

DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD
DASH.HARVARD.EDU

HARVARD LIBRARY


Combining Analytical, Numerical, and AI Models to Improve Cloud and Convection Representation in Climate Simulations

Citation

Hu, Zeyuan. 2025. Combining Analytical, Numerical, and AI Models to Improve Cloud and Convection Representation in Climate Simulations. Doctoral Dissertation, Harvard University Graduate School of Arts and Sciences.

Link

https://dash.harvard.edu/handle/1/42719675

Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material (LAA), as set forth at

https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFlOWI2M2EwYTg

Accessibility

https://accessibility.huit.harvard.edu/digital-accessibility-policy

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. Submit a story

HARVARD UNIVERSITY
Graduate School of Arts and Sciences



DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the

Department of Earth and Planetary Sciences

have examined a dissertation entitled

"Combining Analytical, Numerical, and AI Models to Improve Cloud and Convection Representation in Climate Simulations"

presented by Zeyuan Hu

candidate for the degree of Doctor of Philosophy and hereby certify that it is worthy of acceptance.

Signature _____

Typed name: Prof. Zhiming Kuang

Signature _____

Typed name: Prof. Peter Huybers

Signature _____

Typed name: Prof. Jerry Mitrovica

Signature _____

Typed name: Prof. Eli Tziperman

Date: April 7, 2025

Combining Analytical, Numerical, and AI Models to Improve Cloud and Convection Representation in Climate Simulations

A DISSERTATION PRESENTED
BY
ZEYUAN HU
TO
THE DEPARTMENT OF EARTH AND PLANETARY SCIENCES

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN THE SUBJECT OF
EARTH AND PLANETARY SCIENCES

HARVARD UNIVERSITY
CAMBRIDGE, MASSACHUSETTS
APRIL 2025

©2024 – Zeyuan Hu
all rights reserved.

Dissertation advisor: Zhiming Kuang Zeyuan Hu

Combining Analytical, Numerical, and AI Models to Improve Cloud and Convection Representation in Climate Simulations

Abstract

Accurate representation of small-scale processes, such as convection and cloud formation, remains one of the greatest challenges in climate modeling, even in kilometer-scale storm-resolving simulations. These processes are essential for determining large-scale atmospheric behavior, but computational constraints prevent their full representation in global climate models. For instance, deep convection may not exhibit convergent behavior with increasing resolution, and significant uncertainties persist in ice microphysics parameterization. This thesis explores these challenges in two parts: the first (Chapters 2 and 3) examines how small-scale convection and cloud processes influence large-scale atmospheric states in idealized radiative-convective equilibrium simulations, while the second (Chapter 4) investigates machine learning (ML) as a tool for efficiently emulating these processes in climate models.

Chapter 2 addresses a fundamental question: what controls the vertical thermal structure of an equilibrium atmosphere? To answer this, I developed a refined zero-buoyancy plume model that analytically solves equilibrium atmospheric profiles given boundary conditions. The model highlights how plume-environment mixing influences vertical temperature profiles, upper-tropospheric convective mass flux, and cloud fraction. These findings align with convection-permitting simulations, which reveal that higher horizontal resolution—acting as a proxy for enhanced plume-environment mixing—leads to increased cloud fraction and mass flux in the upper troposphere.

Chapter 3 explores the impact of microphysics scheme uncertainties on equilibrium atmospheric states, particularly focusing on deep convective overshoots into the tropical tropopause layer (TTL). We find that different microphysics schemes produce distinct heat balance regimes in the TTL. Two schemes lead to a "hard-landing" scenario, where frequent, strong convective overshoots induce significant cooling (~0.2 K day^{-1}), while a third scheme results in a "soft-landing" scenario, with weaker overshoots and minimal cooling (~0.03 K day^{-1}). This difference arises from variations in upper-tropospheric stratification driven by atmospheric cloud radiative effects (ACRE). The scheme producing the soft-landing scenario generates stronger ACRE, leading to a 3-K warmer, more stable upper-tropospheric layer that buffers convective updrafts.

Chapter 4 demonstrates how ML can emulate these small-scale processes efficiently by learning directly from high-resolution simulations. Using data from superparameterized climate simulations, we train ML models to replace the embedded cloud-resolving models. While previous studies show that such hybrid ML-physics simulations can reproduce key climate statistics, they often suffer from online instability, particularly in setups with real geography and explicit cloud condensate coupling. By integrating an expressive U-Net architecture with cloud microphysics constraints, we achieve stable and skillful multi-year hybrid climate simulations with realistic cloud climatology and explicit cloud condensate coupling.

Contents

Title Page **i**

Copyright **ii**

Abstract **iii**

Table of Contents **v**

Author List **vii**

Listing of figures **ix**

Acknowledgements **xvii**

1 Introduction **1**
- 1.1 Challenges in Representing Small-Scale Processes in Climate Models 1
- 1.2 Idealized Simulations as a Testbed for Understanding Model Sensitivity 4
- 1.3 The Potential of Hybrid Physics-ML Approaches 6

2 A Refined Zero-buoyancy Plume Model for Large-scale Atmospheric Profiles and Anvil Clouds in Radiative-Convective Equilibrium **9**
- 2.1 Abstract 10
- 2.2 Introduction 11
- 2.3 Refined zero-buoyancy plume model with a self-consistent solution of mass flux . 15
- 2.4 Insights from the refined zero-buoyancy plume model 18
- 2.5 Sensitivity of small-domain RCE simulations on horizontal resolution 21
  - 2.5.1 Experiment setup 21
  - 2.5.2 Impacts of horizontal resolution on mean atmospheric state 24
  - 2.5.3 Budgets for environmental subsidence 29
- 2.6 Conclusions and Discussion 39

3 Influence of Upper-Troposphere Stratification and Cloud–Radiation Interaction on Convective Overshoots in the Tropical Tropopause Layer **43**
3.1 Abstract . . . 44
3.2 Introduction . . . 45
3.3 Experiment setup . . . 48
3.4 The dependence of convective overshoots on cloud microphysics . . . 52
3.5 Influence of upper-troposphere stratification and atmospheric cloud radiative effects 55
3.6 Understanding the different ACRE associated with different microphysics schemes 65
3.7 Discussion . . . 69
3.7.1 Sensitivity to model resolution . . . 69
3.7.2 Implication for TTL heat balance and cold-point tropopause . . . 70
3.7.3 Potential impact from varying ozone concentration to convective overshoots 73
3.8 Summary . . . 75

4 Stable Machine-Learning Parameterization of Subgrid Processes in a Comprehensive Atmospheric Model Learned From Embedded Convection-Permitting Simulations **78**
4.1 Abstract . . . 79
4.2 Introduction . . . 80
4.3 Methods . . . 85
4.3.1 Dataset . . . 85
4.3.2 Architecture . . . 85
4.3.3 CRM Setup in the E3SM-MMF Model . . . 87
4.3.4 Microphysical Constraints . . . 89
4.3.5 Inputs and Outputs . . . 91
4.3.6 Training . . . 93
4.3.7 Coupling E3SM and NN . . . 93
4.4 Offline results . . . 94
4.5 Online results . . . 97
4.5.1 Online Error Development in the Initial First Year . . . 97
4.5.2 Climatology Evaluation of Stable 5-year Hybrid Simulations . . . 102
4.6 Conclusions . . . 108
4.7 Discussion and Limitations . . . 111
4.7.1 Complexity of the Reference and Hybrid Simulations . . . 111
4.7.2 Improving the Offline and Online Performance . . . 112

5 Conclusion **117**

Appendix A Supplemental Material for Chapter 2 **120**
A.1 Refined solutions of a zero-buoyancy plume model . . . 120

A.2 Relationship between cloud evaporation and perimeter area ratio 130
A.3 Additional Figures in the Supporting Information 134

Appendix B Supplemental Material for Chapter 4 **138**
B.1 Neural Network Model Configuration 138
B.1.1 Input and Output Features 138
B.1.2 NN architecture 139
B.1.3 Normalization 141
B.1.4 Additional Preprocessing 142
B.1.5 Feature Pruning 142
B.2 Training 143
B.3 Online Metric 145
B.4 Microphysics classifier 146
B.5 Online Zonal Mean Climatology Bias with a Different U-Net Checkpoint 148
B.6 Online error growth 148
B.7 Additional Figures in the Supporting Information 148

References **166**

Author List

The following authors contributed to:

Chapter 2: Zeyuan Hu, Nadir Jeevanjee, Zhiming Kuang.

Chapter 3: Zeyuan Hu, Faycal Lamraoui, Zhiming Kuang.

Chapter 4: Zeyuan Hu, Akshay Subramaniam, Zhiming Kuang, Jerry Lin, Sungduk Yu, Walter Hannah, Noah Brenowitz, Josh Romero, Mike Pritchard.

Listing of figures

2.1 Atmosphere profiles in the zero-buoyancy plume model with varying entrainment rates (warmer color represents lower entrainment rate). The upper row has a cloud evaporation parameter μ=1. The variables shown are detrainment rate (a), updraft mass flux (b), relative humidity (c), temperature (d), the vertical gradient of dry static energy (e), and latent heating rate due to cloud evaporation in the environment (f). The dashed lines in panel a are the profiles of the prescribed entrainment rate. The temperature in panel d is shown as deviation to one of the simulations, which is denoted by the red line with zero deviation. The lower row is similar to the upper row but with cloud evaporation parameter μ=0.1. 20

2.2 Similar to Fig. 2.1 but with a fixed entrainment rate of 0.5 km^{-1} and varying cloud evaporation parameter (warmer color represents less efficient cloud evaporation). . 21

2.3 The distribution of convective updraft mass flux in FMSE-height space (left column) and in entrainment-height space (right column) for grid size of 4km (upper row) and 125m (lower row). In the left column, we show the mass flux distribution (with a unit of kg m^{-2} s^{-1} bin^{-1}) binned by their FMSE (in unit of K). There are 50 bins with 0.5K intervals between 325K to 350K. The individual lines represent the FMSE profiles of entraining plumes rising from the cloud base with different entrainment rates, except the black line which represents domain-mean FMSE profiles. In the right column, we show the mass flux distribution (with a unit of kg m^{-2} s^{-1} bin^{-1}) binned by their effective entrainment rate. The bin boundaries have entrainment rates of $2^{(i/2)-4}$, for $i = 0, 1, 2, ..., 16$, with a unit of km^{-1}. We calculated the instantaneous FMSE profiles with these different entrainment rates and binned the mass flux according to these FMSE profiles. For example, all the mass flux between the cyan and blue lines in the left panels will be counted as having entrainment rate between 0.5 to 1 km^{-1} in the right panels. 26

2.4 Domain-mean steady-state profiles of cloud fraction (left column), atmosphere cloud radiative effects (the second column), and relative humidity (the third column). The last column shows the histogram of column integrated ice-water path. The upper row corresponds to the SAM1MOM simulations, while the lower row represents the Morrison et al. simulations. Different colors indicate varying grid sizes, with warmer colors denoting coarser resolutions. Shaded uncertainty range is estimated as the standard deviation of time-mean profiles of each non-overlapping 2.5-day time block. . . . 27
2.5 Domain-mean steady-state profiles of convective updraft mass flux (upper row) and cloud ice sedimentation rate (lower row). The left column corresponds to the SAM1MOM simulations, while the right column represents the Morrison et al. simulations. Different colors indicate varying grid sizes, with warmer colors denoting coarser resolutions. Shaded uncertainty range is estimated as the standard deviation of time-mean profiles of each non-overlapping 2.5-day time block. 30
2.6 Energy budget for environmental subsidence for the SAM1MOM simulations. The first row shows the subsidence contributed by latent heat (panel b), by radiative cooling (panel c), and by both (a). Panel d shows the latent heating rate averaged in environments. Panel e shows the latent heating rate associated with the phase change between clouds and vapor averaged in environments. Panel f normalizes the cloud-related latent heating rate in panel e by the domain-mean cloud mixing ratio. Shaded uncertainty range is estimated as the standard deviation of time-mean profiles of each non-overlapping 2.5-day time block. 32
2.7 The upper row shows the cloud mixing ratio snapshots at z=10km in a 32km$\times$32km subdomain in the SAM1MOM simulations with grid size of 4km (a) and 125m (b). Panel c shows the perimeter area ratio in the SAM1MOM simulations with different grid sizes. Panel d shows the normalized evaporation in Fig. 2.6f at z=10km. Blue bars are direct model diagnostic values, and orange bars are estimated by Equation 2.8. 35
2.8 Energy budget for environmental subsidence for the Morrison et al. simulations. The first row shows the subsidence contributed by latent heat (panel b), by radiative cooling (panel c), and by both (a). Panel d shows the radiative heating rate averaged in environments. Panel e shows the vertical gradient of dry static energy averaged in environments. Panel f shows the absolute temperature profiles as deviation to the 4km Morrison et al. simulation. Shaded uncertainty range is estimated as the standard deviation of time-mean profiles of each non-overlapping 2.5-day time block. 36

2.9 Convective updraft mass flux and energy budget for environmental subsidence in modified Morrison et al. simulations. Solid lines represent simulations where cloud-radiation interactions are deactivated. Dashed lines indicate simulations with both cloud-radiation interactions deactivated and expedited cloud ice sublimation. Dotted lines represent default Morrison et al. simulations. Panel a displays the convective updraft mass flux. Panels b to d present the subsidence contributions from latent heat (panel c), radiative cooling (panel d), and a combination of both (panel b). Shaded uncertainty range is estimated as the standard deviation of time-mean profiles of each non-overlapping 2.5-day time block. 38

3.1 Differences of convective overshoots in the DEFAULT set of simulations: SAM1MOM_DFT (red), Thompson_DFT (orange), and Morrison_DFT (blue). (a) Domain-averaged profiles of convective heating rate, calculated as the negative of the radiative heating rate. The short horizontal lines indicate the cold-point tropopause. (b) Domain-averaged profiles of virtual potential temperature flux ($\rho C_p \overline{w'\theta_v'}$). (c) The complementary cumulative distribution function (CCDF, similar to CDF but integrating from the largest sample value) of vertical velocity at 14 km. (d) Contribution of different vertical velocity to the virtual potential temperature flux at 14 km, defined as $\text{Contr}(W_0) = \sum_{w'>W_0} w'\theta_v' / \sum_{\text{all } w'} w'\theta_v'$. The shadings in panel (a) and panel (b) indicate the 95% confidence interval. The shadings in panel (c) and panel (d) indicate the range based on the five 10-day periods within the total 50 days. 52

3.2 Domain-averaged profiles of potential temperature in the DEFAULT set of simulations: SAM1MOM_DFT (red), Thompson_DFT (orange), and Morrison_DFT (blue). The deviation of potential temperature profile from SAM1MOM_DFT in Thompson_DFT and Morrison_DFT is shown in the embedded figure. For all the lines shown, the shading indicates the 95% confidence interval. Some portions of the uncertainty range are too small to see the shading. 56

3.3 Domain-mean profiles of potential temperature, as the deviation from that in SAM1MOM_DFT, in the switch-profile simulations (solid) and in the default set of simulations (dashed): SAM1MOM_MORprl (solid red), Morrison_SAMprl (solid blue), SAM1MOM_DFT (dashed red), and Morrison_DFT (dashed blue). (b) As in (a), but for the virtual potential temperature flux ($\rho C_p \overline{w'\theta_v'}$). For switch-profile simulations (solid lines), the shading indicates the 95% confidence interval. Some portions of the uncertainty range are too small to see the shading. 58

3.4 Domain-mean profiles of (a) net, (b) longwave, and (c) shortwave atmospheric cloud radiative effects in the default set of simulations: SAM1MOM_DFT (red), Thompson_DFT (orange), and Morrison_DFT (blue). The shading indicates the 95% confidence interval. 59

3.5 Temperature and moisture profiles in the NOCRE simulations and their change from NOCRE to DEFAULT simulations for the three microphysics schemes. (a) As in Figure 3.2, but for the NOCRE set of simulations. (b) The potential temperature changes from NOCRE to DEFAULT simulations. (c) The profiles of water vapor mixing ratio for the NOCRE simulations. (d) The change of water vapor mixing ratio from NOCRE to DEFAULT simulations. (e) The change of moist static energy (solid lines) and the change of saturation moist static energy (dashed lines, only the values in the upper troposphere between 8 and 14 km are shown) from NOCRE to DEFAULT simulations. The shading indicates the 95% confidence interval. 61
3.6 As in Figure 3.1, but the solid lines are for the NOCRE set of simulations. The lines in Figure 3.1c for the DEFAULT set of simulations are overlaid in (c) as dashed lines for reference. 62
3.7 (a) The convective heating rate, (b) virtual potential temperature flux, and (c) potential temperature deviation from the Morrison_DFT simulation for the Morrison_NOCRE (dark blue), Morrison_NOCRESW (light blue), Morrison_NOCRELW (magenta), and Morrison_DFT (dashed black) simulations. The shading for the NOCRESW and NOCRELW simulations indicates the 95% confidence interval. 64
3.8 The relative frequency histograms of (a) the vertically integrated optical depth and (b) water path for the sum of radiatively active frozen hydrometeors in the default set of simulations: SAM1MOM_DFT (red), Thompson_DFT (orange), and Morrison_DFT (blue). The optical depth is for the 625–778-nm band. In the Morrison and Thompson simulations, both cloud ice and snow are radiatively active, while in the SAM1MOM simulation only cloud ice is radiatively active. The frequency is normalized by the number of total sampled grids including both clear-sky and cloudy-sky grid points. (c) The cloud fraction profiles with shading indicating 95% confidence interval. 66
3.9 Vertical distribution of (a)–(c) cloud radiative effect, (d)–(f) cloud ice content, (g),(h) snow content, and (i)–(k) frozen hydrometer effective radius sorted by model columns with different values of frozen water path. (left) Morrison_DFT. (center) SAM1MOM_DFT. The snow content in the SAM1MOM simulations is not radiatively active so we do not show the snow content here for SAM1MOM_DFT. (right) Thompson_DFT. The cloud ice content and snow content (g m^{-3}) is normalized by the total frozen water path (g m^{-2}) is each bin, so the unit becomes km^{-1}. The normalized cloud ice/snow contents also represent the vertical probability distribution of frozen mass in each bin. In (i)–(k), we show the cloud ice effective radius for Morrison_DFT and SAM1MOM_DFT but snow effective radius for Thompson_DFT, because cloud radiative effective is dominated by cloud ice in the SAM1MOM and Morrison simulations but is dominated by snow in the Thompson simulations. 68

3.10 (a) The convective heating rate, (b) virtual potential temperature flux, and (c) cloud fraction for the SAM1MOM_100m (solid red lines) and Morrison_100m (solid blue lines). The SAM1MOM_DFT and Morrison_DFT are represented as long-dashed lines for reference. The short-dashed lines in (c) represent the cloud fraction after re-gridding the 100-m-resolution simulations from 100-m resolution to 1-km resolution. 70
3.11 (a) Large-scale vertical velocity used in the BDC set of simulations. (b) The virtual potential temperature flux for SAM1MOM_BDC (red), Thompson_BDC (orange), and Morrison_BDC (blue). (c) Potential temperature change from the DEFAULT set to the BDC set of simulations. (d) The heat balance in the Morrison_BDC simulation. The solid purple line is the radiative heating rate. The dashed purple line is the adiabatic heating rate due to large-scale vertical velocity. The dotted purple line is the convective heating rate. The horizontal black line indicates the cold-point tropopause. (e) As in (d), but for SAM1MOM_BDC. (f) As in (d), but for Thompson_BDC. . . 72
3.12 (a) Profiles of ozone concentration used in this study, (b) potential temperature deviation from SAM1MOM_DFT, and (c) virtual potential temperature flux for SAM1MOM_FEB (cyan), SAM1MOM_Aug (solid pink), SAM1MOM_Aug+2 (dashed pink), and SAM1MOM_DFT (purple). 74

4.1 The schematic of the U-Net architecture. The U-Net model is composed of multiple ResBlocks, with 4 layers in depth and latent dimension sizes of [128, 256, 256, 256]. The model requires the input and output of each sample to be prepared as 2D arrays (see Appendix B.1.4) with dimensions corresponding to the number of variables times the vertical length. The numbers on top of each tensor indicate the tensor shape for the U-Net model with the microphysical constraint. Each tensor in the encoder has a skip connection to a corresponding tensor in the decoder, illustrated by the red arrows showing some of these skip connections. 86
4.2 Assessment of the microphysical constraints. (a) The fraction of liquid over total cloud water mixing ratios as a function of temperature. This temperature-based phase partitioning holds exactly on each grid in the cloud-resolving model. (b) Histogram of the actual liquid cloud fraction in the E3SM grid (y-axis) versus the predicted liquid fraction based on the E3SM-grid temperature (x-axis). (c) Histogram of the tropopause level versus the cloud top level, where levels correspond to the hybrid pressure level used in E3SM. In each column, the tropopause level is identified as the lowest level where the pressure is less than 400 hPa and the vertical gradient of potential temperature exceeds 10 K/km. A cloud top is defined as the highest level with a total cloud water mixing ratio greater than 1 mg/kg. 89

4.3 Offline R^2 scores across various variables for MLP (blue), U-Net-baseline (cyan), U-Net-expanded (orange), and U-Net-expanded-constrained (red). Variables are the full target variables listed in Table B.2, including temperature tendency ($\frac{dT}{dt}$), water vapor tendency ($\frac{dQ_v}{dt}$), liquid cloud mixing ratio tendency ($\frac{dQ_l}{dt}$), ice cloud mixing ratio tendency ($\frac{dQ_i}{dt}$), zonal wind tendency ($\frac{dU}{dt}$), meridional wind tendency ($\frac{dV}{dt}$), and eight flux variables. 95
4.4 Zonal mean R^2 evaluated on the test set for the U-Net-expanded-constrained model. The five panels denote the R^2 for the tendencies of temperature (a), water vapor (b), total cloud (liquid plus ice) mixing ratio (c), zonal wind (d), and meridional wind (e). 97
4.5 Online monthly and globally averaged (both horizontally and vertically and weighted by mass in each grid) RMSE of temperature (K), water vapor (g/kg), zonal wind (m/s), and total cloud mixing ratio (mg/kg) over a one-year period, comparing baseline MLP (blue), U-Net-baseline (cyan), U-Net-expanded (orange), and U-Net-expanded-constrained (red) against the reference E3SM-MMF simulation. The dashed black lines represent the uncertainty of the reference physical simulation due to internal atmospheric variability. Each architecture was trained with three slightly different configurations (e.g., loss functions and learning rate schedules) to sample ML uncertainty, represented by different markers but with the same line color. 98
4.6 The zonal mean liquid (upper row) and ice (lower row) cloud mixing ratio at month 10 in the reference E3SM-MMF simulation (left) and in the hybrid simulations with the U-Net-expanded-constrained model (middle) and with the unconstrained U-Net-expanded model (right). 100
4.7 Five-year time series of area-weighted mean values for near-surface temperature (at the lowest model level, top row), precipitable water (middle row), and total cloud path (sum of liquid and ice, bottom row). The data is shown across four regions: global mean (first column), 30°N–90°N (second column), 30°S–90°S (third column), and 30°S–30°N (fourth column). The solid black line and dashed black line represent two reference 5-year E3SM-MMF simulations initialized with the same initial conditions to sample internal variability (similar to the dashed lines in Figure 4.5). The solid and dashed red lines represent hybrid simulations using two U-Net-expanded-constrained checkpoints: Huber loss with reduce-on-plateau scheduler (solid red) and Huber loss with step scheduler (dashed red) (see B.2). 102

4.8 Online RMSE of the 5-year time-mean climatology for temperature (a), water vapor (b), liquid cloud mixing ratio (c), ice cloud mixing ratio (d), zonal wind (e), and meridional wind (f). Profiles represent RMSE values per vertical level, while the global RMSE values (integrated across horizontal and vertical space) are noted in the legend. RMSE is defined as in B.3. The black line shows the RMSE between two reference E3SM-MMF simulations with identical initial conditions, serving as a proxy for a perfect emulator (diverging only due to non-bit-for-bit numerical differences). The solid and dashed red lines represent the online RMSE of hybrid simulations using two U-Net-expanded-constrained checkpoints: Huber loss with a reduce-on-plateau scheduler (solid red) and Huber loss with a step scheduler (dashed red) (see B.2). 104
4.9 Five-year zonal mean atmospheric state in the reference E3SM-MMF simulation (left) and in the hybrid simulation with the U-Net-expanded-constrained model (middle). The right column shows the zonal mean bias as the mean state from the hybrid simulation minus that of the reference simulation. The five rows show temperature, water vapor, zonal wind, liquid cloud, and ice cloud. The Black dashed line in panel o indicates the 5-year mean position of the tropopause level used in the microphysical constraints. 106
4.10 Five-year zonal mean atmospheric tendencies in the reference E3SM-MMF simulation (left column), in the hybrid online simulation with the U-Net-expanded-constrained model (middle column), and the difference (hybrid minus MMF run, right column). The four columns correspond to the tendencies of temperature, water vapor, zonal wind, and total cloud (liquid plus ice) mixing ratio. 107
4.11 (a) 5-year zonal mean precipitation in the reference E3SM-MMF simulation (solid lines) versus the online hybrid simulation with the U-Net model incorporating microphysical constraints (dashed lines). Different colors denote samples taken globally (blue), only over land (orange), and only over the ocean (red). (b) Histogram of the hourly instantaneous precipitation. 109
4.12 Horizontal maps of the five-year mean surface precipitation: (a) reference E3SM-MMF simulation, (b) hybrid U-Net-expanded-constrained simulation, and (c) bias (hybrid minus reference). Panel c also documents the RMSE of 0.96 mm/day for the five-year mean precipitation between the hybrid and reference simulations. 109

A.1 Atmosphere profiles in the zero-buoyancy plume model with varying cloud-base relative humidity (blue color represents more moist environment), entrainment rate $\varepsilon = 0.5 km^{-1}$, and cloud evaporation parameter $\mu = 1$. The variables shown are detrainment rate (a), updraft mass flux (b), relative humidity (c), temperature (d), the vertical gradient of dry static energy (e), and latent heating rate due to cloud evaporation in the environment (f). Dashed lines in panel a are the profiles of the prescribed entrainment rate. The temperature in panel d is shown as deviation to one of the simulations, which is denoted by the red line with zero deviation. 128

A.2 Similar to Fig. A.1 but with sea surface temperature (blue color represents colder surface temperature), entrainment rate $\varepsilon = 0.5km^{-1}$, and cloud evaporation parameter $\mu = 1$. 129
A.3 Similar to Fig. 2.1 but with non-zero atmospheric cloud radiative effects. Panel e shows the net radiative heating rate profiles. We use a fixed cloud evaporation parameter $\mu = 1$, increasing the entrainment rate ε from 0.4 to 0.8 km^{-1}, and increasing peak atmospheric cloud radiative effects H_0 from 0 to 0.6 K/day. 130
A.4 Panel a shows a schematic of cloud evaporation due to resolved horizontal advection and the following saturation adjustment in the SAM1MOM scheme. We set some clouds initially in the grids with blue shading and advect the cloud by the horizontal wind with $U_x = U_y = U_{\text{adv}}$. $q_{\text{c,edge}}$ is the cloud mixing ratio near the cloud edge. $q^*_{\text{v,env}}$ represents the environmental saturation specific humidity, and RH denotes relative humidity in the environment near the cloud. After a small timestep dt, some cloud mass is advected into environment grids (yellow shading), and some environmental mass is advected into cloudy grids (orange shading). Circular arrows represent the saturation adjustment in each grid due to the microphysics scheme. Panel b shows the relative value of different terms in Equation A.34 in the SAM1MOM simulations with different resolutions. Each term is standardized (divided by the maximum value across the simulations with different resolutions) to have a value between 0 and 1. 131
A.5 Panel a compares the domain-mean radiative heating rate in the model simulations and the prescribed radiative heating rate used in the zero-buoyancy plume model as a function of temperature. Panel b shows the absolute temperature anomaly averaged in convective updrafts in the model simulations. Panel c shows the hydrostatic balance ratio, defined as $-(dp/dz)/(\rho g)$, in the model simulations. The orange line is for the 1km-resolution simulation, and the blue line is for the 250m-resolution simulation. 135
A.6 Atmosphere profiles cloud mixing ratios in the SAM1MOM simulations (upper row) and Morrison et al. simulations (lower row). The left column is the domain mean total (liquid plus ice) cloud mixing ratio. The middle column is the liquid cloud mixing ratio averaged within cloudy-grids (grids with total cloud mixing ratio $> 10^{-5}$ kg/kg). The right column is the ice cloud mixing ratio averaged within cloudy-grids. . . . 136
A.7 Domain-mean averaged mass flux defined with different averaging conditions. Panel a shows the mass flux averaged in grids with vertical velocity W>1 m/s and cloud mixing ratio $> 10^{-5}$ kg/kg. Panel b shows the mass flux averaged in grids with vertical velocity W>0 m/s and cloud mixing ratio $> 10^{-5}$ kg/kg. Panel c shows the net convective mass flux. Net convective mass flux in panel c is defined as cloudy mass flux (grids with cloud mixing ratio $> 10^{-5}$ kg/kg, including both updraft and downdraft) plus precipitating downdraft (precipitating water $> 10^{-4}$ kg/kg and W<0 m/s). 137

B.1 (a) The distribution of the three classes at each atmospheric level. Only the levels below level 15 are shown since it is mostly Class 0 above level 15. (b) The prediction accuracy of the classifier for each class. Prediction accuracy is defined as the fraction of samples of a given class that are correctly identified by the classifier. (c) The confusion matrix of the classifier for all grid samples at all levels. 146
B.2 Similar to Figure 4.9, but using the U-Net-expanded-constrained model with Huber loss and a different learning rate schedule, as described in B.2. 149
B.3 Time evolution of the zonal mean bias averaged within 30S-30N for water vapor (a) and liquid cloud mixing ratio (b) over the first 5 days in the hybrid simulation with the U-Net model incorporating the microphysical constraints. 150
B.4 Online monthly and globally averaged (a) near-surface temperature at the lowest model level (K), (b) precipitable water (kg/m^2), and (c) total (liquid+ice) cloud path (kg/m^2) over a one-year period, comparing baseline MLP (blue), U-Net-baseline (cyan), U-Net-expanded (orange), and U-Net-expanded-constrained (red) against the reference E3SM-MMF simulations (dashed black lines). The dashed black lines construct an ensemble of reference physical simulations, and their spread represents internal atmospheric variability. Each architecture was trained with three slightly different configurations (e.g., loss functions and learning rate schedules) to sample ML uncertainty, represented by different markers but with the same line color. 150
B.5 Five-year time series of spatial standard deviation of near-surface temperature at the lowest model level (top row) and precipitable water (bottom row). The standard deviation is computed using daily-averaged fields across all grid points within three latitude bands: 40°N–50°N (left column), 40°S–50°S (middle column), and 5°S–5°N (right column). A 15-day running average is applied to smooth the time series. The black line represents the reference E3SM-MMF simulation. The solid and dashed red lines represent the hybrid simulations using two U-Net-expanded-constrained checkpoints: Huber loss with a reduce-on-plateau scheduler (solid red) and Huber loss with a step scheduler (dashed red). 151
B.6 Five-year zonal mean relative humidity in the reference E3SM-MMF simulation (left) and in the hybrid simulation with the U-Net-expanded-constrained model (middle). The right column shows the zonal mean bias as the mean state from the hybrid simulation minus that of the reference simulation. The contouring interval in the right column is 4%. 151

Acknowledgments

Completing this thesis would not have been possible without the generous help, guidance, and support of so many people. I am deeply grateful to everyone who has contributed to my academic journey, both personally and professionally.

First and foremost, I would like to express my heartfelt gratitude to my advisor, Zhiming Kuang, for his unwavering support, guidance, and mentorship. His role as both an advisor and a role model has been invaluable to me. I am particularly grateful for the freedom and encouragement he provided, which allowed me to explore my research interests and grow as a scientist. I feel incredibly fortunate to have had the opportunity to engage in countless insightful conversations with him, often during impromptu visits to his office. His vision, critical thinking, and exceptional ability to teach and communicate have been a constant source of inspiration, helping me become not only a better researcher but also a better person. Additionally, I deeply appreciate his generous support in encouraging me to explore different career paths, including internships and full-time industry opportunities, and for always prioritizing what is best for my career development.

I also extend my sincere thanks to my undergraduate advisor, Yongyun Hu at Peking University, for nurturing my early interest in atmospheric sciences. His encouragement to pursue excellence and his generous support for research and educational opportunities, including two summer undergraduate research exchanges at UCLA and Harvard, laid the foundation for my academic journey. I am forever grateful for his mentorship and belief in my potential.

My gratitude also goes to my thesis committee members, Eli Tziperman, Peter Huybers, and Jerry Mitrovica, for their invaluable advice and guidance throughout my PhD. I am deeply appreciative of their time and effort in supporting my academic development.

I am also very grateful to my colleagues and collaborators during my research internship at NVIDIA. Special thanks go to Mike Pritchard and Akshay Subramaniam for their guidance on my internship project, which forms Chapter 3 of this thesis, as well as for their continued support of my career. I would also like to thank Sungduk Yu, Jerry Lin, Walter Hannah, Noah Brenowitz, Jaideep Pathak, Yair Cohen, Suman Ravuri, Chenggong Wang, and Dale Durran for their helpful discussions and career advice during my time at NVIDIA.

I am deeply grateful to Nadir Jeevanjee and Andrea Jenney for being wonderful collaborators and for their support and guidance throughout my career. Chapter 1 of this thesis originates from my collaboration with Nadir, and I am also thankful for the opportunity to work with Andrea on a related paper, even though it is not included in this thesis. Their generosity in exchanging ideas,

collaborating on research, and sharing career opportunities has been invaluable.

I would also like to thank Eli Tziperman, Tim Cronin, Aixue Hu, and Gang Chen for their guidance and support during my undergraduate research projects. Their mentorship taught me how to conduct research and gave me the confidence and skills necessary to pursue a career in this field.

To my fellow members of the Kuang group—Lei Wang, Jiahua Guo, Packard Chan, Yang Tian, Xin Wei, Ned Kleiner, Fayçal Lamraoui, Ding Ma, Qiyu Song, Nathanial Wong, Sarah Weidman, and Walter Shen—thank you for creating such a supportive and collaborative environment. I will always cherish the fruitful group meetings, the insightful discussions, and the many enjoyable activities we shared together. From the very first day I arrived at Harvard, I felt welcomed and inspired by the intelligent conversations and camaraderie within the group. Your friendship and encouragement have made my time here truly memorable.

I am also incredibly fortunate to have had wonderful friends and colleagues at Harvard EPS. Beyond those already mentioned, I would like to thank my amazing officemates—Duo Chan, Teddy Amdur, Aleyda Trevino, Alex Berry, Caro Park, and Greta Berendes—for creating a warm and supportive office environment filled with so many good memories. My thanks also go to EPS friends, including Yaowei Li, Hongwei Sun, Hairuo Fu, Elise Penn, Xiaoting Yang, Wanying Kang, Makoto Kelp, Camille Hankel, Kara Hartig, Stephen Bourguet, Jennifer Sun, Ben Geyman, Kaitlyn Loftus, Natasha Toghramadjian, Congcong Yuan, Mingzhen Yu, Gong Zheng, and Ziwei Li (from MIT), for their friendship and camaraderie.

A special thank you goes to my PhD roommates—Yu Cheng, Feicheng Wang, and Yaowei Li—for making my PhD journey much less stressful and more enjoyable. Their companionship and support have been a constant source of comfort and joy.

I would also like to acknowledge my ping pong buddies, photography buddies, and dinner buddies—Qiyu Song, Yu Cheng, Boer Zhang, Xingjie Pan, Buyu Lin, Hanwen Cui, Tian Lu, Liyuan Cheng, Tianle Liu, Jingcheng Tan, Mingrui Wu, and Jingnan Du—for adding so much fun and relaxation to my life during the long PhD years.

Above all, I am eternally grateful to my Wuhan and New York families for their unconditional love. I am deeply thankful for the countless experiences, opportunities, and education they have provided me throughout my life. Their unwavering support has been my anchor during this journey, giving me the courage and strength to face any challenge that came my way.

Finally, I would like to express my deepest gratitude to my partner Yanwen Si (and our lovely cat Qiqi) for her love, companionship, and endless encouragement. Her optimism, kindness, and unwavering belief in me have been a constant source of strength, filling my life with so much joy and meaning. She has not only inspired me to pursue a better life and career but has also helped me grow into a better person.

1

Introduction

1.1 Challenges in Representing Small-Scale Processes in Climate Models

Accurately capturing clouds and convection in climate models remains one of the most enduring challenges in the field. These processes, which operate on scales ranging from micrometers (for cloud droplets) to hundreds of kilometers (for organized mesoscale convection), are critical drivers of climate dynamics. However, their spatial and temporal scales far exceed the resolution capabilities

of conventional global climate models (GCMs). Consequently, clouds, convection, and associated processes such as cloud microphysics, turbulence, and boundary-layer dynamics must be parameterized, introducing significant uncertainties into climate projections (e.g., Bony et al., 2015; Sherwood et al., 2014; Zelinka et al., 2020).

To overcome these limitations, alternative modeling frameworks that explicitly resolve convection have been developed. Notable among these are the Multiscale Modeling Framework (MMF), also referred to as superparameterization (SP), and global storm-resolving models (GSRMs). MMF (e.g., Grabowski & Smolarkiewicz, 1999; Benedict & Randall, 2009; Randall, 2013; Hannah et al., 2020; Norman et al., 2022) replaces traditional parameterizations of clouds and convection with embedded cloud-resolving models (CRMs) within each grid column of the host model, typically employing two-dimensional (2D) domains with resolutions on the order of 1 km. This approach starts to allow for the direct simulation of deep convection and its interactions with large-scale atmospheric dynamics.

Over the past two decades, MMF simulations have demonstrated significant advancements in capturing key climate phenomena, such as the Madden–Julian Oscillation (MJO) (Khairoutdinov et al., 2008; Benedict & Randall, 2009), the diurnal cycle of precipitation (Pritchard & Somerville, 2009b,a), and the statistics of extreme rainfall (Li et al., 2012). Additionally, SP-based models have been successfully applied to coupled climate simulations, improving representations of monsoons, tropical variability, and aerosol–cloud interactions (e.g., Stan et al., 2010; Goswami et al., 2011; Wang et al., 2011).

Despite these achievements, MMF is not without its limitations. The original SP framework assumes scale separation, which can restrict the propagation of convective systems across grid columns. The use of 2D embedded models inherently neglects critical aspects of convective momentum transport due to the absence of a third horizontal dimension. Furthermore, CRMs with resolutions on the order of 1 km may not fully capture shallow convection and turbulence dynamics. Efforts to

enhance CRM resolution (Grabowski, 2016; Parishani et al., 2017; Terai et al., 2020; Peng et al., 2024) and transition from 2D to 3D frameworks (e.g., Khairoutdinov et al., 2005) hold promise for improving the representation of these processes, albeit at a substantially higher computational cost.

Meanwhile, GSRMs (e.g., Satoh et al., 2019; Stevens et al., 2019) have pushed global model resolutions to the order of 1–4 km, enabling the explicit simulation of deep convection without the need for arbitrary scale separation. These models have shown notable advantages, such as improved statistics of heavy precipitation (Ma et al., 2022). However, GSRMs remain computationally very expensive for decadal-scale climate simulations or extensive parameter tuning. Moreover, like MMF, GSRMs must contend with the challenge of accurately representing subgrid-scale processes.

Both MMF and GSRMs typically operate at resolutions on the order of 1 km, a regime where deep convective properties have not yet fully converged (Jeevanjee, 2017). Understanding the sensitivity of these high-resolution simulations to model resolution is therefore crucial. Resolution dependence in atmosphere models that explicitly resolve deep convection has been extensively studied in various simulation setup. By changing horizontal grid size from 80km to 2.5km in a GSRM, Hohenegger et al. (2020) showed that many 40-day mean, global mean climate statistics, such as precipitation, sensible heat flux, and outgoing longwave radiation, exhibit weak resolution dependence compared with the uncertainties across different GSRMs. However, Hohenegger et al. (2020) also showed some convection and cloud properties, such as the width of the Intertropical Convergence Zone and the fraction of deep convective clouds, have not converged even at 2.5km resolution. Miyamoto et al. (2013) also examined the sensitivity of deep convection to resolution at around the order of 1km in global simulations. They showed that deep convective cores start to occupy more than one grid point at around 2km and have stronger upward velocity with finer resolution. In idealized squall line simulations, Bryan et al. (2003) showed decreasing grid size from the order of 1km to the order of 100m tends to give more turbulent flow with resolved entrainment and overturning within clouds. In limited-area 24-hour simulations with tropical maritime large-scale

forcing, Khairoutdinov et al. (2009) found low sensitivity of quantities such as cloud fraction, relative humidity, and precipitation rate to grid size ranging from 100m to 1600m, but updraft core statistics are senstive to resolution, with finer resolution showing larger upward velocity and more total water in updraft core. From limited-area radiative-convective equilibrium (RCE) simulation studies, Jeevanjee (2017) showed that the updraft velocity can keep increasing with finer resolution until grid size is at the order of 100m. Jeevanjee & Zhou (2022) found that, in RCE simulation, high cloud fraction exhibits strong resolution dependence, with finer resolution leading to higher anvil cloud fraction. Previous studies (e.g., Ohno & Satoh, 2018; Ohno et al., 2019; Jenney et al., 2023) have also shown that the vertical resolution can also influence the mass flux and anvil cloud fraction.

Beyond resolution constraints, another significant challenge lies in the parameterization of subgrid processes at the scale of 100m or smaller. Cloud microphysics, in particular, remains a major source of uncertainty in simulated climate states and sensitivity (e.g., Grabowski, 2003; Hu et al., 2021; Zhao et al., 2016).

1.2 Idealized Simulations as a Testbed for Understanding Model Sensitivity

Radiative-convective equilibrium (RCE) simulations have long served as an idealized framework for investigating fundamental atmospheric processes and model sensitivities. In RCE, the atmosphere reaches a statistical equilibrium where radiative cooling is balanced by convective and surface heating, providing a simplified environment to study the interactions between radiation, convection, and large-scale atmosphere states. Early one-dimensional models (e.g., Manabe & Strickler, 1964) laid the groundwork for understanding tropical thermodynamic structures, while more recent three-dimensional cloud-resolving simulations have enabled the exploration of emergent phenom-

ena such as convective self-aggregation and anvil cloud feedbacks (e.g., Bretherton et al., 2005; Wing et al., 2018a; Kuang & Hartmann, 2007; Seeley et al., 2019b).

The Radiative-Convective Equilibrium Model Intercomparison Project (RCEMIP; Wing et al., 2018b, 2020) further institutionalized RCE as a community standard for intercomparing models ranging from cloud-resolving models (CRMs) and large-eddy simulations (LES) to general circulation models (GCMs). RCEMIP aims to investigate the robustness of the RCE state across model hierarchies, the response of clouds to warming, and the dependence of convective aggregation on surface temperature. Among its key findings is that models show significant spread in the mean state and cloud profiles, and yet many agree on a decrease in anvil cloud area and an increase in aggregation with warming (Wing et al., 2020). These findings underscore the utility of RCE simulations as a controlled environment to explore structural model uncertainties and their implications for climate sensitivities.

This thesis leverages small-domain RCE simulations to explore model sensitivities arising from grid configuration (Chapter 2) and the choice of microphysics schemes (Chapter 3). While the spatial domain and boundary conditions preclude the formation of large-scale circulations and convective self-aggregation, the simulations remain a valuable testbed for studying the interaction between cloud, deep convection, radiation and vertical mean thermodynamic structures.

Chapter 2 investigates the mechanisms underlying resolution-dependent anvil cloud coverage, building on the work of Jeevanjee & Zhou (2022). As the project evolved, we developed a refined zero-buoyancy plume model capable of analytically solving equilibrium atmospheric profiles based on boundary conditions. This model, originally conceptualized by Singh & O'Gorman (2013); Romps (2014); Singh et al. (2019); Romps (2021); Singh & Neogi (2022), assumes that the equilibrium atmosphere can be represented by a single plume encompassing all updrafts and a clear-sky environment. Our work addresses a self-inconsistency in the original model, enhancing its reliability for predicting mass flux profiles. Using this refined model, we explore how vertical mass flux, anvil

cloud amount, and atmospheric profiles respond to model parameters such as plume-environment mixing strength. These insights are validated against numerical RCE simulations, which found that increasing model resolution—as a proxy of increasing plume-environment mixing—leads to enhanced convective updraft mass flux and anvil cloud fraction.

Chapter 3 shifts focus to the sensitivity of equilibrium atmospheric states to microphysics assumptions, particularly the role of deep convective overshoots in the tropical tropopause layer (TTL). By conducting RCE simulations with different microphysics schemes, we found that swapping the microphysics schemes can lead to different heat balance regimes, with deep convective overshoots can often versus rarely reaching the TTL. We illustrate that the microphysics scheme can impact the simulated cloud coverage and the associated cloud radiative effects, which can further influence the stratification and the convection properties.

1.3 The Potential of Hybrid Physics-ML Approaches

Machine learning has the advantage of learning from high-fidelity data and has shown the promise of improving the traditional climate models. One active research area is to use ML models to emulate and replace computationally expensive parts of a climate model. Machine learning models can learn the effects of small-scale physical processes directly from high-fidelity simulation data, such as those from MMF climate simulations (Gentine et al., 2018; Rasp et al., 2018; Brenowitz et al., 2020; Han et al., 2020; Mooers et al., 2021; Wang et al., 2022b; Bhouri et al., 2023; Lin et al., 2024; Iglesias-Suarez et al., 2024; Kühbacher et al., 2024), globally uniform km-resolution simulations (Brenowitz & Bretherton, 2018, 2019; Yuval & O'Gorman, 2020; Yuval et al., 2021; Wang et al., 2022a), and reanalysis products (Kochkov et al., 2024). Once trained, an ML emulator can be coupled with a host climate model to perform hybrid physics-ML simulations, potentially enhancing simulation accuracy for a given computational budget when comparing to traditional climate simu-

lations using empirical convection parameterizations (Rasp et al., 2018; Yuval & O'Gorman, 2020). Besides ML parameterization, other hybrid application includes bias correction (Bretherton et al., 2022; Clark et al., 2022; Kwa et al., 2023; Sanford et al., 2023b), i.e., train a ML-based bias corrector that can be coupled into the climate model to provide interactive correction of the atmosphere states.

A significant challenge in hybrid ML-climate modeling is to ensure that models trained offline can perform reliably in online settings, where they dynamically interact with the host climate model (Yu et al., 2024b, e.g.,). Offline training involves traditional supervised learning, where ML models map input features to target features using a fixed dataset. In contrast, online evaluation assesses the performance of hybrid simulations where ML models replace high-resolution physics solvers and interact with resolved climate dynamics. This requires evaluating the accuracy of climate statistics, such as mean atmospheric states, over extended periods.

When evaluating online performance, one critical aspect is the online stability. Optimizing a ML model's offline skill (by optimizing some loss function of the ML model's prediction at one single time step) does not necessarily lead to optimized or stable online performance (Ott et al., 2020; Wang et al., 2022b). Many previous studies have demonstrated proof-of-concept success in achieving stable online simulations using idealized setups such as aquaplanets. A few recent works (Wang et al., 2022b; Han et al., 2023; Sanford et al., 2023a; Kochkov et al., 2024; Behrens et al., 2024) have demonstrated the promise of stable online simulations incorporating real geographic features. Kochkov et al. (2024) achieved mostly stable (though occasional instability was observed) hybrid simulations using end-to-end online training, developing a fully differentiable hybrid model and directly optimizing multi-step rollout loss. Nevertheless, making existing Fortran-based climate models differentiable is challenging.

