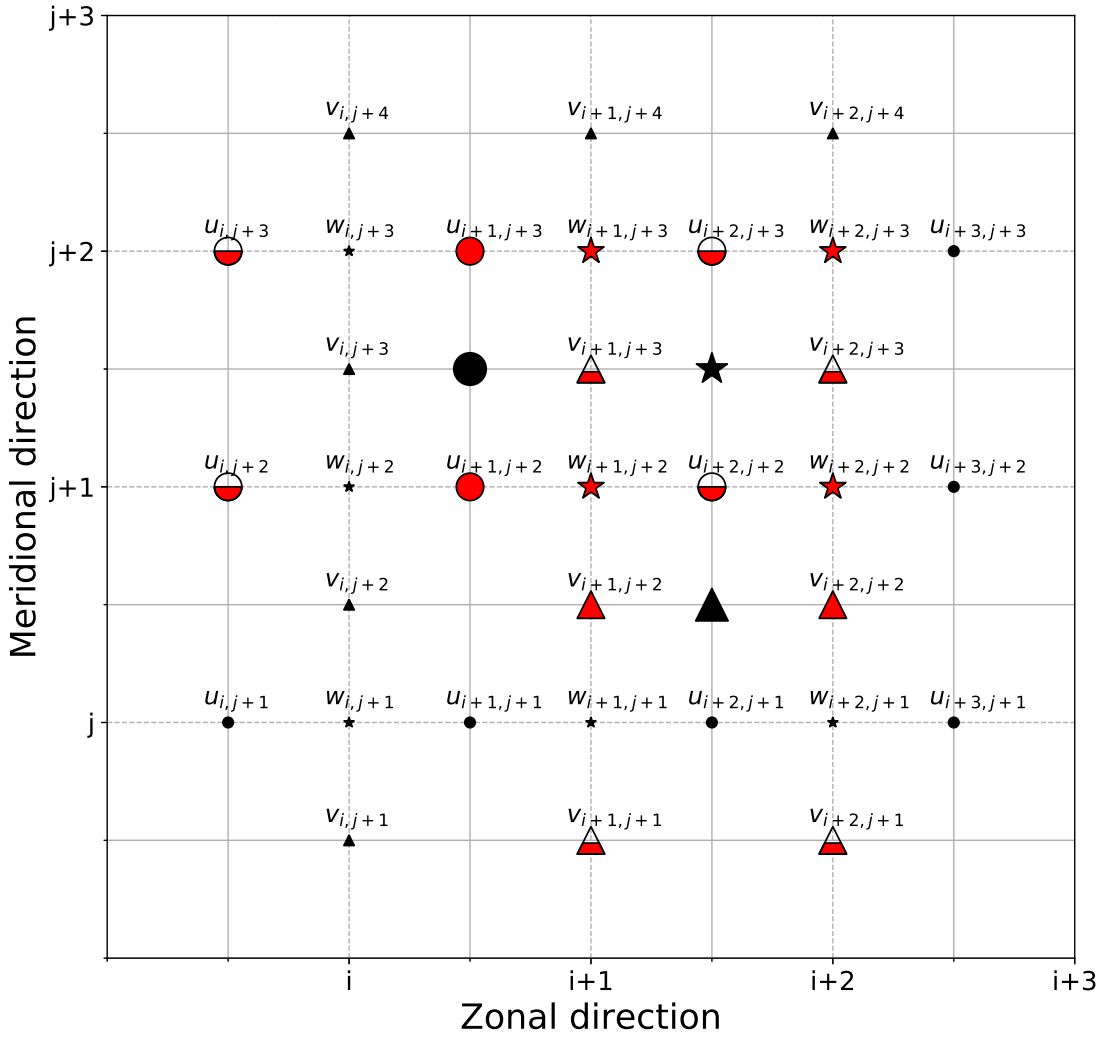

Figure S2. Illustration of the C-grid used in SAM and the coarse graining procedure done to achieve a **C-grid** for the coarse-grained variables using a coarse-graining factor of 2 for simplicity. Red and smaller black circles, triangles and stars represent the high-resolution grids of u, v, and w, 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).