In the context of emulating the CRM in MMF simulations with a ML parameterization, no previous work had succeeded in completely coupling all input and output variables of the CRM

(e.g., tendencies of cloud condensates, surface fluxes for atmosphere-surface coupling) with real geography via ML. For example, Behrens et al. (2024) had to sidestep the condensate coupling to maintain a 5-month online stability by using a partial coupling approach, where condensate prediction is provided by a simultaneously running CRM. Han et al. (2023) predicted cloud species and achieved stable online simulations with real geography, but they did not emulate radiation (and thus the radiative fluxes required for atmosphere-land coupling) and showed significant online biases in high-latitude temperature and humidity.

Chapter 4 addresses these challenges by developing a stable ML parameterization for a near-operational setup with real geography and explicit cloud condensate coupling. Using an expressive U-Net architecture and incorporating microphysics constraints, we achieve skillful and stable multi-year hybrid simulations, with zonal mean tropospheric temperature biases within 2 K and water vapor biases within 1 g kg^{-1}.

2

A Refined Zero-buoyancy Plume Model for Large-scale Atmospheric Profiles and Anvil Clouds in Radiative-Convective Equilibrium

2.1 Abstract

A simple analytical model, the zero-buoyancy plume (ZBP) model, has been proposed to understand how small-scale processes such as plume-environment mixing and evaporation affect the steady-state structure of the atmosphere. In this study, we refine the ZBP model to achieve self-consistent analytical solutions for convective mass flux, addressing the inconsistencies in previous solutions. Our refined ZBP model reveals that increasing plume-environment mixing can increase upper-troposphere mass flux through two pathways: increased cloud evaporation or reduced atmospheric stability. To validate these findings, we conducted small-domain convection-permitting Radiative-Convective Equilibrium (RCE) simulations with horizontal resolutions ranging from 4km to 125m. As a proxy for plume-environment mixing strength, the diagnosed entrainment rate increases with finer resolution. Consistent with a previous study, we observed that both anvil cloud fraction and upper-troposphere mass flux increase with higher resolution. Analysis of the clear-sky

energy balance in the simulations with two different microphysics schemes identified both pathways proposed by the ZBP model. The dominant pathway depends on the relative strengths of evaporation cooling and radiative cooling in the environment. Our work provides a refined simple framework for understanding the interaction between small-scale convective processes and large-scale atmospheric structure.

2.2 Introduction

Uncertainty associated with cloud and convection remains one of the most significant factors contributing to climate feedback uncertainties in future climate change projections (e.g., Bony et al., 2015; Zelinka et al., 2020; Sherwood et al., 2014). Representing such small-scale processes in the coarse-grid global climate models requires parameterization of convection and cloud microphysics, which have become a major source of uncertainty. In recent years, the scientific community has made significant strides in developing and examining global storm-resolving models (GSRM) with grid sizes of 1-5km (e.g., Satoh et al., 2019; Stevens et al., 2019). By explicitly resolving deep convection, GSRMs do not require the use of convective parameterization, but the actual convective mixing strength can still be sensitive to horizontal and vertical resolution, sub-grid turbulence models, and advection schemes. It remains unclear how the large-scale atmosphere state in km-scale simulations can be influenced by the representation of the small-scale mixing processes.

A simple analytical model that can illustrate the effect of small-scale processes on large-scale tropical atmospheric state is the zero-buoyancy plume (ZBP) model (Singh & O'Gorman, 2013; Romps, 2014; Singh et al., 2019; Romps, 2021; Singh & Neogi, 2022). This model can analytically solve for the lapse rate and relative humidity profiles of an equilibrium atmosphere when small-scale processes like plume-environment mixing strength and cloud evaporation rate are specified. The model makes the approximation that the ensemble of convection can be treated as a single neurally

buoyant plume (Singh & O'Gorman, 2013), which is justified by the fact that the buoyancy of convective plumes is relatively small due to mixing and convective adjustment (usually less than 1K in observations and simulations, e.g., Lawson & Cooper, 1990; Romps & Charn, 2015). The zero-buoyancy plume model has been used to explain many aspects of the equilibrium atmosphere, especially the balance and sensitivities of radiative-convective equilibrium (RCE) simulations. Romps (2014) shows that the tropical relative humidity can be analytically expressed as a balance between convective detrainment and environment subsidence, and is independent of surface temperature. Singh et al. (2019) introduced the large-scale vertical motion into the model and can reproduce the temperature and moisture mean state response to large-scale forcing in RCE simulations. Romps (2021) expanded the model to account for convective aggregation and can reproduce how thermal profiles respond to the degree of aggregation. Singh & Neogi (2022) coupled the zero-buoyancy plume model with a dynamical solver for large-scale vertical motion through weak temperature gradients. It was able to reproduce a top-heavy large-scale upward motion like the large-scale circulation pattern in tropical convection-active ascent regions which is warmer than the tropical mean. Singh & Neogi (2022) also showed that entrainment mixing helps amplify this top-heavy response. In summary, the ZBP model can be a useful framework for understanding the fundamental aspects of the large-scale atmosphere state and the influence of small-scale processes.

One caveat of the solutions of the ZBP model provided in Singh & Neogi (2022) and Romps (2021) is the solution of the convective mass flux profile. Convective mass flux is of great importance because it drives precipitation and controls cloud fraction through mass detrainment. In their solutions, Singh & Neogi (2022) and Romps (2021) assumed equal fractional entrainment rate and detrainment rate, which in principle should suggest no vertical change in the convective updraft mass flux through the mass conservation. However, their solution of mass flux profile, e.g., Fig. 7 in Singh & Neogi (2022), does not follow this assumption, especially in the upper troposphere where mass flux rapidly decreases. Singh & Neogi (2022) acknowledged this caveat as an approximation

for simplicity and examined at which heights this approximation in their solutions is reasonable.

In km-scale convection-permitting equilibrium simulations, it remains unclear how small-scale mixing can influence convective mass flux and cloud fraction. At shorter time scales, mixing can influence individual convective updraft through entrainment and detrainment. At longer time scales, time-averaged mass flux is required to provide enough latent heating to balance the atmospheric radiative cooling through condensation. The processes involved in this energy balance can be influenced by the small-scale mixing through changing thermal structure and evaporation efficiency. In km-scale or hm-scale convection-permitting simulations, although there are no direct parameters like entrainment rate that control convective mixing, some model parameters, e.g., model resolution, could potentially change mixing strength. In idealized squall line simulations, Bryan et al. (2003) showed decreasing grid size from the order of 1km to the order of 100m tends to give more turbulent flow with resolved entrainment and overturning within clouds. By perturbing horizontal resolution in small-domain RCE simulations, Jeevanjee & Zhou (2022) (hereafter, JZ22) showed a striking dependence of anvil cloud fraction and mass flux on horizontal resolution. In their simulations with finer resolution, they observed that the updraft mass flux and anvil cloud fraction rise dramatically, with peak anvil cloud fraction from approximately 5% at the coarsest 16km grid size to over 40% at the finest 62.5m resolution. They argued that finer horizontal resolution corresponds to stronger mixing which can enhance cloud evaporation and lower precipitation efficiency. A smaller precipitation efficiency would then lead to greater cloud-base mass flux, which would lead to more mass flux reaching the upper troposphere and producing more anvil clouds. However, given the unconstrained effects of entrainment and detrainment, Jeevanjee (2022) showed that the relation between cloud base mass flux and precipitation efficiency is not necessarily tightly constrained; furthermore, and it is unclear whether changes in cloud base mass flux can consistently project to the upper troposphere.

In this work, we study and illustrate how small-scale mixing can influence the mass flux and anvil

cloud fraction. We first refine the ZBP model to get self-consistent solutions of mass flux profiles. We show that the inconsistency in previous studies comes from the ill-posed equations due to the mismatch in the numbers of equations and unknown variables. We provide a well-posed solution by adjusting unknown variables and incorporating a boundary condition involving RCE. In this refined ZBP model, we find that increasing mixing strength can lead to an increase of upper troposphere mass flux through two pathways, either more cloud evaporation or weaker stability in the upper troposphere. Furthermore, following JZ22, we test the dependence of mass flux and cloud fraction on mixing strength in small-domain RCE simulations by changing the horizontal resolutions with a fixed domain size. Consistent with JZ22, we found anvil cloud fraction and upper troposphere mass flux increase with increasing horizontal resolution. Through clear-sky energy balance analysis, our RCE simulations with two different microphysics schemes show two pathways as the ZBP model suggests. Compared to JZ22, our ZBP model and RCE simulations reveal an additional mechanism pathway in which the stability change in the upper troposphere can dominate the mass flux change. This pathway dominates when evaporative cooling is much smaller than the radiative cooling in the upper troposphere. However, it should be noted that the ZBP model primarily serves to illustrate these pathways and mechanisms, rather than directly emulate the simulation results. In addition to the sensitivity of anvil cloud fraction to resolution, we also noticed a large sensitivity of anvil cloud fraction to the microphysics scheme, which is also found in previous studies (e.g., Powell et al., 2012; Igel et al., 2015; Hu et al., 2021) and deserves further study.

The rest of the manuscript is structured as follows: in Section 2.3 we briefly describe the setup of the refined ZBP model with the complete derivation in Appendix A.1. Section 2.4 describes the insight from the solutions of the ZBP model. We show how large-scale equilibrium state mass flux and thermal profiles can respond to small-scale processes. Section 2.5 shows the results from the numerical RCE simulations. Section 2.6 is the discussion and summary.

2.3 Refined zero-buoyancy plume model with a self-consistent solution of mass flux

In this section, we describe the setup of our refined ZBP model. We briefly cover the most essential ingredients here, while the readers can find the complete derivation with a more detailed explanation in Appendix A.1. The ZBP model we use here is adapted from the ZBP model in Singh & Neogi (2022). The model presented in Singh & Neogi (2022) includes a thermodynamic component, i.e., the ZBP model, and a dynamical solver that uses the assumption of weak-temperature gradients to solve large-scale vertical motion. In this study, we utilize only the ZBP model without the dynamical solver to examine radiative-convective equilibrium with no large-scale vertical velocity.

The ZBP model assumes that the steady state of the atmosphere can be represented by a single updraft plume and downdrafts in the environment. The updraft and environment can exchange mass, water, and heat via entrainment and detrainment. The model assumes that the steady state of the atmosphere is neutrally buoyant with respect to the entraining plume (Singh & O'Gorman, 2013), i.e., the temperature in the updrafts is the same as the temperature in the environment at the same level. The virtual effect of vapor is ignored. Fig. A.5 shows that the time-averaged temperature anomaly (relative to the domain-mean) in the convective updraft in our cloud-resolving numerical simulations is indeed within 1K except at the top of troposphere above 14km. The model further presumes that the radiative cooling rate is a function of temperature. We choose to specify that the radiative cooling rate is $-1\,\mathrm{K\,day^{-1}}$ when the temperature is above 250K, and decreases to 0 at 200K through a cosine function. By combining conservation equations (for mass, water vapor, and moist static energy) and the hydrostatic balance equation, this ZBP model can be written as a system of ordinary differentiate equations (ODE):

$$\frac{\partial}{\partial z}(M_c, p, q, T) = F(M_c, p, q, T|\varepsilon, \mu) \tag{2.1}$$

where M_c, p, q, T represent the convective mass flux, pressure, specific humidity, and temperature. ε and μ are the entrainment rate and the evaporation parameter, which are the model parameters that we need to specify and can adjust. Given boundary conditions (we will discuss this later), we can solve Equation 2.1 to get the vertical profiles of mass flux, pressure, specific humidity, and temperature.

The major difference to the solution in Singh & Neogi (2022) is the treatment of mass flux and detrainment rate. Unlike our ODE Equation 2.1, in Singh & Neogi (2022), the mass flux is not included in their ODE equation. They assume the detrainment rate δ equals the entrainment rate ε. This would imply no vertical change in the convective updraft mass flux through the mass conservation equation:

$$\frac{\partial M_c}{\partial z} = M_c(\varepsilon - \delta) \tag{2.2}$$

However, since Singh & Neogi (2022) have more equations than unknown variables, they can still diagnose a mass flux profile by using the rest of their equations. This diagnosed mass flux profile will be inconsistent with the assumption of equal entrainment and detrainment rates. Singh & Neogi (2022) acknowledged this caveat as an approximation for simplicity. In this study, we relax this approximation and use a self-consistent method of solving the equations by allowing the difference between the fractional detrainment rate and the fractional entrainment rate to vary vertically. The shape of the mass flux profile is partially constrained by energetics, as the mass flux needs to diminish where radiative cooling starts to rapidly decrease in the upper troposphere. Therefore, the entrainment rate and detrainment rate cannot be completely independent. To specify the strength of plume-environment mixing in the model, one could specify either the entrainment rate, detrainment rate, or some other variable, such as the mixing rate in Bretherton et al. (2004). Here, for simplicity, we choose to specify the entrainment rate to impose the strength of plume-environment mixing. In this way, we have the same number of equations as the number of unknown variables.

Following Singh & Neogi (2022), in this model, cloud evaporation is parameterized as:

$$s_{\text{evap}} = \mu d(q^* - q) \tag{2.3}$$

where d is mass detrainment, q^* and q is the specific humidity in the saturated updraft and in the environment. A larger evaporation parameter μ tends to produce more cloud evaporation in the environment. This equation has two underlying assumptions. First, this equation assumes that the detrained flux of condensate equals $\mu_1 dq^*$, i.e., proportional to the detrained flux of water vapor with a scaling factor μ_1. Second, it assumes that the fraction of detrained condensate that evaporates (as opposed to precipitating to the ground) is $\mu_2(1 - q/q^*)$, where $(1 - q/q^*)$ approximately equals $1 - \mathcal{R}$ ($\mathcal{R}$ represents relative humidity RH) and μ_2 is another scaling factor. The μ in equation 2.3 is simply $\mu_1\mu_2$. It is likely that the ratio of condensate evaporation versus conversion to rain is less sensitive to RH when RH is far less than 1. We also explored a different parameterization defined by $s_{\text{evap}} = \mu dq^*(1 - \mathcal{R})^{0.5}$, which yielded results that are qualitatively similar (not shown).

Once we specify the boundary conditions, we can solve the system of ODEs. Here we specify the temperature, pressure, and specific humidity directly at the cloud base level at z=500m, but we cannot directly specify the mass flux at the cloud base. In principle, the amount of convection that can happen is determined by the amount of latent heating required to balance the radiative cooling in the whole atmosphere. That is, the mass flux at the cloud base is not determined by boundary processes but by the energy balance in the whole free troposphere. To satisfy radiative-convective equilibrium at all model levels, we require the mass flux to diminish to zero at the same level where the radiative cooling rate goes to zero (for the simple radiation we specify, this is the level where the temperature goes to 200K). We can search different cloud base mass flux values and find a unique value that satisfies this RCE condition. We can then use this cloud base mass flux value to integrate the ODE equation upward to get the full atmosphere profiles. The full details of the model equa-

tions, derivation of the solution, and some sensitivity tests are documented in Appendix A.1.

2.4 Insights from the refined zero-buoyancy plume model

With this refined ZBP model, we now test the sensitivity of the steady-state atmosphere profiles to the entrainment rate and the evaporation rate. One should focus primarily on the profile behavior below around 12-14km, above which the neural buoyancy assumption is no longer justified. First, we test the sensitivity to the fractional entrainment rate ε for two different values of the evaporation parameter μ (Fig. 2.1, upper row for $\mu = 1$ and lower row for $\mu = 0.1$). In both cases, with an increase in the entrainment rate, we observe an increase in the detrainment rate, mass flux, relative humidity, and latent cooling in the environment. The temperature in the upper troposphere is colder with a higher entrainment rate, and the dry stratification (ds/dz, where s is dry static energy) is lower. Fig. 2.1 suggests that increasing the entrainment rate can lead to a relatively uniform increase in mass flux from cloud base to anvil level.

Considering the dry static energy balance $M_c = (Q_{rad} + Q_{lat})/(\frac{\partial s}{\partial z})$, increasing entrainment rate leads to both increasing cloud evaporation and less stratified upper troposphere. Both these factors can contribute to an increasing mass flux. When cloud evaporation is efficient (Fig. 2.1 with $\mu = 1$), the change of latent cooling can dominate the change in mass flux. However, when cloud evaporation is weak (Fig. 2.1 with $\mu = 0.1$), the absolute latent cooling and the change of latent cooling are small compared to the prescribed radiative cooling rate. The change in stratification due to increasing entrainment rate is larger with small μ and can dominate the change of mass flux. Environmental relative humidity is important in determining the sensitivity of stability to changing the entrainment rate. The relative humidity is lower with μ=0.1 than with μ=1 (Fig. 2.1c and i). In this zero-buoyancy plume model, the lapse rate in the environment equals to the lapse rate of the updraft plume. The impact of the entrainment dilution on the updraft lapse rate depends on the

environmental saturation deficit (Eq. (A.16)). Therefore, a small μ tends to make stability more sensitive to the change of entrainment (Fig. 2.1e and k).

In Fig. 2.2, we maintain a constant fractional entrainment rate as 0.5 km^{-1} and test the sensitivity of atmosphere profiles to the evaporation parameter μ. As the evaporation strength increases, the free troposphere is warmer, deeper, and more moist (Fig. 2.2b,2.2c,2.2d). The increased relative humidity is a direct result of the enhanced efficiency of cloud evaporation. This is consistent with JZ22 which shows that the evaporation efficiency plays an important role for the relative humidity, especially in the upper troposphere. Consequently, with a more moist atmosphere, the entrainment dilution has a weaker impact on the temperature of the updraft plume, resulting in a warmer and deeper troposphere. In the lower troposphere, we see a clear increase in mass flux with increasing cloud evaporation. However, in the upper troposphere, the mass flux adjustment is more akin to an upward shift with weak change in magnitude. The peak mass flux near the anvil level remains largely unchanged, suggesting a minor change in the convergence of mass flux at higher altitudes. From the perspective of the energy budget, an increase in the evaporation rate could induce greater latent cooling. However, this is offset by an increase in stability in the upper troposphere, effectively suppressing the change in mass flux (Fig. 2.2e and 2.2f).

In addition to the energy budget, a different way to understand the mass flux response to entrainment rate change in this analytical model is through the Betts's rule described in Jeevanjee (2022). Consider the water vapor budget for the atmosphere above a certain level z. The mass flux at z satisfies:

$$M_c q^* (1 - \mathcal{R}) = \int_z^{z_{\text{top}}} -c_p \rho Q_{\text{rad}} dz / \ell_v \tag{2.4}$$

where z_{top} is the top of the atmosphere, c_p is the isobaric specific heat, ρ is air density, ℓ_v is the latent heat of vaporization. The left-hand side (LHS) represents the net water vapor transported upward across level z by saturated updraft and unsaturated subsidence. In a steady state, this transport of

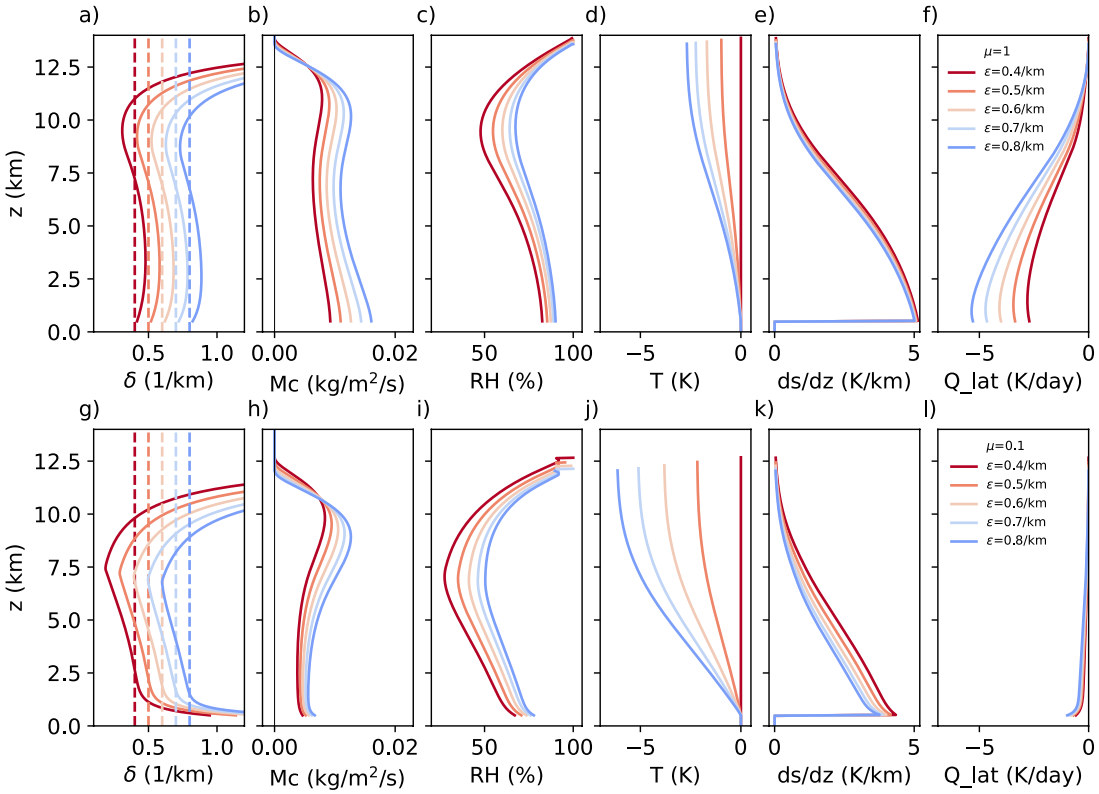


Figure 2.1: Atmosphere profiles in the zero-buoyancy plume model with varying entrainment rates (warmer color represents lower entrainment rate). The upper row has a cloud evaporation parameter μ=1. The variables shown are detrainment rate (a), updraft mass flux (b), relative humidity (c), temperature (d), the vertical gradient of dry static energy (e), and latent heating rate due to cloud evaporation in the environment (f). The dashed lines in panel a are the profiles of the prescribed entrainment rate. The temperature in panel d is shown as deviation to one of the simulations, which is denoted by the red line with zero deviation. The lower row is similar to the upper row but with cloud evaporation parameter μ=0.1.

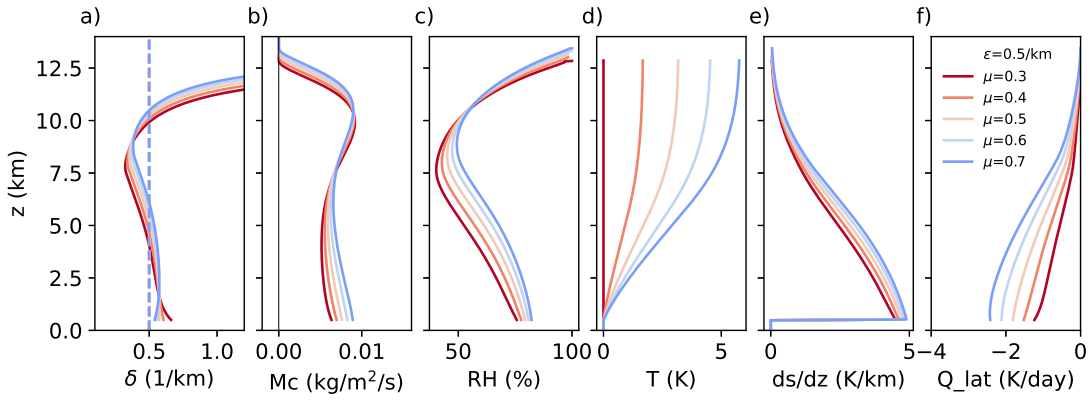


Figure 2.2: Similar to Fig. 2.1 but with a fixed entrainment rate of 0.5 km^{-1} and varying cloud evaporation parameter (warmer color represents less efficient cloud evaporation).

vapor must be balanced by the net condensation, which is required to balance the total radiative cooling above level z (the right-hand side, RHS). Since in the model, the prescribed radiative cooling is constant at $-1\,\mathrm{K\,day^{-1}}$ for the troposphere where the temperature is larger than 250K, the change in the RHS is relatively small, especially for the lower and middle troposphere. When we increase the horizontal mixing (Fig. 2.1), the relative humidity increases, and more clouds get detrained. The temperature through the whole troposphere also decreases, leading to a decreasing saturation vapor specific humidity q^*. To satisfy the equation, the mass flux on the LHS has to increase to provide sufficient upward vapor transport.

2.5 Sensitivity of small-domain RCE simulations on horizontal resolution

2.5.1 Experiment setup

We use the System for Atmosphere Modeling (SAM; Khairoutdinov & Randall, 2003), version 6.10.6, configured as a cloud-resolving model. We run three-dimensional RCE simulations using the same domain size of 128km×128km with different horizontal resolutions of 4km, 2km, 1km, 500m, 250m, and 125m. All simulations use 60 vertical levels with the model top located at 26km and a

rigid-lid top boundary condition. The vertical grid spacing increases from 75m near the surface to a constant of 500m throughout the whole free troposphere and above. The model time step is specified as 15 seconds but will automatically reduce by half iteratively when the Courant–Friedrichs–Lewy (CFL) condition is not satisfied. The effective time step for finer resolutions will be shorter than that for coarser resolutions to satisfy the CFL condition. A sponge layer is located in the upper 30% of the model domain (i.e., above 18km). The radiation scheme used is the Rapid and Accurate Radiative Transfer Model for General Circulation Models (RRTMG) (Iacono et al., 2008). A simple Smagorinsky-type scheme (Khairoutdinov & Randall, 2003) is used for the effect of subgrid-scale motion. In this subgrid-scale scheme, the horizontal grid size does not explicitly change the mixing length, which is set to the vertical grid size in unstable atmosphere and is modified by the local stability in stable atmosphere. When doing the horizontal scalar diffusion, the eddy diffusivity is scaled by a grid factor that depends on horizontal grid size. However, the resolution dependence shown in the later sections is largely unchanged when we turn off the horizontal scalar diffusion (not shown). We use a constant solar insolation (no diurnal cycle) with a fixed solar constant of 683.5 W m^{-2} and a zenith angle of 50.5°. The domain-averaged horizontal wind is nudged to zero at each vertical level with a nudging time scale of one hour. Sea surface temperature is uniformly fixed at 303K.

For the simulations with horizontal resolution ranging from 4km to 250m, the first 50 days are taken as the model spinup and considered long enough for the model to reach equilibrium. After the 50-day spinup, a 20-day post-equilibrium period is used for analysis. The 30-samples-per-hour data are then averaged to obtain an hourly output of domain-mean statistics. For the 125m-resolution simulation, we initialize the simulation with the equilibrium temperature and moisture profiles from the 500m-resolution simulation. Then we run only 30 days for spinup and 5 days for analysis.

We use two different microphysics schemes: SAM single-moment scheme (SAM1MOM, Khairout-

dinov & Randall, 2003) and a double-moment Morrison et al. scheme (Morrison et al., 2005). The only prognostic water species in the SAM1MOM scheme are the total water (vapor plus non-precipitating condensate) and precipitating water. The SAM one-moment scheme uses an instantaneous saturation adjustment to determine the partition of water vapor, generate and remove cloud condensate. Cloud condensate is then divided into liquid and/or ice phases based on temperature. Between 273.16 K and 253.16 K, the partitioning of cloud condensate into cloud ice and liquid water depends linearly on the temperature (at 253.16 K, all condensate is ice; at 273.16 K, all condensate is liquid water). Similarly, precipitating water is diagnostically partitioned into rain, snow, and graupel based on the temperature. The saturation adjustment occurs at each effective time step. It should be noted that with finer resolution, saturation adjustment occurs faster given the effective time step is smaller. The Morrison et al. scheme includes many more pathways for conversion between different hydrometeors. The Morrison et al. scheme prognostically calculates the mixing ratio and number concentration of cloud water, cloud ice, rain, snow, and graupel. Regarding cloud-radiation coupling, for frozen water species, only cloud ice is radiatively active in the SAM1MOM scheme, while both cloud ice and snow are radiatively active in the Morrison et al. scheme. In addition, different codes are used to calculate cloud optical depth for the two microphysics schemes before the radiation calculation. We tried turning off the cloud-radiative interactions for both microphysics schemes, assuming clear-sky conditions in all radiation calculations. The atmosphere mean state and resolution sensitivity for the SAM1MOM scheme remains qualitatively the same. For the Morrison et al. scheme, the impact of turning off cloud-radiation interactions on the resolution dependence is discussed in a later section.

The Morrison et al. scheme tends to produce more optically thin ice clouds in the upper troposphere (e.g., Powell et al., 2012; Hu et al., 2021) and consequently strong atmospheric cloud radiative heating in the middle and upper troposphere. This stronger atmospheric cloud radiative heating can stabilize the upper troposphere and weaken the convective updraft reaching the upper tropo-

sphere (Hu et al., 2021). It is unclear why the Morrison et al. scheme tends to produce excessive anvil clouds compared to other microphysics schemes including the SAM1MOM scheme. Powell et al. (2012) argued that the excessive anvil clouds are due to the large cap value for ice number concentration in the Morrison et al. scheme. However, even with a much smaller cap value in the latest version of the SAM, the Morrison et al. scheme still tends to produce more anvil clouds (e.g., Hu et al., 2021). Unlike the SAM1MOM scheme, the Morrison et al. scheme does not use saturation adjustment for cloud ice (but has saturation adjustment for liquid cloud), potentially leading to slower ice sublimation. In addition to sublimation, the Morrison et al. scheme also seems to have slower cloud ice sedimentation (see later in Fig. 2.5) compared with the SAM1MOM scheme. Other microphysical processes that convert ice clouds to snow are also important sinks of cloud ice (Fig. 8 in Bao et al., 2019). It is worth studying in the future to understand which of these microphysical processes are responsible for the excessive anvil clouds in the Morrison et al. simulations.

2.5.2 Impacts of horizontal resolution on mean atmospheric state

To confirm that changing horizontal resolution is indeed an effective way of changing mixing strength, we first examined how the entrainment rate varies across different simulations. Here, we choose to illustrate the entrainment change with a method of a spectrum of entraining plumes, following Kuang & Bretherton (2006). In this plume spectrum calculation, we use environmental profiles from each simulation to infer the entrainment rate for updrafts. We take the Morrison et al. simulations for illustration, while the SAM1MOM simulations have very similar results (not shown). In Fig. 2.3a and 2.3c, we show the convective updraft mass flux distribution in the space of frozen moist static energy (FMSE) and height. FMSE is defined as $c_p T + gz + L_v q - L_f q_i$. The individual lines represent the FMSE profiles of entraining plumes rising from the cloud base at different entrainment rates. Here, these entraining plumes are assumed to follow $dh/dz = -\varepsilon(h - h_{env})$, where h_{env} is the time-mean, domain-mean FMSE profile and ε is the fractional entrainment rate.

The convective updrafts in the 125m-resolution simulation (Fig. 2.3c) shift towards FMSE profiles with higher entrainment rate compared to the updrafts in the 4km-resolution simulation (Fig. 2.3a). Once we have computed the FMSE profiles with varying entrainment rates, we can measure the amount of mass flux allocated to each entrainment rate bin. Subsequently, we can represent the updraft mass flux in the space of height and entrainment rate. As shown in Fig. 2.3b and 2.3d, it is apparent that the mass flux distribution shifts towards higher entrainment rates with finer resolution. There is also a shift towards lower entrainment rates with higher altitudes, with roughly a factor of 3 difference in mode entrainment rate between lower tropospheric and upper tropospheric plumes (Kuang & Bretherton (2006), see also Zhou & Xie (2019)). It is important to note that the sensitivity of the entrainment rate on grid size could be model-dependent. In the SAM model, the entrainment mixing seems to be contributed mainly by numerical diffusion, while sub-grid diffusion is very weak in the free troposphere. Whether the resolution dependence of the entrainment rate would hold with other models using different advection and sub-grid diffusion schemes needs to be further tested.

In our small-domain RCE simulations, increasing horizontal resolution leads to increased relative humidity and convective updraft mass flux (Fig. 2.4 and Fig. 2.5), which is consistent with the response of the analytical zero-buoyancy plume model to increasing mixing strength. Fig. 2.4 illustrates the resolution-dependent behavior of cloud fraction, atmospheric cloud radiative effects (ACRE), relative humidity, and the distribution of column-integrated ice-water path. ACRE is defined as the difference between the actual radiative heating rate profile and the hypothetical clear-sky radiative heating rate profile which assumes no cloud condensate. Both the SAM1MOM and Morrison et al. simulations exhibit a rise in relative humidity throughout the entire free troposphere and an increase of high cloud fraction with finer resolution. A grid is classified as cloudy if the cloud mass (the sum of ice and liquid water) mixing ratio exceeds 10^{-5} kg kg^{-1}. As the grid spacing decreases from 4km to 125m, the peak anvil cloud fraction increases from 7.5% to 13% in

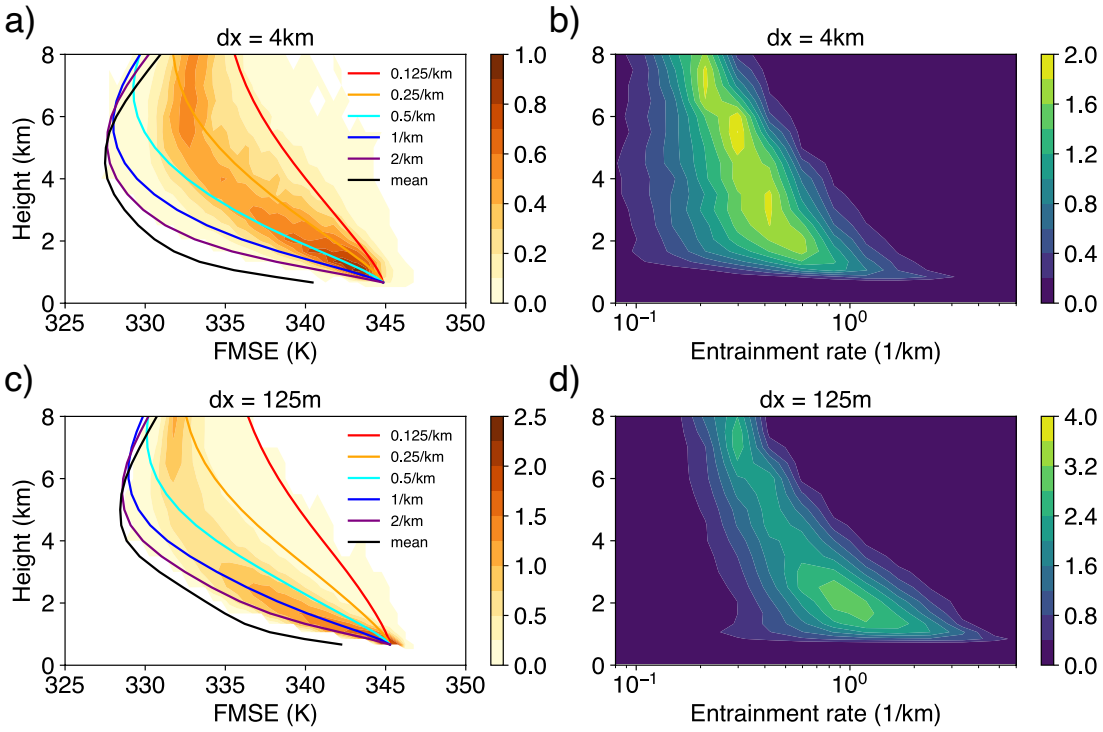


Figure 2.3: The distribution of convective updraft mass flux in FMSE-height space (left column) and in entrainment-height space (right column) for grid size of 4km (upper row) and 125m (lower row). In the left column, we show the mass flux distribution (with a unit of kg m^{-2} s^{-1} bin^{-1}) binned by their FMSE (in unit of K). There are 50 bins with 0.5K intervals between 325K to 350K. The individual lines represent the FMSE profiles of entraining plumes rising from the cloud base with different entrainment rates, except the black line which represents domain-mean FMSE profiles. In the right column, we show the mass flux distribution (with a unit of kg m^{-2} s^{-1} bin^{-1}) binned by their effective entrainment rate. The bin boundaries have entrainment rates of $2^{(i/2)-4}$, for $i = 0, 1, 2, \ldots, 16$, with a unit of km^{-1}. We calculated the instantaneous FMSE profiles with these different entrainment rates and binned the mass flux according to these FMSE profiles. For example, all the mass flux between the cyan and blue lines in the left panels will be counted as having entrainment rate between 0.5 to 1 km^{-1} in the right panels.

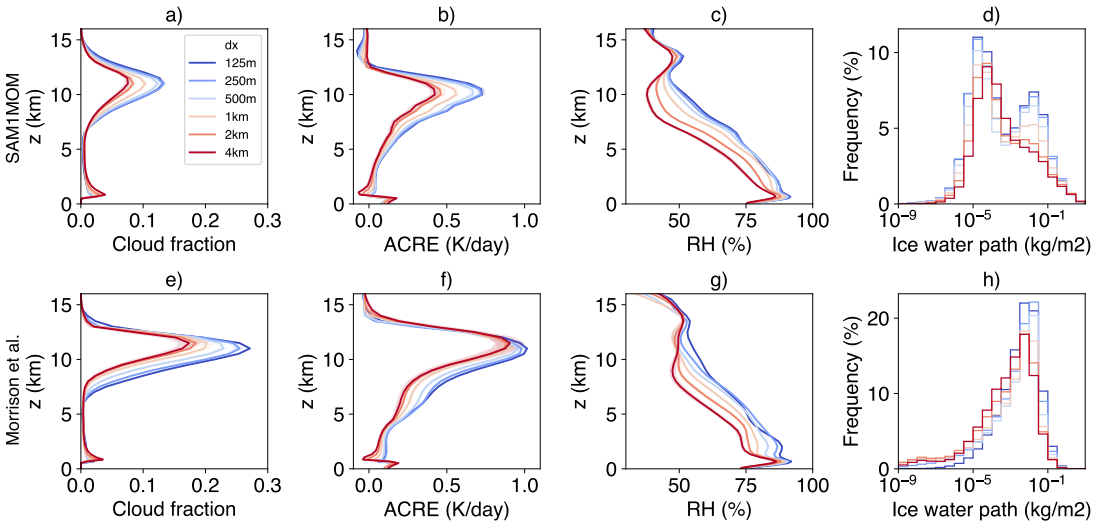


Figure 2.4: Domain-mean steady-state profiles of cloud fraction (left column), atmosphere cloud radiative effects (the second column), and relative humidity (the third column). The last column shows the histogram of column integrated ice-water path. The upper row corresponds to the SAM1MOM simulations, while the lower row represents the Morrison et al. simulations. Different colors indicate varying grid sizes, with warmer colors denoting coarser resolutions. Shaded uncertainty range is estimated as the standard deviation of time-mean profiles of each non-overlapping 2.5-day time block.

the SAM1MOM simulations (Fig. 2.4a) and from 17% to 27% in the Morrison et al. simulations (Fig. 2.4e). This amplified cloud fraction subsequently leads to increased atmospheric cloud radiative heating (ACRE) throughout most of the free troposphere (Fig. 2.4b and 2.4f). The histogram of ice-water path shows that the cloud amount increase is primarily contributed by the more frequent occurrence of columns with ice-water path of near 0.01-0.1 kg m^{-2} (Fig. 2.4d and 2.4h). Hu et al. (2021) (their Fig. 9) showed that this range of ice-water path corresponds to optically thin clouds and has a heating effect for the troposphere, consistent with the ACRE changes. In Fig. A.6, we also present cloud mixing ratio profiles, showing both domain-mean values and averages over cloudy grids only. For both microphysics schemes, increasing the model resolution tends to decrease the liquid cloud mixing ratio, both in the domain mean and in the cloudy-grid mean, while simultaneously reducing the low cloud fraction (Fig. 2.4a and 2.4e).

The increase in anvil cloud fraction can be related to the increasing convective updraft mass flux. Anvil cloud fraction can be diagnosed as the product of mass detrainment and cloud lifetime (e.g., Seeley et al., 2019a; Beydoun et al., 2021). In Fig. 2.5, we present profiles of convective updraft mass flux and in-cloud sedimentation flux divergence to look at the change of mass detrainment and lifetime change. The convective updraft is characterized by grids with a vertical velocity greater than 1 m s^{-1} and a cloud mixing ratio exceeding 10^{-5} kg kg^{-1}. Changing the velocity threshold to be greater than 0 or further including convective downdraft still shows a qualitatively similar trend of increasing mass flux (Fig. A.7). The in-cloud sedimentation flux divergence is defined as dqc_{sed}/qc averaged over cloudy grids, where qc is the cloud mass (ice plus liquid water) mixing ratio and dqc_{sed} is the tendency of qc due to sedimentation of cloud ice. This sedimentation flux divergence is the major term of the net removal rate in Beydoun et al. (2021) and could be interpreted as one over lifetime. Sedimentation flux divergence is positive above approximately 10 km and negative below, representing cloud ice falling from the detraining level downwards. In both the SAM1MOM and Morrison et al. simulations, the convective updraft mass flux at above 11km increases with higher

resolution, signifying an increased vertical mass convergence above this altitude. By mass continuity, the increase of vertical convective mass flux convergence corresponds to an increase of mass detrainment and an increase of horizontal mass convergence in the clear-sky region. The convective updraft mass flux in the middle and lower troposphere shows non-monotonic change. This may be partly due to increased ACRE with finer resolution (e.g., see Fig. A.3 and the corresponding discussion in Appendix A.1). The sedimentation rate is weaker at finer resolutions in the SAM1MOM scheme but is slightly stronger in the Morrison et al. scheme. Hence, the observed increase in cloud fraction with finer resolution in both SAM1MOM and Morrison et al. simulations is predominantly driven by the amplification of mass detrainment. The contribution from lifetime changes is less certain and could be contingent on the microphysics schemes employed.

2.5.3 Budgets for environmental subsidence

In this section, we investigate the mechanisms responsible for the increase in convective updraft mass flux associated with finer resolutions. Because of mass continuity, the mass flux in convective updrafts must be in balance with subsidence in the surrounding environment, which we define as grids possessing a cloud mixing ratio less than 10^{-5} kg kg^{-1}. We shall use dry static energy budget to gain some understanding of changes to environmental subsidence. By employing the dry static energy budget of the environment, we decompose the subsidence and will demonstrate that modifications to mass flux profiles could be attributed to changes in both cloud evaporation rate and entrainment rate. In reality, the changes in cloud evaporation and in entrainment rate are not purely independent as the change of horizontal mixing can influence both of them.

The dry static energy is defined as $s = c_p T + gz$. The budget equation for dry static energy is given by

$$\frac{\partial s}{\partial t} + \vec{u} \cdot \nabla_{\mathrm{h}} s + w \frac{\partial s}{\partial z} = Q_{\mathrm{rad}} + Q_{\mathrm{lat}} \tag{2.5}$$

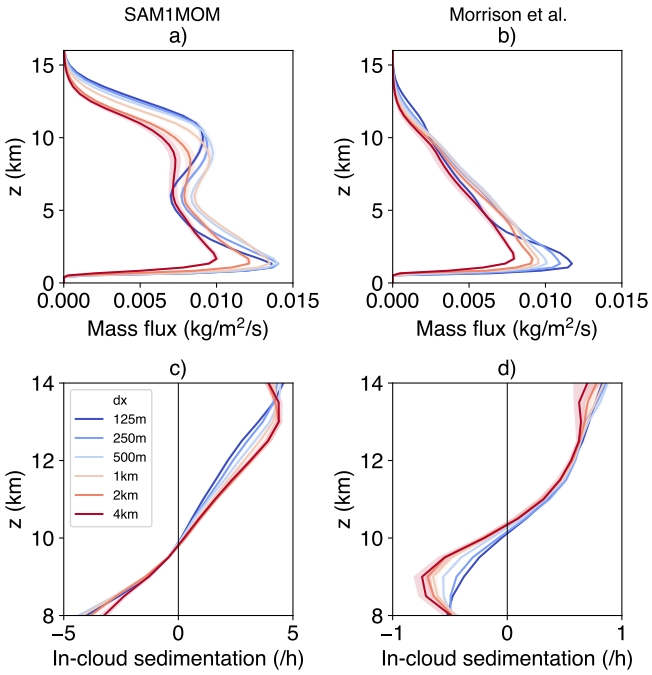


Figure 2.5: Domain-mean steady-state profiles of convective updraft mass flux (upper row) and cloud ice sedimentation rate (lower row). The left column corresponds to the SAM1MOM simulations, while the right column represents the Morrison et al. simulations. Different colors indicate varying grid sizes, with warmer colors denoting coarser resolutions. Shaded uncertainty range is estimated as the standard deviation of time-mean profiles of each non-overlapping 2.5-day time block.

where Q_{rad} is the radiative heating, and Q_{lat} is the latent heating in the environment. By averaging over all environmental grids and time, and ignoring the time tendency, we obtain:

$$\langle \vec{u} \cdot \nabla_{\text{h}} s \rangle + \langle w \frac{\partial s}{\partial z} \rangle = \langle Q_{\text{rad}} \rangle + \langle Q_{\text{lat}} \rangle \tag{2.6}$$

After further decomposition of $\langle w \frac{\partial s}{\partial z} \rangle = \langle w \rangle \langle \frac{\partial s}{\partial z} \rangle + \langle w' \frac{\partial s'}{\partial z} \rangle$, the averaged environmental subsidence can be expressed as:

$$\langle w \rangle = \frac{\langle Q_{\text{rad}} \rangle}{\langle \frac{\partial s}{\partial z} \rangle} + \frac{\langle Q_{\text{lat}} \rangle}{\langle \frac{\partial s}{\partial z} \rangle} - \frac{\langle \vec{u} \cdot \nabla_{\text{h}} s \rangle}{\langle \frac{\partial s}{\partial z} \rangle} - \frac{\langle w' \frac{\partial s'}{\partial z} \rangle}{\langle \frac{\partial s}{\partial z} \rangle} \tag{2.7}$$

This equation encapsulates the energy balance within the environment, implying that the subsidence heating is counterbalanced by the cooling induced by radiation and phase changes in water. In Fig. 2.6, we show the profiles of latent-driven and radiation-driven subsidence for the SAM1MOM simulations. The combined effect of latent- and radiation-driven subsidence closely mirrors the subsidence deduced from model output, and the contribution of advection terms appears minor in comparison to the contribution of radiation and latent heating (not shown).

The subsidence near the anvil level increases with finer resolution (Fig. 2.6a), which is consistent with the change of convective updraft mass flux. In the SAM1MOM simulations, a large portion of the increasing subsidence is counteracted by the negative latent heating in the environment due to evaporation and sublimation of clouds (Fig. 2.6b). Conversely, negative radiative heating accounts for a relatively smaller portion of this balance (Fig. 2.6c). The relative contribution of latent and radiative heating in the Morrison et al. scheme is somewhat different. We will probe into the nuances of the Morrison et al. simulations later in this section. It is important to underscore that the role of latent heating can be influenced by the specific definition of "environment". In our study, the environment, defined as grids with a cloud mixing ratio less than 10^{-5} kg kg^{-1}, incorporates grids distanced from clouds as well as those near clouds, which experience evaporation and sublimation. Results in the following paragraphs are not sensitive to the choice of cloud threshold. Changing the

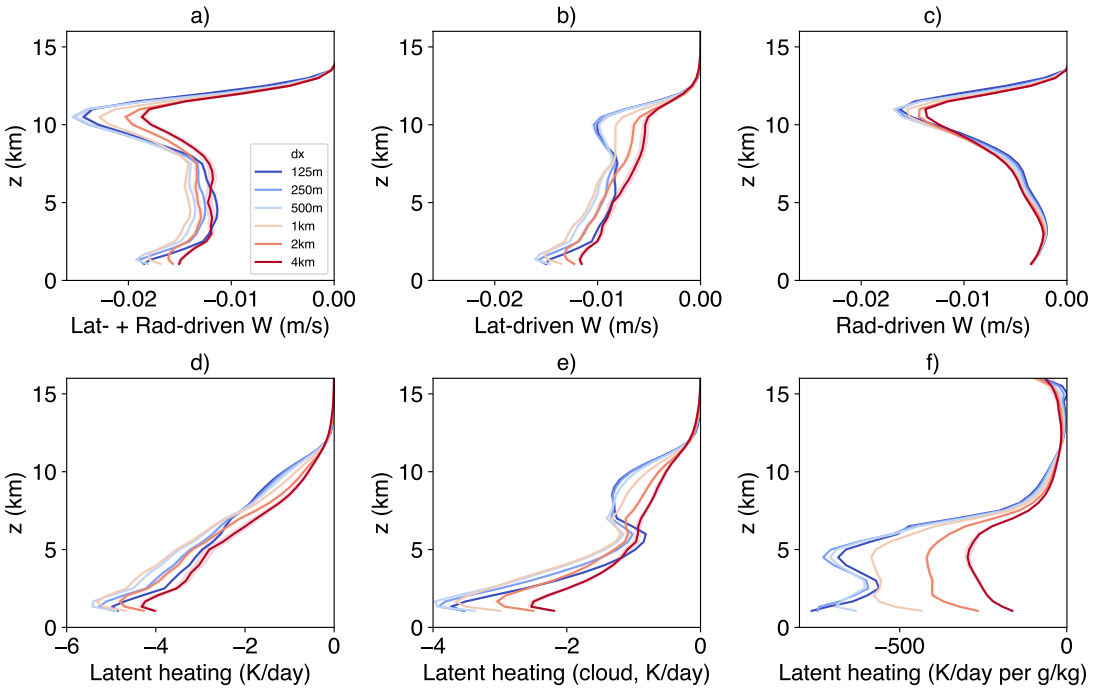


Figure 2.6: Energy budget for environmental subsidence for the SAM1MOM simulations. The first row shows the subsidence contributed by latent heat (panel b), by radiative cooling (panel c), and by both (a). Panel d shows the latent heating rate averaged in environments. Panel e shows the latent heating rate associated with the phase change between clouds and vapor averaged in environments. Panel f normalizes the cloud-related latent heating rate in panel e by the domain-mean cloud mixing ratio. Shaded uncertainty range is estimated as the standard deviation of time-mean profiles of each non-overlapping 2.5-day time block.

threshold from 10^{-5} kg kg^{-1} to 10^{-7} kg kg^{-1} results in minimal change. Such insensitivity might be attributed to the model's time-splitting procedure, where evaporation is calculated before the output of the cloud mixing ratio. Consequently, grid cells can reflect marginal cloud mixing ratios while still indicating evaporation in the resultant data.

The change of latent-driven subsidence is consistent with the change of latent heating in the environment (Fig. 2.6d). In the upper troposphere, the cooling from phase change is primarily associated with cloud evaporation/sublimation (Fig. 2.6e). For simplicity, we will henceforth use the term "evaporation" to refer to both the evaporation of cloud water and the sublimation of cloud

ice. The cooling due to re-evaporation of precipitation, which is not displayed here, is less significant than that of clouds in the upper troposphere, although it presents a similar strength in the lower troposphere. We have shown that a finer resolution model tends to generate more clouds and updraft mass flux. Therefore, the observed increase in latent cooling might be simply a consequence of the larger amount of clouds available for evaporation. However, an interesting observation arises when we normalize the cooling due to cloud evaporation by the domain mean cloud mass mixing ratio (Fig. 2.6f). Domain mean cloud mixing ratio is proportional to the total cloud mass in each layer. It becomes evident that, per unit mass, clouds tend to induce a greater amount of cooling in the environment when modeled at finer resolution.

The observed enhancement in evaporation could be associated with the model resolution through the geometric representation of cloud boundaries. We will use clouds at the anvil level as an example, but we assume the intuition behind should apply to clouds at all levels. Horizontal snapshots of the cloud mixing ratio at an altitude of z=10km are depicted in Fig. 2.7a and 2.7b. These images represent two 32km$\times$32km subdomains in the 4km-resolution and 125m-resolution simulations respectively. When compared to the clouds in the coarser 4km-resolution simulation, the clouds in the 125m-resolution simulation exhibit more complex boundary structures and tend to be more dispersed. As a result, clouds modeled at finer resolutions exhibit a higher perimeter area ratio (Fig. 2.7c). In other words, for a cloud patch of the same area, the total perimeter will be longer in the simulation with finer resolution. This effect is somewhat analogous to the coastline paradox in fractal geometry, where the measured length of a coastline varies depending on the scale of measurement. The increased perimeter-to-area ratio exposes a larger mass of the cloud to the environment, potentially leading to greater evaporation near the cloud edges.

Siebesma & Jonker (2000) explored the fractal nature of cumulus clouds in Large-Eddy Simulations. They argued that while a coarse grid will underestimate cloud surface area, the total sub-grid turbulent transport could become resolution-independent if the grid size is within the inertia sub-

range. However, in our simulations with a grid size on the order of 1km, sub-grid diffusion in the free troposphere is minimal. We observed that turning off horizontal sub-grid diffusion of scalars (such as energy and water) resulted in only minor changes to the profiles of cloud fraction and environmental evaporation (not shown). The cloud evaporation of deep convection is substantially influenced by numerical diffusion and can be enhanced by a larger perimeter-to-area ratio. To illustrate this point, similar to Equation 8 in Jeevanjee & Zhou (2022), we derived an equation (see Appendix A.2 for complete derivation) that describes the relationship between cloud evaporation in relation to resolved advection and the perimeter-to-area ratio:

$$\frac{Q_{\text{lat,env}}}{q_{\text{m}}} = \frac{L}{A} U_{\text{adv}} \frac{q_{\text{c,edge}} + q^*_{\text{v,env}}(1-\mathcal{R})}{q_{\text{cld}}} \frac{L_{\text{v}}}{2c_p f_{\text{env}}} \tag{2.8}$$

The Equation 2.8 indicates that the evaporation due to horizontal mixing at cloud edges is dependent on several factors. These include the perimeter-to-area ratio (L/A), the resolved horizontal velocity near the cloud edge (U_{adv}), the cloud mixing ratio near the cloud edge ($q_{\text{c,edge}}$), the saturation deficit in the environment ($q^*_{\text{v,env}}(1-\mathcal{R})$), and the average cloud mixing ratio within cloudy grids (q_{cld}). We verify this equation at the anvil level, characterized by relatively weak vertical motion near the cloud edge, hence, making cloud evaporation predominantly attributable to horizontal mixing. Fig. 2.7d demonstrates that the diagnosed evaporation using Equation 2.8 qualitatively aligns with the direct model output. From this equation, it is evident that an increased perimeter-to-area ratio can positively contribute to enhanced cloud evaporation. In Appendix A.2, we delve into how other terms in Equation 2.8 vary with model resolution. It is more difficult to validate Equation 2.8 at lower levels. In the middle troposphere, clouds are typically very close to the convective core, and evaporation/condensation associated with vertical motion may not be neglected. However, we assume the enhanced horizontal mixing and larger perimeter area ratio should still positively contribute to the enhanced evaporation we show in Fig. 2.6f. It is important to note the

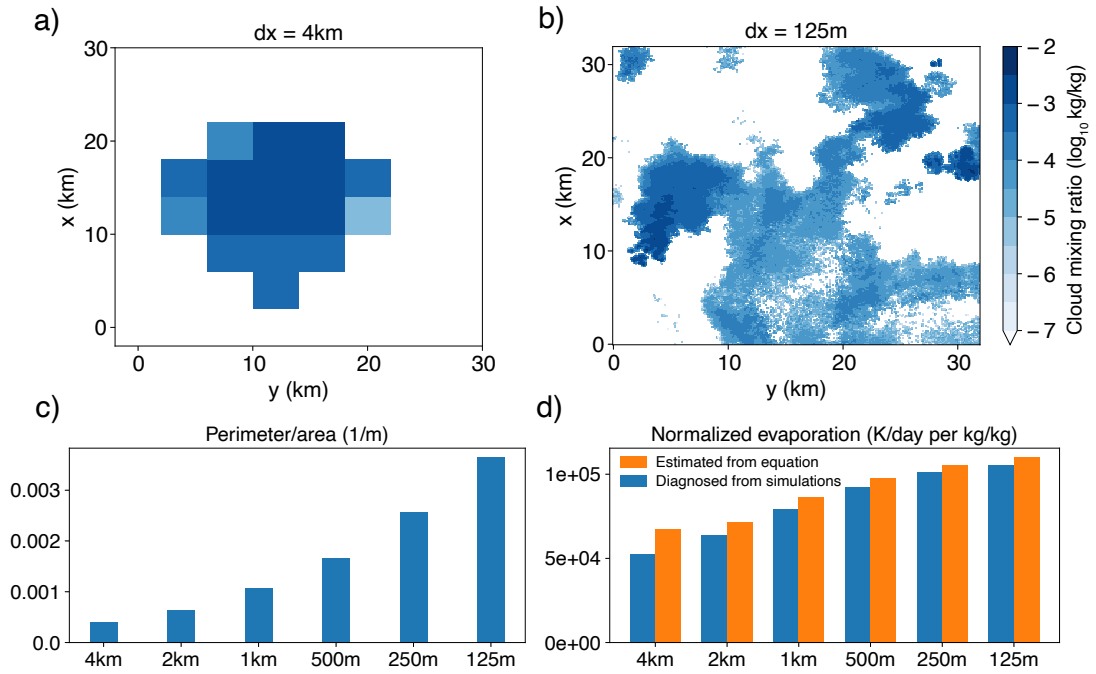


Figure 2.7: The upper row shows the cloud mixing ratio snapshots at z=10km in a 32km×32km subdomain in the SAM1MOM simulations with grid size of 4km (a) and 125m (b). Panel c shows the perimeter area ratio in the SAM1MOM simulations with different grid sizes. Panel d shows the normalized evaporation in Fig. 2.6f at z=10km. Blue bars are direct model diagnostic values, and orange bars are estimated by Equation 2.8.

importance of enhanced horizontal mixing occurring at all levels, not solely at the anvil level. More efficient evaporation at lower levels could contribute to mass flux increase at those levels and, by mass continuity, should have a continuing influence on mass flux at higher levels (Jeevanjee & Zhou, 2022).

In the Morrison et al. simulations, consistent with the mass flux change, we also observe an enhancement in subsidence near the anvil level, as shown in Fig. 2.8a. However, the contribution from latent-driven subsidence (Fig. 2.8b) is weaker in the Morrison et al. scheme compared to the SAM1MOM scheme. Primarily, the subsidence change near the anvil level is dominated by radiation-driven subsidence (see Fig. 2.8c). We further dissect the radiation-driven subsidence into radiation and stability components. The radiative cooling shows slight non-monotonic changes

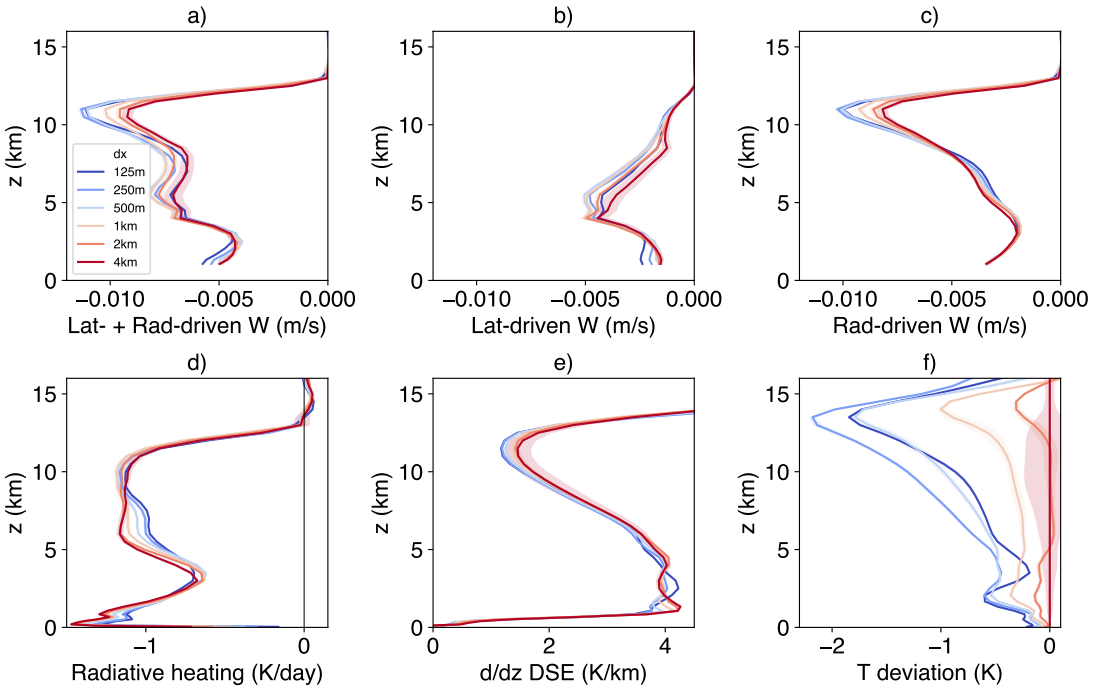


Figure 2.8: Energy budget for environmental subsidence for the Morrison et al. simulations. The first row shows the subsidence contributed by latent heat (panel b), by radiative cooling (panel c), and by both (a). Panel d shows the radiative heating rate averaged in environments. Panel e shows the vertical gradient of dry static energy averaged in environments. Panel f shows the absolute temperature profiles as deviation to the 4km Morrison et al. simulation. Shaded uncertainty range is estimated as the standard deviation of time-mean profiles of each non-overlapping 2.5-day time block.

(Fig. 2.8d), while the upper troposphere is less stable with finer resolution (Fig. 2.8e and 2.8f). The change in stability could be associated with the shift in the entrainment rate (Fig. 2.3), which tends to increase with finer resolution.

The Morrison et al. simulations reveal the stability pathway that is similar to the behavior of the ZBP model with a small evaporation parameter μ. The similarity here primarily comes from the diminishing evaporation in the upper troposphere. However, given the complexity of the SAM1MOM and Morrison microphysics schemes, it is hard to attribute their difference to a single parameter like μ or a single microphysical process. It is important to clarify that we do not intend to use the low

evaporation parameter case in the ZBP model to directly explain the resolution sensitivity observed in the Morrison et al. simulations. Rather, the ZBP model results are meant to illustrate potential pathways and mechanisms by which entrainment can influence upper-level clear-sky convergence and net detrainment.

Fig. 2.9 explores the reasons behind the distinctive environmental energy balance regime observed in the Morrison et al. simulations compared to the SAM1MOM simulations. In the Morrison et al. simulations, the cooling effect from evaporation in the upper troposphere is notably weaker than that from radiation. Two factors could contribute to this subdued evaporation: diminished updrafts and a slower evaporation rate. As previously noted, the Morrison et al. scheme tends to generate more anvil clouds, likely due to the significantly slower ice sedimentation removal rate and prolonged lifetime (refer to Fig. 2.5). The enhanced cloud radiative heating in the Morrison et al. simulations could stabilize the upper troposphere, thereby reducing the intensity of updrafts. When we disable cloud-radiation interactions in the Morrison et al. simulations (represented by solid lines in Fig. 2.9), we observe an increase in upper troposphere convective updrafts and stronger latent-driven subsidence, compared to the default Morrison et al. simulations (dotted lines in Fig. 2.9). Additionally, the Morrison et al. scheme does not employ saturation adjustment for cloud ice, potentially slowing evaporation compared to the SAM1MOM scheme. When we deactivate the cloud-radiation interactions and accelerate the cloud ice sublimation rate 100 times to mimic the saturation adjustment (dashed lines in Fig. 2.9), the result is faster evaporation and intensified updrafts. Consequently, latent-driven subsidence now contributes comparably to radiation-driven subsidence in modifying total subsidence near the anvil level as resolution becomes finer (see Fig. 2.9b to d).

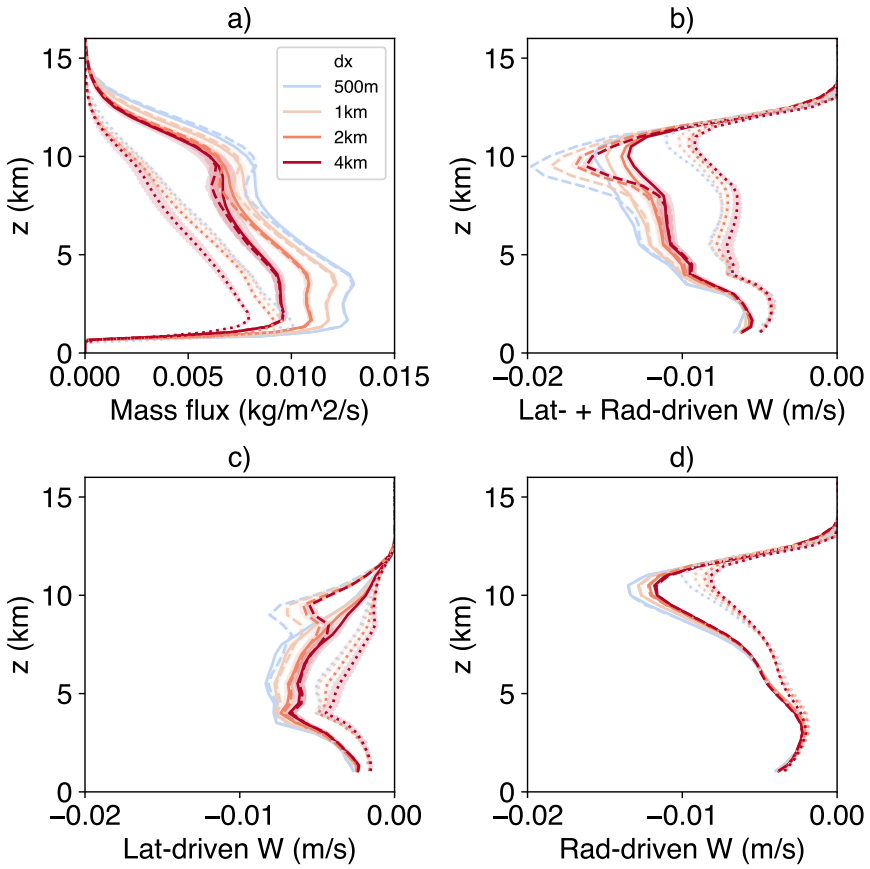


Figure 2.9: Convective updraft mass flux and energy budget for environmental subsidence in modified Morrison et al. simulations. Solid lines represent simulations where cloud-radiation interactions are deactivated. Dashed lines indicate simulations with both cloud-radiation interactions deactivated and expedited cloud ice sublimation. Dotted lines represent default Morrison et al. simulations. Panel a displays the convective updraft mass flux. Panels b to d present the subsidence contributions from latent heat (panel c), radiative cooling (panel d), and a combination of both (panel b). Shaded uncertainty range is estimated as the standard deviation of time-mean profiles of each non-overlapping 2.5-day time block.

2.6 Conclusions and Discussion

In this work, we study how mixing between convective updraft plumes and the environment can influence the macro-scale equilibrium atmospheric state, especially upper troposphere convective mass flux and anvil clouds. We first use an analytical simple model called the zero-buoyancy plume (ZBP) model to illustrate qualitatively how small-scale processes like plume-environment mixing and evaporation rate can influence the large-scale vertical temperature and moisture structure and convective mass flux. With the ZBP model, we find that increasing mixing strength can lead to an increase in upper-troposphere mass flux through two pathways based on clear-sky energy balance, either through more cloud evaporation cooling in the environment or through weaker atmospheric stability. We further tested these pathways in numerical simulations of radiative-convective equilibrium (RCE) by increasing the horizontal resolution from 4km to 125m with a fixed domain size. We found both pathways in the RCE simulations with two different microphysics schemes, a single-moment scheme and a double-moment Morrison et al. scheme. Which pathway dominates may depend on the relative magnitude of environmental cloud evaporation cooling versus radiative cooling in the upper troposphere. When the evaporation cooling is much weaker than the radiative cooling (as in the Morrison et al. simulations or with small evaporation parameter in the ZBP model), the stability pathway dominates, which is a new finding compared to a previous study (Jeevanjee & Zhou, 2022). It should be noted that the ZBP model serves to highlight potential pathways and mechanisms rather than directly explaining the simulation results. The different dominant pathways observed in the simulations using the two different microphysics schemes should not be simply attributed to the evaporation parameter in the ZBP model. Instead, the ZBP model results are meant to illustrate the potential mechanisms by which entrainment can influence upper-level clear-sky convergence and net detrainment. Given the complexity of the two microphysics schemes, attributing their differences to a single evaporation parameter or a single microphysical process is

challenging.

One innovation of this study is the refinement of the ZBP model to get self-consistent mass flux solutions. As described in Section 2.3, previous studies on the ZBP model such as Singh & Neogi (2022) and Romps (2021) have an inconsistency issue of their mass flux solution and assumptions of equal entrainment rate and detrainment rate. By treating the detrainment rate as an unknown variable instead of a parameter and utilizing the radiative-convective equilibrium concept to determine the boundary condition for mass flux, we can achieve a self-consistent mass flux solution in our refined ZBP model. This ZBP can provide a simple framework to understand the equilibrium state atmosphere and the interaction between small-scale processes to large-scale atmosphere properties, as illustrated in this study and many previous studies (Singh & Neogi, 2022; Romps, 2021; Singh et al., 2019; Romps, 2014). While we have only used this refined ZBP model with zero large-scale vertical motion, one could extend this refined ZBP model to include nonzero vertical motion to account for the effect of large-scale circulation and interaction between tropical ascent region and dry subsidence region as in Singh & Neogi (2022) and Romps (2021). Another future direction would be to extend the ZBP model to include the boundary layer. Currently, the ZBP model only solves the atmospheric profiles in the free troposphere and requires boundary conditions at the cloud base. Potentially, it is possible to use boundary layer equilibrium assumptions so that the model only requires surface boundary conditions. Additionally, the microphysics included in the ZBP model are highly simplified. Incorporating the effects of sedimentation, precipitation re-evaporation, and atmospheric cloud radiative effects could allow for a more realistic representation, even for an idealized RCE atmosphere.

In this study, we use varying horizontal resolutions as a proxy for different plume-environment mixing strengths. One needs to be more careful to infer the general resolution dependence of km-scale or hm-scale convection-permitting simulations. First, the dependence of mixing strength on model resolution can depend on the specific subgrid-scale scheme and advection scheme used in a

model. In the case where parameterized sub-grid diffusion is much weaker than the numerical diffusion (as in our simulations), the choice of numerical scheme and the representation of numerical diffusion could potentially matter for large-scale simulation properties and may warrant further study. Second, our study did not consider any large-scale convective organization, which is common in more realistic large-domain RCE simulations or global storm-resolving simulations. It is unclear how small-scale mixing could change the large-scale convective organization, which can be important for the sensitivity of large-scale thermal properties (e.g., Becker et al., 2017). In addition, we only changed the horizontal resolution in this study and fixed the vertical resolution. Previous studies (e.g., Ohno & Satoh, 2018; Ohno et al., 2019; Jenney et al., 2023) have shown that the vertical resolution can also influence the mass flux and anvil cloud fraction. In the numerical model we use, some preliminary experiments of increasing vertical resolution (not shown) yielded results similar to Jenney et al. (2023) and suggest a counter-effect on anvil cloud fraction and mass flux compared to increasing horizontal resolution. How the cloud and mass flux respond to increasing both horizontal and vertical resolution may depend on how the strength of mixing changes, which could again depend on the subgrid-scale scheme and advection scheme. Besides spatial resolution change, temporal resolution could also play a role, but we have not fully quantified its impact. The effective time step is smaller with finer resolution to satisfy the Courant–Friedrichs–Lewy condition. For the one-moment microphysics scheme, a smaller time step means that the instantaneous saturation adjustment also happens faster. Lastly, this study focuses on the equilibrium state of an atmosphere. Sensitivities of simulations of transient convective systems on resolution or the strength of mixing could be very different.

In addition to the sensitivity of mass flux and anvil cloud fraction on mixing strength and resolution, we also noticed a large sensitivity of anvil cloud fraction on the microphysics scheme. Previous studies (e.g., Powell et al., 2012; Hu et al., 2021) have found that the Morrison et al. microphysics scheme tends to produce excessive anvil cloud fraction. It is worth further future study to better un-

derstand the sensitivity of anvil cloud amount on microphysics schemes and different microphysical processes.

3

Influence of Upper-Troposphere Stratification and Cloud–Radiation Interaction on Convective Overshoots in the Tropical Tropopause Layer

3.1 Abstract

It is still debated whether radiative heating observed in the tropical tropopause layer (TTL) is balanced primarily by cooling from convective overshoots, as in an entrainment layer, or by adiabatic cooling from large-scale eddy-driven upwelling. In this study, three-dimensional cloud-resolving model simulations of radiative–convective equilibrium were carried out with three different cloud microphysics schemes and 1-km horizontal resolution. We demonstrate that overshooting cooling in the TTL can be strongly modulated by upper-troposphere stratification. Two of the schemes produce a hard-landing scenario in which convective overshoots reach the TTL with frequent large vertical velocity leading to strong overshooting cooling (~0.2 K day^{-1}). The third scheme produces a soft-landing scenario in which convective overshoots rarely reach the TTL with large vertical velocity and produce little overshooting cooling (~0.03 K day^{-1}). The difference between the two scenarios is attributed to changes in the upper-troposphere stratification related to different atmo-

spheric cloud radiative effects (ACRE). The microphysics scheme that produces the soft-landing scenario has much stronger ACRE in the upper troposphere leading to a ~3-K-warmer and more stable layer that acts as a buffer zone to slow down the convective updrafts. The stratification mechanism suggests the possibility for the ozone variation or eddy-driven upwelling in the TTL to modulate convective overshoots. We further test the sensitivity of overshooting cooling to changes in model resolution by increasing the horizontal resolution to 100 m. The corresponding change of overshooting cooling is much smaller compared with the difference between the hard-landing and soft-landing scenarios.

3.2 Introduction

The tropical tropopause layer (TTL), which is located approximately between 14 and 18.5 km, is a gradual transition layer between tropical troposphere and stratosphere (e.g., Fueglistaler et al., 2009; Randel & Jensen, 2013), and the air enters the stratosphere in the tropics through the TTL (e.g., Brewer, 1949; Holton et al., 1995). The associated dehydration processes in the TTL (e.g., freeze-drying as air passes the cold-point tropopause) are of critical importance on the global stratospheric water vapor content (e.g., Mote, 1996; Holton et al., 1995; Randel & Park, 2019). Changes in stratospheric water vapor can substantially influence ozone chemistry, stratospheric radiation, and even tropospheric climate (e.g., Anderson, 2017; Dessler et al., 2013). Atmospheric properties such as lapse rate, heat balance, and circulation pattern gradually transit through the TTL from their tropospheric to stratospheric characteristics. From the perspective of the heat budget, the dominant process in the troposphere is convection that balances tropospheric negative radiative heating predominantly through the latent heat release. In the stratosphere, the adiabatic cooling associated with the ascent of the eddy-driven Brewer–Dobson circulation (e.g., Butchart, 2014) balances the radiative heating. In the TTL, both deep convection and eddy-driven circulation could contribute

to balancing the observed radiative heating (e.g., Fueglistaler et al., 2009). The relative importance of deep convection and eddy driven circulation in the heat balance is still unclear in the TTL (e.g., Fueglistaler et al., 2009; Lin et al., 2017).

Two main processes related to deep convection have been proposed to balance the radiative heating in the TTL. The first process is the diabatic, irreversible mixing between convective overshoots and the environment. Since the convective overshoots above their levels of neutral buoyancy are negatively buoyant, the irreversible mixing of colder overshooting air with the warmer environment causes a downward buoyancy flux in the TTL (Kuang & Bretherton, 2004; Robinson & Sherwood, 2006; Hartmann et al., 2019), as in an entrainment layer (Deardorff et al., 1980), capping the turbulently mixed troposphere below. At high altitudes in the upper limit of deep convection, the buoyancy flux fades out and the resulting divergence leads to cooling in the TTL. The second process is the adiabatic ascent in response to the convective heating in the upper troposphere. Sherwood et al. (2003) found that adiabatic ascent contributes partly to the TTL cooling above convective systems, which is further supported by numerical simulations in Robinson & Sherwood (2006). Holloway and Neelin (2007) showed that adiabatic ascent could be a hydrostatic adjustment of the large-scale circulation in response to convective heating and high pressure at the upper troposphere.

On the other hand, large-scale eddy-driven upwelling also substantially influences the TTL thermal structure. The annual cycle of tropopause temperature is mainly related to the annual cycle of the adiabatic upwelling (e.g., Jucker & Gerber, 2017; Yulaeva et al., 1994), while the radiative effect associated with the change of ozone and water vapor also plays a role (Ming et al., 2017). The residual-mean upwelling velocity in the TTL estimated from different reanalysis data (Kobayashi & Iwasaki, 2016; Yoshida et al., 2018) is on the same order as the upwelling velocity needed to balance the observed TTL radiative heating (Yang et al., 2008). However, certain discrepancies in upwelling velocity exists between different reanalysis data. It is an ongoing debate whether the observed negative radiative heating rate in the TTL is balanced primarily by cooling from convective overshoots

or by adiabatic cooling from hydrostatic adjustment to tropospheric heating and/or large-scale eddy-driven upwelling.

Near-global cloud-resolving simulations (e.g., Bretherton & Khairoutdinov, 2015; Stevens et al., 2019) that resolve both deep convection and eddy-driven circulation may provide compelling results of the TTL heat budget. However, this type of simulation has just become possible in recent years due to increasing computational power. To our knowledge, such global high-resolution simulations have not yet been designed to study the TTL, which generally needs high vertical resolution near the TTL to prevent artificial wave breaking (e.g., Kuang & Bretherton, 2004; Kubokawa et al., 2012). Before running these computationally expensive runs, it is important to understand whether the processes in the TTL could be sensitive to certain experimental settings. Some previous studies involving 3D small-domain cloud-resolving simulations (e.g., Kuang & Bretherton, 2004; Robinson & Sherwood, 2006) found significant cooling related to convective overshoots in the TTL, while Küpper et al. (2004) did not. Hereafter, we will simply use the term "convective cooling" to describe the cooling related to convective overshoots and irreversible mixing. It is proper in RCE simulations because domain-mean upwelling, and therefore the hydrostatic balance mechanism, is inhibited. Fueglistaler et al. (2009) argued that the convective cooling in the TTL could depend on the prescribed boundary conditions. In a 1D radiative–convective equilibrium (RCE) model with parameterized temperature tendency due to convection, Dacie (2019) showed that the convective cooling could become weaker when the prescribed surface temperature is increased.

In this study, we ran small-domain RCE simulations in a 3D cloud-resolving model to examine what experimental settings could potentially influence the strength of convective cooling in the TTL. We found that cloud microphysics (i.e., the choice of microphysical scheme) can strongly influence the strength of convective overshoots through modifying the upper-troposphere stratification. Two microphysics schemes we used produce a hard-landing scenario in which convective overshoots reach the TTL with high velocity and relatively high frequency and lead to strong

overshooting cooling (~0.2 K day^{-1}) in the TTL. On the other hand, the third scheme produces a soft-landing scenario in which convective overshoots rarely reach the TTL with large velocity and produce little overshooting cooling (~0.03 K day^{-1}) in the TTL. The microphysical scheme in the soft-landing scenario produces much stronger atmospheric cloud radiative effects (ACRE) leading to a ~3-K-warmer and more stable upper troposphere, which acts as a buffer zone to weaken convective updrafts. We will also demonstrate the implication of the microphysics dependence to the question of which processes dominate the heat budget in the TTL, although our experiments cannot directly address this question.

The rest of this paper is organized as follows: section 2 describes the experiment design. Section 3 shows the hard-landing and soft-landing scenarios of convective overshoots produced by the different cloud-microphysics schemes. Section 4 links the different strengths of convective overshoots to the difference in upper-troposphere stratification and ACRE. In section 5, we explore why the three microphysics schemes produce different ACRE. In section 6, we discuss our results. In section 6a, we discuss the influence of increasing model resolution from 1 km to 100 m on convective cooling. In section 6b, we discuss the heat balance in the TTL in hard-landing and soft-landing scenarios when a large-scale upwelling velocity is imposed to mimic eddy-driven upwelling in the TTL. In section 6c, we explore the possibility for ozone variation to modulate convective overshoots through modifying the upper-troposphere stratification. In section 7, we summarize the results.

3.3 Experiment setup

We use the System for Atmosphere Modeling (SAM, Khairoutdinov & Randall, 2003), version 6.10.6, configured as a cloud-resolving model. We run three-dimensional RCE simulations using the same domain size of 64 × 64 grid points and 1-km horizontal grid spacing with periodic lateral boundary condition. All simulations use 170 vertical levels with model top located at 40 km and a

rigid-lid top boundary condition. The vertical grid spacing increases from 50 m near the surface to 500 m at the midtroposphere and then decreases to reach 100 m near the TTL between 14.5 and 19.5 km. The small vertical grid spacing near the TTL is crucial to reducing numerical dissipation caused by artificial gravity wave breaking (e.g., Kuang & Bretherton, 2004). A sponge layer is located in the upper 30% of the model domain (i.e., above 28 km). The radiation scheme is Rapid and Accurate Radiative Transfer Model for General Circulation Models (RRTMG) (Iacono et al., 2008). A simple Smagorinsky-type scheme (Khairoutdinov & Randall, 2003) is used for the effect of subgrid-scale motion. We use a constant solar insolation (no diurnal cycle) with fixed solar constant of 683.5 W m^{-2} and zenith angle of 50.5°. Domain-averaged horizontal wind is nudged to zero at each vertical level with a nudging time scale of 1 h. Sea surface temperature (SST) is fixed uniformly at 30.5°C. The ozone concentration profile is specified and cannot interact with convection (shown and discussed later in section 6c).

A set of three simulations was carried out with three different cloud-microphysics schemes: SAM single-moment scheme (SAM1MOM; Khairoutdinov & Randall, 2003), a double-moment scheme (Morrison et al., 2005), and a hybrid one- and double-moment hydrometeor species (double-moment only for cloud ice and rain) scheme (Thompson et al., 2008). This set of three simulations is hereafter called the default set (DEFAULT), and the three simulations are called herein SAM1MOM_DFT, Morrison_DFT, and Thompson_DFT. Table 3.1 summarizes all the experiment sets that will be discussed later and illustrates the abbreviation of individual simulations and their different treatment of large-scale forcing and radiation.

Significant differences exist between the three microphysics schemes. For example, the SAM one-moment scheme uses an instantaneous supersaturation adjustment to generate cloud condensate. Between 0° and −20°C, partitioning of cloud condensate into cloud ice and liquid water depends linearly on temperature (at −20°C, all condensate is ice; at 0°C, all condensate is liquid water). More pathways for conversion between different hydrometeors are included in the Morrison and Thomp-

Experiment set	Simulation abbreviation	Large-scale forcing	Radiation
DEFAULT	SAM1MOM_DFT Thompson_DFT Morrison_DFT	No forcing	Default RRTMG
Switch-profile	SAM1MOM_MORprl Morrison_SAMprl	Add tuned tendency of temperature and water vapor	Default RRTMG
NOCRE	SAM1MOM_NOCRE Thompson_NOCRE Morrison_NOCRE Morrison_NOCRESW Morrison_NOCRELW	No forcing	No cloud radiative effect No cloud radiative effect No cloud radiative effect No SW cloud radiative effect No LW cloud radiative effect
100-m resolution	SAM1MOM_100m Morrison_100m	No forcing	Default RRTMG
BDC	SAM1MOM_BDC Thompson_BDC Morrison_BDC	W to mimic Brewer–Dobson circulation	Default RRTMG
OZONE	SAM1MOM_Feb SAM1MOM_Aug SAM1MOM_Aug+2	No forcing	ERA5 February ozone profile ERA5 August ozone profile ERA5 August ozone profile +2 mg kg^{-1} between 10 and 14 km

Table 3.1: Summary of experiment setup for different sets of simulations.

son schemes (e.g., see Table 2 in Bao et al., 2019). The cloud ice tendency due to autoconversion from cloud ice to snow is more than 30 times faster in the Thompson scheme than in the Morrison scheme in the upper troposphere (e.g., see Figure 3.8 in Bao et al., 2019). This fast autoconversion from cloud ice to snow can make anvil clouds dissipate faster in the Thompson scheme than in the Morrison scheme (e.g., Hartmann et al., 2018; Feng et al., 2018). Powell et al. (2012) argued that the Morrison scheme tends to overproduce ice cloud due to a large cap on ice number concentration (10 cm^{-3}, two orders of magnitude larger than that in the Thompson scheme, 0.25 cm^{-3})

and consequentially a slower sedimentation rate. Although the cap on ice number concentration in the Morrison scheme has been lowered (by model developers) to 0.3 cm^{-3} as the default value in the version of SAM model that we use, as we will show later, the Morrison scheme still produces much more cloud ice in the upper troposphere. We still do not have a complete understanding of why different microphysics schemes produce different cloud properties, for example, different cloud fraction and mixing ratio of different hydrometers. However, we take this uncertainty as a possible range for the realistic climate on Earth and explore how the convective overshoots in the TTL could behave differently within this range.

For all the simulations listed in Table 3.1, the first 100 days are abandoned for the model spinup and considered long enough to reach equilibrium. This 100-day period is the typical time determined by eye when the time series of domain-averaged temperature at any level below 16 km get flattened and do not exhibit a clear trend. The temperature at ~18 km still has a small trend that has little effect on the convective overshoots below. After the 100-day spinup, a 50-day postequilibrium period is used for most simulations. The domain-mean statistics are sampled every 2 min. The 30-samples-per-hour data are then averaged to get an hourly output of domain-mean statistics. For the analysis period, the linear regression trend of temperature at any level below 16 km is on the order of 0.001 K day^{-1} across all the simulations and is below 0.02 K day^{-1} at 18 km. The DEFAULT simulations were run for 270 days to examine whether the temperature trend in the upper TTL can influence the results. Analyzing days 100–150 and 220–270 will not change the results in the following section. For the 100-m-resolution simulations in Table 3.1, only a 20-day postequilibrium period is used due to computational limitations. When we examine other simulations with a 50-day postequilibrium period, averaging over 20 days gets sufficiently similar and stable domain-mean statistics compared to that obtained from averaging over the full 50 days.

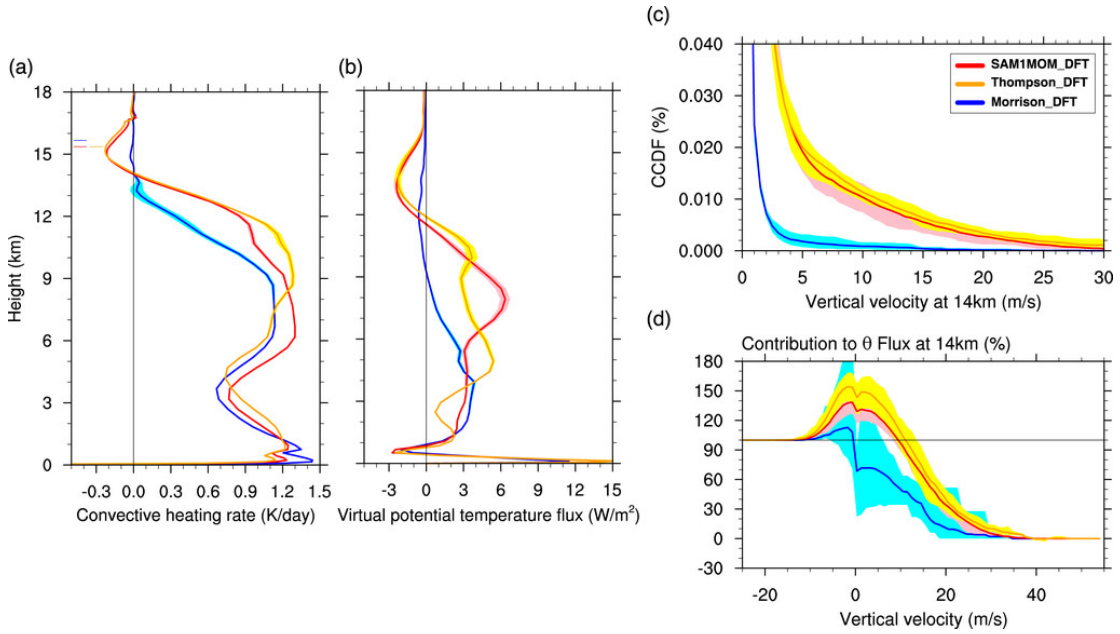


Figure 3.1: Differences of convective overshoots in the DEFAULT set of simulations: SAM1MOM_DFT (red), Thompson_DFT (orange), and Morrison_DFT (blue). (a) Domain-averaged profiles of convective heating rate, calculated as the negative of the radiative heating rate. The short horizontal lines indicate the cold-point tropopause. (b) Domain-averaged profiles of virtual potential temperature flux ($\rho C_p \overline{w'\theta_v'}$). (c) The complementary cumulative distribution function (CCDF, similar to CDF but integrating from the largest sample value) of vertical velocity at 14 km. (d) Contribution of different vertical velocity to the virtual potential temperature flux at 14 km, defined as $\text{Contr}(W_0) = \sum_{w'>W_0} w'\theta_v' / \sum_{\text{all } w'} w'\theta_v'$. The shadings in panel (a) and panel (b) indicate the 95% confidence interval. The shadings in panel (c) and panel (d) indicate the range based on the five 10-day periods within the total 50 days.

3.4 The dependence of convective overshoots on cloud microphysics

To compare the convective influence on TTL heat budget, we show the domain-averaged profiles of convective heating rate in Figure 3.1a. The convective heating is calculated by reversing the sign of the radiative heating rate. As shown in Figure 3.1a, both the SAM1MOM_DFT and Thompson_DFT simulations show a convective cooling between 14 and 17 km with a minimum convective heating rate of ~−0.2 K day^{-1} at around 15 km. On the other hand, the Morrison_DFT simulation shows a weaker convective cooling in the TTL, with a minimum cooling rate of ~−0.03 K day^{-1} at around 15 km and near zero above 15.5 km.

In our way of defining the convective heating rate by reversing radiative heating rate, we assume that the temperature has reached equilibrium and all other nonconvective processes, such as subgrid-scale diffusion or gravity wave processes, are sufficiently small. As we mentioned in the previous section, the maximum temperature drift below 18 km across all the simulations listed in Table3.1 is between 0.001 and 0.02 K day^{-1}. The maximum time-mean temperature tendency due to subgrid-scale diffusion in the TTL across all the simulations is between 0.001 and 0.015 K day^{-1}. In the TTL, there is no direct temperature tendency related to the sponge layer, which is above 28 km. The temperature tendency associated with gravity wave processes, such as wave breaking and physical or numerical damping, is difficult to quantify accurately. However, we will show later in Figure 3.1d that, at least for the SAM1MOM_DFT and Thompson_DFT simulations that have strong convective cooling, the resolved advective temperature tendency in the TTL is dominated by strong convective overshoots rather than gravity waves in the environment.

Consistent with convective cooling, the domain-averaged profile of virtual potential temperature flux shows a clear difference in the Morrison_DFT simulation compared to the other two DEFAULT simulations (Figure 3.1b). Here, the virtual potential temperature flux ($\overline{w'\theta_v'}$) is expressed in the form of energy flux ($\rho C_p \overline{w'\theta_v'}$ with units W m^{-2}), where ρ is the air density, C_p is the specific heat capacity at constant pressure, w' is the perturbation in vertical velocity, and θ_v' is the perturbation in virtual potential temperature. The virtual potential temperature, $\theta_v' = (1 + 0.61r - r_c)$, includes the effect of both water vapor mixing ratio r and the mixing ratio of total (precipitating plus nonprecipitating) cloud condensate r_c. Near the TTL, because the air is quite dry (both r and r_c are on the order of 10^{-5} kg kg^{-1}), the virtual potential temperature flux profile shape is comparable to those of buoyancy flux and turbulent heat flux. For the SAM1MOM_DFT and Thompson_DFT simulations, the domain-mean level of neutral buoyancy (LNB, defined here as the level where the virtual potential temperature flux reaches zero) is near 12 km. The downward virtual potential temperature flux above the LNB is due to the temperature difference between colder convective overshoots and

the warmer environment and the irreversible mixing between them. The virtual potential temperature flux reaches its minimum value at near 14 km, and the divergence of heat flux above leads to the convective cooling in Figure 3.1a. By comparison, the Morrison_DFT simulation generates lower LNB, which indicates that most of the convective updrafts lose their positive buoyancy at a much lower altitude of ~9 km. Furthermore, the corresponding minimum value of the negative virtual temperature flux above the LNB is very small, leading to a weaker heat flux divergence in the TTL.

The differences in the convective cooling in the TTL and the virtual potential temperature flux are associated with weaker convective overshoots (i.e., rare occurrence of strong updrafts at a given height) in the Morrison_DFT simulation. The complementary cumulative distribution function (CCDF) of the vertical velocity [i.e., $CCDF(w_0) = P(w > w_0)$, the probability of w greater than w_0] at 14 km is illustrated in Figure 3.1c. To calculate the CCDF, we output the instantaneous hourly 3D vertical velocity field during the 50-day time period. At the level of ~14 km, the vertical velocity is sampled at each grid point. As shown in Figure 3.1c, the Morrison_DFT simulation indicates an occurrence of strong updraft (e.g., greater than 10 m s^{-1}) one order of magnitude smaller than that in the other two DEFAULT simulations. The CCDF of the vertical velocity at 13 or 15 km also shows much rarer strong updraft in the Morrison_DFT simulation (not shown).

To illustrate which updraft velocity contributes most to the virtual potential temperature flux, we show in Figure 3.1d the integrated contribution of different vertical velocities to the virtual potential temperature flux at 14 km. The integrated contribution is defined as $\mathrm{Contr}(W_0) = \sum_{w'>W_0} w'\theta_v' / \sum_{\text{all } w'} w'\theta_v'$. For the SAM1MOM_DFT and Thompson_DFT simulations, updraft velocities greater than ~10 m s^{-1} contribute almost 100% to the total virtual potential temperature flux, while the contribution from updraft velocities smaller than 10 m s^{-1} is roughly canceled out by downdrafts. Almost all updrafts that exceed 10 m s^{-1} are related to convective overshoots originated from the boundary layer rather than gravity wave processes in the environment (which is supported by tracer analysis, not shown). In the Morrison_DFT simulation, updrafts greater than

10 m s^{-1} only explain part of the total virtual potential temperature flux, while downdrafts smaller than about 1 m s^{-1} also contributes a lot. For the contribution from small velocities, it is hard to cleanly separate convective overshoots and gravity wave processes. Therefore, convective overshoots dominate the virtual potential temperature flux at 14 km in the SAM1MOM_DFT and Thompson_DFT simulations but could have a comparable contribution to gravity wave processes in the Morrison_DFT simulation.