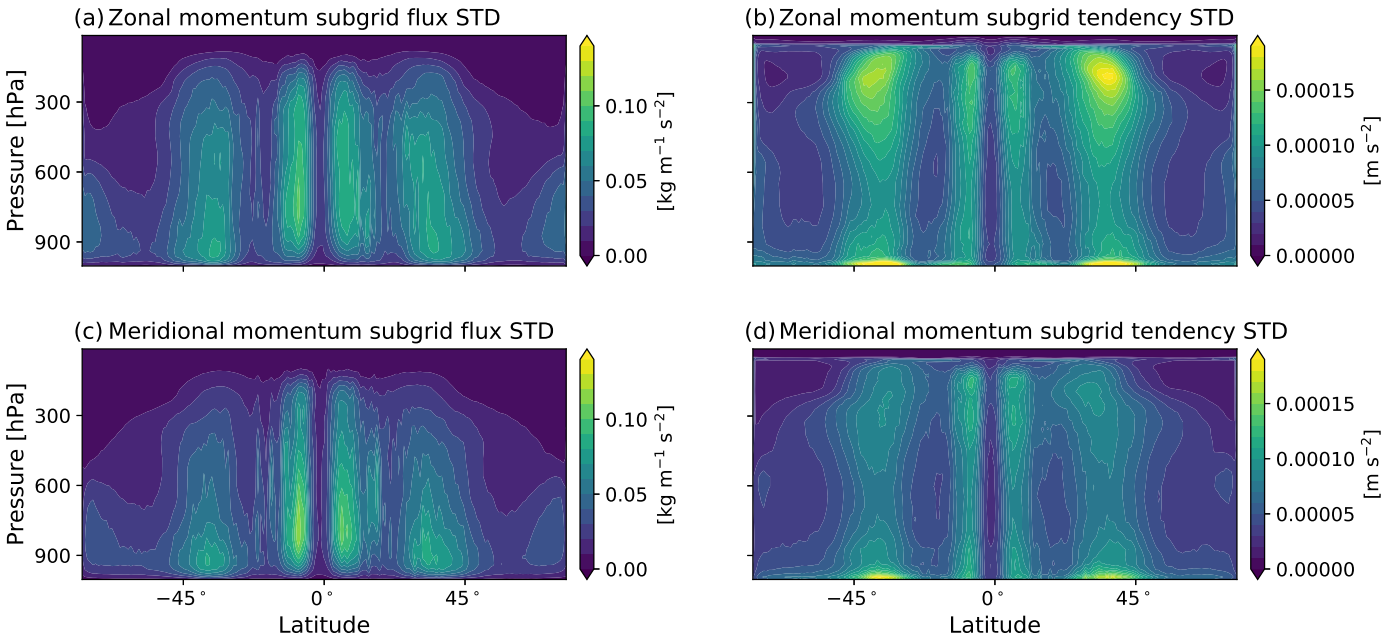

Figure S3. The standard deviation of subgrid terms due to vertical advection as a function of pressure and latitude when using coarse graining on a **collocated grid** for (a) zonal momentum flux, (b) zonal wind tendency, (c) meridional momentum flux, and (d) meridional wind tendency, as calculated from 501 snapshots of x8 test data set.