The difference between the convective overshoots in the SAM1MOM_DFT or Thompson_DFT simulations and those in the Morrison_DFT simulation may be viewed in terms of hard landing versus soft landing. In the hard-landing scenario (for the SAM1MOM_DFT and Thompson_DFT simulations), convective overshoots reach the TTL with frequent strong updrafts, leading to strong downward heat flux and a significant cooling in the TTL. In the soft-landing scenario (for the Morrison_DFT simulation), convective updrafts start to decelerate at lower heights and convective overshoots rarely reach the TTL with high vertical velocity, producing weak downward heat flux and little cooling in the TTL. In the next section, we explore what processes contribute to the difference between the two scenarios.

3.5 Influence of upper-troposphere stratification and atmospheric cloud radiative effects

Why does the Morrison_DFT simulation have much weaker convective overshoots? As shown in Figure 3.2, the Morrison_DFT simulation has a much warmer (~3 K) upper troposphere while the temperature in the lower troposphere is similar to that in the other two simulations. A natural hypothesis would be that the warmer and more stable upper troposphere in the Morrison_DFT simulation lowers the LNB and reduces the buoyancy, leading to much smaller updraft velocity associated with convective overshoots near the TTL.

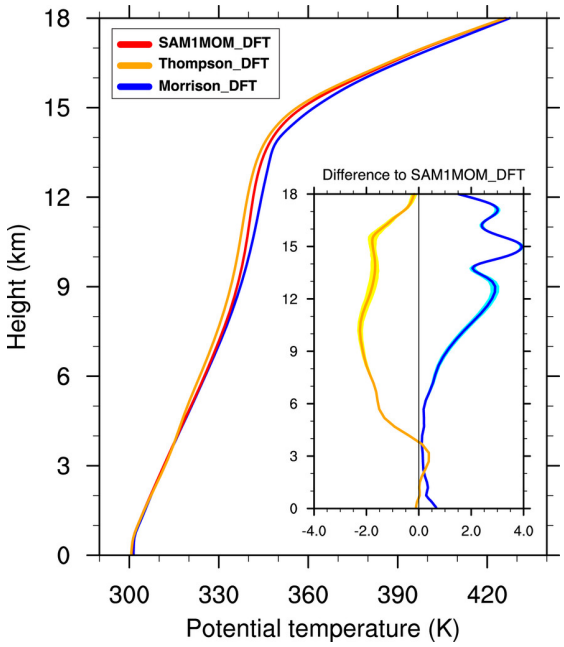


Figure 3.2: Domain-averaged profiles of potential temperature in the DEFAULT set of simulations: SAM1MOM_DFT (red), Thompson_DFT (orange), and Morrison_DFT (blue). The deviation of potential temperature profile from SAM1MOM_DFT in Thompson_DFT and Morrison_DFT is shown in the embedded figure. For all the lines shown, the shading indicates the 95% confidence interval. Some portions of the uncertainty range are too small to see the shading.

To test this hypothesis, we performed two additional switch-profile simulations (Table 3.1). In the first simulation, by adding tuned and time-invariant forcing of temperature and moisture tendencies, we force the model using the Morrison scheme to maintain a temperature and moisture profile similar to the SAM1MOM_DFT simulation ("Morrison_SAMprl" herein). Similarly, in the second simulation, we force the model using the SAM one-moment scheme to maintain a temperature and moisture profile similar to the Morrison_DFT simulation ("SAM1MOM_MORprl" herein). The time-invariant temperature and moisture tendencies are obtained by iteratively running simulations in which we both added time-invariant tendencies (which are zero in the first iteration) and nudged the domain-averaged temperature and moisture profiles to the target profiles. The nudging time scale is 10 h. In each iteration, we ran 50 days, saved the nudging temperature and moisture tendencies (averaged over the last 20 days), and added them into the time-invariant tendencies. After four iterations, we obtained the final time-invariant tendencies and ran the two switch-profile simulations without nudging temperature and moisture. The potential temperature profiles in the switch-profile simulations shown in Figure 3.3a are very close to the target profiles derived from the DEFAULT simulations.

These two new simulations clearly support that perturbation of upper-troposphere temperature profile can impact the virtual potential temperature flux and the LNB (Figure 3.3b). With a colder upper troposphere than the Morrison_DFT simulation, the Morrison_SAMprl simulation now has an LNB that is slightly higher than that in the SAM1MOM_DFT simulation. The inferred convective cooling in the TTL in the Morrison_SAMprl simulation is slightly stronger but on the same order of magnitude as in the SAM1MOM_DFT simulation. The SAM1MOM_MORprl simulation indicates a much lower LNB that is closer to that in the Morrison_DFT simulation. There is no indication of downward virtual potential temperature flux above LNB, suggesting strong suppression of deep convection and convective overshoots in the SAM1MOM_MORprl simulation. Similar switch-profile simulations are performed between the SAM1MOM and Thompson schemes and

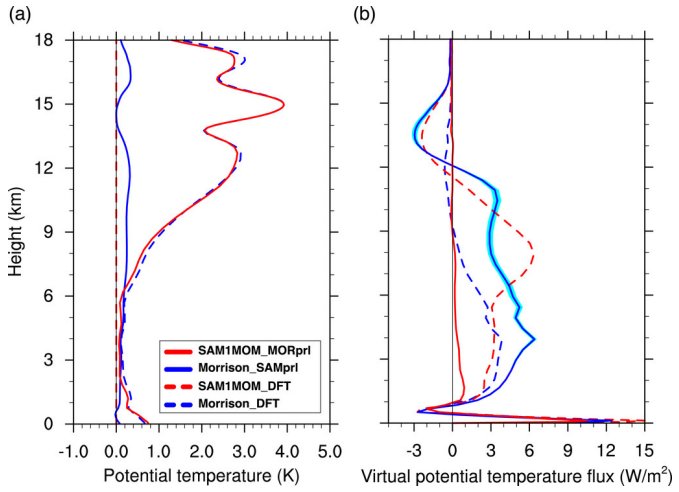


Figure 3.3: Domain-mean profiles of potential temperature, as the deviation from that in SAM1MOM_DFT, in the switch-profile simulations (solid) and in the default set of simulations (dashed): SAM1MOM_MORprl (solid red), Morrison_SAMprl (solid blue), SAM1MOM_DFT (dashed red), and Morrison_DFT (dashed blue). (b) As in (a), but for the virtual potential temperature flux ($\rho C_p \overline{w' \theta_v'}$). For switch-profile simulations (solid lines), the shading indicates the 95% confidence interval. Some portions of the uncertainty range are too small to see the shading.

obtain similar results. The SAM1MOM_THOMprl simulation has strongly enhanced convective overshoots, while the Thompson_SAMprl simulation has strongly suppressed deep convection and convective overshoots (not shown, abbreviations follow the same rule as SAM1MOM_MORprl and Morrison_SAMprl). The switch-profile simulations suggest that the perturbation of the upper-troposphere stratification can strongly modulate the convective overshoots and LNB.

The above switch-profile results should not be interpreted as that environmental profile can uniquely determines LNB and other convective properties. The SAM1MOM_DFT and Thompson_DFT simulations have a temperature difference of about 2 K in the upper troposphere but with similar overshooting properties. The buoyancy of convective updrafts depends not only on the environmental virtual temperature but also on factors such as condensate loading in the updrafts, which decreases buoyancy and the entrainment history of the updrafts. These factors can be very different for simulations using different microphysics schemes. One should also be cautious in interpreting convective properties as a consequence of environmental profiles, because environmental

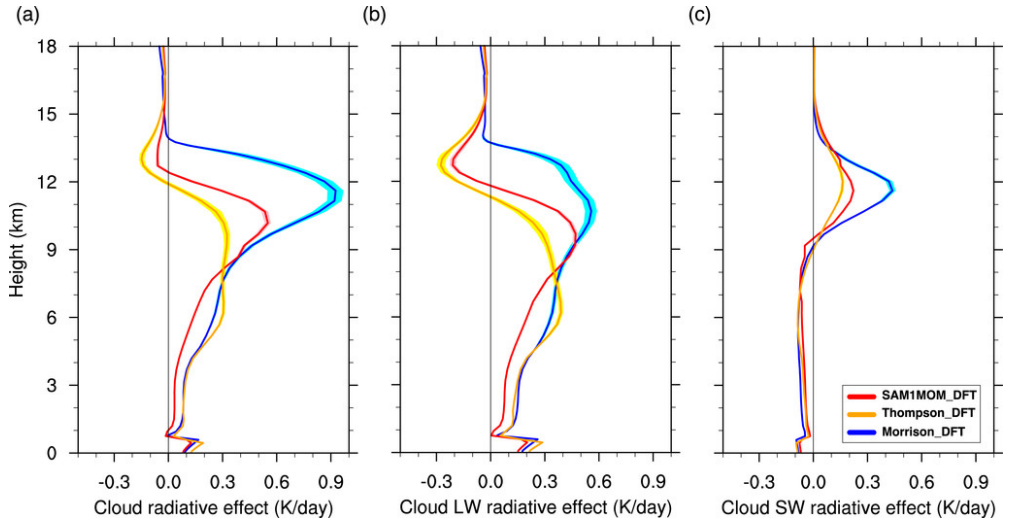


Figure 3.4: Domain-mean profiles of (a) net, (b) longwave, and (c) shortwave atmospheric cloud radiative effects in the default set of simulations: SAM1MOM_DFT (red), Thompson_DFT (orange), and Morrison_DFT (blue). The shading indicates the 95% confidence interval.

profiles and convection mutually interacts with each other and are closely coupled. The mechanism we would like to explore is that processes not directly related to convection, such as large-scale circulation or radiative effect from clouds or tracer gas, can influence the upper-troposphere temperature, which can then modify the convective overshoots.

Because the only difference in the experimental setting of the three default simulations is the cloud microphysics scheme, next we explore whether the different ACRE lead to different environmental temperature profiles. The ACRE in the Morrison_DFT simulation are indeed larger in the upper troposphere than in the other two DEFAULT simulations (Figure 3.4a). Both the longwave and shortwave cloud radiative effects contribute to this difference (Figures 3.4b,c). The SAM1MOM_DFT and Thompson_DFT simulations have a strong longwave cloud cooling at ~13 km while the Morrison_DFT simulation fails to produce such a cooling structure. The shortwave cloud radiative heating at ~12 km in the Morrison_DFT simulation is much stronger than those in the other two simulations.

To provide causal evidence that differences in ACRE lead to different upper-troposphere stratification and different convective overshooting velocities, we ran a new set of three simulations with the three microphysics schemes with the cloud radiative effects turned off. We call this set of simulations the NOCRE set (Table 3.1), in contrast with the DEFAULT set, and we refer to individual NOCRE simulations as SAM1MOM_NOCRE, Thompson_NOCRE, and Morrison_NOCRE. In the NOCRE simulations, the radiation model ignores the presence of any cloud condensate (both precipitating and nonprecipitating) and returns the clear-sky radiation. After turning off the cloud radiative effect, the Morrison_NOCRE simulation experiences the largest temperature decrease and becomes even colder in most of the troposphere compared to the SAM1MOM_NOCRE simulation (Figures 3.5a,b). The Morrison scheme still produces warmer TTL than the SAM1MOM scheme, which is likely related to the radiative effect of different water vapor mixing ratio in the TTL (Figure 3.5c).

The properties of the convective overshoots in the NOCRE simulations are illustrated in Figure 3.6. After turning off the cloud radiative effect, the Morrison_NOCRE simulation shows a significantly stronger convective cooling in the TTL (~-0.15 K day^{-1}), although the convective cooling in the TTL is still somewhat weaker than the SAM1MOM_NOCRE simulation and the Thompson_NOCRE simulation (Figure 3.6a). The LNB in the Morrison_NOCRE simulation converges to those in the other two simulations (Figure 3.6b). Also, the shift from the DEFAULT set to the NOCRE set in the Morrison simulation indicates a significant increase in the frequency of occurrence of strong updraft (Figure 3.6c). Therefore, the NOCRE simulations provide evidence of the causal effect between the stronger ACRE and the weaker convective overshoots in the default Morrison simulation.

One may notice that, in the SAM1MOM_NOCRE and Thompson_NOCRE simulations, the LNB is in fact slightly lower than in their DEFAULT counterpart (comparing Figure 3.1b and Figure 3.6b). The Thompson_NOCRE simulation also produces lower occurrence of strong updraft

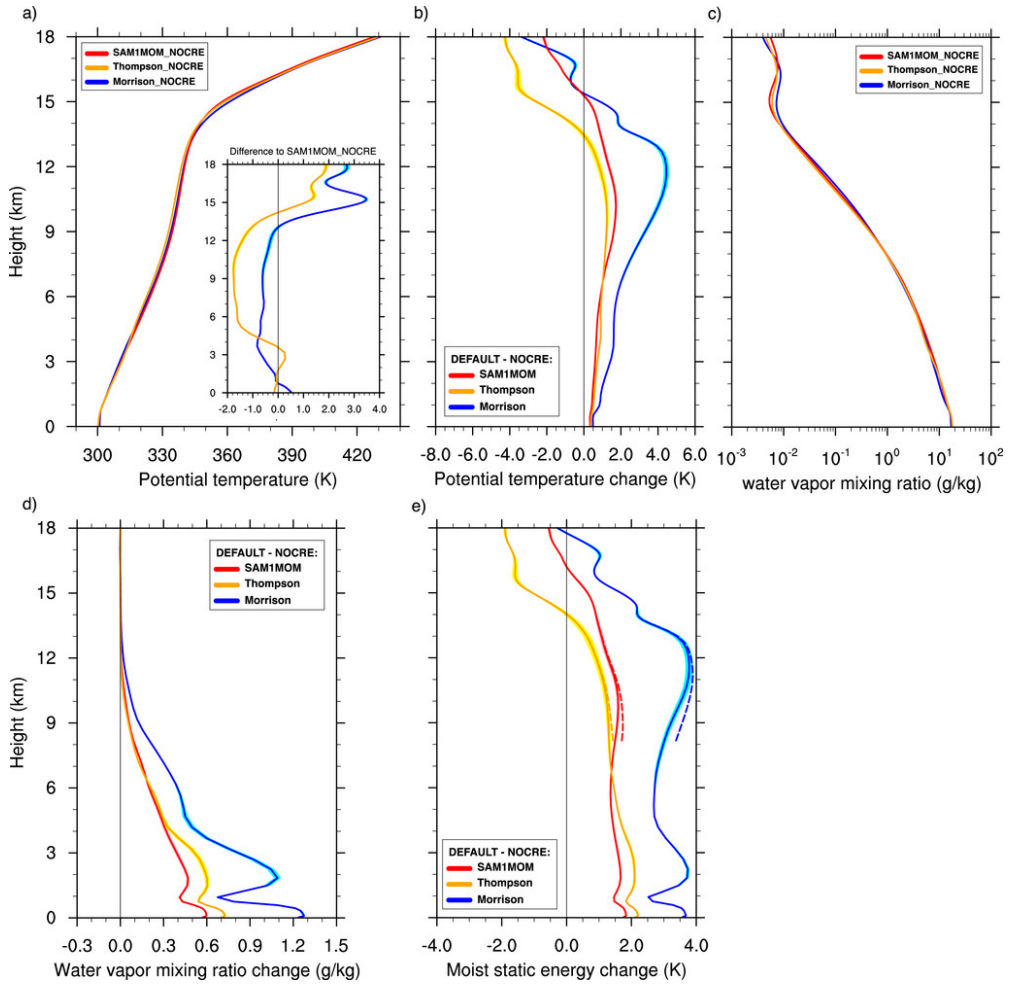


Figure 3.5: Temperature and moisture profiles in the NOCRE simulations and their change from NOCRE to DEFAULT simulations for the three microphysics schemes. (a) As in Figure 3.2, but for the NOCRE set of simulations. (b) The potential temperature changes from NOCRE to DEFAULT simulations. (c) The profiles of water vapor mixing ratio for the NOCRE simulations. (d) The change of water vapor mixing ratio from NOCRE to DEFAULT simulations. (e) The change of moist static energy (solid lines) and the change of saturation moist static energy (dashed lines, only the values in the upper troposphere between 8 and 14 km are shown) from NOCRE to DEFAULT simulations. The shading indicates the 95% confidence interval.

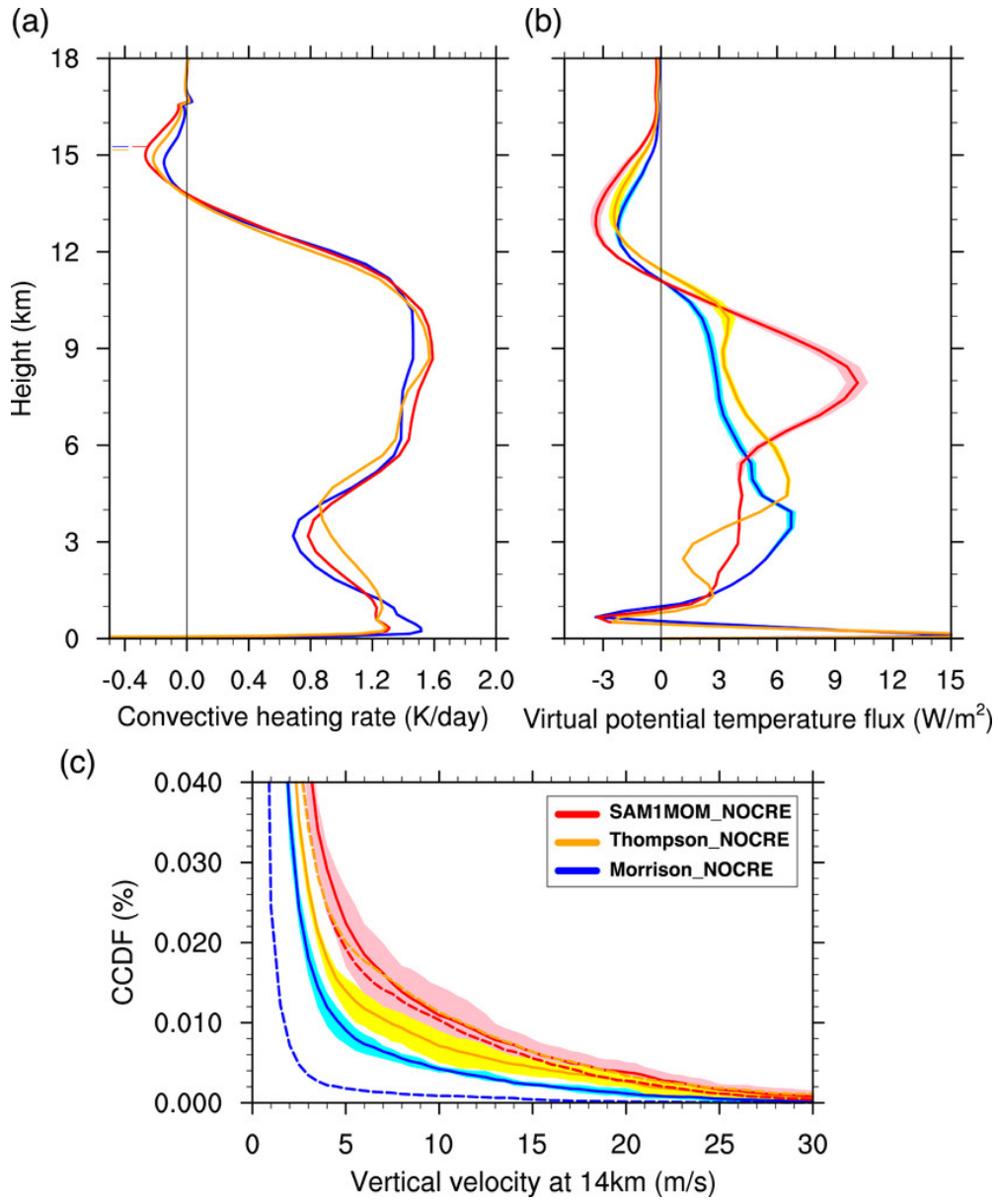


Figure 3.6: As in Figure 3.1, but the solid lines are for the NOCRE set of simulations. The lines in Figure 3.1c for the DEFAULT set of simulations are overlaid in (c) as dashed lines for reference.

than the Thompson_DFT simulation (Figure 3.6c). The different behavior to the Morrison scheme is related to the water vapor change in the lower troposphere in the NOCRE runs (Figure 3.5d). If the temperature profile remains unchanged and only the lower troposphere becomes drier, the convective updraft will become less buoyant and the LNB will become lower. The combined effect of changes in temperature and moisture can be seen in the change of the moist static energy (MSE; shown as MSE/C_p in Figure 3.5e with the unit of K), which is roughly conserved in moist adiabatic processes. To connect MSE change with the change of LNB, let us consider the buoyancy change of a saturated convective updraft plume at a fixed height, for example, the LNB in the DEFAULT run. The buoyancy is proportional to the virtual temperature of the plume minus the virtual temperature of the environment. The temperature difference at a given pressure level can be indicated by the saturation MSE difference. The saturation MSE change in the saturated updraft plume equals to the mixture of MSE change at a range of lower levels where the rising plume originates and where the unsaturated environmental air was entrained into the rising plume. The saturation MSE change in the environment in the upper troposphere (Figure 3.5e dashed lines) is very close to MSE change. For the Thompson_DFT and Thompson_NOCRE simulations (same for the SAM1MOM simulations) at 12 km, which is roughly the LNB in these runs, the saturation MSE change (DEFAULT minus NOCRE) is smaller than the MSE change below. Therefore, an updraft at 12 km tends to be less buoyant in the Thompson_NOCRE simulation than in the Thompson_DFT simulation. That is, the LNB will become lower in the Thompson_NOCRE simulation than in the Thompson_DFT simulation. On the contrary, for the Morrison simulations, the cloud radiative heating is so strong in the upper troposphere that the saturation MSE change at 9–12 km is larger than MSE change at 4–9 km. Therefore, the LNB becomes higher in the Morrison_NOCRE simulation than in the Morrison_DFT simulation.

To investigate the separate impact of longwave and shortwave cloud radiative effect, we perform two additional simulations using Morrison scheme in which we turn off only longwave or short-

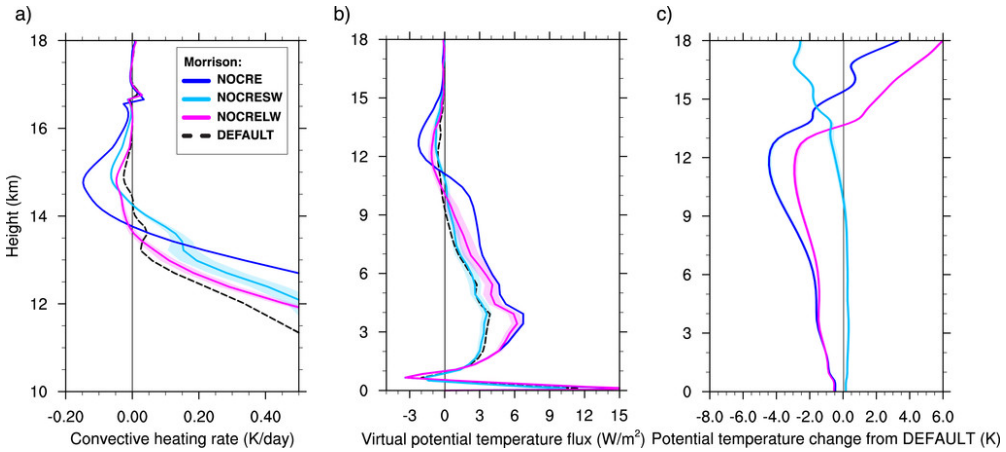


Figure 3.7: (a) The convective heating rate, (b) virtual potential temperature flux, and (c) potential temperature deviation from the Morrison_DFT simulation for the Morrison_NOCRE (dark blue), Morrison_NOCRESW (light blue), Morrison_NOCRELW (magenta), and Morrison_DFT (dashed black) simulations. The shading for the NOCRESW and NOCRELW simulations indicates the 95% confidence interval.

wave ACRE (Morrison_NOCRELW and Morrison_NOCRESW). Because both the longwave and shortwave ACRE have a positive peak in the upper troposphere (Figures 3.4b,c), turning off either longwave or shortwave ACRE weakens the stratification between 8 and 11 km (Figure 3.7c) and therefore can lead to enhanced overshooting cooling and lifted LNB (Figures 3.7a,b). The overall tropospheric longwave ACRE is stronger than the shortwave ACRE, which is mainly concentrated in the upper troposphere. Therefore, the Morrison_NOCRELW simulation has stronger upper and overall troposphere cooling than the Morrison_NOCRESW simulation and has smaller buoyancy frequency (not shown) between 6 and 10 km. However, the lower troposphere drying, which by itself can weaken convective updrafts, in the Morrison_NOCRE simulation (Figure 3.5d) almost all comes from turning off the longwave ACRE. After combining the effect of both stratification and water vapor change, the Morrison_NOCRELW and Morrison_NOCRESW simulations produce a comparable enhancement of convective cooling in the TTL.

3.6 Understanding the different ACRE associated with different microphysics schemes

To understand the ACRE difference in the upper troposphere across the DEFAULT simulations, we show in Figure 3.8a the probability distribution of vertically integrated optical depth of the sum of radiatively active and frozen hydrometeors. For all the frozen hydrometeors, only cloud ice is radiatively active in the SAM one-moment scheme, while both the cloud ice and snow are radiatively active in the Morrison and Thompson schemes. The optical depth is for the 625–778-nm band. For a longwave band near 10.5 μm, the optical depth distribution shown in Figure 3.8a slightly shifts to smaller optical depths but retains the shape (not shown). The cloud optical depth distribution can be largely explained by the distribution of vertically integrated ice (plus snow for the Morrison and Thompson simulations) water path (Figure 3.8b). For simplicity, we will use RFWP to refer to the radiatively active frozen water path for all three schemes. The Morrison_DFT simulation shows a clear distinction from the other two simulations, with much higher frequency of occurrence for clouds with optical depths in the range of 0.01–1, roughly corresponding to RFWP in the range of 10^{-4} to 10^{-2} kg m^{-2}. Figure 3.8c also shows that the Morrison_DFT simulation has a larger cloud fraction than the other two simulations in the upper troposphere.

To further relate the distribution of RFWP to the domain-mean vertical distribution of ACRE, we show in Figure 3.9 the vertical distribution of ACRE, cloud ice and snow content and effective radius sorted by different RFWP bins. That is, we sort all the columns according to their RFWP and average over the columns in the same bins. The ACRE sorted by RFWP shares some common feature across the three schemes. For RFWP greater than 10^{-1} kg m^{-2}, which corresponds to convective cores, there is strong cloud-top cooling in the upper troposphere. For smaller RFWP corresponding to anvil clouds or thin cirrus, there is generally cloud radiative heating in the upper troposphere. Therefore, the much larger occurrence of RFWP between 10^{-4} to 10^{-2} kg m^{-2} for

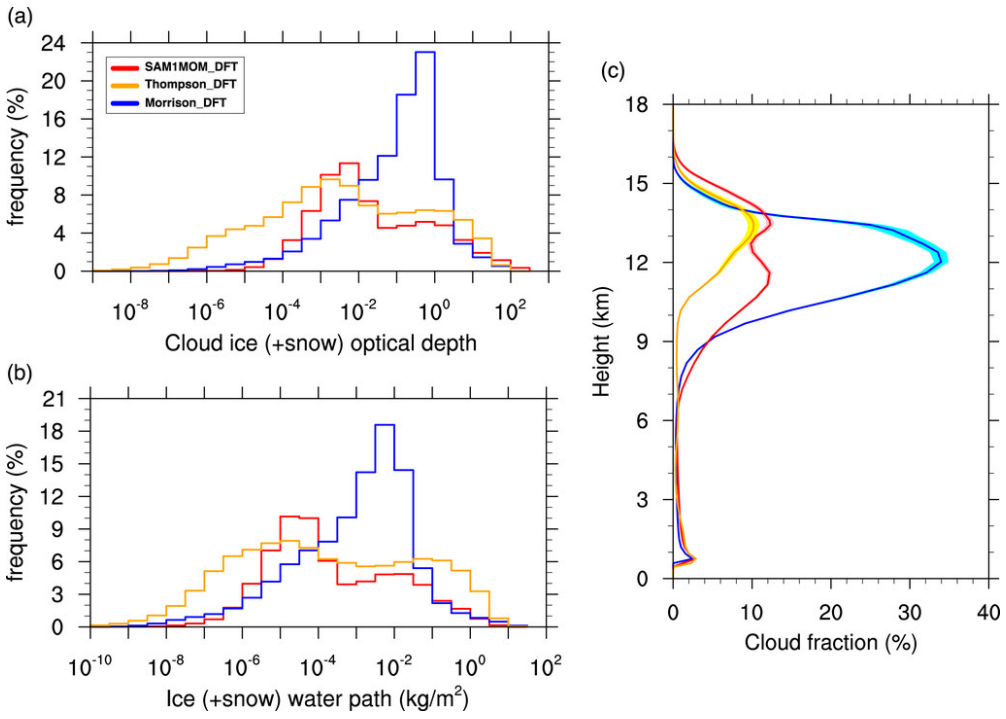


Figure 3.8: The relative frequency histograms of (a) the vertically integrated optical depth and (b) water path for the sum of radiatively active frozen hydrometeors in the default set of simulations: SAM1MOM_DFT (red), Thompson_DFT (orange), and Morrison_DFT (blue). The optical depth is for the 625–778-nm band. In the Morrison and Thompson simulations, both cloud ice and snow are radiatively active, while in the SAM1MOM simulation only cloud ice is radiatively active. The frequency is normalized by the number of total sampled grids including both clear-sky and cloudy-sky grid points. (c) The cloud fraction profiles with shading indicating 95% confidence interval.

the Morrison_DFT simulation could indeed contribute to the larger total ACRE in the upper troposphere. The Thompson_DFT simulation has a higher occurrence of RFWP above 10^{-2} kg m^{-2}, which contributes to the stronger cloud cooling above 12 km in the Thompson_DFT simulation than the other two simulations. The Thompson_DFT simulation also produces a large amount of very thin clouds with RFWP less than 10^{-6} kg m^{-2}, but these clouds are too optically thin and contributes little to the ACRE.

In addition to cloud cover, the difference in vertical distribution of cloud condensate and effective radius could also contribute to the ACRE difference. For RFWP between 10^{-4} to 10^{-2} kg m^{-2}, the ACRE in the upper troposphere in the Morrison_DFT simulation is similar to that of the SAM1MOM_DFT simulation and is much stronger than that of the Thompson_DFT simulation (note that the color bar is not linear). This is related to the difference in the vertical distribution of frozen mass and effective radius. The frozen mass is mostly cloud ice for the Morrison_DFT and SAM1MOM_DFT simulations (Figures 3.9d,e,g) but is primarily snow for the Thompson_DFT simulation (Figures 3.9f,h). The vertical distribution of cloud ice in the Morrison_DFT and SAM1MOM_DFT simulations is also located at higher levels but with a narrower range of height than the vertical distribution of snow in the Thompson_DFT simulation, which could explain the sharper vertical gradient of ACRE in the midtroposphere in the Morrison_DFT and SAM1MOM_DFT simulations (Figure 3.4a). The effective radius for the majority of snow in the Thompson_DFT simulation is greater than 150 μm, while the effective radius for the majority of the cloud ice in the other two simulations is below 50 μm (Figures 3.9i,j,k). Therefore, the same amount of radiatively active frozen mass will have smaller total cross section and thus smaller optical depth in the Thompson scheme.

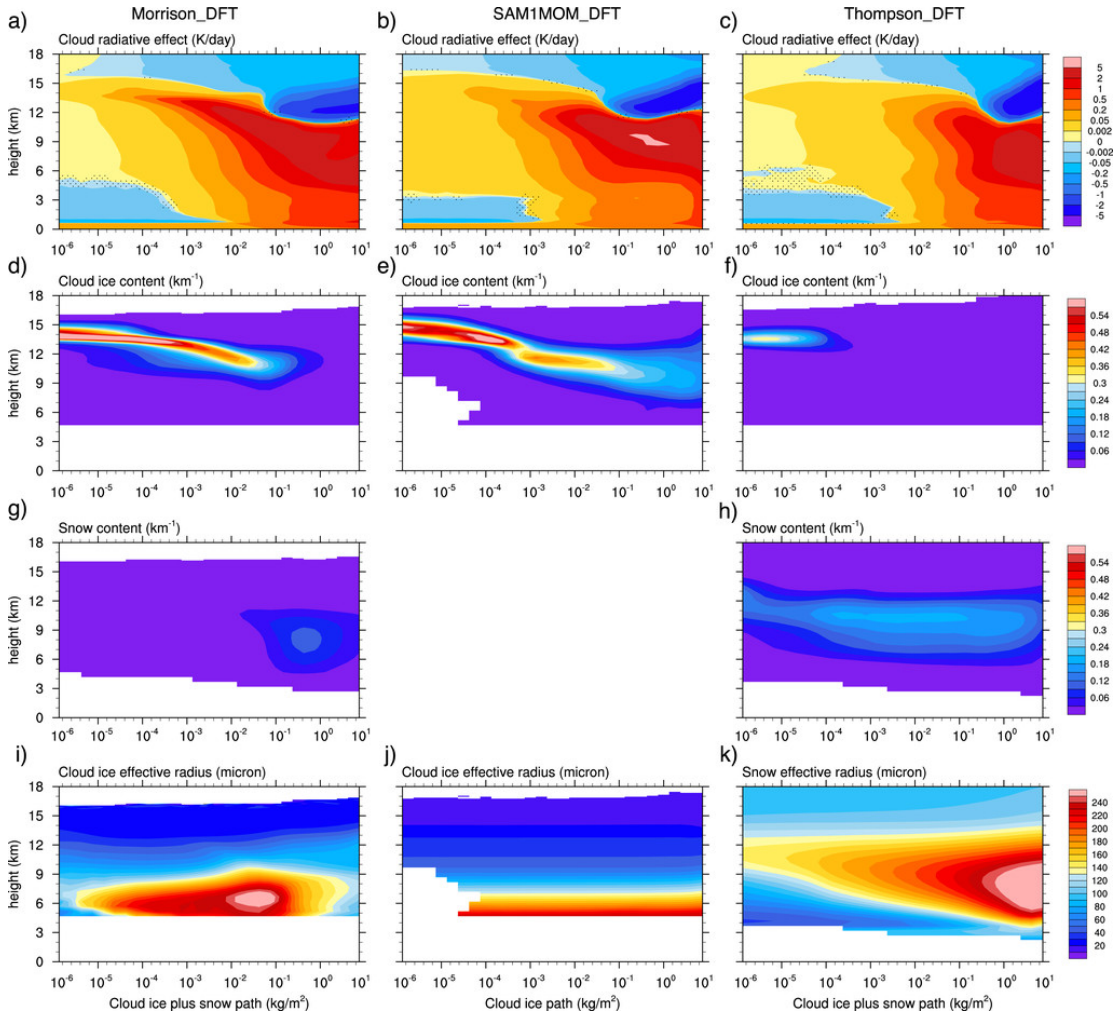


Figure 3.9: Vertical distribution of (a)–(c) cloud radiative effect, (d)–(f) cloud ice content, (g),(h) snow content, and (i)–(k) frozen hydrometer effective radius sorted by model columns with different values of frozen water path. (left) Morrison_DFT. (center) SAM1MOM_DFT. The snow content in the SAM1MOM simulations is not radiatively active so we do not show the snow content here for SAM1MOM_DFT. (right) Thompson_DFT. The cloud ice content and snow content (g m^{-3}) is normalized by the total frozen water path (g m^{-2}) is each bin, so the unit becomes km^{-1}. The normalized cloud ice/snow contents also represent the vertical probability distribution of frozen mass in each bin. In (i)–(k), we show the cloud ice effective radius for Morrison_DFT and SAM1MOM_DFT but snow effective radius for Thompson_DFT, because cloud radiative effective is dominated by cloud ice in the SAM1MOM and Morrison simulations but is dominated by snow in the Thompson simulations.

3.7 Discussion

3.7.1 Sensitivity to model resolution

The above simulations use horizontal resolution of 1 km. However, turbulence below 1-km scale may have a significant impact on processes like irreversible mixing of convective overshoots, convective entrainment, or the behavior of microphysics schemes. To better resolve these processes, we increase the horizontal resolution to 100 m for the SAM1MOM and Morrison simulations (herein SAM1MOM_100m and Morrison_100 m, Table 3.1). The domain size is reduced to 32 km × 32 km to optimize the computational cost. A 20-day period after spinup is used for the analysis. Other than the horizontal domain size and the horizontal resolution, all settings are identical to those of the DEFAULT simulations.

The dependence of convective overshoots on microphysics scheme in 100-m-resolution simulations is consistent with the DEFAULT simulations. Compared with the SAM1MOM_100m simulation, the Morrison_100m simulation still produces a much weaker overshooting cooling, weaker downward heat flux, and lower LNB (Figures 3.10a,b). Within the TTL, the difference in heating rate and heat flux between 100-m- and 1-km-resolution runs is much smaller than the difference between two microphysics schemes.

It is interesting to notice that in the 100-m-resolution simulations, the cloud fraction in the upper troposphere is significantly larger than in the 1-km-resolution simulations (Figure 3.10c). The maximum cloud fraction in the upper troposphere increases from 12% in the DEFAULT simulation to 21% in the 100-m-resolution simulation for SAM1MOM scheme, and from 34% to 49% for the Morrison scheme. The higher cloud fraction leads to stronger cloud radiative heating in the upper troposphere, which explains the weaker convective heating in the upper troposphere in Figure 3.10a. To make an apples-to-apples comparison, the 100-m-resolution output fields (e.g., 3D temperature, cloud water and cloud ice mixing ratio) are regridded to match the coarser 1-km grid, by

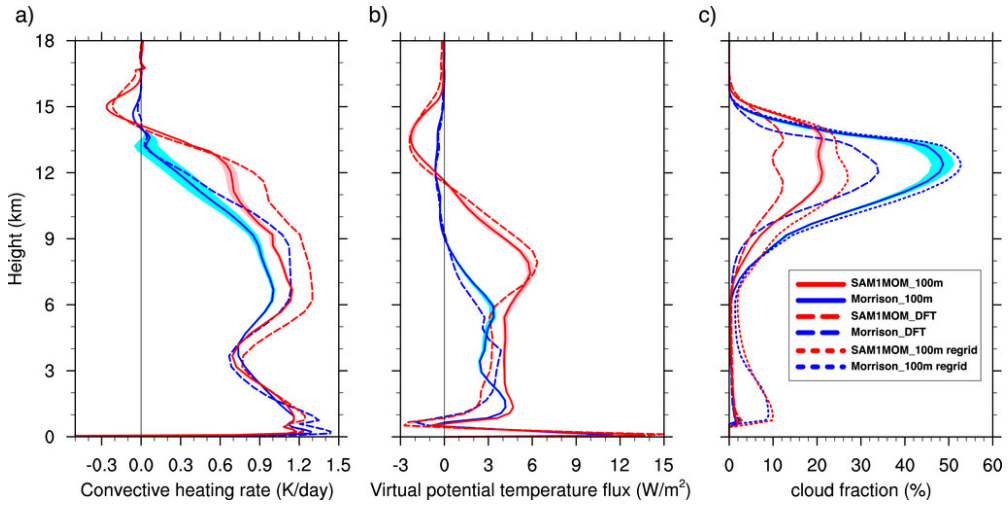


Figure 3.10: (a) The convective heating rate, (b) virtual potential temperature flux, and (c) cloud fraction for the SAM1MOM_100m (solid red lines) and Morrison_100m (solid blue lines). The SAM1MOM_DFT and Morrison_DFT are represented as long-dashed lines for reference. The short-dashed lines in (c) represent the cloud fraction after regridding the 100-m-resolution simulations from 100-m resolution to 1-km resolution.

averaging fine-grid fields in 1 km × 1 km chunks. The cloud fraction based on regridded output is even larger than the cloud fraction based on the raw 100-m-resolution output fields (Figure 3.10c). This result suggests a significant enhancement of cloud fraction by resolved processes at the 100-m scale, although further study is needed to narrow down the exact mechanisms. It is worth noting that Chapter 2 was partially motivated by this finding and explored this resolution dependence in more detail.

3.7.2 Implication for TTL heat balance and cold-point tropopause

Our results have implications for the debate, as we mentioned in the introduction, of whether the observed radiative heating in the TTL is mainly balanced by convective cooling or by adiabatic cooling due to eddy-driven circulation. We perform a set of simulations in which we impose a fixed large-scale vertical velocity mimicking the upwelling of Brewer–Dobson circulation (Figure 3.11a). This set of simulations is called the BDC set, and the individual simulations are SAM1MOM_BDC,

Thompson_BDC, and Morrison_BDC (Table 3.1). In the BDC simulations, the three microphysics schemes continue to produce the hard-landing and soft-landing scenarios. Similar to the Morrison_DFT simulation, the Morrison_BDC simulation still produces a soft-landing scenario, with a much lower LNB and weaker downward heat flux associated with convective overshoots (Figure 3.11b) than the other two schemes. The imposed upwelling makes the atmosphere significantly colder above 13 km than the DEFAULT simulations but has little impact on temperature below 13 km (Figure 3.11c). Furthermore, the imposed upwelling has a very limited impact on the strength of downward heat flux associated with overshoots (comparing Figure 3.11b with Figure 3.1b). For the heat balance in the TTL, in the soft-landing scenario, the radiative heating is dominantly balanced by the adiabatic heating from large-scale forcing representing the eddy-driven ascent, with little contribution from deep convection (Figure 3.11d). In the hard-landing scenario, convective cooling and large-scale forcing make comparable contributions to balance the radiative heating (Figures 3.11e,f).

Another relevant debate is whether the height of the cold-point tropopause (CPT) is coupled to deep convection (e.g., Thuburn & Craig, 2002; Kuang & Bretherton, 2004; Fueglistaler et al., 2009; Fu et al., 2018). In the hard-landing scenario (the DEFAULT and NOCRE simulations with the single-moment scheme and the Thompson scheme and without any large-scale forcing), the cold-point tropopause is collocated with the height of the strongest convective cooling (red and orange horizontal lines in Figures 3.1a and 3.6a), which is consistent with Kuang & Bretherton (2004). It is reasonable to imagine that a small change in the height of the strongest convective cooling could directly influence the CPT by changing the shape of the temperature profile. With the large-scale upwelling, the cold-point tropopause is shifted higher and somewhat decoupled from the height of the strongest convective cooling (Figures 3.11e,f). In the soft-landing scenario, the cold-point tropopause shows a clear separation from deep convection (Figure 3.11d and blue horizontal lines in Figures 3.1a and 3.6a). In this case, a small and smooth change in the convective cooling profile should not directly affect the CPT.

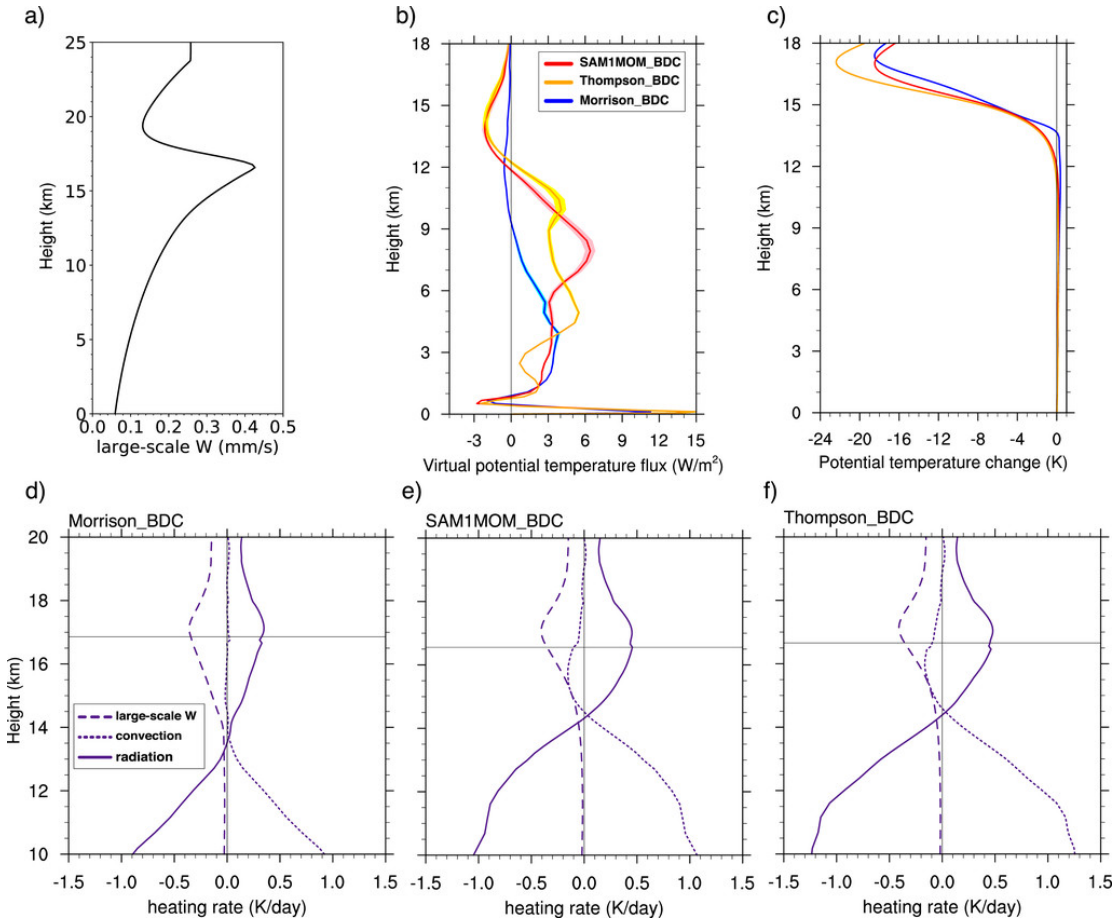


Figure 3.11: (a) Large-scale vertical velocity used in the BDC set of simulations. (b) The virtual potential temperature flux for SAM1MOM_BDC (red), Thompson_BDC (orange), and Morrison_BDC (blue). (c) Potential temperature change from the DEFAULT set to the BDC set of simulations. (d) The heat balance in the Morrison_BDC simulation. The solid purple line is the radiative heating rate. The dashed purple line is the adiabatic heating rate due to large-scale vertical velocity. The dotted purple line is the convective heating rate. The horizontal black line indicates the cold-point tropopause. (e) As in (d), but for SAM1MOM_BDC. (f) As in (d), but for Thompson_BDC.

3.7.3 Potential impact from varying ozone concentration to convective overshoots

In this section we explore whether the variation of ozone concentration could impact the strength of convective overshoots. The seasonal cycle of ozone can lead to significant temperature variations near and above the tropopause (e.g., Fueglistaler et al., 2011). Therefore, one may wonder whether ozone has the potential to impact the strength of convective overshoots through modifying the stratification below the TTL.

To get the seasonal cycle of ozone concentration in the tropics, we use monthly ozone concentration from ERA5 data (Hersbach, 2020) and average them within 30°S–30°N over the period of 2010–19. We choose two ozone profiles for February and August to represent the largest ozone variation in the TTL. These two profiles, in addition to the ozone profile used in the DEFAULT simulations, are shown in Figure 3.12a. Then, we perform a set of ozone simulations (Table 3.1), in which we use SAM one-moment scheme with the February and August ozone profiles (SAM1MOM_Feb and SAM1MOM_Aug).

Figure 3.12b shows that the change in temperature due to the ozone seasonal cycle is very small below 12 km (<0.3 K), although there is a large temperature variation in the TTL. This small temperature difference near the LNB is not large enough to significantly impact convective overshoots. As shown in Figure 3.12c, the profiles of virtual potential temperature flux of the SAM1MOM_DFT, SAM1MOM_Feb, and SAM1MOM_Aug simulations almost overlap on each other. To get a temperature difference comparable to the temperature difference between the SAM1MOM_DFT and Morrison_DFT, we try to uniformly increase the ozone concentration in the upper troposphere between 10 and 14 km. The orange line in Figure 3.12a shows an increase of ozone concentration of 2 mg kg^{-1} (1207 ppmv), which warms the upper troposphere by ~1 K (Figure 3.12b) and weakens the maximum downward virtual potential temperature flux from 2.5 to 1.6 W m^{-2} (Figure 3.12c).

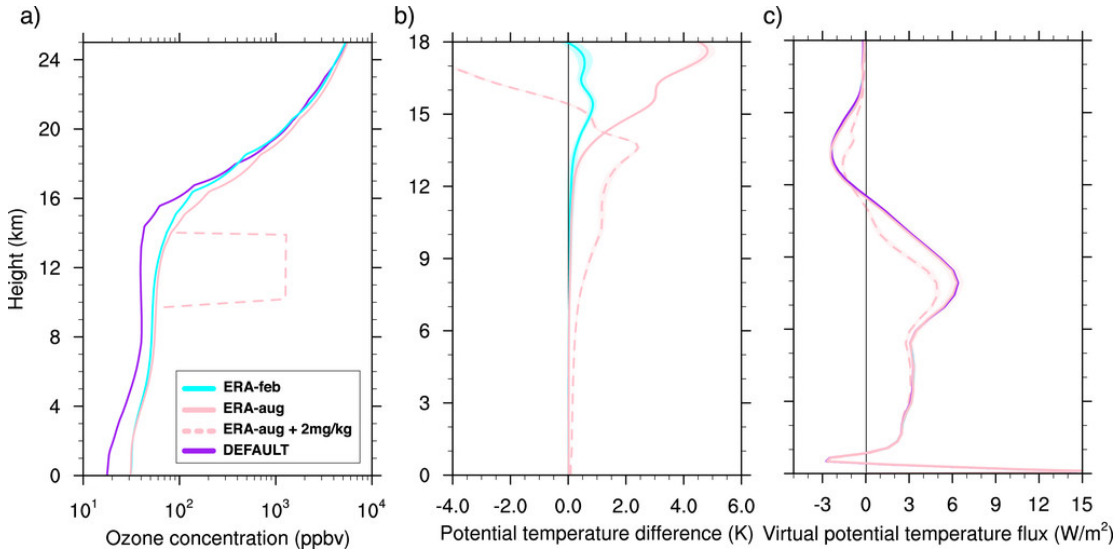


Figure 3.12: (a) Profiles of ozone concentration used in this study, (b) potential temperature deviation from SAM1MOM_DFT, and (c) virtual potential temperature flux for SAM1MOM_FEB (cyan), SAM1MOM_Aug (solid pink), SAM1MOM_Aug+2 (dashed pink), and SAM1MOM_DFT (purple).

Our limited simulations suggest that, under the current RCE framework and SST of 30.5°C, the seasonal cycle of tropical ozone concentration may not have a significant impact on convective overshoots. To warm the upper troposphere by a comparable amount to the cloud radiative effect, we need to increase the ozone concentration in the upper troposphere to an unrealistic high value. However, our results do not exclude the possibility for ozone (and similarly eddy-driven upwelling mentioned in the previous section) to influence convective overshoots. Even though the SST we used is similar to that of the western Pacific warm pool, the LNB in the RCE simulations may not represent a realistic situation. Including large-scale Hadley-like or Walker-like circulation might lift the LNB to higher levels. If the LNB can be lifted by including large-scale circulation to a level where the ozone variation and eddy-driven upwelling can drive significant temperature change (e.g., on the order of 1 K), the variation in ozone concentration and eddy-driven upwelling could then easily modify the strength of convective overshoots. It is also unknown whether in a warmer climate, as troposphere gets deeper, the LNB becomes higher and could be more easily influenced by the

variation of ozone or eddy-driven upwelling. Future study is needed to explore these possibilities.

3.8 Summary

In this study, we sought to study what processes can influence the strength of cooling in the TTL by convective overshoots. We carried out RCE simulations with a three-dimensional cloud-resolving model SAM. To this aim, we explored three distinct microphysics schemes. We found that the choice of cloud-microphysics scheme can substantially influence the strength of convective cooling in the TTL through modifying the upper-troposphere stratification. Two of the microphysics schemes (SAM single-moment scheme and Thompson double-moment scheme) produce a hard-landing scenario: frequent occurrence of convective overshoots reaching the TTL with strong updrafts, high LNB (~12 km), and a substantial convective cooling in the TTL (with a minimum heating rate of ~ -0.2 K day^{-1}). On the other hand, the other double-moment scheme (Morrison scheme) produces a soft-landing scenario: rare occurrence of strong convective overshoots reaching the TTL, low LNB (~9.5 km), and a much weaker convective cooling in the TTL (with a minimum heating rate of ~ -0.03 K day^{-1}). The difference between the hard-landing and soft-landing scenarios largely results from the difference in the upper-troposphere stratification, which is related to different cloud–radiation interaction across schemes. The stronger ACRE produced by the Morrison scheme generates a warmer and more stable upper troposphere, which acts as a buffer zone below the TTL and consequently slows down the convective updrafts and lowers the LNB. This stratification mechanism also suggests the possibility for ozone variation or eddy-driven upwelling in the TTL to modulate the strength of convective overshoots. We further test the sensitivity of overshooting cooling to changes in model resolution by increasing the horizontal resolution to 100 m. The corresponding change of overshooting cooling is much smaller compared with the difference between the hard-landing and soft-landing scenarios. These results highlight the importance of cloud

microphysics and ACRE to accurately simulate the convective influence on the TTL. Also, this study suggests that the treatment of cloud microphysics may be one of the reasons for the disagreement on the importance of convective cooling in the TTL by previous numerical studies (Kuang & Bretherton, 2004; Küpper et al., 2004; Robinson & Sherwood, 2006). Although this study only focuses on the heat budget, the difference in the strength of convective overshoots could also potentially have a major impact on the water budget in the TTL.

However, it is still unclear which scenario (hard-landing or soft-landing) and which cloud-microphysics scheme is more realistic when compared to the observations. Tropical ACRE show large discrepancies between observation, reanalysis, and climate models (Voigt et al., 2019; Fu et al., 2018) Even in our RCE simulations, increasing the horizontal resolution from 1 km to 100 m can significantly increase cloud fraction and ACRE. In reality, ACRE also have large spatial contrast between large-scale subsidence region and convective region. It is therefore hard to simply compare ACRE in RCE simulations to observation to infer which microphysics scheme produces more realistic scenario on convective overshoots. For the study of convective overshoots in the TTL, one possible way to reduce the uncertainty from ACRE can be to nudge the domain mean cloud radiative heating rate toward ACRE estimated from observational cloud fields (e.g., Fu et al., 2018) in the RCE and more complicated experiment setups. To do this, one should be careful about the uncertainties in cloud observations and possible mismatches of the vertical cloud height between observation and simulations due to surface conditions or other experimental setup details.

One limitation of this study is the idealized experimental settings (e.g., small-domain radiative–convective equilibrium simulation, no large-scale circulation, no horizontal wind shear, fixed ozone profile). Many interesting questions can be explored with a systematic increase in the complexity of the experimental setup. What is the impact of including large-scale tropospheric circulation on the height of LNB and convective overshoots? What is the relative role of deep convection and eddy-driven circulation in TTL heat budget in a near-global or a long tropical-channel cloud-resolving

simulation? Could seasonal cycle of ozone concentration and eddy-driven upwelling in the TTL significantly modify the temperature near the LNB in more realistic simulations? Further research to address these questions can potentially improve the current understanding of the role of deep convection in the heat balance in the TTL and may help predict the responses of TTL in a changing climate.

4

Stable Machine-Learning Parameterization of Subgrid Processes in a Comprehensive Atmospheric Model Learned From Embedded Convection-Permitting Simulations

4.1 Abstract

Modern climate projections often suffer from inadequate spatial and temporal resolution due to computational limitations, resulting in inaccurate representations of sub-grid processes. A promising technique to address this is the Multiscale Modeling Framework (MMF), which embeds a kilometer-resolution cloud-resolving model within each atmospheric column of a host climate model to replace traditional convection and cloud parameterizations. Machine learning (ML) offers a unique opportunity to make MMF more accessible by emulating the embedded cloud-resolving model and reducing its substantial computational cost. Although many studies have demonstrated proof-of-concept success of achieving stable hybrid simulations, it remains a challenge to achieve near operational-level success with real geography and comprehensive variable emulation that in-

cludes, for example, explicit cloud condensate coupling. In this study, we present a stable hybrid model capable of integrating for at least 5 years with near operational-level complexity, including coarse-grid geography, seasonality, explicit cloud condensate and wind predictions, and land coupling. Our model demonstrates skillful online performance, achieving a 5-year zonal mean tropospheric temperature bias within 2K, water vapor bias within 1 g/kg, and a precipitation RMSE of 0.96 mm/day. Key factors contributing to our online performance include an expressive U-Net architecture and physical thermodynamic constraints for microphysics. With microphysical constraints mitigating unrealistic cloud formation, our work is the first to demonstrate realistic multi-year cloud condensate climatology under the MMF framework. Despite these advances, online diagnostics reveal persistent biases in certain regions, highlighting the need for innovative strategies to further optimize online performance.

4.2 Introduction

Accurately representing cloud and convection has long been a major challenge for accurate climate model simulations due to the relatively coarse resolution (>50km) in traditional climate models (Eyring et al., 2016). Uncertainty associated with representing these subgrid processes remains one of the most significant factors contributing to climate feedback uncertainties in future climate change projections (IPCC, 2021; Sherwood et al., 2020). To better represent these processes, the community has developed global storm-resolving simulators with <5km resolution. These models are actively being tested for their use in long-term climate simulations, but they are still computationally too demanding (e.g., Taylor et al., 2023; Hohenegger et al., 2023).

Another approach, known as the Multiscale Modeling Framework (MMF)—also commonly referred to as super-parameterization—explicitly embeds a small-domain, km-resolution cloud-resolving model (CRM) in each atmosphere column of a host climate model to replace the tra-

ditional convection and cloud parameterization (Grabowski & Smolarkiewicz, 1999; Benedict & Randall, 2009; Randall, 2013; Hannah et al., 2020; Norman et al., 2022). The embedded cloud-resolving model can directly simulate the atmospheric processes at the km-scale with time steps of a few seconds and feedback to the host climate model the effect of these small-scale processes on the coarser grids. However, such MMF models can still be computationally very demanding, for example, 100-10,000 times more expensive than conventional global climate models depending on the dimensionality and resolution of the interior resolved scale (e.g., Khairoutdinov & Randall, 2001; Parishani et al., 2018; Terai et al., 2020; Hannah et al., 2020; Peng et al., 2024). To enhance computational efficiency, the embedded cloud-resolving model is often configured with a two-dimensional domain.

A hybrid approach that combines traditional climate models with machine learning is a promising solution for high-fidelity climate simulations without incurring additional computational costs. Machine learning models can learn the effects of small-scale physical processes directly from high-fidelity simulation data, such as those from MMF climate simulations (Gentine et al., 2018; Rasp et al., 2018; Brenowitz et al., 2020; Han et al., 2020; Mooers et al., 2021; Wang et al., 2022b; Bhouri et al., 2023; Lin et al., 2024; Iglesias-Suarez et al., 2024; Kühbacher et al., 2024), globally uniform km-resolution simulations (Brenowitz & Bretherton, 2018, 2019; Yuval & O'Gorman, 2020; Yuval et al., 2021; Wang et al., 2022a), and reanalysis products (Kochkov et al., 2024). Once trained, an ML emulator can be coupled with a host climate model to perform hybrid physics-ML simulations, potentially enhancing simulation accuracy for a given computational budget when comparing to traditional climate simulations using empirical convection parameterizations (Rasp et al., 2018; Yuval & O'Gorman, 2020).

The hybrid approach of learning a ML paramterization from MMF simulations has its own advantages and limitations compared to other hybrid methods. The MMF simulations might be considered less physical compared to global storm-resolving simulations due to certain idealizations,

such as the scale separation between host model and the embedded CRM, and the 2D representation of clouds and convection in the 2D CRM. However, the MMF approach effectively captures the essence of the challenge of learning a ML parameterization, e.g, stochasticity, multi-physics, and coarse-graining in time. The scale separation also brings the convenience of locality, as CRMs in different columns do not directly communicate. Therefore, each CRM can be treated as an independent sample, and emulating each CRM would only require the input and output in the same column.

Conversely, using global km-resolution simulation data requires coarse-graining the raw data, and the choice of coarse-graining scale can be challenging, potentially leading to issues when the ML parameterization learned from the coarse-graining data is coupled with a host model (e.g, Ross et al., 2023). The defined machine learning problem of emulating coarse-graining targets may also involve non-locality (Wang et al., 2022a), where the state tendencies of each column can be dependent on the states of adjacent columns. In addition to learning from simulation data, Kochkov et al. (2024) trains their hybrid models, called Neural General Circulation Model (NeuralGCM), directly from reanalysis data. They achieved this by building a fully differentiable model, which includes a differentiable dynamical core and a neural network-based parameterization for physics in each column. The NeuralGCM has the advantage of optimizing climate metrics directly against reanalysis data of historical climate but faces challenges of extrapolating to substantially different future climate. In contrast, the MMF hybrid approach has a code base that is fit for climate prediction, with all necessary components and modules for simulating future climate beyond the historical record.

In addition to advantages and limitations mentioned in the previous paragraph, these hybrid physics-ML climate simulations face challenges related to online instability and online error. Optimizing a ML model's offline skill (by optimizing some loss function of the ML model's prediction at one single time step) does not necessarily lead to optimized online performance when a ML model is coupled with a dynamical core and integrates for years (Ott et al., 2020; Wang et al.,

2022b). Many previous studies have demonstrated proof-of-concept success in achieving stable online simulations using idealized setups such as aquaplanets. A few recent works (Wang et al., 2022b; Han et al., 2023; Sanford et al., 2023a; Kochkov et al., 2024; Behrens et al., 2024) have demonstrated the promise of stable online simulations incorporating real geographic features. Kochkov et al. (2024) achieved mostly stable (though occasional instability was observed) hybrid simulations using end-to-end online training, developing a fully differentiable hybrid model and directly optimizing multi-step rollout loss. Nevertheless, making existing Fortran-based climate models differentiable is challenging.

In the context of emulating the CRM in MMF simulations with a ML parameterization, no previous work has succeeded in completely coupling all input and output variables of the CRM (e.g., tendencies of cloud condensates, surface fluxes for atmosphere-surface coupling) with real geography via ML. For example, Behrens et al. (2024) had to sidestep the condensate coupling to maintain a 5-month online stability by using a partial coupling approach, where condensate prediction is provided by a simultaneously running CRM. Han et al. (2023) predicted cloud species and achieved stable online simulations with real geography, but they did not emulate radiation (and thus the radiative fluxes required for atmosphere-land coupling) and showed significant online biases in high-latitude temperature and humidity. Additionally, Han et al. (2023) predicted cloud species in a semi-implicit manner, diagnosing current time step cloud states based on current temperature and moisture states (and many other input features) without interacting with the dynamical core, as indicated in Table 1 of Han et al. (2020). Conventional cloud parameterizations were also retained in Han et al. (2023) to supply inputs required by their radiative transfer scheme—such as cloud liquid and ice number concentrations and cloud fraction—that were not predicted by their neural network.

In the long term, it is crucial for ML parameterizations to explicitly and accurately predict cloud condensate tendencies and radiative transfer, which are vital for handling aerosol cloud interaction,

pollutant transport, and climate change. A recent open-source dataset, ClimSim (Yu et al., 2024a), provides a valuable opportunity to train ML parameterizations with operational-level complexity. This dataset represents the largest-ever open-source collection tailored for operation-ready hybrid physics-ML climate simulations. It provides a comprehensive set of inputs and outputs for downstream coupling and contains an extensive number of samples with high sampling frequency at multiple spatial resolutions. It includes a coarse-resolution version (11.5°×11.5°, 384 grid columns, $<$ 1 TB), suitable for cost-effective prototyping, and a high-resolution version (1.5°×1.5°, 21,600 grid columns, 42 TB), which enables evaluation of performance under operational-level conditions with more detailed orography and land-sea contrasts.

In this work, we used the coarse-resolution version of the ClimSim dataset to train ML models capable of predicting a complete set of output variables. Our hybrid physics-ML simulations are stable under near-operational-level complexity (with a real but coarse geography, seasonality, explicit prediction of cloud condensate and wind tendencies, and surface coupling), outperforming previous studies that emulate the CRM in MMF climate simulations. Our 5-year stable hybrid simulation demonstrates less than 2K zonal mean troposphere temperature bias, less than 1g/kg zonal mean troposphere humidity bias, a root mean square error of 0.96 mm/day for mean precipitation, and realistic zonal mean cloud distribution. To inform others' practice, we show how this online performance is achieved as the result of several key efforts: employing an expressive U-Net architecture and incorporating cloud microphysical constraints.

The rest of the manuscript is structured as follows: Section 4.3 documents the methods, including details on model architecture, microphysical constraints, data, and training. Section 4.4 presents the offline skill improvement from using the more expressive U-Net architecture and expanded input features compared to a baseline multi-layer perceptron (MLP) model, a fully connected neural network architecture. Section 4.5 demonstrates the online performance of our hybrid simulations. Finally, Section 4.6 provides the discussion and summary.

4.3 Methods

4.3.1 Dataset

The data we used is the low-resolution real-geography dataset in the ClimSim dataset (Yu et al., 2024a). This low-resolution real-geography data consists of 10-year simulation from the Energy Exascale Earth System Model (E3SM)-multiscale modeling framework (MMF) (Hannah et al., 2020), configured with the predefined "F2010-MMF1" compset with "ne4pg2" grid and a model time step of 20 minutes. This low-resolution simulation contains 384 columns in the globe using an unstructured cubed-sphere grid, with an effective resolution of approximately 11.5°×11.5°. The E3SM-MMF model employs 60 levels for global dynamics, extending up to approximately 65 km in altitude. The E3SM-MMF simulation uses prescribed sea surface temperature, sea ice coverage, and aerosols based on present-day climatology. The prescribed sea surface temperature follows a fixed recurring annual cycle. The aerosols are configured to be transparent to radiation code. The first seven years are used for the training set without subsampling. The eighth year is used for the validation set, while the last two years are used for the test set. Both the validation and test sets (not the training set) are subsampled in time using a stride of 7 time steps to be consistent with the constructed validation and test sets in Yu et al. (2024a). Each simulation day has 72 time steps, so the stride of 7 adequately samples the full diurnal cycle. This whole low-resolution dataset provides around 100 million samples with a total data size of 744GB, of which 521GB is used for the training set.

4.3.2 Architecture

Two architectures were used and compared in this study: a multi-layer perceptron (MLP) model and a U-Net model. The first architecture is the MLP model, which is a fully connected neural network with one or more hidden layers. We adopted the hyperparameters recommended by (Yu

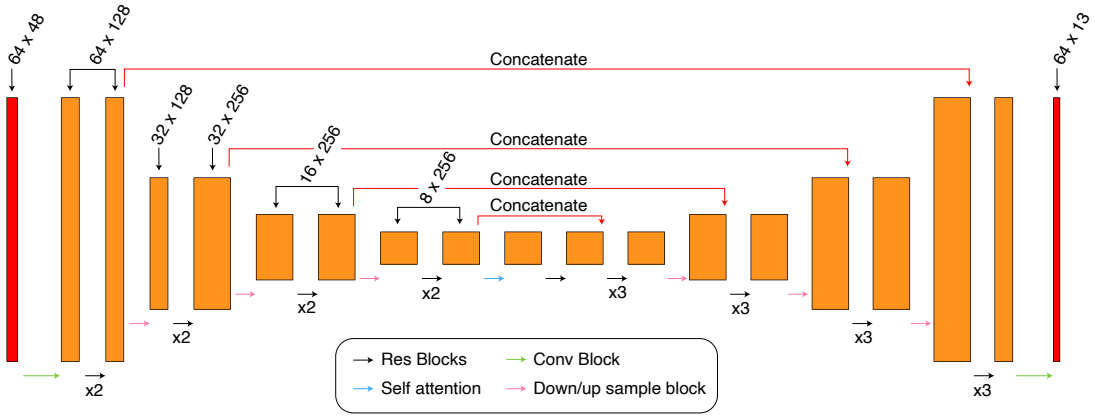


Figure 4.1: The schematic of the U-Net architecture. The U-Net model is composed of multiple ResBlocks, with 4 layers in depth and latent dimension sizes of [128, 256, 256, 256]. The model requires the input and output of each sample to be prepared as 2D arrays (see Appendix B.1.4) with dimensions corresponding to the number of variables times the vertical length. The numbers on top of each tensor indicate the tensor shape for the U-Net model with the microphysical constraint. Each tensor in the encoder has a skip connection to a corresponding tensor in the decoder, illustrated by the red arrows showing some of these skip connections.

et al., 2024a), which were informed by their ambitious hyperparameter search comprising 11,851 trials. MLP models are widely used in previous related studies (Gentine et al., 2018; Rasp et al., 2018; Brenowitz & Bretherton, 2019; Ott et al., 2020; Yuval et al., 2021; Beucler et al., 2024; Lin et al., 2024; Iglesias-Suarez et al., 2024) and serve well as a baseline for comparison.

The second architecture is a U-Net model (Song et al., 2020), suitable for tasks where the output targets share the same spatial dimensions as the input features (here, the number of vertical levels in each atmospheric column). Other architectures have also been used in previous studies, such as Residual Convolutional Neural Networks (ResCu) in Han et al. (2020, 2023) and Residual Deep Neural Networks (ResDNNs) in Wang et al. (2022b); Han et al. (2023). Heuer et al. (2023) recommended the U-Net architecture after comparing it with MLPs, Convolutional Neural Networks, and Residual Neural Networks, demonstrating that U-Net achieved the highest fitting score. We adapted the 2D U-Net model from Song et al. (2020) into a 1D version (Figure 4.1) by replacing 2D convolution layers with 1D convolution layers. The U-Net model has an encoder-decoder structure,

where the encoder compresses the input data and the decoder reconstructs it, with skip connections that retain important details from the input. This setup enables the model to capture both short-range details, such as information from adjacent vertical levels, and long-range patterns, such as vertical structure of the entire column. Compared to the MLP model, the U-Net model can better learn the vertical structures through its convolutional layers and optional self-attention layers. Full details of the model parameters for both the MLP and U-Net models are documented in Appendix B.1.2.

4.3.3 CRM Setup in the E3SM-MMF Model

Here we provide a brief description of the multi-scale modeling framework setup in the E3SM-MMF to help clarify the underlying emulation targets for machine learning models.

The E3SM-MMF climate model embeds a cloud-resolving model (CRM) in each atmospheric model column to replace parameterizations of microphysics, convection, radiation and turbulence. The embedded CRM is the System for Atmosphere Modeling model (SAM, Khairoutdinov & Randall, 2003). The CRM model is configured with a 2D domain of 64 columns and 1km horizontal resolution. While the E3SM has 60 vertical levels, the CRM only operates on the lowest 50 levels, excluding the uppermost 10 levels to circumvent issues associated with applying the anelastic approximation in areas of extremely low density. Therefore, the CRM output has zero tendencies for water species and momentum in the top 10 E3SM levels, but it still has non-zero tendency for temperature since the radiation calculation is conducted through the whole 60 levels.

The CRM model uses a one-moment microphysics scheme (Khairoutdinov & Randall, 2003). The only prognostic water species in this one-moment scheme are the total water (vapor plus non-precipitating condensate) and total precipitating water. The one-moment scheme uses an instantaneous saturation adjustment to generate and remove cloud condensate on the internal CRM grid. Cloud condensate is then partitioned into liquid, ice, or mixed phase based on a diagnostic func-

tion of temperature as follows. Between 273.16K and 253.16K, partitioning of cloud condensate into cloud ice and liquid water depends linearly on temperature (at 253.16K, all condensate is ice; at 273.16K, all condensate is liquid water). Precipitating water is partitioned into rain, snow, and graupel in the same way. It is worth to note that the E3SM grid only tracks water vapor, cloud liquid, and cloud ice and does not track precipitating water.

The CRM in E3SM-MMF uses the Explicit Scalar Momentum Transport (ESMT) method (Tulich, 2015) to represent convective momentum transport (CMT) for large-scale winds. ESMT treats the zonal and meridional components of large-scale momentum as non-conserved scalars, which are advected by resolved fine-scale flows within the 2D CRM. Since the 2D CRM lacks a third spatial dimension, the ESMT scheme implicitly assumes that the advective flow field has identical spatial structures in the zonal and meridional directions. Despite this limitation, Yang et al. (2022) demonstrated that the E3SM-MMF with a 2D CRM and ESMT produces a reasonable global distribution of CMT, comparable to simulations using a 3D CRM with fully explicit convective momentum transport. While the qualitative patterns of CMT tendencies are similar between the 2D and 3D approaches, the ESMT-derived tendencies are consistently weaker in magnitude compared to those produced by the 3D CRM.

In E3SM-MMF, radiation transfer calculations are performed at every GCM timestep (20 minutes) after the CRM module, separately on individual sub-columns of the CRM state. The calculated radiative heating tendencies are applied as constant forcings to the next timestep's CRM calculation. The radiation scheme from the GCM is used, taking the time-averaged state of each CRM column from the previous CRM calculation as input (Hannah et al., 2020).

The CRM module replaces several parameterizations in conventional GCMs, including those for convection, cloud fraction, and microphysics. The CRM module also calculates the surface precipitation fluxes, which are passed to the land component. Our NN emulators aim to replace the entire CRM including its radiation calculation.

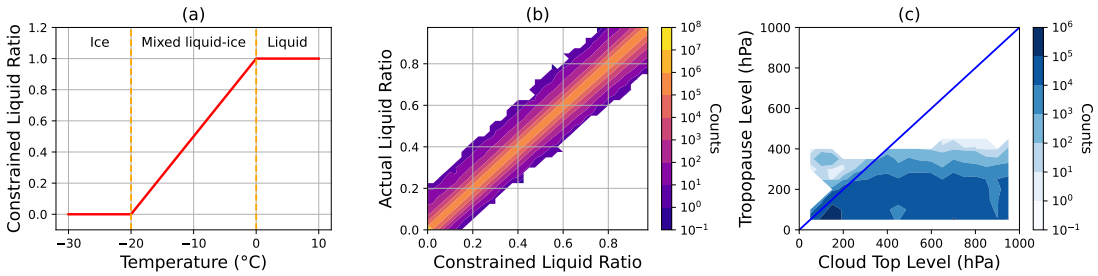


Figure 4.2: Assessment of the microphysical constraints. (a) The fraction of liquid over total cloud water mixing ratios as a function of temperature. This temperature-based phase partitioning holds exactly on each grid in the cloud-resolving model. (b) Histogram of the actual liquid cloud fraction in the E3SM grid (y-axis) versus the predicted liquid fraction based on the E3SM-grid temperature (x-axis). (c) Histogram of the tropopause level versus the cloud top level, where levels correspond to the hybrid pressure level used in E3SM. In each column, the tropopause level is identified as the lowest level where the pressure is less than 400 hPa and the vertical gradient of potential temperature exceeds 10 K/km. A cloud top is defined as the highest level with a total cloud water mixing ratio greater than 1 mg/kg.

Exterior to the CRM, on large-scales, the host atmosphere model in the E3SM-MMF retains conventional treatments of advection physics within the dynamical core, as well as additional physics modules that handle vertical diffusion and gravity wave drag. The dynamical core uses a spectral element method on a cubed-sphere geometry (Hannah et al., 2020), updating all state variables including temperature, winds, water vapor, cloud liquid, and cloud ice.

4.3.4 Microphysical Constraints

Liquid-ice Cloud Partition

It is tempting to think of the emulation problem from the perspective of the host model, focusing on what the Global Climate Model (GCM) typically receives from the CRM. The ClimSim dataset (Yu et al., 2024a) was designed with this view in mind, providing the necessary data for seamless integration. However, our work demonstrates that it is especially important to consider the CRM and its own fundamental prognostics. In particular, the one-moment microphysics scheme in the CRM only predicts the total cloud (ice plus liquid) mixing ratio and uses temperature to diagnose the

mixing ratios of liquid cloud and ice cloud on each CRM grid. According to the thermodynamic principles and CRM model code design, only liquid cloud is allowed to exist when the temperature is above 273.16K, and only ice cloud is allowed when the temperature is below 253.16K (Figure 4.2). In between, the fraction of liquid cloud over total cloud mixing ratio follows a linear function of temperature.

While this temperature-based partition relationship between liquid and ice clouds is strictly enforced in every CRM grid, Figure 4.2b shows that this relationship also holds on the E3SM grids which reflect the horizontally averaged values in the CRM domain. The exact fraction of liquid cloud over total cloud condensate on each E3SM grid is well predicted by the temperature-based function, as most of the samples fall into either the lower left corner (pure ice cloud), the upper right corner (pure liquid cloud), or the diagonal (mixed-phase cloud).

To leverage this microphysical constraint, instead of predicting liquid and ice separately, as the host GCM is accustomed to receiving its information from its embedded CRM in an MMF, we can predict only the change of total (i.e., liquid plus ice) cloud condensate at each step, a more fundamental total water prognostic of the CRM itself. After the neural network updates the temperature and total cloud mixing ratio in each grid, the updated temperature is then used to decompose the total into the amounts of liquid and ice clouds. This separation is necessary because the dynamical core advects liquid and ice clouds independently. This approach ensures that the correct cloud phase appears at the appropriate temperature range.

Removing clouds above stable tropopause level

In the hybrid online simulations, we implement a constraint of removing all the cloud above the tropopause to improve simulation stability. The tropopause is identified as the lowest level where the pressure is less than 400 hPa and the vertical gradient of potential temperature exceeds 10 K/km. The intuition of this constraint is that the tropopause acts as a stable barrier preventing deep con-

vection from penetrating and forming clouds above it. Any clouds above the tropopause, whether formed by deep penetrating convection in previous time steps or advected from lower levels, tend to be removed by sedimentation processes. The empirically chosen thresholds for identifying the tropopause ensure that the tropopause generally lies above the cloud top in most cases (Figure 4.2c). The cloud top is defined as the highest level with a total cloud mixing ratio greater than 1 mg/kg. Most residual clouds above the defined tropopause after invoking the CRM have mixing ratios on the order of 0.1 mg/kg or smaller, while the zonal mean cloud mixing ratio in the upper troposphere is at the order of 10 mg/kg (see Section 4.5). This constraint does not alter the model architecture or training process but is implemented during online inference to enhance stability. Implementing this constraint allowed us to easily test its impact and test different thresholds for the defined tropopause without needing to retrain the model. However, it is also possible to incorporate this constraint directly into the ML model itself and both methods are equivalent.

4.3.5 Inputs and Outputs

The input features for the MLP model include GCM-scale state variables (e.g., temperature, relative humidity, cloud liquid mixing ratio, cloud ice mixing ratio, zonal wind, and meridional wind), atmospheric gas components, and boundary conditions (e.g., solar insolation, surface heat fluxes, surface stress, albedo, and geographic information). These GCM state variables are saved immediately prior to a code region containing the CRM and radiation physics. The output targets include the CRM's time tendencies of temperature, specific humidity, cloud liquid and ice mixing ratios, zonal and meridional winds, as well as radiative fluxes and surface precipitation rates. That is, the ML targets encompass all the forcings that the CRM transmits to its host GCM in the MMF approach to climate simulation, including challenging condensate tendencies.

Expanding Inputs to Include Large-scale Forcings and Convection Memory

The U-Net model uses expanded input features beyond ClimSim's standard input that also include large-scale forcings (primarily the advective tendencies of temperature, total water, and zonal wind from the dynamical core), convection memory (CRM tendencies from previous steps), and latitudes.

Large-scale forcings refer to the tendencies of temperature, total water (water vapor plus liquid and ice cloud condensates), and zonal wind outside of the CRM in each time step, primarily from the dynamical core. During the CRM calculation, these large-scale forcings are applied uniformly in time and at each vertical level in the CRM domains. Such large-scale forcings can influence convection; for example, a large-scale forcing that stabilizes the atmosphere could make the convection more difficult to trigger, making them causally relevant input features. However, most previous studies of hybrid MMF simulations failed to include these features, with only a few exceptions (Han et al., 2020; Wang et al., 2022b; Han et al., 2023).

Convection memory refers to the internal convection state within the CRM that is not adequately represented in the GCM grid state (which is the averaged value of the CRM domain). The internal convection state, including aspects such as internal cloud structure and convection organization, can be useful predictors for convective tendencies and precipitation (Shamekh et al., 2023; Kuang, 2024). The CRM tendencies at previous time steps have been used in a few studies (Han et al., 2020, 2023; Lin et al., 2024) as a proxy to represent aspects of convection memory.

Due to the physical relevance, we expect that adding these expanded input features can help with both offline and online skill. Both inputs and outputs are normalized to be on the order of 1. Detailed descriptions of the full list of input and output features and their normalization methods can be found in the Appendix (Sections B.1.1 and B.1.3).

4.3.6 Training

We used and compared four sets of offline model designs: a baseline MLP architecture without expanded input features such as large-scale forcings or convection memory, a U-Net architecture without expanded input features (hereafter U-Net-baseline), a U-Net architecture with expanded input features (hereafter U-Net-expanded), and a U-Net architecture with both expanded input features and the cloud microphysics constraint of the temperature-based liquid-ice partition (hereafter U-Net-expanded-constrained).

To account for stochasticity in ML optimization, we used three different checkpoints for each model architecture. These checkpoints were obtained using varying loss functions (either mean-absolute-error loss or Huber loss) and learning rate schedules, creating a small ensemble to account for some of the uncertainty in online performance stemming from offline ML optimization variability. All the models were trained in PyTorch (Paszke et al., 2019) using the NVIDIA Modulus framework (NVIDIA Modulus Team, 2025). Please see B.2 for the detailed training hyperparameters for these different checkpoints. In the subsequent sections and plots where only one of the three checkpoints is shown, it is the one that uses the Huber loss and a reduce-on-plateau learning rate scheduler. Choosing other checkpoints does not change the following results qualitatively.

4.3.7 Coupling E3SM and NN

The trained PyTorch neural networks are integrated with the Fortran-based E3SM codes using an open-source library called Pytorch-Fortran (Alexeev, 2023). This library requires saving the PyTorch model in TorchScript (PyTorch Contributors, 2024) serialized format, allowing the Fortran-based E3SM to directly read the TorchScript model and perform efficient batch inference within the Fortran process using bindings to the TorchScript C++ API. This coupling approach is compatible with most machine learning models written in PyTorch, allowing for extensive testing of diverse ML

models. Furthermore, this E3SM-NN coupling workflow is integrated into a containerized, end-to-end framework called ClimSim-Online (Yu et al., 2024b) which allows researchers to run and test such hybrid E3SM simulations on multiple platforms.

4.4 Offline results

Figure 4.3 shows the comparison of R^2 across all target variables for the MLP and U-Net models. R^2 for each target variable at each vertical level is calculated using data across all the columns and time steps in the test set. For the U-Net-expanded-constrained model, the R^2 values reflect only the constraint involving liquid-ice cloud partitioning, as the constraint for removing stratospheric clouds is applied solely in online simulations. These R^2 results are from one of the checkpoints using the Huber loss function.

Our baseline MLP model shows similar R^2 of temperature and lower tropospheric water vapor tendencies compared to the MLP models in Yu et al. (2024a). However, our MLP model demonstrates significantly better R^2 of cloud condensates and wind tendencies. Figure 8 in the Supplementary Information in Yu et al. (2024a) shows non-positive R^2 at most levels for winds and cloud ice tendencies, as opposed to our MLP results in Figure 4.3.

This improvement in MLP offline skills is likely primarily due to differences in normalization strategies. We used a per-level standard deviation that varies with altitude to make each target variable at each level at the order of 1. The normalization in Yu et al. (2024a) used a scaling factor that does not vary vertically, which may prevent effective optimization of targets at all levels. In addition, the choice of the output scaling factors in Yu et al. (2024a) made the loss from wind and cloud tendencies too small from the optimizer's perspective compared to the loss from temperature and water vapor tendencies, making their MLP model difficult to learn the wind and cloud tendencies. Additionally, many input features, such as cloud water mixing ratios, have highly skewed distributions

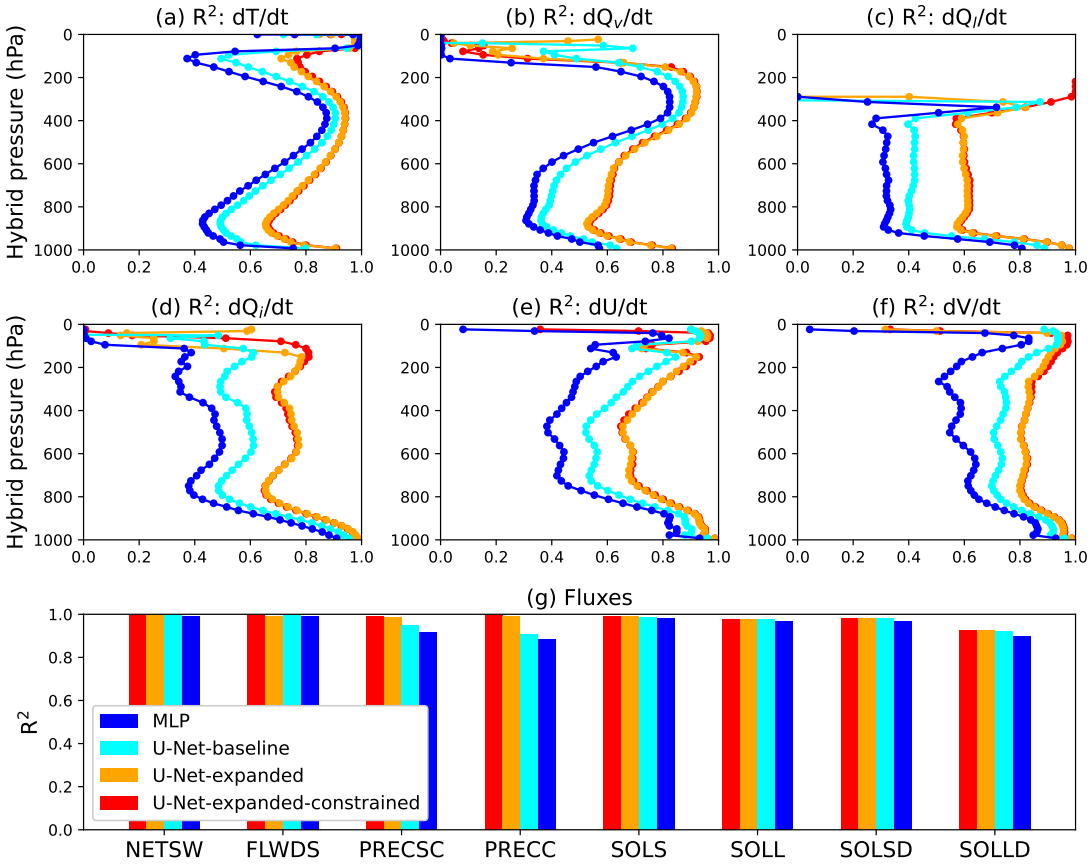


Figure 4.3: Offline R^2 scores across various variables for MLP (blue), U-Net-baseline (cyan), U-Net-expanded (orange), and U-Net-expanded-constrained (red). Variables are the full target variables listed in Table B.2, including temperature tendency ($\frac{dT}{dt}$), water vapor tendency ($\frac{dQ_v}{dt}$), liquid cloud mixing ratio tendency ($\frac{dQ_l}{dt}$), ice cloud mixing ratio tendency ($\frac{dQ_i}{dt}$), zonal wind tendency ($\frac{dU}{dt}$), meridional wind tendency ($\frac{dV}{dt}$), and eight flux variables.

with long tails above 10 standard deviations. To address this, we transformed the exponential-like cloud input features into more uniform distributions, as documented in Appendix B.1.3. Although we did not conduct ablation tests on different normalization strategies, we believe that properly transforming skewed distributions with extreme tails to less skewed distributions (ideally normal distribution) can enhance training dynamics and improve model performance.

Switching from MLP to U-Net-baseline significantly improves the R^2 for almost all target variables at all levels, suggesting that the U-Net architecture is more expressive and better suited to capture the data compared to the baseline MLP model. Expanded input features further improve the R^2 values in the U-Net-expanded model for almost all target variables at all levels when compared to the U-Net-baseline model. Although we do not present a formal ablation of the individual effects of large-scale forcing and convection memory, preliminary tests indicate that both components contribute to improving the R^2 scores.

When comparing the U-Net models with and without the microphysical constraint, the R^2 values are quite similar overall. However, the U-Net model with the liquid-ice partitioning constraint shows a marked improvement in predicting liquid cloud tendencies at several levels above about 350 hPa (Figure 4.3c). At these levels, the constrained U-Net shows very high R^2 values, while the unconstrained U-Net shows little skill. These levels correspond to temperatures too cold for liquid clouds to exist, where the CRM tends to remove any liquid clouds that may have been advected there by the dynamical core. We also observe a decrease in R^2 for water vapor tendencies above the tropopause in the U-Net-expanded-constrained model (Figure 4.3b), though the cause remains unclear. One possible explanation is that using only the total cloud mixing ratio as input may omit relevant information needed to accurately predict water vapor tendencies at these levels, although it is impossible to rule out that other factors are at play.

Figure 4.4 shows the zonal mean R^2 for different variables for the U-Net-expanded-constrained model. Zonal mean R^2 is evaluated by first calculating the R^2 for each column across samples at

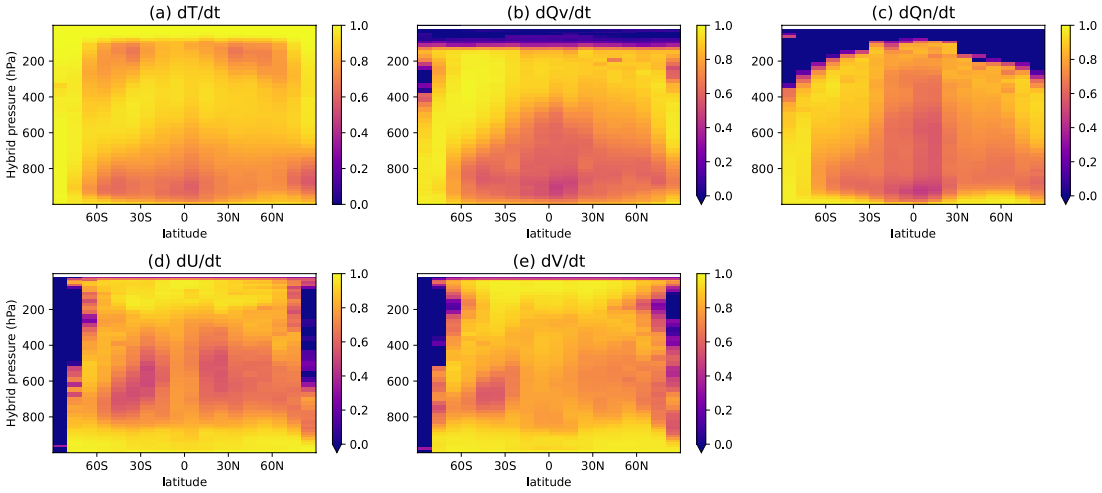


Figure 4.4: Zonal mean R^2 evaluated on the test set for the U-Net-expanded-constrained model. The five panels denote the R^2 for the tendencies of temperature (a), water vapor (b), total cloud (liquid plus ice) mixing ratio (c), zonal wind (d), and meridional wind (e).

different times and then averaging the R^2 zonally. The negative R^2 of vapor and cloud at the stratosphere is likely because the vapor and cloud tendencies at those high levels are too small and noisy to predict. Our U-Net model also poorly predicts wind tendencies at the north and south poles. More careful normalization or more training data might be required to improve the R^2 there.

4.5 Online results

4.5.1 Online Error Development in the Initial First Year

While the offline skill is important, the more critical aspect of the hybrid simulations is the online performance. No previous work under the MMF framework has achieved good online error with full condensate coupling. Figure 4.5 summarizes the online performance of the baseline MLP models and various U-Net models. All the hybrid simulations and the reference E3SM-MMF simulations in Section 4.5 use an identical initial condition.

For each hybrid simulation, we evaluate the monthly evolution of the globally averaged RMSE of

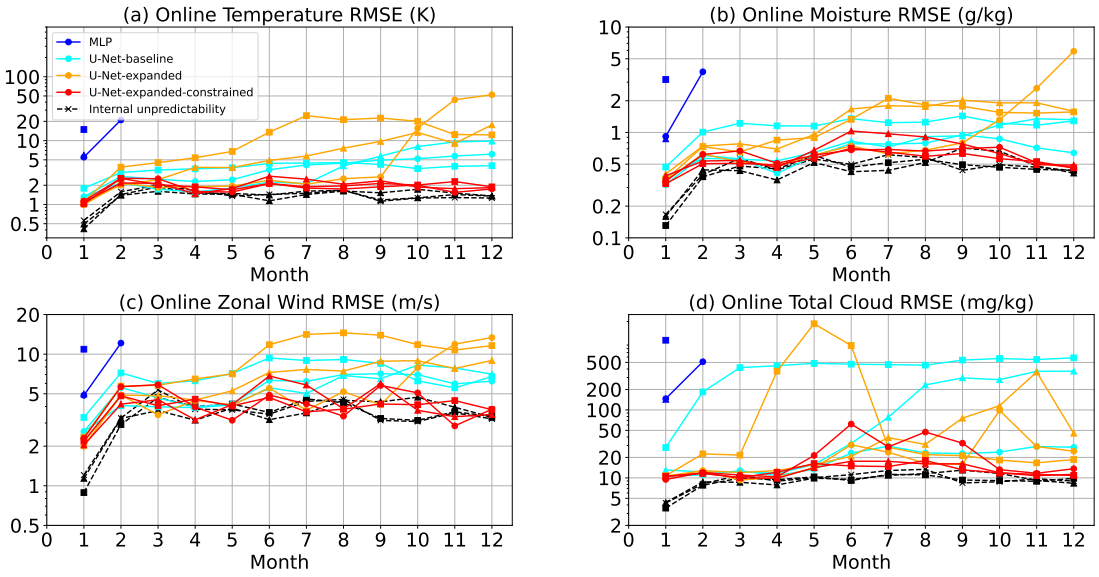


Figure 4.5: Online monthly and globally averaged (both horizontally and vertically and weighted by mass in each grid) RMSE of temperature (K), water vapor (g/kg), zonal wind (m/s), and total cloud mixing ratio (mg/kg) over a one-year period, comparing baseline MLP (blue), U-Net-baseline (cyan), U-Net-expanded (orange), and U-Net-expanded-constrained (red) against the reference E3SM-MMF simulation. The dashed black lines represent the uncertainty of the reference physical simulation due to internal atmospheric variability. Each architecture was trained with three slightly different configurations (e.g., loss functions and learning rate schedules) to sample ML uncertainty, represented by different markers but with the same line color.

monthly-mean temperature, water vapor, zonal wind, and total cloud mixing ratio over the course of one-year hybrid simulations. The RMSE quantifies the discrepancy of monthly-mean fields between the hybrid simulations and the reference E3SM-MMF simulation. The exact definition of the monthly RMSE is in B.3. Here we focus on monthly-mean values rather than instantaneous values because the instantaneous weather pattern is chaotic and becomes largely unpredictable beyond the initial autocorrelation timescale. In contrast, monthly-mean patterns associated with seasons and geography remain more predictable.