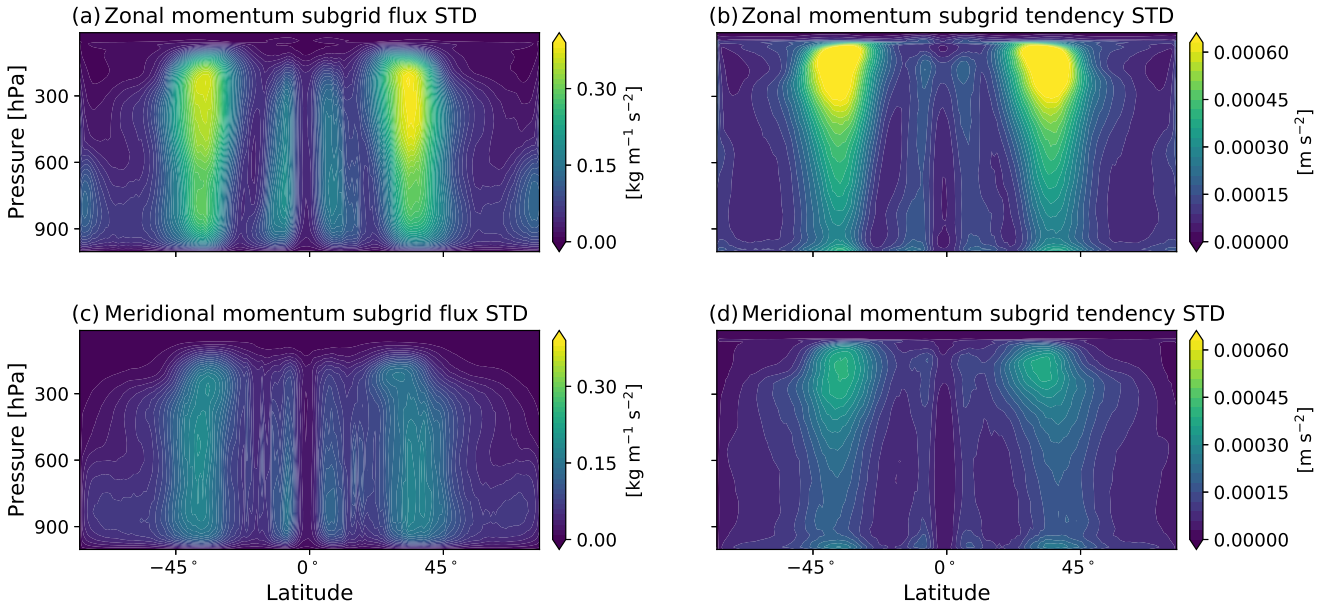

Figure S4. The standard deviation of subgrid terms due to vertical advection as a function of pressure and latitude when using coarse graining on a **staggered C-grid** for (a) zonal momentum flux, (b) zonal wind tendency, (c) meridional momentum flux, and (d) meridional wind tendency, as calculated from 501 snapshots of x8 test data set. Colors are saturated in panels a-b to highlight fluxes and tendencies outside the jet region.

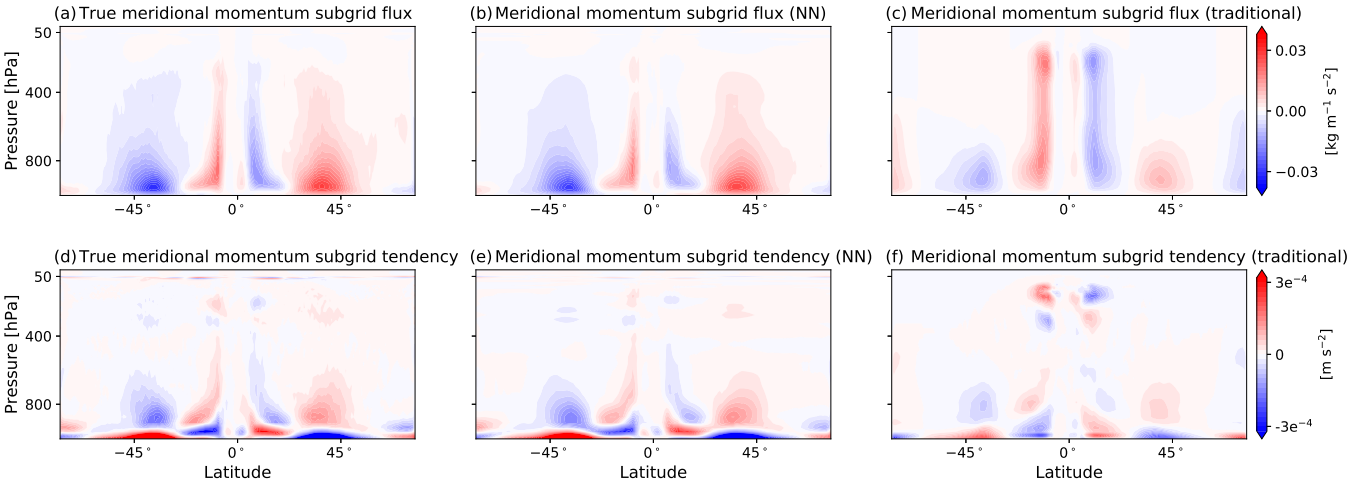

Figure S5. The time- and zonal-mean subgrid meridional momentum flux (a,b,c) and meridional wind tendency (d,e,f) from vertical advection as calculated from hires (a,d), and as estimated by (b,e) NN-MOM and (c,f) the conventional convection scheme (Section 5). Colors are saturated in panels d-f to highlight fluxes and tendencies outside of the boundary layer. The sponge layer is active above 50hPa. All quantities are calculated from 501 snapshots of the x8 test data set.