We ran a small ensemble of E3SM-MMF simulations to estimate the internal atmosphere unpredictability (dashed black lines in Figure 4.5), representing an error floor that the hybrid simulations should not be expected to surpass (i.e., the performance of a perfect emulator). The E3SM-MMF climate model runs on a hybrid CPU/GPU system and is not bit-for-bit reproducible. Tiny numerical differences due to the non-associativity of floating point arithmetic and the use of atomic operations (specific to how the current E3SM-MMF version implements parallel reductions) in the calculations can grow over time, leading to small but non-zero uncertainty even in globally averaged metrics of monthly atmospheric states in pure physical simulations. Such numerical differences, when accumulated over a few steps, naturally act as perturbations to the initial conditions that generate ensemble simulations. To create the ensemble, we ran the reference E3SM-MMF climate model three additional times using identical initial conditions and compared the resulting differences to the reference simulation.

The simulations using MLP models all crashed within the first three months and exhibited large RMSE in the first month. In contrast, all the U-Net simulations allowed for more stable integrations, with all U-Net simulations completing the full year. For the first simulated month, where all hybrid runs finished, the U-Net simulations demonstrated significantly lower monthly RMSE of all variables compared to the hybrid simulations using the MLP models.

However, the U-Net models can still experience large online error growth. For the U-Net models

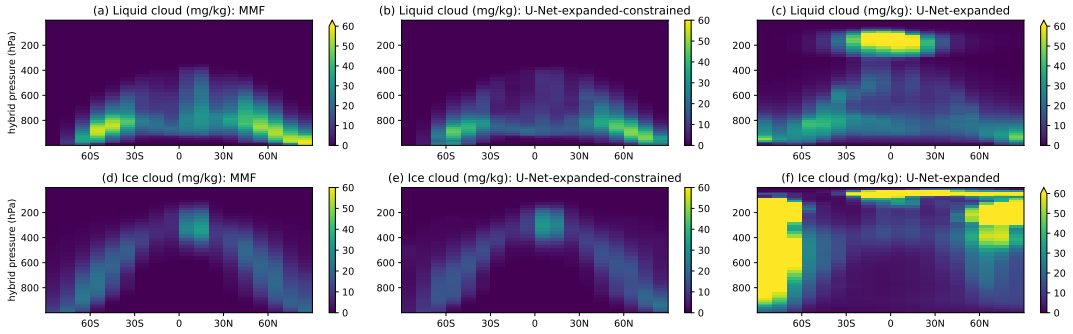


Figure 4.6: The zonal mean liquid (upper row) and ice (lower row) cloud mixing ratio at month 10 in the reference E3SM-MMF simulation (left) and in the hybrid simulations with the U-Net-expanded-constrained model (middle) and with the unconstrained U-Net-expanded model (right).

without microphysical constraints (U-Net-baseline and U-Net-expanded), while their first-month RMSE is reasonably satisfactory, the online error becomes unacceptably large even a few months into the hybrid simulation (orange and cyan lines in Figure 4.5), e.g., with temperature RMSE exceeding 10 K—far surpassing the level of internal unpredictability. Such large RMSE suggests a systematic bias in the hybrid simulation that is significantly larger than the variability of the monthly climatology. Indeed, this trend is reflected in Figure B.5, which shows the monthly evolution of global mean near-surface temperature, precipitable water, and total cloud path in the hybrid and reference simulations. Both the U-Net-baseline and U-Net-expanded simulations significantly diverge from the reference E3SM-MMF simulations after the first year, consistent with the observed RMSE growth.

We hypothesize that a root cause of the error growth is unrealistic and insufficiently constrained cloud formation. The first line of evidence is that the unconstrained U-Net models can have greater than 100 mg/kg RMSE of total clouds in some months (Figure 4.5d). Figure 4.6 compares the zonal mean liquid and ice cloud mixing ratios between the reference pure physics simulation and the U-Net-expanded-constrained and U-Net-expanded simulations at month 10, i.e. when one of the U-Net-expanded simulations (the orange line with the circle markers in Figure 4.5) begins to show

highly unrealistic cloud formation and quick growth of the RMSE of temperature and water vapor.

The second, most telling, line of evidence is that the unphysical cloud phase is immediately apparent as a pathology in the hybrid condensate climatology. The unconstrained U-Net hybrid run shows an excessive amount of liquid cloud near the tropical tropopause (Figure 4.6c), where the temperature is too cold to allow any liquid water. We also observe excessive cloud ice in the stratosphere (Figure 4.6f). These unrealistic cloud distributions are clearly out-of-sample conditions that never existed in the training data, and the unconstrained U-Net fails to remove them, continuing to make extrapolated predictions based on these out-of-sample input features. This out-of-distribution challenge only becomes apparent in the online evaluation of the hybrid models and is difficult to reveal with just offline evaluation.

Conversely, when we directly enforce the microphysical constraints on the liquid-ice partition and constrain the existence of deep penetrating overshoots, the zonal mean cloud distribution appears more consistent with the pure physics simulation (Figure 4.6b and e vs. a and d). With these constraints, the downstream RMSE stabilized after an initial rise within the first two months. The global temperature RMSE remained around 2K, and the global water vapor RMSE stayed below 1 g/kg (Figure 4.5a and b). The global zonal wind RMSE remains around 5 m/s, and the global total cloud RMSE remains within 20 mg/kg except for one of the checkpoints showing an RMSE of 50 mg/kg in summer months.

As a substantial advancement, the U-Net-expanded-constrained models begin to match the monthly global mean climatology as well as the global variability in the reference E3SM-MMF simulations (Figures B.4 and B.5). Although the online RMSE of the hybrid simulations is generally slightly higher than the internal unpredictability (roughly within 1K temperature RMSE and 0.5 g/kg water vapor RMSE), in some months, the RMSE of the hybrid simulations overlaps with the internal unpredictability levels (Figure 4.5), suggesting the systematic bias of the U-Net-expanded-constrained simulations becomes comparable to the variability of the monthly climatology.

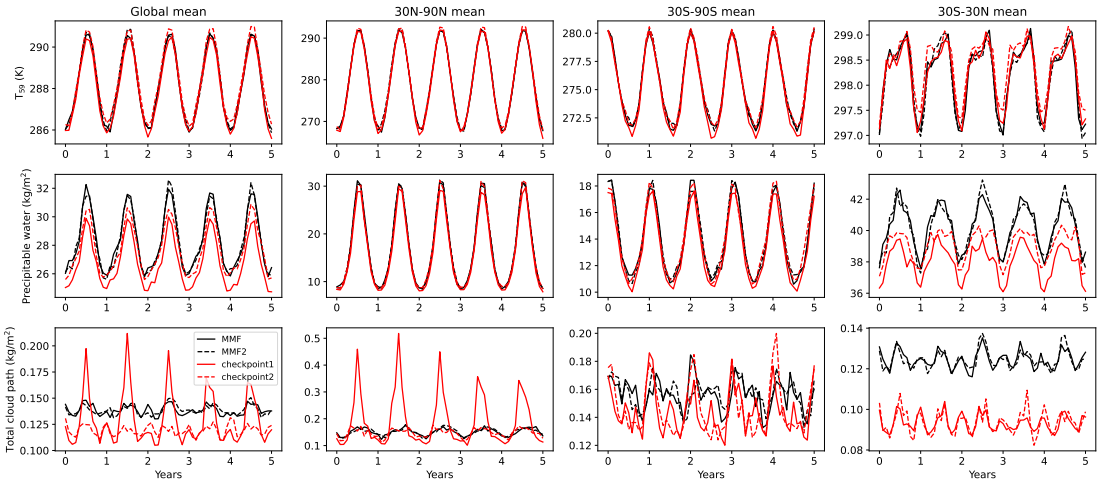


Figure 4.7: Five-year time series of area-weighted mean values for near-surface temperature (at the lowest model level, top row), precipitable water (middle row), and total cloud path (sum of liquid and ice, bottom row). The data is shown across four regions: global mean (first column), 30°N–90°N (second column), 30°S–90°S (third column), and 30°S–30°N (fourth column). The solid black line and dashed black line represent two reference 5-year E3SM-MMF simulations initialized with the same initial conditions to sample internal variability (similar to the dashed lines in Figure 4.5). The solid and dashed red lines represent hybrid simulations using two U-Net-expanded-constrained checkpoints: Huber loss with reduce-on-plateau scheduler (solid red) and Huber loss with step scheduler (dashed red) (see B.2).

4.5.2 Climatology Evaluation of Stable 5-year Hybrid Simulations

The U-Net-expanded-contrained simulations can be stably integrated for at least 5 years (Figure 4.7). Due to the maximum job runtime of 24 hours on the computing cluster we used, we only ran the hybrid simulations for five years plus one month and evaluated the 5-year mean climatology after dropping the first month for the spin-up period. Restart capabilities are not yet engineered and are nontrivial to implement given that the ML model collects a non-standard history from multiple previous time steps. However, we expect our hybrid simulations can integrate longer when there is no maximum job runtime limit.

During the 5-year simulations, the global mean near-surface air temperature, precipitable water, and total cloud (liquid plus ice) path reasonably reproduce the seasonal cycle of the reference E3SM-MMF simulation without any significant drift (Figure 4.7). Two hybrid simulations using

different U-Net-expanded-constrained checkpoints are shown to distinguish between systematic and checkpoint-dependent errors. The reference physics simulation was run twice for five years using identical initial conditions (black solid and dashed lines) to sample internal unpredictability and represent what a perfect emulation can achieve. Figure B.5 further demonstrates that the U-Net-expanded-constrained simulation generally captures the seasonal variation of spatial variability of near-surface air temperature and precipitable water seen in the reference E3SM-MMF simulation. However, during the northern hemisphere summer, the hybrid simulation exhibits slightly higher variance in precipitable water for the region spanning 40°N–50°N compared to the reference E3SM-MMF simulation.

Systematic biases still remain. The hybrid U-Net-expanded-constrained simulations displayed systematic underestimations of precipitable water and total cloud path, primarily in the tropics (Figure 4.7). One checkpoint exhibited a substantial overestimation of the total cloud path in the northern hemisphere (solid red line), while the other did not (dashed red line).

Figure 4.8 summarizes the RMSE of the 5-year time-mean climatology for state variables across vertical levels, along with the 3D global mean values (noted in the legend). Both U-Net-expanded-constrained checkpoints achieved a time-mean temperature RMSE below 2K and water vapor RMSE below 1 g/kg throughout the troposphere. The global mean RMSE for temperature and water vapor was 0.99K and 0.33 g/kg for one checkpoint, and 1.21K and 0.25 g/kg for the other. The checkpoint with better cloud condensate performance (dashed red line) exhibited a peak liquid cloud RMSE of 10 mg/kg near 800 hPa and an ice cloud RMSE of 4 mg/kg near 200 hPa. RMSE values for the zonal and meridional winds remained within 2 m/s and 1 m/s, respectively, across most levels, with exceptions at the uppermost level.

Figure 4.9 shows the zonal mean bias of the 5-year hybrid simulation climatology, with a tropospheric zonal mean temperature bias below 2K and a water vapor bias below 1 g/kg. Figure B.6 shows that the zonal mean relative humidity bias is generally within 4% across most of the tro-

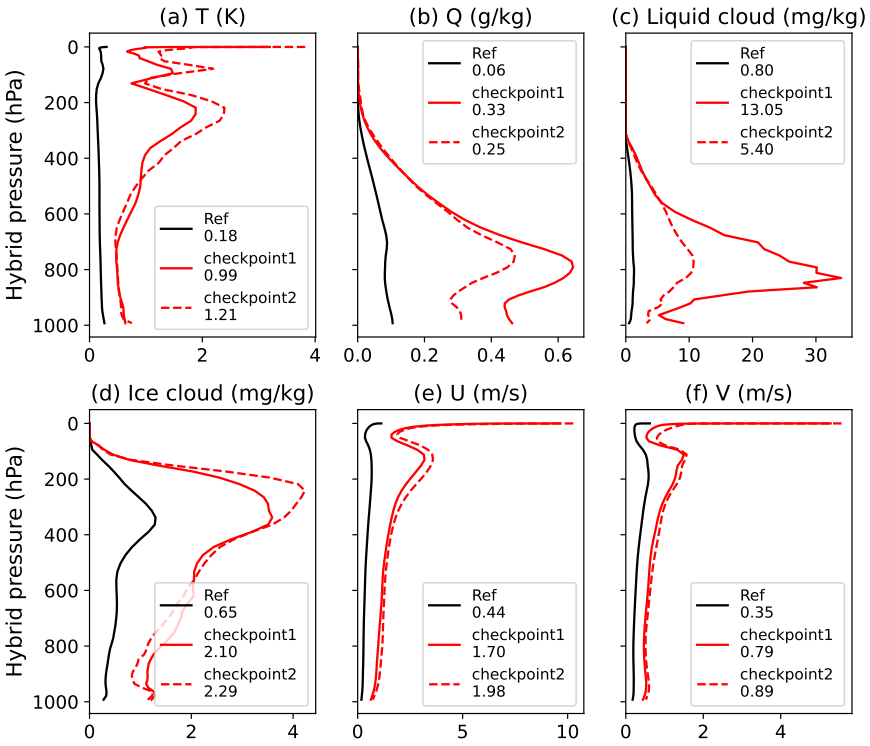


Figure 4.8: Online RMSE of the 5-year time-mean climatology for temperature (a), water vapor (b), liquid cloud mixing ratio (c), ice cloud mixing ratio (d), zonal wind (e), and meridional wind (f). Profiles represent RMSE values per vertical level, while the global RMSE values (integrated across horizontal and vertical space) are noted in the legend. RMSE is defined as in B.3. The black line shows the RMSE between two reference E3SM-MMF simulations with identical initial conditions, serving as a proxy for a perfect emulator (diverging only due to non-bit-for-bit numerical differences). The solid and dashed red lines represent the online RMSE of hybrid simulations using two U-Net-expanded-constrained checkpoints: Huber loss with a reduce-on-plateau scheduler (solid red) and Huber loss with a step scheduler (dashed red) (see B.2).

posphere, with a peak bias of 8% in the tropical regions. However, larger relative humidity biases are observed in extremely cold environments, such as near the surface at the South Pole and in the stratosphere. Additionally, we control the global zonal mean zonal wind bias within 5 m/s and maintain a reasonable distribution of liquid and ice clouds. While we have a near 40 g/kg cloud water bias near 60N with this model checkpoint, another different checkpoint has only within 10 g/kg cloud water bias in most places (Figure B.2 in B.5).

Our performance surpasses the online zonal mean temperature and water vapor biases reported by Han et al. (2023). Han et al. (2023) showed greater than 10K temperature bias and near 1.5 g/kg water vapor bias in the polar regions in their hybrid simulations (see Figure S9 in their Supplementary Information). Han et al. (2023) did not provide the online bias of their wind and cloud climatology so we cannot make a comparison on these variables.

Figure 4.10 presents the five-year averaged, zonal mean subgrid tendencies of temperature, water vapor, zonal wind, and total cloud mixing ratio for both the reference E3SM-MMF simulation and the hybrid run with the constrained U-Net model. The hybrid simulation effectively reproduces the patterns of these tendencies. The temperature and water vapor tendencies in the reference simulation are well captured by the hybrid simulation, showing similar vertical and meridional distributions, such as the two heating and drying bands in the tropics corresponding to regions with frequent deep convection and precipitation. Additionally, the hybrid simulation successfully replicates the zonal wind and total cloud mixing ratio tendencies, demonstrating its ability to correctly capture the large-scale dynamics in the E3SM-MMF physical simulation. However, some discrepancies exist between the hybrid simulation and the reference E3SM-MMF simulation. For instance, the hybrid simulation underestimates the heating and drying tendency band near 15N, underestimates the moistening tendency in the tropical lower troposphere, and produces anomalous heating in the polar stratosphere near 200 hPa.

The online precipitation statistics in the 5-year hybrid simulation reasonably reproduced the

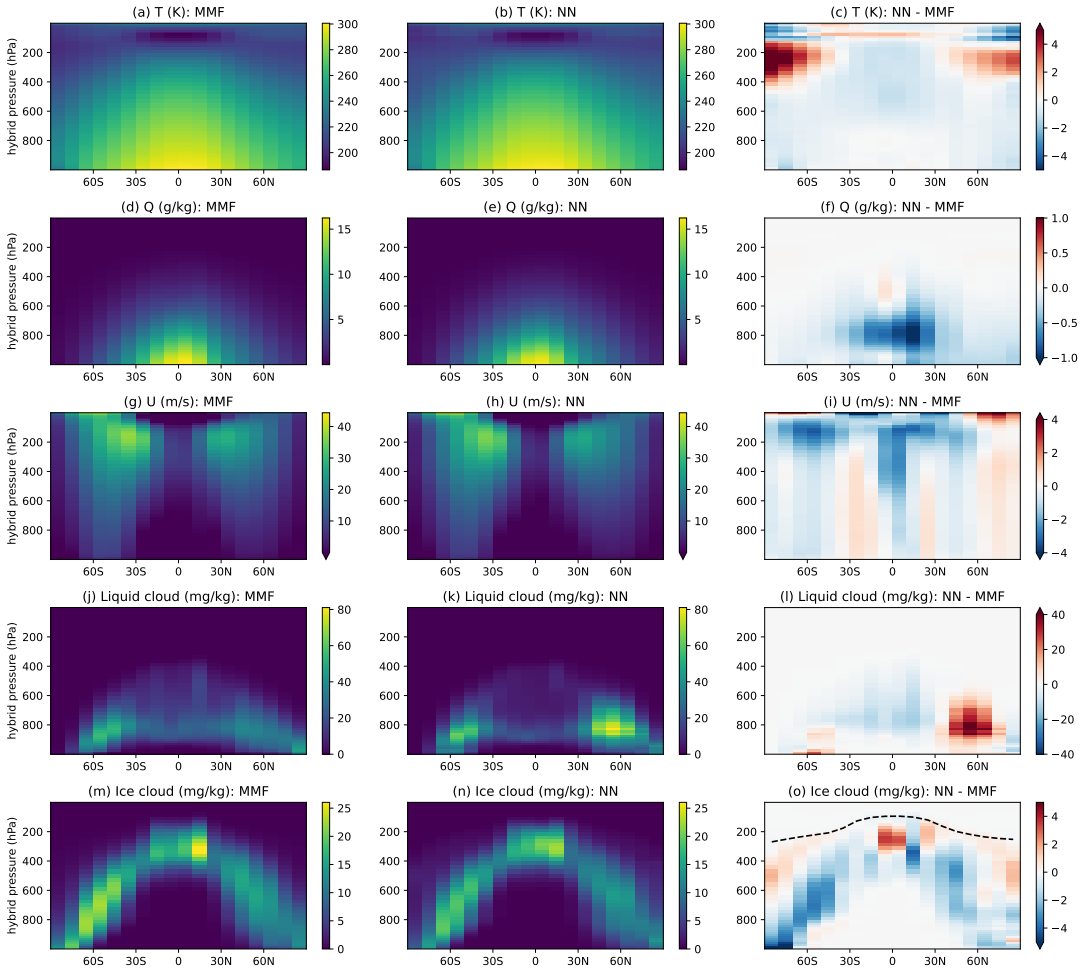


Figure 4.9: Five-year zonal mean atmospheric state in the reference E3SM-MMF simulation (left) and in the hybrid simulation with the U-Net-expanded-constrained model (middle). The right column shows the zonal mean bias as the mean state from the hybrid simulation minus that of the reference simulation. The five rows show temperature, water vapor, zonal wind, liquid cloud, and ice cloud. The Black dashed line in panel o indicates the 5-year mean position of the tropopause level used in the microphysical constraints.

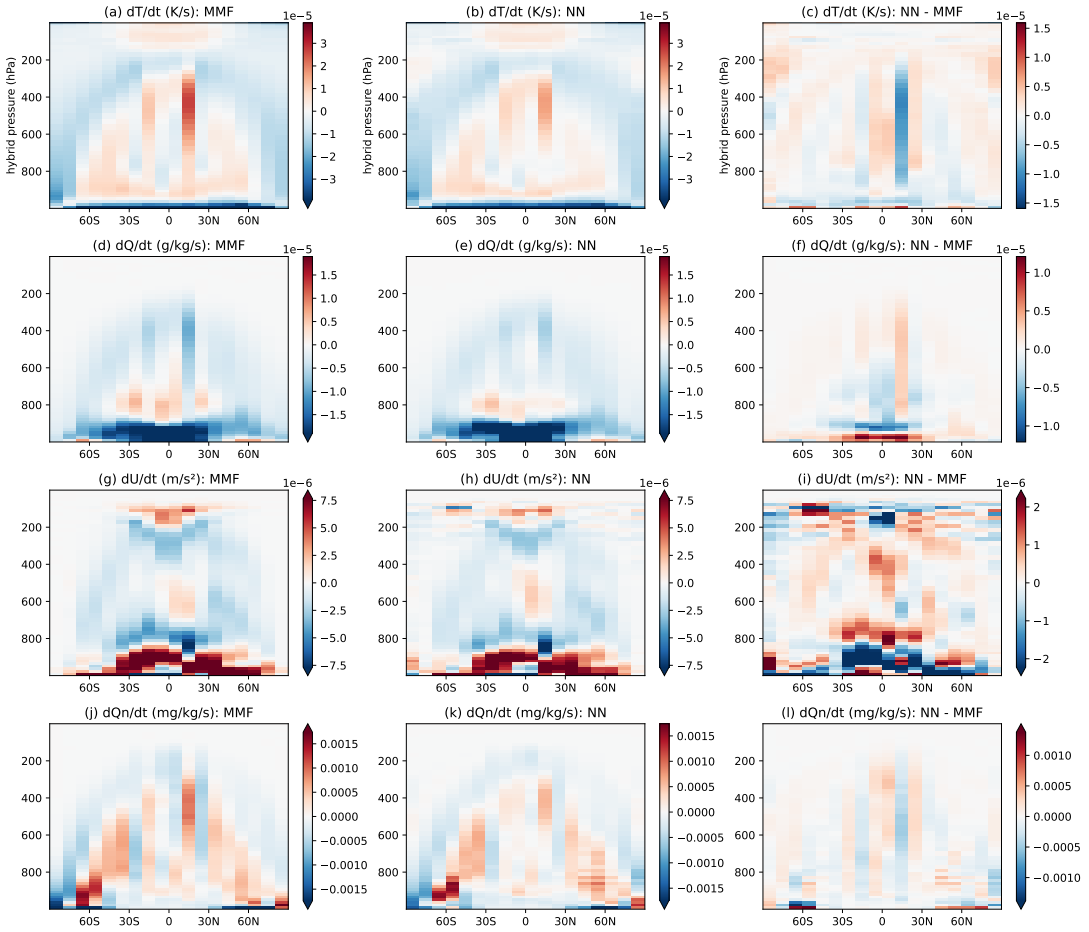


Figure 4.10: Five-year zonal mean atmospheric tendencies in the reference E3SM-MMF simulation (left column), in the hybrid online simulation with the U-Net-expanded-constrained model (middle column), and the difference (hybrid minus MMF run, right column). The four columns correspond to the tendencies of temperature, water vapor, zonal wind, and total cloud (liquid plus ice) mixing ratio.

reference E3SM-MMF simulation (Figure 4.11). The zonal mean, 5-year average precipitation distribution in the hybrid simulation matches that of the reference E3SM-MMF simulation (Figure 4.11a). However, the precipitation peak in the northern hemisphere is underestimated, primarily over the ocean. Figure 4.12 presents the horizontal map of the 5-year mean precipitation pattern. The hybrid simulation captures the spatial distribution of global precipitation reasonably well but underestimates the rain belt on the northern side of the equator. Quantitatively, the hybrid simulation exhibits an RMSE of 0.96 mm/day for the 5-year mean precipitation.

The hybrid run also reasonably reproduces the extreme tail of the PDF of the instantaneous precipitation sampled every hour (Figure 4.11b). The frequency in the hybrid run aligns well with the reference simulation down to a frequency of 10^{-6}, demonstrating similar skills that match previous studies (Rasp et al., 2018; Yuval & O'Gorman, 2020; Wang et al., 2022b). However, beyond the frequency of 10^{-6}, the hybrid simulations failed to reproduce the most extreme and rarest of precipitation events.

4.6 Conclusions

The Multi-scale Modeling Framework (MMF) offers a promising solution to better represent subgrid processes that are poorly parameterized in traditional climate models. By integrating machine learning (ML) to emulate the embedded cloud-resolving models, we can potentially reduce computational costs substantially while maintaining high fidelity in the resulting hybrid climate simulations. Compared to other hybrid approaches, the MMF framework provides the convenience of locality due to its scale separation and retains all necessary components (e.g., ocean and land modules, aerosol module) for future climate change projections.

Here, we used the coarse-resolution version (with an effective resolution of $11.5° \times 11.5°$) of the ClimSim data (Yu et al., 2024a) to train a stable ML parameterization that emulates the full physics

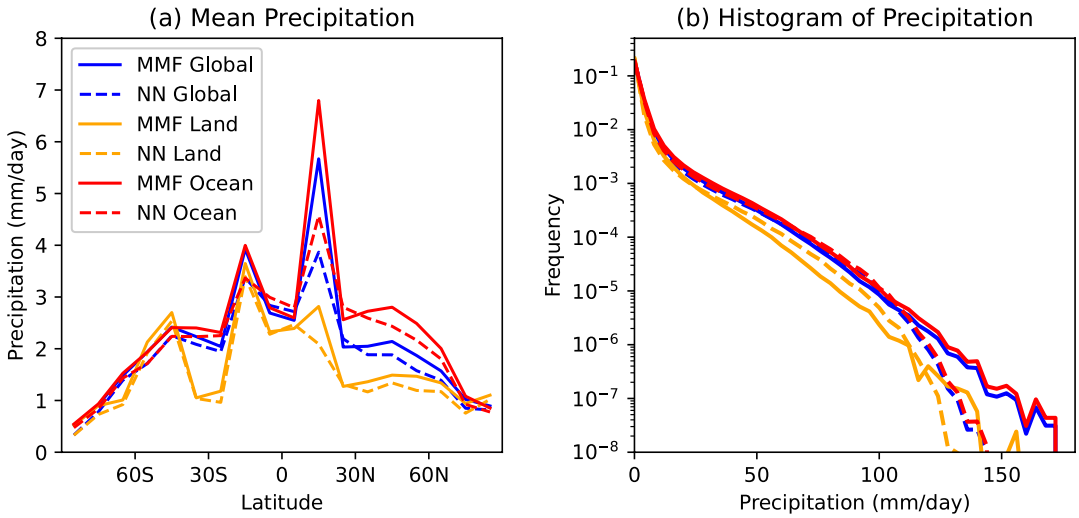


Figure 4.11: (a) 5-year zonal mean precipitation in the reference E3SM-MMF simulation (solid lines) versus the online hybrid simulation with the U-Net model incorporating microphysical constraints (dashed lines). Different colors denote samples taken globally (blue), only over land (orange), and only over the ocean (red). (b) Histogram of the hourly instantaneous precipitation.

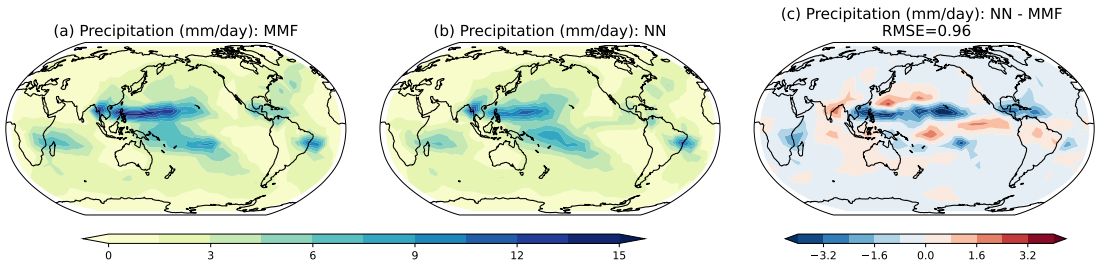


Figure 4.12: Horizontal maps of the five-year mean surface precipitation: (a) reference E3SM-MMF simulation, (b) hybrid U-Net-expanded-constrained simulation, and (c) bias (hybrid minus reference). Panel c also documents the RMSE of 0.96 mm/day for the five-year mean precipitation between the hybrid and reference simulations.

(convection plus radiation) of the cloud-resolving model in the E3SM-MMF climate model. In the hybrid MMF simulation studies, to our knowledge, this is the first work to achieve a multi-year stable ML parameterization that emulates the full set of variables required to couple with the dynamical core and surface models with geography. Our hybrid simulation achieved state-of-the-art online 5-year zonal-mean biases in temperature and water vapor climatology compared to previous hybrid MMF studies with similar complexities (Han et al., 2023; Behrens et al., 2024). Our hybrid simulation also achieved realistic multi-year cloud climatology with explicit cloud condensate coupling.

We utilized an improved U-Net architecture with expanded input features, such as large-scale forcings and convective memory, which significantly improved offline fitting skills compared to the baseline multi-layer perceptron architecture used in many previous studies. While offline skill is not completely representative of online performance, we found that incorporating microphysical constraints in the ML prediction significantly enhanced the online performance of the hybrid simulations. These constraints were informed by the initial analysis of hybrid errors in the simulated condensate. We found that they help maintain a reasonable liquid and ice cloud distribution in the hybrid simulation. They also significantly benefited other state variables such as temperature and water vapor bias through associated dynamical couplings. Our microphysics-constrained U-Net model also reasonably reproduced the spatial distribution of five-year mean precipitation and the frequency of extreme precipitation compared to the reference E3SM-MMF simulation, with a root mean square error of 0.96 mm/day for five-year mean precipitation.

Despite these advancements, our multi-year online performance reveals systematic biases that remain to be addressed. For example, we observe a persistent low-bias in tropical lower troposphere water vapor and liquid cloud mixing ratio, as well as a vertical dipole pattern in temperature biases within the polar stratosphere. Additionally, the hybrid simulation underestimates precipitation rates over the northern flank of the equatorial rain belt. Addressing these biases presents significant

challenges, underscoring the need for innovative approaches to optimizing online performance and improving the fidelity of hybrid climate simulations.

Our work demonstrates the potential of achieving stable and skillful ML parameterization for operational-level hybrid climate simulations. We illustrate the importance and potential ways of incorporating domain knowledge and physical constraints into the ML parameterization to improve the online bias and stability of hybrid physics-ML climate simulations. We prove that including explicit cloud condensate coupling in our ML parameterization is achievable while maintaining satisfactory multi-year climatology, which would be crucial to further increase the complexity of the ML parameterization such as including pollutant transport and cloud-aerosol interaction.

4.7 Discussion and Limitations

4.7.1 Complexity of the Reference and Hybrid Simulations

Our online results establish a robust baseline that approaches operational-level complexity, demonstrating substantial progress, but also highlighting areas needing further development. Looking forward, additional complexity is required in the ML parameterization and the training data to achieve complete compatibility for the purpose of climate projection. For example, our online simulations used prescribed sea surface temperature and sea ice. To simulate future climate change, it will be necessary to train and evaluate the hybrid simulations with a fully coupled ocean module that allows for interactions between the atmosphere, ocean, and sea ice. Aerosol-cloud interaction is also a critical process for climate projection but is not included in the simulations that generate the ClimSim data. To develop ML parameterization that includes aerosol-cloud interaction, one can create a new training dataset that includes aerosol-related inputs and outputs. This can be done using simulations that implement the Explicit-Cloud Parameterized-Pollutant (ECPP) scheme (Gustafson et al., 2008).

For computational expedience, our ML parameterization is trained on data from the coarse-resolution E3SM-MMF simulations in the ClimSim dataset (Yu et al., 2024a), with an effective resolution of 11.5°×11.5°, and tested in a similarly coarse prognostic setting. On the one hand, this serves as a simple, cost-efficient starting point to rapidly iterate on different ideas to improve offline and online performance. However, this resolution poorly simulates tropical precipitation, synoptic scale flows, fronts, and large-scale convective systems like Madden-Julian Oscillation (MJO). The coarse resolution introduces challenges for the parameterization problem, as it requires the parameterization to account for not only unresolved convection and clouds but also a substantial portion of the large-scale circulation typically resolved in an atmospheric model. However, it may also simplify the problem by avoiding the "gray zone" issues encountered in models with resolutions of 50–100 km, where some gravity waves, fronts, and large-scale convective systems are partially resolved but poorly simulated.

Future work should thus extend this study to the most operationally relevant, highest-resolution data in the ClimSim dataset (effective resolution of 1.5°×1.5°), which better resolves synoptic-scale flows and contains more detailed topography. However, higher-resolution data may shift input distributions significantly, including changes in mean atmospheric states, large-scale advective forcings, and the emergence of previously unencountered extrema, such as newfound peak topography and new land surface properties. These differences will likely require retraining the ML parameterizations to handle the altered data characteristics and ensure reliable performance in the high-resolution setting.

4.7.2 Improving the Offline and Online Performance

Our results are consistent with the view of Lin et al. (2024) that pursuing higher offline skills can lead to improved online performance in hybrid simulations, suggesting exploring advanced architectures might be beneficial. MLPs with deeper architectures (10+ layers) and residual connections

(Kochkov et al., 2024; Wang et al., 2022b) may offer greater expressive power compared to the MLP model evaluated in this study and may exhibit different online performance. Transformer models or Long Short-Term Memory (LSTM) models might effectively learn vertical structure of the input features. A Recurrent Neural Network (RNN)-style model in the time dimension, like a state-space model or LSTM, could better represent the convection memory.

In addition to exploring advanced architectures, more careful normalization in the upper troposphere and the stratosphere or location-dependent normalization might further improve offline skills. While these advancements hold promise in general, our current U-Net model and normalization strategies continue to face significant challenges in specific areas. It struggles to predict stratospheric cloud and water vapor tendencies as well as wind tendencies in the polar regions.

For further insights into improving offline performance, readers are encouraged to review the solutions from a recent Kaggle competition, "LEAP-Atmospheric Physics AI (ClimSim)" (`https://www.kaggle.com/competitions/leap-atmospheric-physics-ai-climsim`). The participants worked on the same ClimSim dataset and explored various innovative architectures, data preprocessing techniques, and training strategies to optimize the offline skill of their models. For example, we speculate that the low offline R^2 of water vapor, ice cloud, and wind tendencies in the tropopause and stratosphere (Figure 4.3) could be improved by better normalization and architectures.

It could be viewed as surprising that the expanded input features (large-scale advective forcing plus previous steps' convection memory) did not improve the online performance (Figure 4.5) given that some previous studies, such as Lin et al. (2024), have suggested adding convection memory can improve stability and online performance. A potential explanation is that a larger sample size of architectures may be necessary to robustly assess the impact of these expanded input features. We encourage future studies to further test and ablate these features to better understand their effects on model performance.

Exploring the pathologies of bias and instability in hybrid climate simulations is highly beneficial.

In our study, the observation of unrealistic cloud formation was crucial in motivating our implementation of microphysical constraints. These unrealistic clouds (clearly out-of-sample simulation states) highlight a broader challenge in hybrid physics-ML simulations: ensuring that ML-generated physics do not drive the simulation into non-physical states. Addressing this issue is difficult because training data sourced solely from physically accurate numerical simulations may not prepare the ML model to correct such errors. Our work demonstrates that addressing such out-of-sample issues can benefit from in-depth error diagnosis and constraints informed by domain knowledge to prevent un-physical situations from happening in the first place.

Our hybrid simulations exhibit certain bias patterns, such as the vertical dipole temperature bias in the polar stratosphere (Figure 4.9c), a drying water vapor bias, and a liquid cloud low bias in the tropical lower troposphere (Figure 4.9f and 4.9l). These bias patterns also exist in the hybrid simulations across all the checkpoints (e.g., Figure B.2) of the U-Net model with microphysical constraints. B.6 shows that the tropical lower troposphere bias patterns of too little water vapor and liquid cloud develop very quickly within the first few days and maintain their magnitude for years.

Some other bias patterns vary with different checkpoints. The constrained U-Net model checkpoint used to create Figure 4.9 corresponds to the red circle line in Figure 4.5, which exhibits slightly better temperature and water vapor zonal mean biases but worse biases for online cloud mixing ratio than other U-Net checkpoints (Figure 4.5d). The positive bias of the lower troposphere liquid cloud (Figure 4.9l) can be significantly improved with other checkpoints (Figure B.2). A more extensive hyperparameter search could potentially further optimize some of the observed online bias.

Further optimizing the online bias of our constrained U-Net model can be challenging once an optimal architecture is decided upon. Unlike offline errors, online errors are non-differentiable and cannot be directly optimized using gradient descent. Once the model architecture is fixed, we currently rely on checkpoint searches (e.g., by perturbing training hyperparameters) to identify the checkpoint that yields the best online performance.

When evaluating online error growth, we used a fixed initial condition for all hybrid and reference simulations. In future work, it would be valuable to better isolate the development of online biases from the sensitivity of reference simulations to initial conditions. For instance, conducting an ensemble of simulations with perturbed initial conditions could provide a clearer measure of the system's chaotic behavior. Additionally, using different initial conditions across various times of the year could improve the robustness of the evaluation. Analyzing the bias development on daily or even shorter timescales could also offer insights into the unsteady onset of these biases.

It remains an open question how to use gradient-free methods to optimize these online errors. Gradient-free methods such as imitation learning and Ensemble Kalman Inversion could be promising, as they have shown success in other applications (Ross et al., 2011; Rasp, 2020; Kelp et al., 2022; Lopez-Gomez et al., 2022; Pahlavan et al., 2024; Christopoulos et al., 2024). Another possible yet highly non-trivial approach is to develop a fully differentiable climate model, as demonstrated for the atmospheric primitive equation solver subcomponent by (Kochkov et al., 2024), although this is a highly nontrivial effort for a fully featured climate model compatible for future climate projection. With a model that can be integrated over time in a differentiable way, it would be possible to optimize multi-step losses using gradient descent. Some of the online error patterns develop very quickly within the first day of the hybrid simulation, and optimizing multi-step losses could potentially help address these error patterns.

There is room for adjustment and improvement in the microphysical constraints we used. The temperature-based liquid-ice partition constraint is not directly applicable to reconfigurations of the CRM that use more advanced microphysics schemes of the sort required for fully explicit aerosol-cloud interaction science. However, some soft constraints can still be applied, such as prohibiting liquid clouds below a certain temperature and similarly prohibiting ice clouds above a warm temperature threshold. The constraint of removing all clouds in the stratosphere eliminates deep penetrating clouds, which are infrequent but not entirely absent. A more careful, data-driven approach

could be used to define the tropopause, requiring fewer approximations and better capturing the rare but non-zero deep convective overshoots. Looking ahead, now that this baseline problem that intentionally used simplistic microphysics is proving tractable, extensions to more modern microphysical representations would be timely.

5

Conclusion

This thesis addresses critical challenges in representing small-scale processes, such as convection and cloud formation, in climate models. These processes are essential for accurately simulating large-scale atmospheric behavior, yet computational constraints and unresolved physical complexities continue to limit their representation. Through a combination of idealized modeling, numerical simulations, and machine learning (ML) approaches, this work advances our understanding of how small-scale processes influence atmospheric states and explores innovative methods to improve their

representation in climate models.

In Chapter 2, we investigated how small-scale processes, particularly plume-environment mixing, influence large-scale atmospheric states, such as upper-tropospheric convective mass flux and anvil cloud fraction. Using a refined zero-buoyancy plume (ZBP) model, we demonstrated that increased mixing strength enhances upper-tropospheric mass flux through two pathways: (1) increased cloud evaporation cooling in the environment and (2) reduced atmospheric stability. These findings were validated using high-resolution radiative-convective equilibrium (RCE) simulations, which revealed that the dominant pathway depends on the relative magnitudes of evaporation cooling and radiative cooling in the environment. This work highlights the importance of small-scale mixing in shaping large-scale atmospheric properties and provides a framework for understanding resolution-dependent behaviors in convection-permitting models. The refined ZBP model, which resolves self-consistency issues in earlier formulations, offers a valuable tool for exploring the interactions between small-scale processes and large-scale atmospheric states.

Chapter 3 extended this exploration by examining the sensitivity of tropical tropopause layer (TTL) heat balance to microphysics parameterizations. When swapping different microphysics schemes in RCE simulations, we identified two distinct regimes: a ”hard-landing” scenario, where frequent and strong convective overshoots induce significant cooling in the TTL, and a ”soft-landing” scenario, where weaker overshoots result in minimal cooling. These regimes are driven by differences in upper-tropospheric stratification, which are modulated by atmospheric cloud radiative effects (ACRE). Our results underscore the critical role of microphysics schemes in determining convective behavior and TTL heat balance, highlighting the need for accurate representation of high cloud coverage and cloud-radiation interactions in climate models.

In Chapter 4, we demonstrated the potential of machine learning (ML) to emulate small-scale processes in climate models efficiently. By training ML models on high-resolution data from the Multi-scale Modeling Framework (MMF) climate simulations, we developed a stable and skillful

ML parameterization that replaces the embedded cloud-resolving models in the MMF. Our hybrid ML-physics simulations achieved multi-year stability and realistic cloud climatology, even with explicit cloud condensate coupling—a significant advancement over previous efforts. The integration of microphysical constraints into the ML architecture proved crucial for maintaining online stability and reducing biases in temperature, water vapor, and cloud condensate. Despite these successes, systematic biases in tropical water vapor and polar stratospheric temperature remain, pointing to areas for future improvement. This work represents a major step toward operationalizing hybrid ML-physics climate models, offering a pathway to reduce computational costs while preserving simulation fidelity.

A

Supplemental Material for Chapter 2

A.1 Refined solutions of a zero-buoyancy plume model

Here we document the details of how we solve the zero-buoyancy plume model to get self-consistent solutions about steady-state mass flux, detrainment rate, and other atmosphere profiles. Following

Singh & Neogi (2022), the equations we solve are:

$$\frac{\partial M_c}{\partial z} = e - d \tag{A.1}$$

$$M_c + M_e = 0 \tag{A.2}$$

$$\frac{\partial (M_c q^*)}{\partial z} = eq - dq^* - s_{cond} \tag{A.3}$$

$$\frac{\partial (M_e q)}{\partial z} = dq^* - eq + s_{evap} \tag{A.4}$$

$$\frac{\partial (M_c h^*)}{\partial z} = eh - dh^* \tag{A.5}$$

$$\frac{\partial (M_e h)}{\partial z} = dh^* - eh + Q_{rad} \tag{A.6}$$

$$h^* - h = L_v(q^* - q) \tag{A.7}$$

$$s_{evap} = \mu d(q^* - q) \tag{A.8}$$

$$\frac{\partial p}{\partial z} = -\frac{pg}{R_d T} \tag{A.9}$$

Equation A.1 and A.2 are mass conservation equations. M_c is the mass flux of convective updrafts, and M_e is mass flux in the environment. We assume there is no large-scale advection, so the net mass flux in updrafts and the environment is 0. e is mass entrainment, and d is mass detrainment. Fractional entrainment rate ε and fractional detrainment rate δ are defined as:

$$\varepsilon = e/M_c \tag{A.10}$$

$$\delta = d/M_c \tag{A.11}$$

Equation A.3 and A.4 describe the water vapor conservation in the updraft plume and the environment separately. q is the specific humidity in the environment. q^* is the saturation specific

humidity in the updraft plume, which is simply a function of temperature and pressure:

$$q^* = 0.622 p_{\mathrm{v}}^* / p = 0.622 \frac{p_0}{p} e^{-\frac{L_{\mathrm{v}}}{R_{\mathrm{v}} T}} \tag{A.12}$$

where $p_{\mathrm{v}}^* = p_0 exp(-L_{\mathrm{v}}/(R_{\mathrm{v}} T))$ is the saturation vapor pressure, L_{v}=2.51e6 J kg^{-1} is the latent heat of condensation, R_{v}= 461 J $kg^{-1} K^{-1}$ is gas constant for water vapor, p_0=2.69e11 Pa is a constant.

Equation A.5 and A.6 describe the conservation of moist static energy in the updraft plume and the environment. $h = c_p T + gz + L_{\mathrm{v}} q$ is the moist static energy in the environment, and $h^* = c_p T + gz + L_{\mathrm{v}} q^*$ is the saturation moist static energy in the updraft plume. We specify radiative heating rate to be simply a function of temperature,

$$Q_{\mathrm{rad}}/(c_p \rho) = \begin{cases} Q_0, & \text{if } T > 250K \\ Q_0(0.5 + 0.5 cos(\pi(250 - T)/(250 - 200)), & \text{if } 250K > T > 200K \\ 0, & \text{if } T < 200K \end{cases} \tag{A.13}$$

where $Q_0 = -1\mathrm{K\,day}^{-1}$. Radiative heating rate is constantly $-1\mathrm{K\,day}^{-1}$ in the lower and middle troposphere and gradually decays to 0 from T=250K to T=200K. $\rho = p/R_{\mathrm{d}} T$ is the air density. It is worth noting that Equation A.13 ignores the positive radiative heating often observed in the tropopause layer (e.g., Fig. S1a) associated with deep convective overshoots.

Equation A.7 implies the zero-buoyancy assumption that the temperature in updrafts is the same as the temperature in the environment at the same height. Fig. S1b in the Supporting Information shows that the time-averaged buoyancy in the convective updraft in our cloud-resolving numerical simulations is within 1K except at the top of the troposphere above 14km. At higher altitudes, the Equation A.7 is not a good assumption as deep convective overshoots are infrequent and the stratification becomes gradually decoupled from deep convection.

Equation A.8 is the parameterization of cloud evaporation in the environment, following the definition in the Singh & Neogi (2022). μ is a unitless parameter that controls the speed of cloud evaporation. We assume cloud evaporation happens at the level where clouds are condensed, and we assume there is no evaporation of precipitation. Equation A.9 is the hydrostatic balance, and $R_\mathrm{d} = 287 \mathrm{J\,kg^{-1}K^{-1}}$ is the gas constant for dry air. Fig. S1c demonstrates that the domain-mean profiles in cloud-resolving numerical simulations follow the hydrostatic balance well.

For Equation A.1 to A.9, there are 9 equations but 11 unknown variables: M_c, M_e, ε, δ, q^*, q, h, s_cond, s_evap, μ, and p. We have excluded h^* and Q_rad from unknown variables since they can be expressed using T and p through Equation A.12 and A.13. We take ε and μ to be the free parameters that we can specify, and the rest of the equations is just enough to get a self-consistent solution. If one further specifies δ, then there will be more equations than unknown variables, in which case there cannot be the self-consistent solution. Next, we will describe how we solve these equations as an ODE problem and express the equations as $\frac{\partial}{\partial z}(M_\mathrm{c}, p, q, T) = F(M_\mathrm{c}, p, q, T)$.