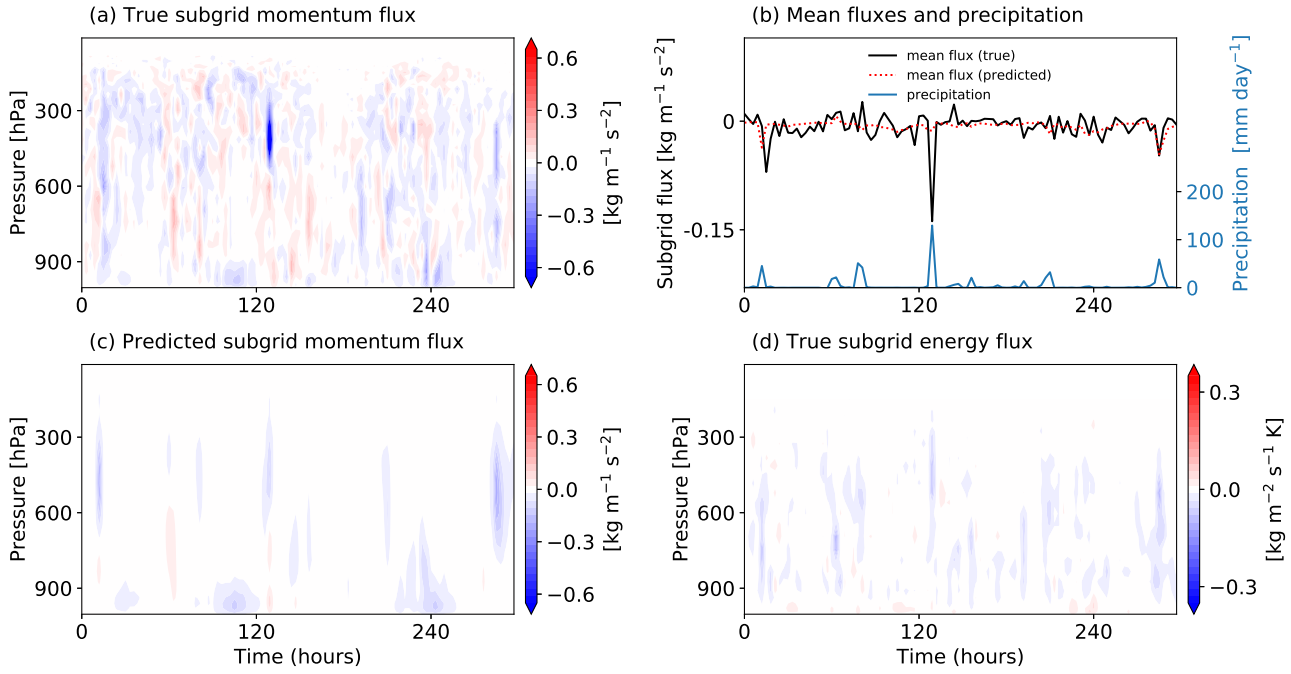

Figure S6. Time series of subgrid fluxes due to vertical advection for an extratropical column at latitude 35.3° and a coarse-graining factor of 8: (a) true zonal momentum flux, (b) vertical-mean true (black) and NN-MOM predicted (dotted red) zonal momentum flux, (c) NN-MOM predicted zonal momentum flux, (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 as in Figure 3 in the main paper.