Replacing Equation A.1 into Equation A.3 to A.7, we can get:

$$M_\mathrm{c}\frac{\partial q^*}{\partial z} = -\varepsilon(1-\mathcal{R})M_\mathrm{c}q^* - s_\mathrm{cond} \tag{A.14}$$

$$M_\mathrm{e}\frac{\partial q}{\partial z} = \delta(1+\mu)(1-\mathcal{R})M_\mathrm{c}q^* \tag{A.15}$$

$$M_\mathrm{c}\frac{\partial h^*}{\partial z} = -\varepsilon L_\mathrm{v}(1-\mathcal{R})M_\mathrm{c}q^* \tag{A.16}$$

$$M_\mathrm{e}\frac{\partial h}{\partial z} = \delta L_\mathrm{v}(1-\mathcal{R})M_\mathrm{c}q^* + Q_\mathrm{rad} \tag{A.17}$$

$\mathcal{R} = q/q^*$ is the relative humidity in the environment.

Equation A.1 can be rewritten as:

$$\frac{\partial M_\mathrm{c}}{\partial z} = M_\mathrm{c}(\varepsilon - \delta) \tag{A.18}$$

Using Equation A.2 and A.15, we get:

$$\frac{\partial q}{\partial z} = -\delta(1+\mu)(1-\mathcal{R})q^* \tag{A.19}$$

Equation A.16 can be used to express the temperature lapse rate $\Gamma = -\frac{\partial T}{\partial z}$. From the definition of h^*, we have:

$$\frac{\partial h^*}{\partial z} = -c_p\Gamma + g + L_\mathrm{v}\frac{\partial q^*}{\partial z} \tag{A.20}$$

Using Equation A.9 and A.12 and defining $\gamma = -(1/q^*)\frac{\partial q^*}{\partial z}$, we can get:

$$\gamma = \frac{L_\mathrm{v}\Gamma}{R_\mathrm{v}T^2} - \frac{g}{R_\mathrm{d}T} \tag{A.21}$$

Replacing Equation A.20 and A.21 into Equation A.16, we can get:

$$\frac{\partial T}{\partial z} = -\Gamma_m - \frac{\varepsilon L_\mathrm{v}(1-\mathcal{R})q^*}{c_p + q^*L_\mathrm{v}^2/(R_\mathrm{v}T^2)} \tag{A.22}$$

where

$$\Gamma_m = \frac{g(1+\frac{L_\mathrm{v}q^*}{R_\mathrm{d}T})}{c_p + q^*L_\mathrm{v}^2/(R_\mathrm{v}T^2)}$$

is the moist adiabatic lapse rate when the fractional entrainment rate $\varepsilon = 0$.

By replacing the Γ in Equation A.21 with Equation A.22, we can also rewrite γ as:

$$\gamma = \gamma_m + \varepsilon b(1-\mathcal{R}) \tag{A.23}$$

where

$$\gamma_m = \frac{L_\mathrm{v}\Gamma_m}{R_\mathrm{v}T^2} - \frac{g}{R_\mathrm{d}T}$$

is the fractional saturation specific humidity gradient along a moist adiabat, and

$$b = \frac{q^* L_\mathrm{v}^2}{q^* L_\mathrm{v}^2 + c_p R_\mathrm{v} T^2}$$

is a nondimensional residual term.

When we sum Equation A.3 and A.4, sum Equation A.5 and A.6, and use Equation A.2 and A.7, we can get the energy balance equation:

$$Q_\mathrm{rad} = -L_\mathrm{v}(s_\mathrm{cond} - s_\mathrm{evap}) \tag{A.24}$$

Replacing Equation A.8 and A.14 into Equation A.24, we can get the expression of δ:

$$\delta = -\frac{\varepsilon(1-b)}{\mu} + \frac{\gamma_m}{\mu(1-\mathcal{R})} + \frac{Q_\mathrm{rad}}{\mu(1-\mathcal{R})q^* L_\mathrm{v} M_\mathrm{c}} \tag{A.25}$$

Now with Equation A.9, A.18, A.19, A.22, and A.25, we have the closed form expression for our ODE problem:

$$\frac{\partial}{\partial z}(M_\mathrm{c}, p, q, T) = F(M_\mathrm{c}, p, q, T) \tag{A.26}$$

where the right-hand side only depends on M_c, p, q, and T. Given boundary conditions at cloud base (we use z=500m), Equation A.26 can integrate upwards and get the full atmosphere profiles. It is worth noting that equation A.25 is singular if $\mu = 0$. We will discuss the solution for that case later in this section.

For boundary conditions, we specify a surface temperature of 303K and surface pressure of 10^5 Pa. We assume a dry adiabatic lapse rate of $g/c_p = 9.8Kkm^{-1}$ below the cloud base, and we can use Equation A.9 to integrate pressure p from the surface to the cloud base. For environmental specific humidity q at the cloud base, we do not have a solid constraint. If one assumes $\frac{\partial q}{\partial z} \approx \mathcal{R}\frac{\partial q^*}{\partial z}$ (the vertical gradient of RH is much smaller than the vertical gradient of q^*), Equation A.15 can reduce to:

$$\mathcal{R} = \frac{\delta(1+\mu)}{\delta(1+\mu)+\gamma} \tag{A.27}$$

We determine our cloud base q using Equation A.27, and the value of δ in Equation A.27 is taken from the ε. In this way, we implicitly assume that increasing ε or μ can have a moistening effect at the cloud base, which intuitively makes sense. We will show later that the sensitivity of the solution to the value of cloud base q is relatively small above 5km except for absolute temperature.

For M_c, we do not have a direct cloud base constraint. However, we assume our solution is in radiative-convective equilibrium (RCE), which says radiative cooling must be balanced by latent heat release at all levels. The RCE condition requires that M_c reaches 0 exactly at the level where the radiative cooling rate becomes 0, i.e., at T=200K (Equation A.13). If cloud base M_c is too large, M_c will still be positive where T=200K. If cloud base M_c is too small, M_c will go to 0 before radiative cooling decays to 0. We can have a random initial guess of cloud base M_c and change our guess based on this RCE condition. If this guess makes the M_c solution never go to 0 or go to 0 at higher and colder altitudes than T=200K, we need to reduce the guess. On the contrary, if the M_c solution goes to 0 at lower and warmer altitudes than T=200K, then we need to increase the guess. A solution is also not valid if the detrainment rate δ becomes negative during the upward integration. When this occurs, we examine the last level where δ remains positive. If $\delta > \varepsilon$ at that level, it indicates that M_c is decaying too rapidly, resulting in a M_c value too small, causing the last term in

Equation A.25 to make δ turn negative. In this case, the initial guess for M_c is likely too small and should be increased. Conversely, if $\delta < \varepsilon$ at that level, we choose to reduce our guess. Although this last criterion does not guarantee an upper bound, it works across a range of parameter regimes to find valid solutions in practice. Further investigation is warranted to establish a more robust criterion for the upper and lower bound. Once we find lower and upper bounds of the cloud base M_c, we use binary search to iteratively guess between the bounds and narrow the bounds until we find the M_c that satisfies the RCE condition.

In the case where $\mu = 0$ (i.e., no condensate evaporation and $s_{\text{evap}} = 0$ given the Equation A.8), the Equation A.25 becomes singular. We need to solve the equations in a slightly different way. By subtracting Equation A.14 multiplied by L_v from Equation A.16 and then using Equation A.24, we can express M_c as a function of p, q, and T:

$$M_c(p, q, T) = -Q_{\text{rad}} / \frac{\partial s}{\partial z} \tag{A.28}$$

where $s = c_p T + gz$ is the dry static energy in the environment, and $\partial s / \partial z = -c_p \Gamma + g$ is known from Equation A.22. This expression for M_c automatically guarantees the RCE condition at the top boundary (M_c reaches 0 at the same level where radiative cooling becomes 0).

Using Equation A.18, we can get an expression for δ:

$$\delta = \varepsilon - \frac{1}{M_c} \frac{\partial M_c}{\partial z} \tag{A.29}$$

Now with Equation A.9, A.19, A.22, we have come up with a slightly different ODE problem:

$$\frac{\partial}{\partial z}(p, q, T) = G(p, q, T, \delta) \tag{A.30}$$

where the δ on the right-hand side depends on the left-hand side through Equation A.28 and A.29.

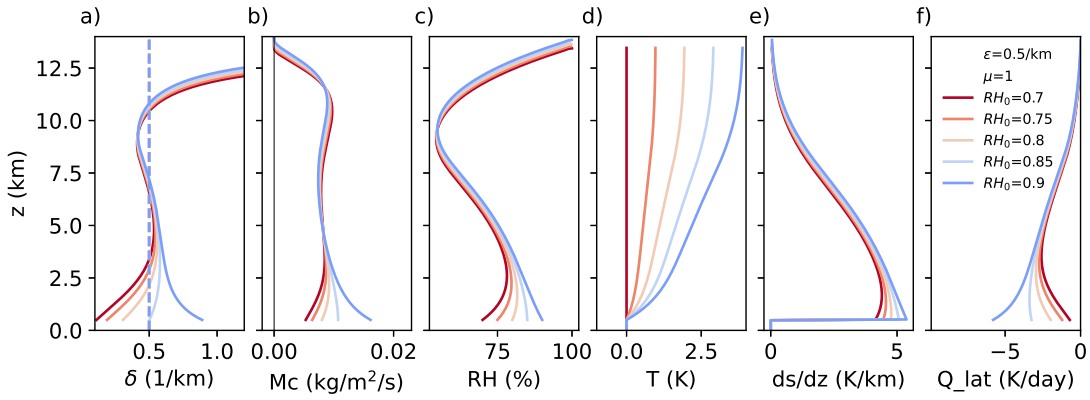


Figure A.1: Atmosphere profiles in the zero-buoyancy plume model with varying cloud-base relative humidity (blue color represents more moist environment), entrainment rate $\varepsilon = 0.5km^{-1}$, and cloud evaporation parameter $\mu = 1$. The variables shown are detrainment rate (a), updraft mass flux (b), relative humidity (c), temperature (d), the vertical gradient of dry static energy (e), and latent heating rate due to cloud evaporation in the environment (f). Dashed lines in panel a are the profiles of the prescribed entrainment rate. The temperature in panel d is shown as deviation to one of the simulations, which is denoted by the red line with zero deviation.

This ODE problem should in principle be solvable using an implicit method, although we will leave this to future work.

In Fig. A.1 we test the sensitivity of the atmospheric profiles to the cloud base specific humidity (or equivalently RH). We change the cloud base RH from 70% to 90%. Except temperature profile, the influence of cloud base RH on other variables is primarily within the lower 5km and does not have a big impact on the upper troposphere. The temperature becomes warmer through the whole troposphere with a moister cloud base environment. For cloud base mass flux, it strongly depends on the RH based on Equation 2.4 in the main text. The way we determine the cloud base RH using Equation A.27 will implicitly lead to the sensitivity that cloud base mass flux increases when ε or μ increases. Since our main focus in this paper is the upper troposphere mass flux, the uncertainty in how we determine the cloud base RH will likely not change our results. We also tested fixing the relative humidity at the cloud base. The sensitivities regarding mixing strength and evaporation rate remain qualitatively the same. In future research, it would be beneficial to integrate considerations

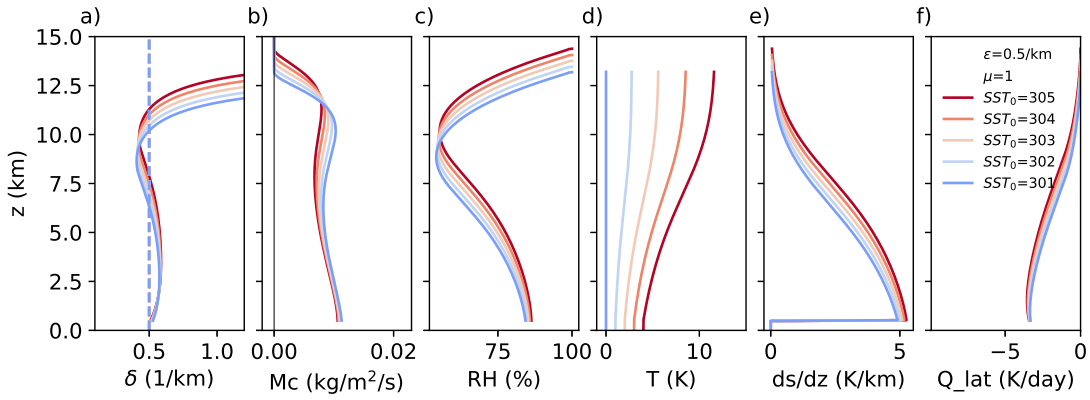


Figure A.2: Similar to Fig. A.1 but with sea surface temperature (blue color represents colder surface temperature), entrainment rate $\varepsilon = 0.5km^{-1}$, and cloud evaporation parameter $\mu = 1$.

of energy and water conservation in the sub-cloud layer, along with surface flux parameterization, to automatically determine the cloud base relative humidity.

In Fig. A.2 we test the sensitivity to different sea surface temperatures. We can see that the whole troposphere becomes higher with the profiles of most quantities shifting upwards. The peak value of mass flux near the anvil level decreases with warmer surface temperature, which will indicate a weaker mass detrainment and likely a decrease of anvil cloud fraction (if the lifetime is assumed to be unchanged with surface warming). The decrease of upper troposphere mass flux is consistent with the stability iris effect proposed in Bony et al. (2016).

In Fig. A.3, we test the sensitivity of atmospheric profiles to varying atmospheric cloud radiative effects (ACRE). Since mass flux change could potentially change the anvil clouds and the associated cloud radiative heating, here we add cloud radiative heating term to explore the potential impact of cloud radiative heating on RCE solutions. The imposed ACRE is:

$$\text{ACRE} = \begin{cases} 0.5H_0(1 + cos(\pi\frac{T-250}{30})), & \text{if } 220K < T < 280K \\ 0, & \text{if } T < 220K \text{ or } T > 280K \end{cases}$$

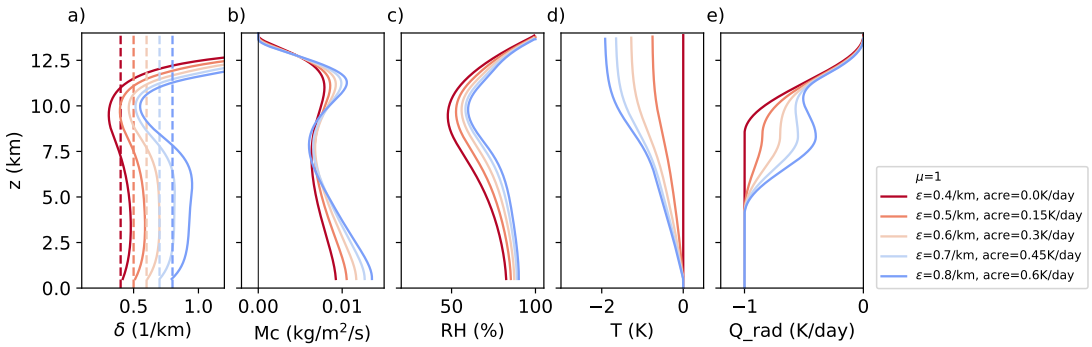


Figure A.3: Similar to Fig. 2.1 but with non-zero atmospheric cloud radiative effects. Panel e shows the net radiative heating rate profiles. We use a fixed cloud evaporation parameter $\mu = 1$, increasing the entrainment rate ε from 0.4 to 0.8 km^{-1}, and increasing peak atmospheric cloud radiative effects H_0 from 0 to 0.6 K/day.

where H_0 is the peak ACRE value at T=250K. By increasing the ACRE together with increasing the entrainment rate, Fig. A.3b exhibits non-monotonic mass flux change. We may interpret this non-monotonic behavior from the energy balance view. On top of the effects of changing the mixing strength, at the heights where ACRE increases, the net radiative cooling becomes weaker (Fig. A.3e), reducing the amount of mass flux needed to supply latent heating. Such interaction between the mixing strength and atmospheric cloud radiative effects could partially explain the non-monotonic convective updraft mass flux change we observed in the numerical model simulations in Fig. 2.5.

A.2 Relationship between cloud evaporation and perimeter area ratio

In the preceding sections, we highlighted the increased perimeter area ratio of cloud mass at higher resolutions, which potentially leads to greater exposure of the cloud mass to an unsaturated environment, thereby amplifying cloud evaporation. In this section, we derive a quantitative relationship between the cloud evaporation rate and the perimeter area ratio.

Consider a specific level with a unit thickness, where the cloud mass has a total area (A) and to-

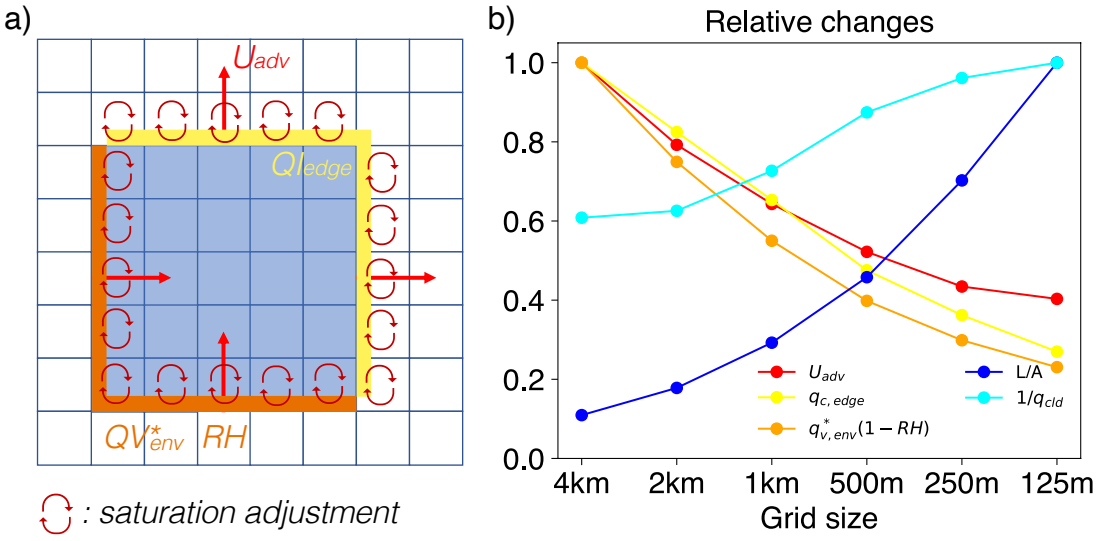


Figure A.4: Panel a shows a schematic of cloud evaporation due to resolved horizontal advection and the following saturation adjustment in the SAM1MOM scheme. We set some clouds initially in the grids with blue shading and advect the cloud by the horizontal wind with $U_x = U_y = U_{\mathrm{adv}}$. $q_{\mathrm{c,edge}}$ is the cloud mixing ratio near the cloud edge. $q^*_{\mathrm{v,env}}$ represents the environmental saturation specific humidity, and RH denotes relative humidity in the environment near the cloud. After a small timestep dt, some cloud mass is advected into environment grids (yellow shading), and some environmental mass is advected into cloudy grids (orange shading). Circular arrows represent the saturation adjustment in each grid due to the microphysics scheme. Panel b shows the relative value of different terms in Equation A.34 in the SAM1MOM simulations with different resolutions. Each term is standardized (divided by the maximum value across the simulations with different resolutions) to have a value between 0 and 1.

tal perimeter (L). The clouds are advected in grid points through resolved horizontal wind with a representative speed of U_{adv}. Approximately half of the cloud boundary exhibits horizontal resolved wind pointing outwards from the cloud, while the other half features wind directed inward (Fig. A.4). After a time step dt, the volume of clouds advected across the boundary amounts to $0.5LU_{\text{adv}}dt$ (represented by the yellow area in Fig. A.4a). An equivalent volume of environmental air is advected into the original cloudy grids (illustrated by the orange area in Fig. A.4a). Following advection, the SAM1MOM scheme performs saturation adjustment. The yellow cloud mass becomes fully mixed with the environmental air in the respective grids, subsequently evaporating. On average, since the cloud mixing ratio near cloud edges is relatively minimal, we assume complete evaporation of the yellow cloud mass. The evaporation associated with this yellow cloud mass should be proportional to the product of the volume and the cloud mixing ratio at the edge $q_{\text{c,edge}}$. Similarly, in the grids containing orange environmental air, a portion of the cloud must evaporate to bring the unsaturated orange environmental air to saturation. The evaporation amount would be the product of the volume and the saturation deficit $q^*_{\text{v,env}}(1-\mathcal{R})$, where $q^*_{\text{v,env}}$ represents the environmental saturation specific humidity and $\mathcal{R}$ is relative humidity. The total evaporation rate associated with saturation adjustment can be expressed as:

$$Evaporation = 0.5LU_{\text{adv}}\rho[q_{\text{c,edge}} + q^*_{\text{v,env}}(1-\mathcal{R})] \tag{A.31}$$

We assume all this evaporation can be counted as in the environment. We further assume the total evaporation in the environment is dominated by this numerical diffusion associated with resolved horizontal wind. This assumption likely works well for high clouds where vertical wind and sub-grid diffusion are weak, but may not work well for low clouds where vertical wind and sub-grid diffusion could be strong. Total air mass in the environment can be written as $\rho A_d f_{\text{env}}$, where A_d is domain area, f_{env} is the fraction of environment. Therefore, the latent heating rate in the environ-

ment can be written as:

$$Q_{\text{lat,env}} = 0.5LU_{\text{adv}}[q_{\text{c,edge}} + q^*_{\text{v,env}}(1-\mathcal{R})]\frac{L_{\text{v}}}{c_p A_d f_{\text{env}}} \tag{A.32}$$

Since more clouds tend to generate more evaporation, we normalize the latent heating by the total cloud mass to get a quantity that reflects evaporation efficiency. Total cloud mass is proportional to the domain-mean cloud mixing ratio q_{m}, which can be further expressed as $q_{\text{m}} = f_{\text{cld}} q_{\text{cld}}$. f_{cld} is cloud fraction, and q_{cld} is the cloud mixing ratio averaged in cloudy grids. The normalized latent heating rate can be expressed as:

$$\frac{Q_{\text{lat,env}}}{q_{\text{m}}} = 0.5LU_{\text{adv}}[q_{\text{c,edge}} + q^*_{\text{v,env}}(1-\mathcal{R})]\frac{L_{\text{v}}}{c_p A_d f_{\text{env}} f_{\text{cld}} q_{\text{cld}}} \tag{A.33}$$

Note that total cloud area can be written as $A = f_{\text{cld}} A_d$, the above equation can be rewritten as:

$$\frac{Q_{\text{lat,env}}}{q_{\text{m}}} = \frac{L}{A} U_{\text{adv}} \frac{q_{\text{c,edge}} + q^*_{\text{v,env}}(1-\mathcal{R})}{q_{\text{cld}}} \frac{L_{\text{v}}}{2c_p f_{\text{env}}} \tag{A.34}$$

We define cloud boundaries as grid interfaces that separate a grid with zero cloud mixing ratio from a grid with non-zero cloud mixing ratio. Subsequently, we evaluate the average values of U_{adv}, $q_{\text{c,edge}}$, $q^*_{\text{v,env}}$, and $\mathcal{R}$ at grids immediately adjacent to the boundaries, either on the inside or the outside. In Fig. A.4b, we demonstrate the variation in different terms of Equation A.34 as the resolution becomes finer. With increased resolution, the perimeter area ratio rises, while advection velocity, cloud mixing ratio, and environmental saturation deficit decrease. The decline in near-edge cloud mixing ratio and environmental saturation deficit could be attributed to the improved representation of the transition between cloudy grids and environmental grids at finer scales. The enhanced transition at cloud boundaries in higher resolutions tends to reduce numerical diffusion and partially counterbalance the effect of the growing perimeter area ratio. The cause of the weakened

advection wind and reduced in-cloud mixing ratio remains unclear and merits further investigation.

Overall, finer resolution enables a better representation of turbulent cloud boundaries, enhancing the interaction between clouds and their environment. However, finer resolution also leads to a reduction in numerical diffusion. The interplay between these two effects may be crucial in determining whether cloud evaporation efficiency converges at a specific resolution. A comprehensive understanding of these factors is essential for improving the accuracy and reliability of Earth system models.

A.3 Additional Figures in the Supporting Information

Figure A.5 checks how well the assumptions of the zero-buoyancy plume model are held in the numerical simulations. We examine the assumptions on the shape of radiative cooling rate, neutral buoyancy, and hydrostatic balance. Figure A.6 shows the profiles of cloud mixing ratios with different model resolutions and microphysics. Figure A.7 shows the profiles of convective mass flux under different definitions, e.g., with different thresholds of vertical velocities or whether to include convective downdraft.

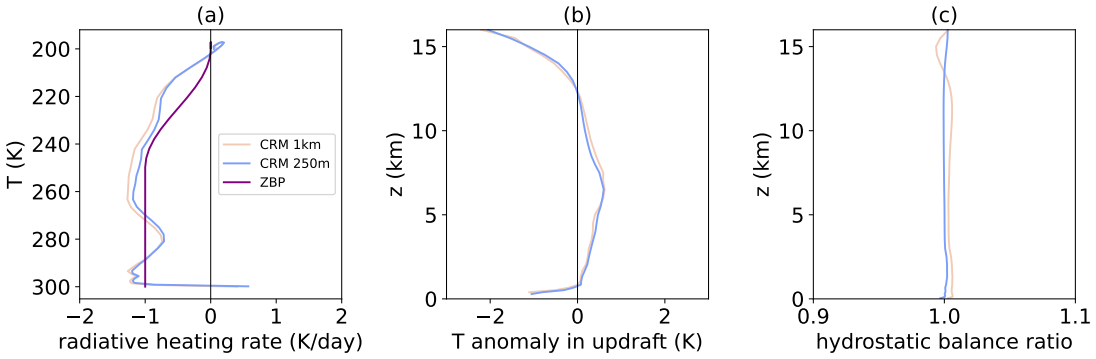


Figure A.5: Panel a compares the domain-mean radiative heating rate in the model simulations and the prescribed radiative heating rate used in the zero-buoyancy plume model as a function of temperature. Panel b shows the absolute temperature anomaly averaged in convective updrafts in the model simulations. Panel c shows the hydrostatic balance ratio, defined as $-(dp/dz)/(\rho g)$, in the model simulations. The orange line is for the 1km-resolution simulation, and the blue line is for the 250m-resolution simulation.

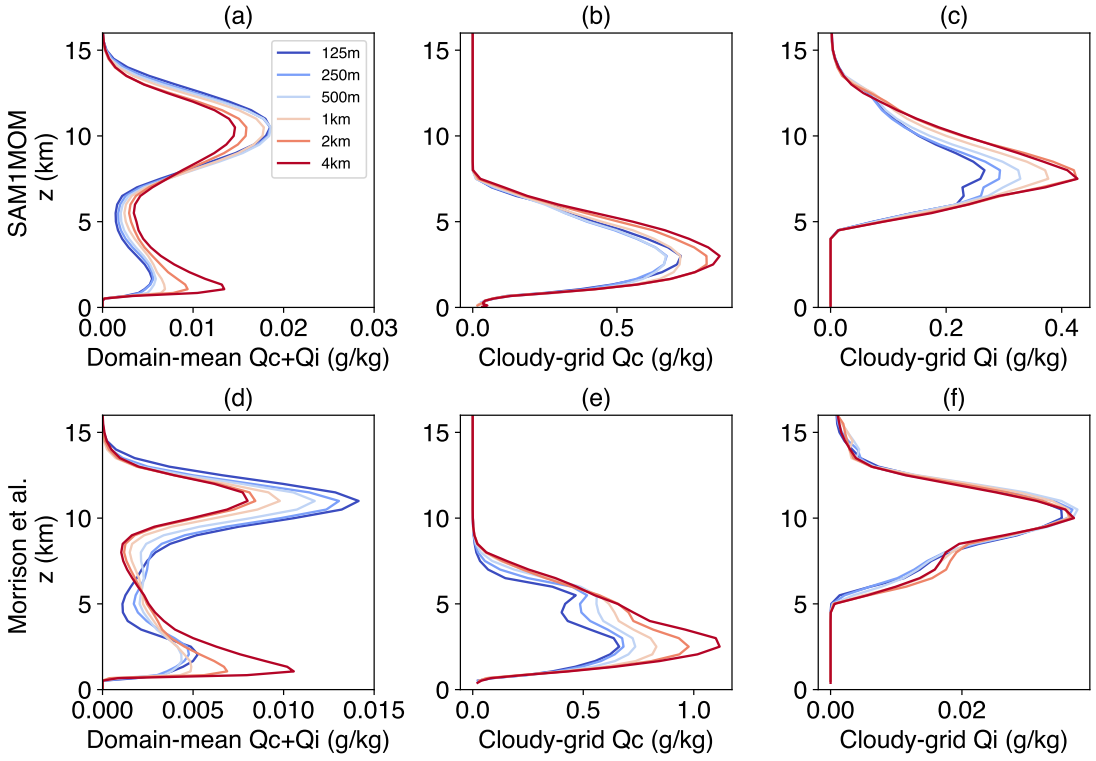


Figure A.6: Atmosphere profiles cloud mixing ratios in the SAM1MOM simulations (upper row) and Morrison et al. simulations (lower row). The left column is the domain mean total (liquid plus ice) cloud mixing ratio. The middle column is the liquid cloud mixing ratio averaged within cloudy-grids (grids with total cloud mixing ratio $> 10^{-5}$ kg/kg). The right column is the ice cloud mixing ratio averaged within cloudy-grids.

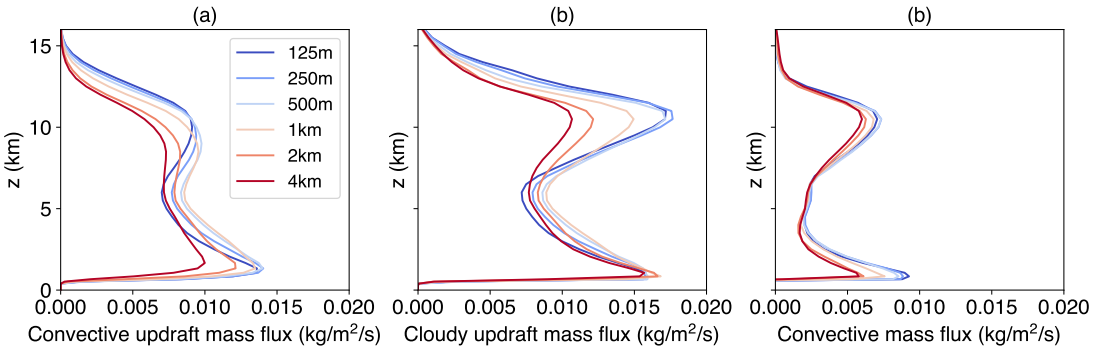


Figure A.7: Domain-mean averaged mass flux defined with different averaging conditions. Panel a shows the mass flux averaged in grids with vertical velocity W$>$1 m/s and cloud mixing ratio $> 10^{-5}$ kg/kg. Panel b shows the mass flux averaged in grids with vertical velocity W$>$0 m/s and cloud mixing ratio $> 10^{-5}$ kg/kg. Panel c shows the net convective mass flux. Net convective mass flux in panel c is defined as cloudy mass flux (grids with cloud mixing ratio $> 10^{-5}$ kg/kg, including both updraft and downdraft) plus precipitating downdraft (precipitating water $> 10^{-4}$ kg/kg and W$<$0 m/s).

B

Supplemental Material for Chapter 4

B.1 Neural Network Model Configuration

B.1.1 Input and Output Features

Table B.1 and B.2 list the complete input and output features for the baseline MLP models. In the inputs, we include the current state variables like temperature, relative humidity, liquid and ice cloud mixing ratio, zonal and meridional wind. In addition, we also include in the input features

the atmosphere gas components (ozone, N2O, and CH4 mixing ratios), boundary conditions (surface pressure, solar insolation and cosine zenith angle, surface heat flux and surface stress, albedo, snow depth), and geography information (land fraction, sea ice fraction). In the output, we predict the time tendencies of temperature, specific humidity, liquid and ice cloud mixing ratio, zonal wind and meridional wind. In addition to the state tendencies, we also predict radiative fluxes and precipitation (snow and rain) at the surface which are required to couple a land model.

Table B.3 lists the additional input features used in the U-Net-expanded and U-Net-expanded-constrained models. These additional inputs include large-scale forcing at the current and previous time steps, and the convective memory (i.e., the target tendencies) at two previous time steps. Large-scale forcing and convective memory are also used in (Han et al., 2020; Wang et al., 2022b; Han et al., 2023). Cosine and sine of latitudes are included as well. When microphysical constraints are added, mixing ratios and previous steps' tendencies of liquid cloud and ice cloud inputs are replaced with the corresponding total cloud condensate inputs (liquid plus ice), along with a diagnosed fraction of liquid cloud based on temperature. That is, the last three rows in the Table B.3 are only used in the U-Net-expanded-constrained model, which also excludes $q_l(z)$, $q_i(z)$, $dq_l(z, t_{-1}, t_{-2})$, and $dq_i(z, t_{-1}, t_{-2})$ from the inputs.

B.1.2 NN architecture

MLP: For the results presented in the main text, We used the architecture parameters recommended from the hyperparameter search in (Yu et al., 2024a) in our MLP model. It contains 3 hidden layers with $N_{\text{nodes}} = [384, 1024, 640]$ and ReLU activation. Additionally, we explored an alternative MLP model with $N_{\text{nodes}} = [256, 256, 256, 256, 256, 256, 256, 256, 256]$ and ReLU activation, as proposed in (Rasp et al., 2018). We observed similar offline and online performance between the two configurations. The results from the second configuration are not shown in the manuscript.

U-Net: We adapted the 2D U-Net model from (Song et al., 2020) into a 1D version for column-

to-column prediction. The architecture schematic is shown in Figure 4.1. Each ResBlock consists of the following operations:

$$y = \text{Conv1D}(\text{GM}(\text{Conv1D}(\text{silu}(\text{GM}(x))))) + x$$

This series of operations includes group normalization (GM), sigmoid linear unit (silu) activation function, 1D convolution (Conv1D) with a kernel size of 3, and a residual connection. The U-Net has 4 layers in depth with latent feature dimensions in each layer $N_{\text{latent}} = [128, 256, 256, 256]$. The model comprises 13 million parameters.

To evaluate the inference speed of these architectures, we configured the reference E3SM-MMF simulation and hybrid simulations on the NERSC Perlmutter system https://www.nersc.gov/systems/perlmutter. The simulations were run on a single GPU node equipped with four A100 GPUs, using 4 MPI tasks and 8 threads per task. The reference E3SM-MMF simulation achieves a throughput of 9.9 simulation years per day (SYPD), leveraging GPU acceleration for cloud-resolving model calculations. In comparison, the hybrid MLP simulation achieves a slightly higher throughput of 10.1 SYPD, while the U-Net-baseline simulation achieves 5.2 SYPD. Notably, the neural network computations in these hybrid simulations were performed entirely on CPUs due to unresolved engineering challenges, and GPU acceleration was not utilized in our hybrid simulations. We anticipate that enabling NN inference on GPUs could significantly enhance the throughput of hybrid simulations, making them more computationally efficient than the reference E3SM-MMF simulations. It is important to note that these timing metrics are not very meaningful as our work is not concerned with unthrottling the maximum throughput of hybrid simulations, but rather proving what it takes to encourage their success from a scientific design perspective. It is logical to expect ML model distillation and hardware acceleration strategies can then recover the high throughput expected of hybridized simulations, which is a separate topic.

B.1.3 Normalization

The normalization methods for each input and output feature are listed in Table B.1, B.2, and B.3. The majority of input features are normalized by subtracting the mean and then dividing by the difference of maximum and minimum values. All output variables are normalized by the standard deviation. Mean, maximum, minimum, and standard deviation values are all evaluated per level. For variables that are already in the range of 0 to 1, no normalization is performed. For the large-scale forcings, solar insolation, and surface latent/sensible heat fluxes, we choose not to subtract the mean to maintain their sign. Convective memory inputs are normalized by the same standard deviation value used in the corresponding output variable.

For input liquid, ice, and total cloud mixing ratio, an exponential transformation ($x' = 1 - exp(-\lambda x)$) is taken to deal with the long positive tail in the distribution of cloud mixing ratio. The rationale is that, if the tail follows an exponential distribution, then an exponential transformation is able to convert the exponential distribution into a 0-1 uniform distribution. The exponential parameter λ is evaluated by one over the conditional mean cloud mixing ratio for grids with mixing ratio greater than $10^{-7} kg/kg$, i.e., $\lambda = 1/mean(q_n(q_n > 10^{-7}))$.

When calculating the standard deviation for the output tendencies of water vapor, total cloud, zonal and meridional winds, an arbitrary lower bound is added and the actual normalization factor is min(st.d., threshold). These thresholds are added to penalize the contribution of stratosphere levels in the total loss function. In the stratosphere, the standard deviation of these tendencies dramatically decreases with height. This threshold is chosen as 3^{-10} kg/kg for tendencies of water vapor and cloud and as 10^{-6} m/s for zonal and meridional winds. The tendencies at those levels are also likely noisy and less physical given they are within the sponge layer of the CRM which damps the perturbation. This penalizes water vapor above level 22 (200 hPa), clouds above level 17 (70 hPa), and winds above level 15 (50h Pa). Correspondingly, we also set the top 15 levels of NN output

tendencies of water vapor, total cloud, zonal and meridional winds to 0.

B.1.4 Additional Preprocessing

After normalization, we concatenate all the input features and output targets into two one-dimensional arrays for each sample and feed them to the MLP model. For the U-Net models, we construct the input features for each sample into a two-dimensional array with a vertical length of 60 and a channel length equal to the number of features. Each profile input feature represents an individual channel, while scalar features such as solar insolation or surface heat fluxes are broadcast as channels with constant values. To facilitate three downsampling steps in the U-Net model, we concatenate four additional layers of zeros at the top of the atmosphere, expanding the vertical column length to 64. The direct output of the U-Net model is a sequence of length 64 with 13 channels. We then drop the top 4 levels in the output. The first 5 output channels correspond to the profile output variables, while the last 8 channels are averaged through the level dimension to obtain the 8 scalar output variables.

B.1.5 Feature Pruning

The top levels in the cloud-resolving model are within the sponge layer, where the tendencies of water vapor, clouds, and winds can be unphysical and unpredictable. Our output normalization approach also tends to penalize the upper-level tendencies in the total loss function. We do not expect our neural networks can well predict those levels. Therefore, we set the top 15 levels to zero for the predicted tendencies of water vapor, liquid and ice clouds, and zonal and meridional winds, excluding them from the loss function. Additionally, we manually prune the liquid, ice, or total cloud mixing ratio in the top 15 levels (above 50 hPa) from the input by setting them to zero before feeding them to the neural network. This pruning decision is somewhat arbitrary, and we did not

test its necessity. A more careful normalization of the output tendencies in the stratosphere could potentially improve their prediction without negatively affecting the fitting skill at lower height levels. In such a case, it would be more appropriate to evaluate the necessity of this pruning.

Previous studies, such as Brenowitz & Bretherton (2018), suggested that excluding stratosphere temperature and water vapor (e.g., above around 200 hPa) could be crucial for achieving stable online simulations. To test this hypothesis, we also experimented with a more aggressive pruning strategy, pruning water vapor and clouds above 170 hPa, as well as pruning large-scale forcing and convective memories above 50 hPa. However, we did not observe significant differences with or without this aggressive pruning strategy. All the results presented in this manuscript do not employ this aggressive pruning.

B.2 Training

Previous studies, such as (Ott et al., 2020; Wang et al., 2022b), have shown that improved offline skill can benefit downstream error, but the downstream error is not fully constrained by offline skill. This means different checkpoints with very similar offline skills can exhibit varying downstream stability and error. A trial-and-error approach is often used to identify the checkpoint with the best downstream performance. For subtle sensitivities, hundreds of trials can become important to detect online signals of hybrid climate error from upstream ML noise (Lin et al., 2024).

To account for some of this downstream uncertainty, we trained three MLP models with different loss functions and learning rate schedules listed below. We trained these MLP models with a batch size of 1024 using the Adam optimizer. The training uses Distributed Data Parallel on 4 GPUs.

- The first model used Mean Absolute Error (MAE) loss, with an initial learning rate of 1×10^{-3}, reduced by a factor of 0.3162 every 7 epochs for a total of 28 epochs.

- The second model used the same learning rate schedule but a standard Huber loss with $\delta = 1$.

- The third model also used Huber loss but with a different learning rate schedule: starting at 1×10^{-3}, using a ReduceOnPlateau scheduler with a patience of 3 epochs and a reduction factor of 0.3162 for a total of 20 epochs. This was followed by manually reducing the learning rate to 1×10^{-4} for fine-tuning over 12 more epochs with a patience of 0 epochs and a reduction factor of 0.5.

Similarly, we trained three versions of the U-Net models for each of the three configurations: without expanded input features (U-Net-baseline), with expanded input features (U-Net-expanded), and with both expanded input features and cloud physics constraints (U-Net-expanded-constrained). We used the same batch size of 1024 and the Adam optimizer. The training uses Distributed Data Parallel on 4 GPUs.

- The first used MAE loss, with an initial learning rate of 1×10^{-4}, reduced by a factor of 0.5 every 3 epochs for a total of 16 epochs.

- The second model used the same learning rate schedule but a standard Huber loss with $\delta = 1$.

- The third used Huber loss with a different learning rate schedule: starting at 1×10^{-4} with a ReduceOnPlateau scheduler with a patience of 3 epochs and a reduction factor of 0.3162 for a total of 20 epochs. For U-Net-expanded and U-Net-expanded-constrained, the learning rate does not drop within the first 20 epochs. Therefore, for these two configurations, we further manually reduced the learning rate to 5×10^{-5} for fine-tuning over 8 more epochs with a patience of 0 epochs and a reduction factor of 0.5.

B.3 Online Metric

In this section, we document how the two main online metrics below are calculated.

Root Mean Square Error: The root mean square error (RMSE) are computed separately for each variable in the online simulations. The goal is to measure the error in simulated climate by analyzing state variables that are sufficiently averaged in space and time. For a given evaluation timescale, RMSE for each variable is calculated as follows:

$$\text{RMSE} = \sqrt{\sum_{i=1}^{S_m} w_i(\hat{y}_m - y_m)^2}$$

where:

- S_m is the number of samples. Each grid cell counts as one sample. For a variable evaluated at a specific vertical level, S_m includes all grid cells at that level. For global evaluations, S_m includes all grid cells in the 3D domain (horizontal and vertical).
- $\hat{y}_m$ represents the fields from the ML-coupled simulation averaged over the given evaluation timescale. In Section 4.5.1, $\hat{y}_m$ represents monthly averaged data. In Section 4.5.2, $\hat{y}_m$ represents five-year mean data.
- y_m represents the mean values from the reference simulation averaged over the same timescale as $\hat{y}_m$.
- $w_1, w_2, \ldots, w_{S_m}$ are mass-weights that sum to 1, proportional to the mass in each grid cell.

Zonal Mean Bias: Additionally, we evaluate the long-term zonal mean bias, which measures the average difference between the online simulation and the reference simulation across various atmospheric variables, such as temperature, water vapor, wind, and liquid and ice clouds. The zonal mean bias is derived by comparing variables averaged over time and longitudes. The E3SM-MMF

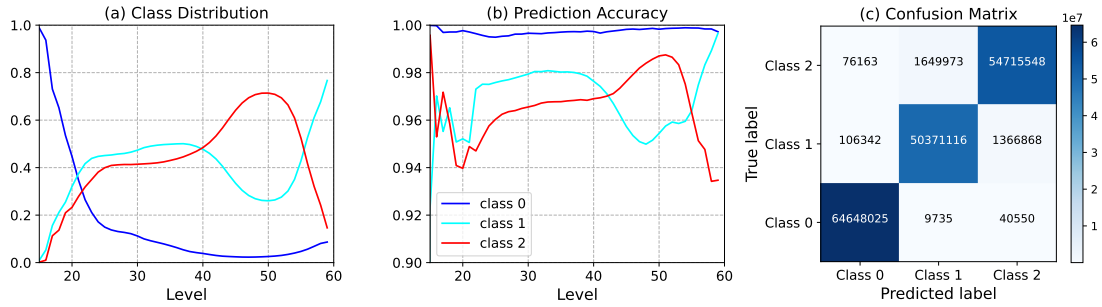


Figure B.1: (a) The distribution of the three classes at each atmospheric level. Only the levels below level 15 are shown since it is mostly Class 0 above level 15. (b) The prediction accuracy of the classifier for each class. Prediction accuracy is defined as the fraction of samples of a given class that are correctly identified by the classifier. (c) The confusion matrix of the classifier for all grid samples at all levels.

climate model uses unstructured grids instead of regular latitude-longitude grids. To average a variable over all longitudes, we first define latitude bins. For the low-resolution version of the ClimSim dataset, we choose 10-degree intervals from 90°S to 90°N. Within each bin, we count all the grid columns that fall into the bin and calculate the horizontal average of those columns weighted by their area.

B.4 Microphysics classifier

The tendencies of liquid or ice cloud mixing ratios follow a mixture of discrete and continuous distributions. Some clear-sky grids may have no clouds and, consequently, zero cloud tendencies. However, a naive neural network regression model is likely to predict very small but non-zero values on these grids. In our hybrid model, a non-negative check sets any negative water vapor or cloud mixing ratios to zero. (Brenowitz et al., 2022) suggested that a separate classifier model can help mitigate the online accumulation of errors in the discrete mode because the non-negative check removes negative errors but retains positive errors. Similarly, (Gettelman et al., 2021) used a classifier as part of their model to emulate a microphysics scheme. Therefore, we trained a microphysics classifier to test if it could help reduce the online cloud bias.

Following Brenowitz et al. (2022), we defined three classes for cloud changes in the reference E3SM-MMF simulations:

- **Class 0**: No cloud change (defined as the tendency of total cloud mixing ratio smaller than 10^{-13} kg/kg/s).
- **Class 1**: Not Class 0, and all cloud is removed (defined as remaining total cloud mixing ratio smaller than 10^{-10} kg/kg).
- **Class 2**: Not Class 0 or Class 1.

Figure B.1a shows that a majority of the grids are in either Class 0 or Class 1. Only less than 50% of grids above level 40 are within Class 2. We trained a classifier model using the same U-Net architecture as before, except with only 1/4 of the latent channel number and 1/2 of the number of Resblocks. The classifier takes the same input features but predicts the logits of each class at each level. We trained the classifier to optimize a cross-entropy loss.

Our classifier model can accurately predict the defined classes but does not significantly change the online error. Figures B.1b and B.1c show that the classifier is very accurate, with per-level accuracy exceeding 94% for all classes across all levels. However, when we implemented the classifier into the hybrid E3SM model and masked the regression model's cloud prediction for Class 0 and Class 1 (by setting cloud tendencies to 0 for Class 0 and cloud mixing ratio to 0 for Class 1), we observed little change in the zonal mean climatology with or without the microphysics classifier. We believe that while the classifier prevents the accumulation of prediction errors in the discrete mode, as suggested by Brenowitz et al. (2022), this error is small compared to the prediction error in grids of Class 2 in our hybrid simulations. This likely explains why adding the microphysics classifier does not significantly improve online performance. All the results presented outside this section do not use a microphysics classifier.