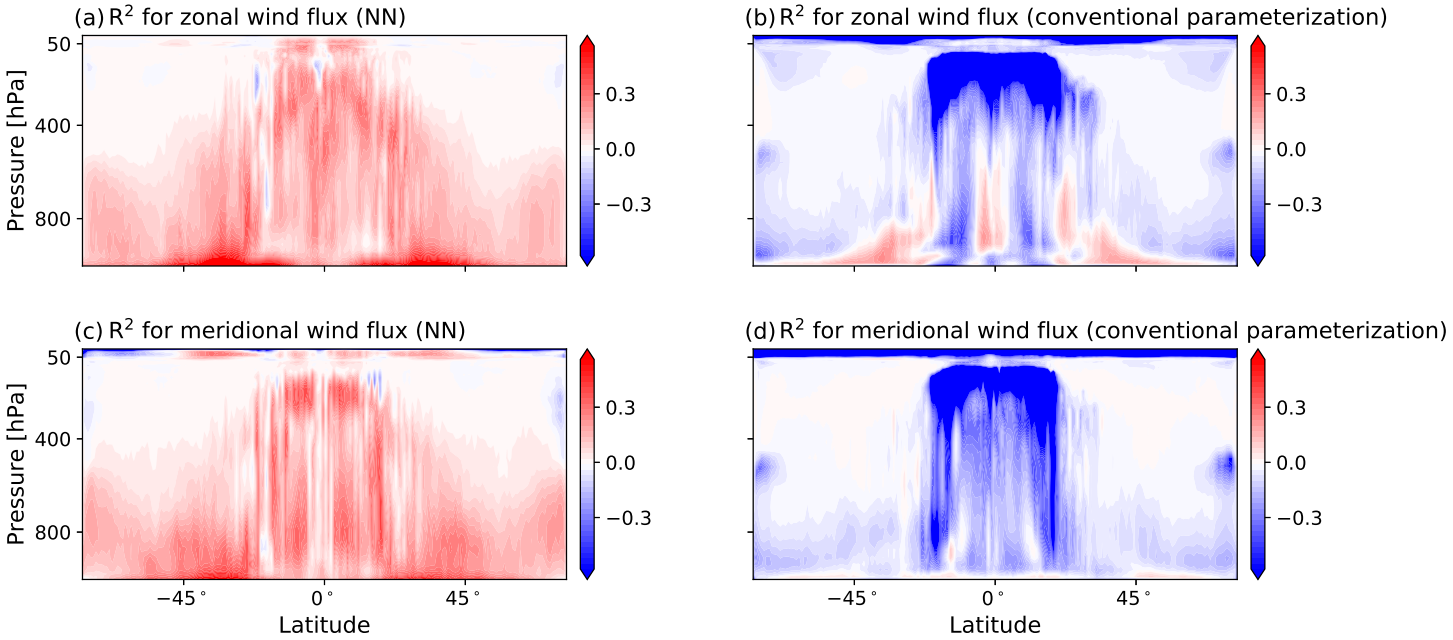

Figure S7. The coefficient of determination (R^2) for offline performance at a coarse graining factor of 8 of NN-MOM (a,c) and of the conventional convective momentum parameterization (b,d; Section 5) for the: (a,b) subgrid zonal momentum fluxes due to vertical advection and (c,d) subgrid meridional momentum fluxes due to vertical advection as a function of pressure and latitude. The coefficient of determination is calculated from 501 snapshots of the x8 test data set.

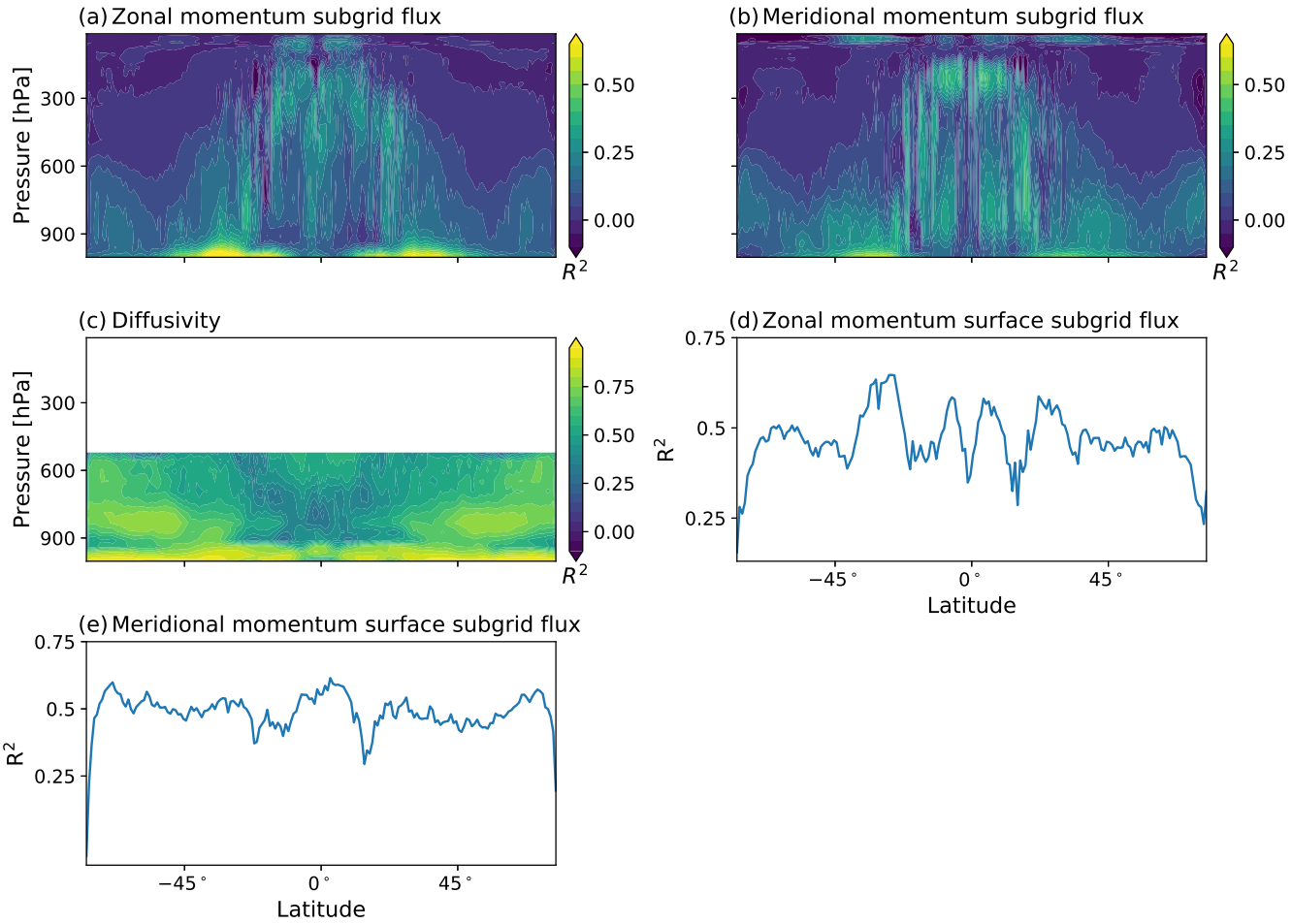

Figure S8. The coefficient of determination (R^2) for offline performance of NN-MOM for the: (a) subgrid zonal momentum fluxes due to vertical advecton (b) subgrid meridional momentum fluxes due to vertical advecton and (c) turbulent diffusivity as a function of pressure and latitude, and of (d) subgrid 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.

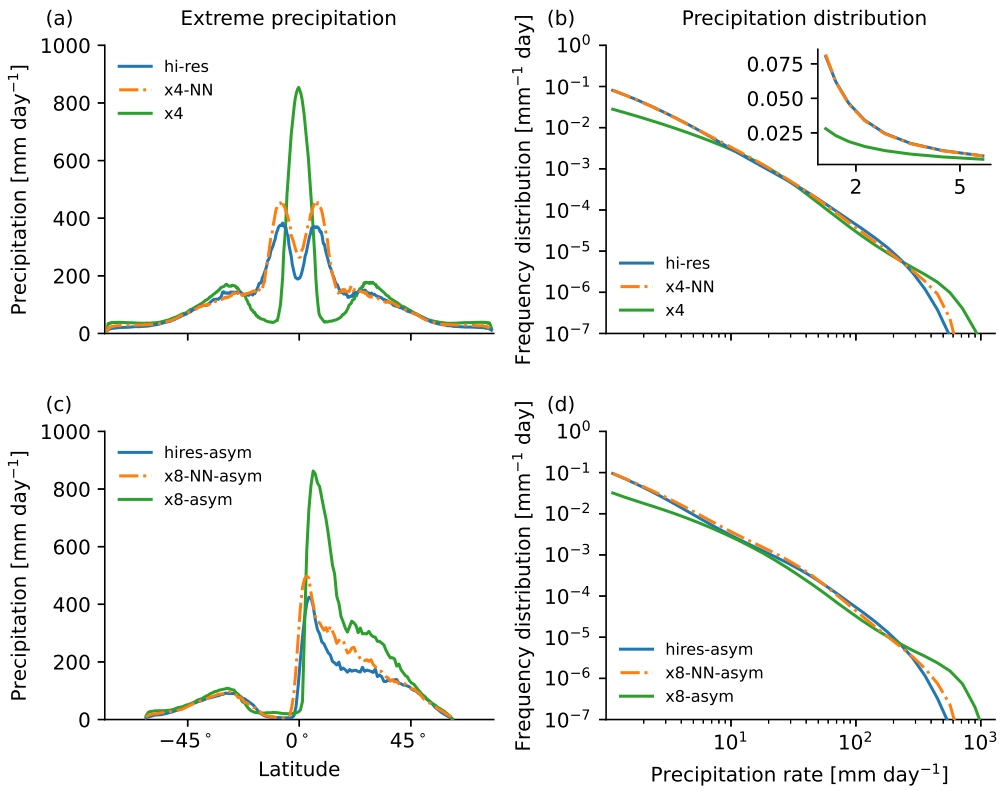

Figure S9. (a,c) Extreme precipitation rate (defined as the 99.9th percentile) as a function of latitude and (b,d) frequency distribution of 3-hourly precipitation rate. Results are shown for the high-resolution simulation (hi-res in blue in panels (a)-(b); hires-asym in blue in panels (c)-(d)), the coarse-resolution simulations (x4 in green in panels (a)-(b); x8-asym in panels (c)-(d)), the coarse-resolution simulations with the NN parameterization (x4-NN in orange dash-dotted in panels (a)-(b); x8-NN-asym in orange dash-dotted in panels (c)-(d)). The frequency distribution is shown for axes with linear scale in the inset of panel (b). For the high-resolution simulations, the precipitation is coarse-grained to the grid spacing of x4 or x8 prior to calculating the frequency distribution or the extreme precipitation.

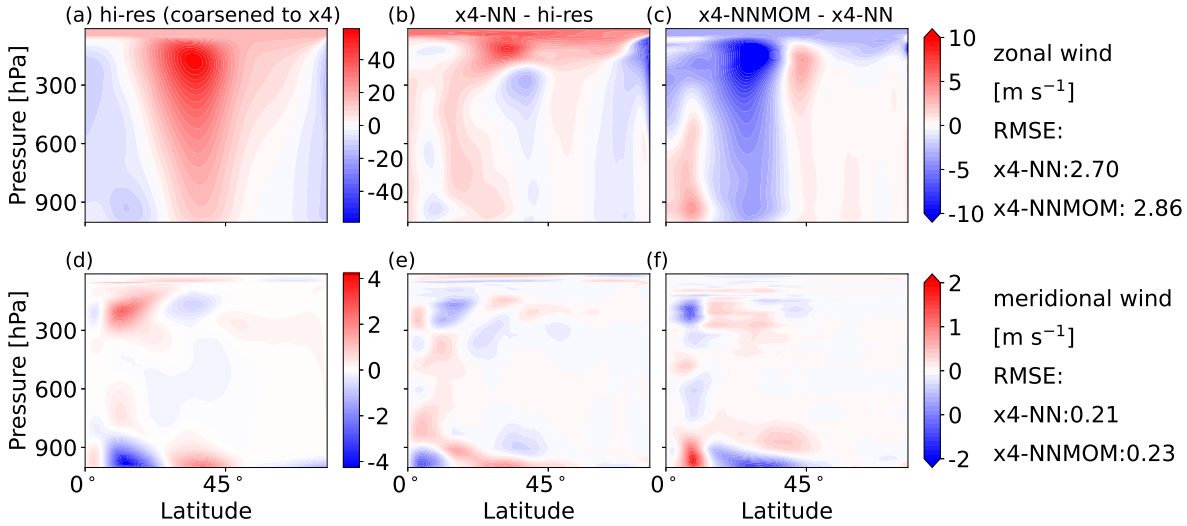

Figure S10. The zonal- and time-mean zonal (a-c) and meridional (d-f) wind as a function of pressure and latitude for simulations with coarser grids by factors of 4 compared to hi-res. This figure is similar to Figure 5 in the manuscript, but for a coarse graining factor of 4 compared-to hi-res.

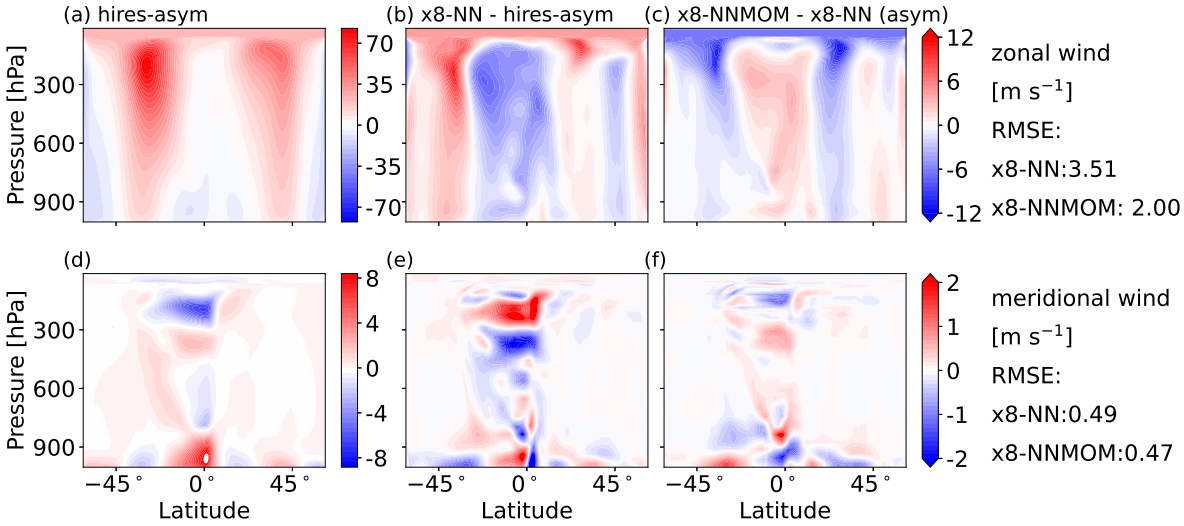

Figure S11. The zonal- and time-mean zonal (a-c) and meridional wind (d-f) as a function of pressure and latitude for simulations in the hemispherically asymmetric configuration with coarser grids by factors of 8 compared to hires-asym. This figure is similar to Figure 5 in the manuscript, but here the hemispherically asymmetric configuration is shown, and the results were not averaged over both hemispheres.

	RMSE relative to hi-res		
	x4	x4-NN	x4-NNMOM
Zonal wind [m s^{-1}]	4.58	2.70	2.85
Meridional wind [m s^{-1}]	0.66	0.21	0.23
EKE [m^2 s^{-2}]	39.77	28.22	34.23
	Correlation (bias in NN vs. correction of NNMOM)		
Zonal wind (r)	-0.47		
Meridional wind (r)	-0.54		
EKE (r)	-0.07		

Table S1. Same as Table 1 but for simulations that are coarser than hires by a factor of 4 rather than 8.

	RMSE relative to hi-res		
	x8-asym	x8-asym-NN	x8-asym-NNMOM
Zonal wind [m s^{-1}]	2.88	3.51	2.00
Meridional wind [m s^{-1}]	0.35	0.49	0.47
EKE [m^2 s^{-2}]	33.94	28.56	24.31
	Correlation (bias in NN vs. correction of NNMOM)		
Zonal wind (r)	-0.83		
Meridional wind (r)	-0.38		
EKE (r)	-0.57		

Table S2. Same as Table 1 but for hemispherically asymmetric simulations.