B.5 Online Zonal Mean Climatology Bias with a Different U-Net Checkpoint

Figure B.2 shows the 5-year zonal-mean climatology using the constrained U-Net model, similar to Figure 4.9, but with a different U-Net checkpoint due to the use of a different learning rate schedule. This checkpoint exhibits a better extratropical cloud water bias, with only around 10 g/kg cloud water bias compared to the 40 g/kg bias near 60N shown in Figure 4.9l. When comparing Figure B.2 to Figure 4.9, many bias patterns remain consistent, such as the vertical dipole temperature bias in the polar stratosphere, a drying water vapor bias and a liquid cloud low bias in the tropical lower troposphere.

B.6 Online error growth

We evaluate the online error growth of the hybrid simulation with a constrained U-Net model in comparison to the reference E3SM-MMF simulation. At each time step, a zonal mean bias is calculated and then averaged within 30S-30N. Figure B.3 visualizes the tropical mean zonal mean bias for the first 5 days using a pressure-versus-time method. The tropical lower troposphere drying bias develops within the first day and is comparable to the 5-year zonal mean bias shown in Figure 4.9f. The tropical lower troposphere bias of reduced liquid cloud develops even faster, within only a few hours, and becomes comparable to the 5-year zonal mean bias shown in Figure 4.9l. This suggests that online biases in the first few simulation days can be indicative of yearly mean online biases. Better understanding of these rapidly developing bias patterns could potentially help in optimizing them.

B.7 Additional Figures in the Supporting Information

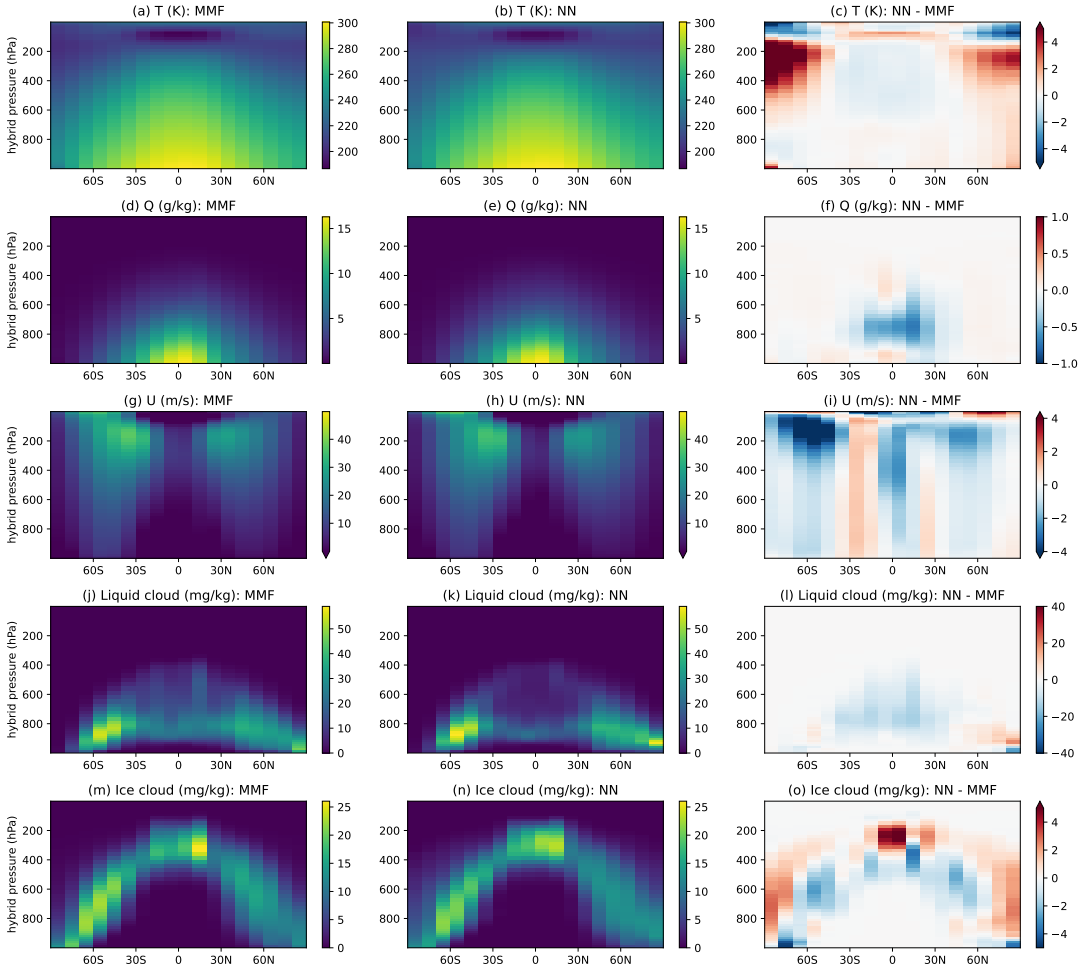


Figure B.2: Similar to Figure 4.9, but using the U-Net-expanded-constrained model with Huber loss and a different learning rate schedule, as described in B.2.

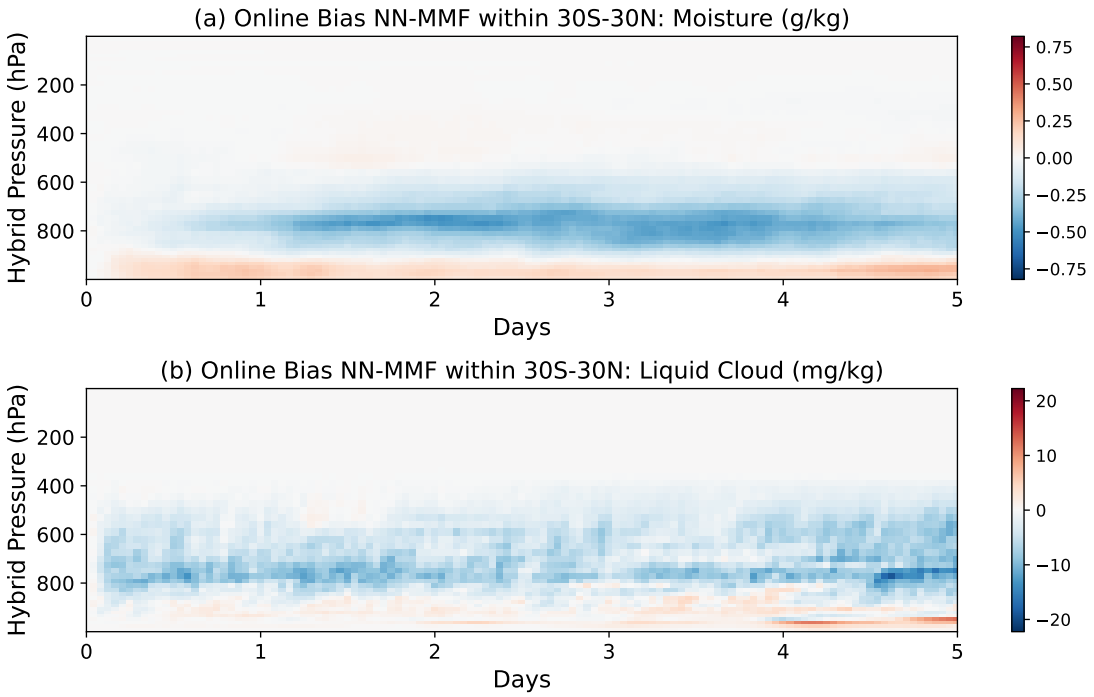


Figure B.3: Time evolution of the zonal mean bias averaged within 30S-30N for water vapor (a) and liquid cloud mixing ratio (b) over the first 5 days in the hybrid simulation with the U-Net model incorporating the microphysical constraints.

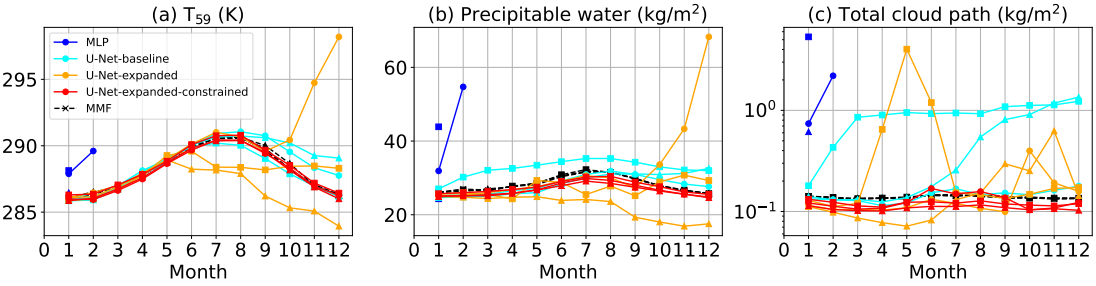


Figure B.4: Online monthly and globally averaged (a) near-surface temperature at the lowest model level (K), (b) precipitable water (kg/m^2), and (c) total (liquid+ice) cloud path (kg/m^2) over a one-year period, comparing baseline MLP (blue), U-Net-baseline (cyan), U-Net-expanded (orange), and U-Net-expanded-constrained (red) against the reference E3SM-MMF simulations (dashed black lines). The dashed black lines construct an ensemble of reference physical simulations, and their spread represents internal atmospheric variability. Each architecture was trained with three slightly different configurations (e.g., loss functions and learning rate schedules) to sample ML uncertainty, represented by different markers but with the same line color.

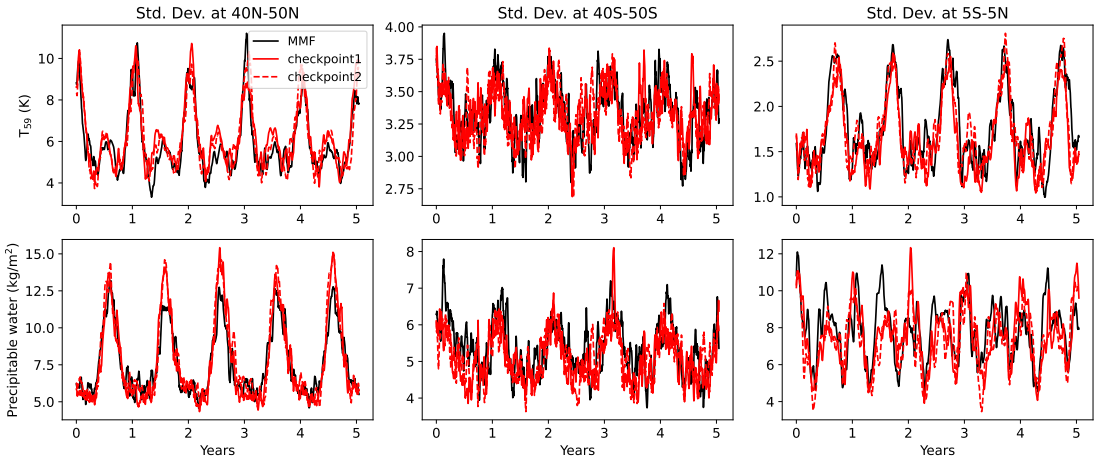


Figure B.5: Five-year time series of spatial standard deviation of near-surface temperature at the lowest model level (top row) and precipitable water (bottom row). The standard deviation is computed using daily-averaged fields across all grid points within three latitude bands: 40°N–50°N (left column), 40°S–50°S (middle column), and 5°S–5°N (right column). A 15-day running average is applied to smooth the time series. The black line represents the reference E3SM-MMF simulation. The solid and dashed red lines represent the hybrid simulations using two U-Net-expanded-constrained checkpoints: Huber loss with a reduce-on-plateau scheduler (solid red) and Huber loss with a step scheduler (dashed red).

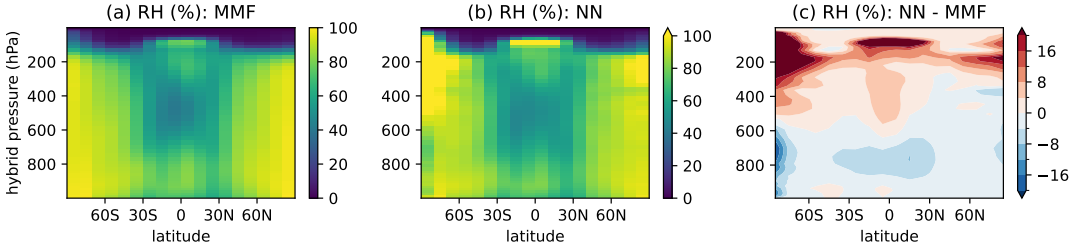


Figure B.6: Five-year zonal mean relative humidity in the reference E3SM-MMF simulation (left) and in the hybrid simulation with the U-Net-expanded-constrained model (middle). The right column shows the zonal mean bias as the mean state from the hybrid simulation minus that of the reference simulation. The contouring interval in the right column is 4%.

Table B.1: List of input features used in the MLP model.

Variable	Units	Description	Normalization
$T(z)$	K	Temperature	(x-mean)/(max-min)
$RH(z)$		Relative humidity	
$q_l(z)$	kg/kg	Liquid cloud mixing ratio	1 - exp(-λx)
$q_i(z)$	kg/kg	Ice cloud mixing ratio	1 - exp(-λx)
$u(z)$	m/s	Zonal wind	(x-mean)/(max-min)
$v(z)$	m/s	Meridional wind	(x-mean)/(max-min)
O3(z)	mol/mol	Ozone volume mixing ratio	(x-mean)/(max-min)
CH4(z)	mol/mol	Methane volume mixing ratio	(x-mean)/(max-min)
N2O(z)	mol/mol	Nitrous volume mixing ratio	(x-mean)/(max-min)
PS	Pa	Surface pressure	(x-mean)/(max-min)
SOLIN	W/m^2	Solar insolation	x/(max-min)
LHFLX	W/m^2	Surface latent heat flux	x/(max-min)
SHFLX	W/m^2	Surface sensible heat flux	x/(max-min)
TAUX	W/m^2	Zonal surface stress	(x-mean)/(max-min)
TAUY	W/m^2	Meridional surface stress	(x-mean)/(max-min)
COSZRS		Cosine of solar zenith angle	(x-mean)/(max-min)
ALDIF		Albedo for diffuse longwave radiation	(x-mean)/(max-min)
ALDIR		Albedo for direct longwave radiation	(x-mean)/(max-min)
ASDIF		Albedo for diffuse shortwave radiation	(x-mean)/(max-min)
ASDIR		Albedo for direct shortwave radiation	(x-mean)/(max-min)
LWUP	W/m^2	Surface upward longwave flux	(x-mean)/(max-min)
ICEFRAC		Sea-ice area fraction	
LANDFRAC		Land area fraction	
OCNFRAC		Ocean area fraction	
SNOWHLAND	m	Snow depth over land	(x-mean)/(max-min)

Table B.2: List of output features used in the MLP and U-Net-baseline model.

Variable	Units	Description	Normalization
$dT(z, t_0)$	K/s	Temperature tendency	x/std
$dq_v(z, t_0)$	kg/kg/s	Water vapor tendency	x/std
$dq_l(z, t_0)$	kg/kg/s	Liquid cloud tendency	x/std
$dq_i(z, t_0)$	kg/kg/s	Ice cloud tendency	x/std
$du(z, t_0)$	m/s^2	Zonal wind tendency	x/std
$dv(z, t_0)$	m/s^2	Meridional wind tendency	x/std
NETSW	W/m^2	Net shortwave flux at surface	x/std
FLWDS	W/m^2	Downward longwave flux at surface	x/std
PRECSC	m/s	Snow rate (liquid water equivalent)	x/std
PRECC	m/s	Rain rate	x/std
SOLS	W/m^2	Downward visible direct solar flux to surface	x/std
SOLL	W/m^2	Downward near-IR direct solar flux to surface	x/std
SOLSD	W/m^2	Downward visible diffuse solar flux to surface	x/std
SOLLD	W/m^2	Downward near-IR diffuse solar flux to surface	x/std

Table B.3: List of additional input features used in the U-Net-expanded model and the U-Net-expanded-constrained model. The last three rows are only used for the U-Net-expanded-constrained model. In addition, the U-Net-expanded-constrained model excludes $q_l(z)$, $q_i(z)$, $dq_l(z, t_{-1}, t_{-2})$, and $dq_i(z, t_{-1}, t_{-2})$ from the inputs.

Variable	Units	Description	Normalization
$dT_{adv}(z, t_0, t_{-1})$	K/s	Large-scale forcing of temperature	x/(max-min)
$dq_{T,adv}(z, t_0, t_{-1})$	kg/kg/s	Large-scale forcing of total water	x/(max-min)
$du_{adv}(z, t_0, t_{-1})$	m/s^2	Large-scale forcing of zonal wind	x/(max-min)
$dT(z, t_{-1}, t_{-2})$	K/s	Temperature tendency	x/std
$dq_v(z, t_{-1}, t_{-2})$	kg/kg/s	Water vapor tendency	x/std
$dq_l(z, t_{-1}, t_{-2})$	kg/kg/s	Liquid cloud tendency	x/std
$dq_i(z, t_{-1}, t_{-2})$	kg/kg/s	Ice cloud tendency	x/std
$du(z, t_{-1}, t_{-2})$	m/s^2	Zonal wind tendency	x/std
cos(lat)		Cosine of latitude	
sin(lat)		Sine of latitude	
$liq_partition(z)$		Fraction of liquid cloud	
$q_n(z)$	kg/kg	Total cloud (liquid + ice) mixing ratio	1 - exp(-λx)
$dq_n(z, t_{-1}, t_{-2})$	kg/kg/s	Total cloud tendency	x/std

References

Alexeev, D. (2023). alexeedm/pytorch-fortran: Version v0.4 [software].

Anderson, J. G. (2017). Stratospheric ozone over the united states in summer linked to observations of convection and temperature via chlorine and bromine catalysis. *Proc. Natl. Acad. Sci. USA*, 114, E4905–E4913.

Bao, J. W., Michelson, S. A., & Grell, E. D. (2019). Microphysical process comparison of three microphysics parameterization schemes in the wrf model for an idealized squall-line case study. *Mon. Wea. Rev.*, 147, 3093–3120.

Becker, T., Stevens, B., & Hohenegger, C. (2017). Imprint of the convective parameterization and sea-surface temperature on large-scale convective self-aggregation. *Journal of Advances in Modeling Earth Systems*, 9(2), 1488–1505.

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

Benedict, J. J. & Randall, D. A. (2009). Structure of the madden–julian oscillation in the superparameterized cam. *J. Atmos. Sci.*, 66(11), 3277–3296.

Beucler, T., Gentine, P., Yuval, J., Gupta, A., Peng, L., Lin, J., Yu, S., Rasp, S., Ahmed, F., O'Gorman, P. A., et al. (2024). Climate-invariant machine learning. *Science Advances*, 10(6), eadj7250.

Beydoun, H., Caldwell, P. M., Hannah, W. M., & Donahue, A. S. (2021). Dissecting anvil cloud response to sea surface warming. *Geophysical Research Letters*, 48(15), e2021GL094049.

Bhouri, M. A., Peng, L., Pritchard, M. S., & Gentine, P. (2023). Multi-fidelity climate model parameterization for better generalization and extrapolation. *arXiv preprint arXiv:2309.10231*.

Bony, S., Stevens, B., Coppin, D., Becker, T., Reed, K. A., Voigt, A., & Medeiros, B. (2016). Thermodynamic control of anvil cloud amount. *Proceedings of the National Academy of Sciences*, 113(32), 8927–8932.

Bony, S., Stevens, B., Frierson, D. M., Jakob, C., Kageyama, M., Pincus, R., Shepherd, T. G., Sherwood, S. C., Siebesma, A. P., Sobel, A. H., et al. (2015). Clouds, circulation and climate sensitivity. *Nature Geoscience*, 8(4), 261–268.

Brenowitz, N. D., Beucler, T., Pritchard, M., & Bretherton, C. S. (2020). Interpreting and stabilizing machine-learning parametrizations of convection. *Journal of the Atmospheric Sciences*, 77(12), 4357–4375.

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(8), 2728–2744.

Brenowitz, N. D., Perkins, W. A., Nugent, J. M., Watt-Meyer, O., Clark, S. K., Kwa, A., Henn, B., McGibbon, J., & Bretherton, C. S. (2022). Emulating fast processes in climate models. *arXiv preprint arXiv:2211.10774*.

Bretherton, C. S., Blossey, P. N., & Khairoutdinov, M. (2005). An energy-balance analysis of deep convective self-aggregation above uniform sst. *Journal of the atmospheric sciences*, 62(12), 4273–4292.

Bretherton, C. S., Henn, B., Kwa, A., Brenowitz, N. D., Watt-Meyer, O., McGibbon, J., Perkins, W. A., Clark, S. K., & Harris, L. (2022). Correcting coarse-grid weather and climate models by machine learning from global storm-resolving simulations. *J. Adv. Model. Earth Syst.*, 14(2), e2021MS002794.

Bretherton, C. S. & Khairoutdinov, M. F. (2015). Convective self-aggregation feedbacks in near-global cloud-resolving simulations of an aquaplanet. *J. Adv. Model. Earth Syst.*, 7, 1765–1787.

Bretherton, C. S., McCaa, J. R., & Grenier, H. (2004). A new parameterization for shallow cumulus convection and its application to marine subtropical cloud-topped boundary layers. part i: Description and 1d results. *Monthly weather review*, 132(4), 864–882.

Brewer, A. W. (1949). Evidence for a world circulation provided by the measurements of helium and water vapour distribution in the stratosphere. *Quart. J. Roy. Meteor. Soc.*, 75, 351–363.

Bryan, G. H., Wyngaard, J. C., & Fritsch, J. M. (2003). Resolution requirements for the simulation of deep moist convection. *Monthly Weather Review*, 131(10), 2394–2416.

Butchart, N. (2014). The brewer-dobson circulation. *Rev. Geophys.*, 52, 157–184.

Christopoulos, C., Lopez-Gomez, I., Beucler, T., Cohen, Y., Kawczynski, C., Dunbar, O. R., & Schneider, T. (2024). Online learning of entrainment closures in a hybrid machine learning parameterization. *Journal of Advances in Modeling Earth Systems*, 16(11), e2024MS004485.

Clark, S. K., Brenowitz, N. D., Henn, B., Kwa, A., McGibbon, J., Perkins, W. A., Watt-Meyer, O., Bretherton, C. S., & Harris, L. M. (2022). Correcting a 200 km resolution climate model in multiple climates by machine learning from 25 km resolution simulations. *Journal of Advances in Modeling Earth Systems*, 14(9), e2022MS003219.

Dacie, S. (2019). A 1d rce study of factors affecting the tropical tropopause layer and surface climate. *J. Climate*, 32, 6769–6782.

Deardorff, J. W., Willis, G. E., & Stockton, B. H. (1980). Laboratory studies of the entrainment zone of a convectively mixed layer. *J. Fluid Mech.*, 100, 41–64.

Dessler, A. E., Schoeberl, M. R., Wang, T., Davis, S. M., & Rosenlof, K. H. (2013). Stratospheric water vapor feedback. *Proc. Natl. Acad. Sci. USA*, 110, 18 087–18 091.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development*, 9(5), 1937–1958.

Feng, Z., Leung, L. R., Houze, R. A., Hagos, S., Hardin, J., Yang, Q., Han, B., & Fan, J. (2018). Structure and evolution of mesoscale convective systems: Sensitivity to cloud microphysics in convection-permitting simulations over the united states. *J. Adv. Model. Earth Syst.*, 10, 1470–1494.

Fu, Q., Smith, M., & Yang, Q. (2018). The impact of cloud radiative effects on the tropical tropopause layer temperatures. *Atmosphere*, 9, 377.

Fueglistaler, S., Dessler, A. E., Dunkerton, T. J., Folkins, I., Fu, Q., & Mote, P. W. (2009). Tropical tropopause layer. *Rev. Geophys.*, 47.

Fueglistaler, S., Haynes, P. H., & Forster, P. M. (2011). The annual cycle in lower stratospheric temperatures revisited. *Atmos. Chem. Phys.*, 11, 3701–3711.

Gentine, P., Pritchard, M., Rasp, S., Reinaudi, G., & Yacalis, G. (2018). Could machine learning break the convection parameterization deadlock? *Geophysical Research Letters*, 45(11), 5742–5751.

Gettelman, A., Gagne, D. J., Chen, C.-C., Christensen, M., Lebo, Z., Morrison, H., & Gantos, G. (2021). Machine learning the warm rain process. *Journal of Advances in Modeling Earth Systems*, 13(2), e2020MS002268.

Goswami, B. B., Mani, N. J., Mukhopadhyay, P., Waliser, D. E., Benedict, J. J., Maloney, E. D., Khairoutdinov, M., & Goswami, B. (2011). Monsoon intraseasonal oscillations as simulated by the superparameterized community atmosphere model. *Journal of Geophysical Research: Atmospheres*, 116(D22).

Grabowski, W. W. (2003). Impact of cloud microphysics on convective–radiative quasi equilibrium revealed by cloud-resolving convection parameterization. *Journal of Climate*, 16(21), 3463–3475.

Grabowski, W. W. (2016). Towards global large eddy simulation: Super-parameterization revisited. *Journal of the Meteorological Society of Japan. Ser. II*, 94(4), 327–344.

Grabowski, W. W. & Smolarkiewicz, P. K. (1999). Crcp: A cloud resolving convection parameterization for modeling the tropical convecting atmosphere. *Phys. D: Nonlinear Phenom.*, 133(1-4), 171–178.

Gustafson, W. I., Berg, L. K., Easter, R. C., & Ghan, S. J. (2008). The explicit-cloud parameterized-pollutant hybrid approach for aerosol–cloud interactions in multiscale modeling framework models: tracer transport results. *Environmental Research Letters*, 3(2), 025005.

Han, Y., Zhang, G. J., Huang, X., & Wang, Y. (2020). A moist physics parameterization based on deep learning. *Journal of Advances in Modeling Earth Systems*, 12(9), e2020MS002076.

Han, Y., Zhang, G. J., & Wang, Y. (2023). An ensemble of neural networks for moist physics processes, its generalizability and stable integration. *Journal of Advances in Modeling Earth Systems*, 15(10), e2022MS003508.

Hannah, W. M., Jones, C. R., Hillman, B. R., Norman, M. R., Bader, D. C., Taylor, M. A., Leung, L., Pritchard, M. S., Branson, M. D., Lin, G., et al. (2020). Initial results from the super-parameterized e3sm. *Journal of Advances in Modeling Earth Systems*, 12(1), e2019MS001863.

Hartmann, D. L., Blossey, P. N., & Dygert, B. D. (2019). Convection and climate: What have we learned from simple models and simplified settings? *Curr. Climate Change Rep.*, 5, 196–206.

Hartmann, D. L., Gasparini, B., Berry, S. E., & Blossey, P. N. (2018). The life cycle and net radiative effect of tropical anvil clouds. *J. Adv. Model. Earth Syst.*, 10, 3012–3029.

Hersbach, H. (2020). The era5 global reanalysis. *Quart. J. Roy. Meteor. Soc.*, 146, 1999–2049.

Heuer, H., Schwabe, M., Gentine, P., Giorgetta, M. A., & Eyring, V. (2023). Interpretable multiscale machine learning-based parameterizations of convection for icon. *arXiv preprint arXiv:2311.03251*.

Hohenegger, C., Korn, P., Linardakis, L., Redler, R., Schnur, R., Adamidis, P., Bao, J., Bastin, S., Behravesh, M., Bergemann, M., et al. (2023). Icon-sapphire: simulating the components of the earth system and their interactions at kilometer and subkilometer scales. *Geoscientific Model Development*, 16(2), 779–811.

Hohenegger, C., Kornblueh, L., Klocke, D., Becker, T., Cioni, G., Engels, J. F., Schulzweida, U., & Stevens, B. (2020). Climate statistics in global simulations of the atmosphere, from 80 to 2.5 km grid spacing. *Journal of the Meteorological Society of Japan. Ser. II*, 98(1), 73–91.

Holton, J. R., Haynes, P. H., McIntyre, M. E., Douglass, A. R., Rood, R. B., & Pfister, L. (1995). Stratosphere-troposphere exchange. *Rev. Geophys.*, 33, 403–439.

Hu, Z., Lamraoui, F., & Kuang, Z. (2021). Influence of upper-troposphere stratification and cloud–radiation interaction on convective overshoots in the tropical tropopause layer. *Journal of the Atmospheric Sciences*, 78(8), 2493–2509.

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., & Collins, W. D. (2008). Radiative forcing by long-lived greenhouse gases: Calculations with the aer radiative transfer models. *J. Geophys. Res.*, 113, D13103.

Igel, A. L., Igel, M. R., & van den Heever, S. C. (2015). Make it a double? sobering results from simulations using single-moment microphysics schemes. *Journal of the Atmospheric Sciences*, 72(2), 910–925.

Iglesias-Suarez, F., Gentine, P., Solino-Fernandez, B., Beucler, T., Pritchard, M., Runge, J., & Eyring, V. (2024). Causally-informed deep learning to improve climate models and projections. *Journal of Geophysical Research: Atmospheres*, 129(4), e2023JD039202.

IPCC (2021). *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*.

Jeevanjee, N. (2017). Vertical velocity in the gray zone. *Journal of Advances in Modeling Earth Systems*, 9(6), 2304–2316.

Jeevanjee, N. (2022). Three rules for the decrease of tropical convection with global warming. *Journal of Advances in Modeling Earth Systems*, 14(11), e2022MS003285.

Jeevanjee, N. & Zhou, L. (2022). On the resolution-dependence of anvil cloud fraction and precipitation efficiency in radiative-convective equilibrium. *Journal of Advances in Modeling Earth Systems*, 14(3), e2021MS002759.

Jenney, A., Ferretti, S., & Pritchard, M. (2023). Vertical resolution impacts explicit simulation of deep convection. *Journal of Advances in Modeling Earth Systems*, 15(10), e2022MS003444.

Jucker, M. & Gerber, E. P. (2017). Untangling the annual cycle of the tropical tropopause layer with an idealized moist model. *J. Climate*, 30, 7339–7358.

Kelp, M. M., Jacob, D. J., Lin, H., & Sulprizio, M. P. (2022). An online-learned neural network chemical solver for stable long-term global simulations of atmospheric chemistry. *Journal of Advances in Modeling Earth Systems*, 14(6), e2021MS002926.

Khairoutdinov, M., DeMott, C., & Randall, D. (2008). Evaluation of the simulated interannual and subseasonal variability in an amip-style simulation using the csu multiscale modeling framework. *Journal of Climate*, 21(3), 413–431.

Khairoutdinov, M., Krueger, S. K., Moeng, C.-H., Bogenschutz, P. A., & Randall, D. A. (2009). Large-eddy simulation of maritime deep tropical convection. *Journal of Advances in Modeling Earth Systems*, 1(4).

Khairoutdinov, M., Randall, D., & DeMott, C. (2005). Simulations of the atmospheric general circulation using a cloud-resolving model as a superparameterization of physical processes. *Journal of the Atmospheric Sciences*, 62(7), 2136–2154.

Khairoutdinov, M. & 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.

Khairoutdinov, M. F. & Randall, D. A. (2001). A cloud resolving model as a cloud parameterization in the ncar community climate system model: Preliminary results. *Geophysical Research Letters*, 28(18), 3617–3620.

Kobayashi, C. & Iwasaki, T. (2016). Brewer-dobson circulation diagnosed from jra-55. *J. Geophys. Res. Atmos.*, 121, 1493–1510.

Kochkov, D., Yuval, J., Langmore, I., Norgaard, P., Smith, J., Mooers, G., Klöwer, M., Lottes, J., Rasp, S., Düben, P., et al. (2024). Neural general circulation models for weather and climate. *Nature*, (pp. 1–7).

Kuang, Z. (2024). Linear time-invariant models of a large cumulus ensemble. *Journal of the Atmospheric Sciences*, 81(3), 605–627.

Kuang, Z. & Bretherton, C. S. (2004). Convective influence on the heat balance of the tropical tropopause layer: A cloud-resolving model study. *J. Atmos. Sci.*, 61, 2919–2927.

Kuang, Z. & Bretherton, C. S. (2006). A mass-flux scheme view of a high-resolution simulation of a transition from shallow to deep cumulus convection. *Journal of the Atmospheric Sciences*, 63(7), 1895–1909.

Kuang, Z. & Hartmann, D. L. (2007). Testing the fixed anvil temperature hypothesis in a cloud-resolving model. *Journal of Climate*, 20(10), 2051–2057.

Kubokawa, H., Fujiwara, M., Nasuno, T., Miura, M., Yamamoto, M. K., & Satoh, M. (2012). Analysis of the tropical tropopause layer using the nonhydrostatic icosahedral atmospheric model (nicam): 2. an experiment under the atmospheric conditions of december 2006 to january 2007.

Kühbacher, B., Iglesias-Suarez, F., Kilbertus, N., & Eyring, V. (2024). Towards physically consistent deep learning for climate model parameterizations. *arXiv preprint arXiv:2406.03920*.

Kwa, A., Clark, S. K., Henn, B., Brenowitz, N. D., McGibbon, J., Watt-Meyer, O., Perkins, W. A., Harris, L., & Bretherton, C. S. (2023). Machine-learned climate model corrections from a global

storm-resolving model: Performance across the annual cycle. *J. Adv. Model. Earth Syst.*, 15(5), e2022MS003400.

Küpper, C., Thuburn, J., Craig, G. C., & Birner, T. (2004). Mass and water transport into the tropical stratosphere: A cloud-resolving simulation. *J. Geophys. Res.*, 109, D10111.

Lawson, R. P. & Cooper, W. A. (1990). Performance of some airborne thermometers in clouds. *Journal of Atmospheric and Oceanic Technology*, 7(3), 480–494.

Li, F., Rosa, D., Collins, W. D., & Wehner, M. F. (2012). "super-parameterization": A better way to simulate regional extreme precipitation? *Journal of Advances in Modeling Earth Systems*, 4(2).

Lin, J., Yu, S., Peng, L., Beucler, T., Wong-Toi, E., Hu, Z., Gentine, P., Geleta, M., & Pritchard, M. (2024). Sampling hybrid climate simulation at scale to reliably improve machine learning parameterization. *arXiv preprint arXiv:2309.16177v2*.

Lin, P., Paynter, D., Ming, Y., & Ramaswamy, V. (2017). Changes of the tropical tropopause layer under global warming. *J. Climate*, 30, 1245–1258.

Lopez-Gomez, I., Christopoulos, C., Langeland Ervik, H. L., Dunbar, O. R., Cohen, Y., & Schneider, T. (2022). Training physics-based machine-learning parameterizations with gradient-free ensemble kalman methods. *Journal of Advances in Modeling Earth Systems*, 14(8), e2022MS003105.

Ma, H.-Y., Klein, S. A., Lee, J., Ahn, M.-S., Tao, C., & Gleckler, P. J. (2022). Superior daily and sub-daily precipitation statistics for intense and long-lived storms in global storm-resolving models. *Geophysical Research Letters*, 49(8), e2021GL096759.

Manabe, S. & Strickler, R. F. (1964). Thermal equilibrium of the atmosphere with a convective adjustment. *Journal of Atmospheric Sciences*, 21(4), 361–385.

Ming, A., Maycock, A. C., Hitchcock, P., & Haynes, P. (2017). The radiative role of ozone and water vapour in the annual temperature cycle in the tropical tropopause layer. *Atmos. Chem. Phys.*, 17, 5677–5701.

Miyamoto, Y., Kajikawa, Y., Yoshida, R., Yamaura, T., Yashiro, H., & Tomita, H. (2013). Deep moist atmospheric convection in a subkilometer global simulation. *Geophysical Research Letters*, 40(18), 4922–4926.

Mooers, G., Pritchard, M., Beucler, T., Ott, J., Yacalis, G., Baldi, P., & Gentine, P. (2021). Assessing the potential of deep learning for emulating cloud superparameterization in climate models with real-geography boundary conditions. *Journal of Advances in Modeling Earth Systems*, 13(5), e2020MS002385.

Morrison, H., Curry, J., & Khvorostyanov, V. (2005). A new double-moment microphysics parameterization for application in cloud and climate models. part i: Description. *Journal of the atmospheric sciences*, 62(6), 1665–1677.

Mote, P. W. (1996). An atmospheric tape recorder: The imprint of tropical tropopause temperatures on stratospheric water vapor. *J. Geophys. Res.*, 101, 3989–4006.

Norman, M. R., Bader, D. C., Eldred, C., Hannah, W. M., Hillman, B. R., Jones, C. R., Lee, J. M., Leung, L., Lyngaas, I., Pressel, K. G., et al. (2022). Unprecedented cloud resolution in a gpu-enabled full-physics atmospheric climate simulation on olcf's summit supercomputer. *Int. J. High Perform. Compu. Appl.*, 36(1), 93–105.

NVIDIA Modulus Team (2025). Nvidia modulus [software]. https://github.com/NVIDIA/modulus/tree/main.

Ohno, T. & Satoh, M. (2018). Roles of cloud microphysics on cloud responses to sea surface temperatures in radiative-convective equilibrium experiments using a high-resolution global nonhydrostatic model. *Journal of Advances in Modeling Earth Systems*, 10(8), 1970–1989.

Ohno, T., Satoh, M., & Noda, A. (2019). Fine vertical resolution radiative-convective equilibrium experiments: Roles of turbulent mixing on the high-cloud response to sea surface temperatures. *Journal of Advances in Modeling Earth Systems*, 11(6), 1637–1654.

Ott, J., Pritchard, M., Best, N., Linstead, E., Curcic, M., & Baldi, P. (2020). A fortran-keras deep learning bridge for scientific computing. arxiv:2004.10652.

Pahlavan, H. A., Hassanzadeh, P., & Alexander, M. J. (2024). Explainable offline-online training of neural networks for parameterizations: A 1d gravity wave-qbo testbed in the small-data regime. *Geophysical Research Letters*, 51(2), e2023GL106324.

Parishani, H., Pritchard, M. S., Bretherton, C. S., Terai, C. R., Wyant, M. C., Khairoutdinov, M., & Singh, B. (2018). Insensitivity of the cloud response to surface warming under radical changes to boundary layer turbulence and cloud microphysics: Results from the ultraparameterized cam. *Journal of Advances in Modeling Earth Systems*, 10(12), 3139–3158.

Parishani, H., Pritchard, M. S., Bretherton, C. S., Wyant, M. C., & Khairoutdinov, M. (2017). Toward low-cloud-permitting cloud superparameterization with explicit boundary layer turbulence. *Journal of Advances in Modeling Earth Systems*, 9(3), 1542–1571.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. (2019). Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32.

Peng, L., Blossey, P. N., Hannah, W. M., Bretherton, C. S., Terai, C. R., Jenney, A. M., & Pritchard, M. (2024). Improving stratocumulus cloud amounts in a 200-m resolution multi-scale modeling framework through tuning of its interior physics. *Journal of Advances in Modeling Earth Systems*, 16(3), e2023MS003632.

Powell, S. W., Houze, R. A., Kumar, A., & McFarlane, S. A. (2012). Comparison of simulated and observed continental tropical anvil clouds and their radiative heating profiles. *Journal of the Atmospheric Sciences*, 69(9), 2662–2681.

Pritchard, M. S. & Somerville, R. C. (2009a). Assessing the diurnal cycle of precipitation in a multi-scale climate model. *Journal of Advances in Modeling Earth Systems*, 1(4).

Pritchard, M. S. & Somerville, R. C. (2009b). Empirical orthogonal function analysis of the diurnal cycle of precipitation in a multi-scale climate model. *Geophysical Research Letters*, 36(5).

PyTorch Contributors (2024). Torchscript documentation [software].

Randall, D. A. (2013). Beyond deadlock. *Geophys. Res. Lett.*, 40(22), 5970–5976.

Randel, W. J. & Jensen, E. J. (2013). Physical processes in the tropical tropopause layer and their roles in a changing climate. *Nat. Geosci.*, 6, 169–176.

Randel, W. J. & Park, M. (2019). Diagnosing observed stratospheric water vapor relationships to the cold point tropical tropopause.

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*, 115(39), 9684–9689.

Robinson, F. J. & Sherwood, S. C. (2006). Modeling the impact of convective entrainment on the tropical tropopause. *J. Atmos. Sci.*, 63, 1013–1027.

Romps, D. M. (2014). An analytical model for tropical relative humidity. *Journal of Climate*, 27(19), 7432–7449.

Romps, D. M. (2021). Ascending columns, wtg, and convective aggregation. *Journal of the Atmospheric Sciences*, 78(2), 497–508.

Romps, D. M. & Charn, A. B. (2015). Sticky thermals: Evidence for a dominant balance between buoyancy and drag in cloud updrafts. *Journal of the Atmospheric Sciences*, 72(8), 2890–2901.

Ross, A., Li, Z., Perezhogin, P., Fernandez-Granda, C., & Zanna, L. (2023). Benchmarking of machine learning ocean subgrid parameterizations in an idealized model. *Journal of Advances in Modeling Earth Systems*, 15(1).

Ross, S., Gordon, G., & Bagnell, D. (2011). A reduction of imitation learning and structured prediction to no-regret online learning. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics* (pp. 627–635).: JMLR Workshop and Conference Proceedings.

Sanford, C., Kwa, A., Watt-Meyer, O., Clark, S. K., Brenowitz, N., McGibbon, J., & Bretherton, C. (2023a). Improving the reliability of ml-corrected climate models with novelty detection. *Journal of Advances in Modeling Earth Systems*, 15(11), e2023MS003809.

Sanford, C. H., Kwa, A., Watt-Meyer, O., Clark, S. K., Brenowitz, N. D., McGibbon, J., & Bretherton, C. S. (2023b). Improving the reliability of ml-corrected climate models with novelty detection. *Authorea Preprints*.

Satoh, M., Stevens, B., Judt, F., Khairoutdinov, M., Lin, S.-J., Putman, W. M., & Düben, P. (2019). Global cloud-resolving models. *Current Climate Change Reports*, 5, 172–184.

Seeley, J. T., Jeevanjee, N., Langhans, W., & Romps, D. M. (2019a). Formation of tropical anvil clouds by slow evaporation. *Geophysical Research Letters*, 46(1), 492–501.

Seeley, J. T., Jeevanjee, N., & Romps, D. M. (2019b). Fat or fitt: Are anvil clouds or the tropopause temperature invariant? *Geophysical Research Letters*, 46(3), 1842–1850.

Shamekh, S., Lamb, K. D., Huang, Y., & Gentine, P. (2023). Implicit learning of convective organization explains precipitation stochasticity. *Proceedings of the National Academy of Sciences*, 120(20), e2216158120.

Sherwood, S., Webb, M. J., Annan, J. D., Armour, K. C., Forster, P. M., Hargreaves, J. C., Hegerl, G., Klein, S. A., Marvel, K. D., Rohling, E. J., et al. (2020). An assessment of earth's climate sensitivity using multiple lines of evidence. *Rev. Geophys.*, 58(4), e2019RG000678.

Sherwood, S. C., Bony, S., & Dufresne, J.-L. (2014). Spread in model climate sensitivity traced to atmospheric convective mixing. *Nature*, 505(7481), 37–42.

Sherwood, S. C., Horinouchi, T., & Zeleznik, H. A. (2003). Convective impact on temperatures observed near the tropical tropopause. *J. Atmos. Sci.*, 60, 1847–1856.

Siebesma, A. & Jonker, H. (2000). Anomalous scaling of cumulus cloud boundaries. *Physical review letters*, 85(1), 214.

Singh, M. S. & Neogi, S. (2022). On the interaction between moist convection and large-scale ascent in the tropics. *Journal of Climate*, 35(14), 4417–4435.

Singh, M. S. & O'Gorman, P. A. (2013). Influence of entrainment on the thermal stratification in simulations of radiative-convective equilibrium. *Geophysical Research Letters*, 40(16), 4398–4403.

Singh, M. S., Warren, R. A., & Jakob, C. (2019). A steady-state model for the relationship between humidity, instability, and precipitation in the tropics. *Journal of Advances in Modeling Earth Systems*, 11(12), 3973–3994.

Song, Y., Sohl-Dickstein, J., Kingma, D. P., Kumar, A., Ermon, S., & Poole, B. (2020). Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*.

Stan, C., Khairoutdinov, M., DeMott, C. A., Krishnamurthy, V., Straus, D. M., Randall, D. A., Kinter III, J. L., & Shukla, J. (2010). An ocean-atmosphere climate simulation with an embedded cloud resolving model. *Geophysical Research Letters*, 37(1).

Stevens, B., Satoh, M., Auger, L., Biercamp, J., Bretherton, C. S., Chen, X., Düben, P., Judt, F., Khairoutdinov, M., Klocke, D., et al. (2019). Dyamond: the dynamics of the atmospheric general circulation modeled on non-hydrostatic domains. *Progress in Earth and Planetary Science*, 6(1), 1–17.

Taylor, M., Caldwell, P. M., Bertagna, L., Clevenger, C., Donahue, A., Foucar, J., Guba, O., Hillman, B., Keen, N., Krishna, J., et al. (2023). The simple cloud-resolving e3sm atmosphere model running on the frontier exascale system. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis* (pp. 1–11).

Terai, C., Pritchard, M., Blossey, P., & Bretherton, C. (2020). The impact of resolving subkilometer processes on aerosol-cloud interactions of low-level clouds in global model simulations. *Journal of Advances in Modeling Earth Systems*, 12(11), e2020MS002274.

Thompson, G., Field, P. R., Rasmussen, R. M., & Hall, W. D. (2008). Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. part ii: Implementation of a new snow parameterization. *Mon. Wea. Rev.*, 136, 5095–5115.

Thuburn, J. & Craig, G. C. (2002). On the temperature structure of the tropical substratosphere. *J. Geophys. Res.*, 107, 4017.

Tulich, S. (2015). A strategy for representing the effects of convective momentum transport in multiscale models: Evaluation using a new superparameterized version of the weather research and forecast model (sp-wrf). *Journal of Advances in Modeling Earth Systems*, 7(2), 938–962.

Voigt, A., Albern, N., & Papavasileiou, G. (2019). The atmospheric pathway of the cloud-radiative impact on the circulation response to global warming: Important and uncertain. *J. Climate*, 32, 3051–3067.

Wang, M., Ghan, S., Ovchinnikov, M., Liu, X., Easter, R., Kassianov, E., Qian, Y., & Morrison, H. (2011). Aerosol indirect effects in a multi-scale aerosol-climate model pnnl-mmf. *Atmospheric Chemistry and Physics*, 11(11), 5431–5455.

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

Wang, X., Han, Y., Xue, W., Yang, G., & Zhang, G. J. (2022b). Stable climate simulations using a realistic general circulation model with neural network parameterizations for atmospheric moist physics and radiation processes. *Geoscientific Model Development*, 15(9), 3923–3940.

Wing, A. A., Emanuel, K., Holloway, C. E., & Muller, C. (2018a). Convective self-aggregation in numerical simulations: A review. *Shallow clouds, water vapor, circulation, and climate sensitivity*, (pp. 1–25).

Wing, A. A., Reed, K. A., Satoh, M., Stevens, B., Bony, S., & Ohno, T. (2018b). Radiative–convective equilibrium model intercomparison project. *Geoscientific Model Development*, 11(2), 793–813.

Wing, A. A., Stauffer, C. L., Becker, T., Reed, K. A., Ahn, M.-S., Arnold, N. P., Bony, S., Branson, M., Bryan, G. H., Chaboureau, J.-P., et al. (2020). Clouds and convective self-aggregation in a multimodel ensemble of radiative-convective equilibrium simulations. *Journal of Advances in Modeling Earth Systems*, 12(9), e2020MS002138.

Yang, Q., Fu, Q., Austin, J., Gettelman, A., Li, F., & Vömel, H. (2008). Observationally derived and general circulation model simulated tropical stratospheric upward mass fluxes. *J. Geophys. Res.*, 113.

Yang, Q., Hannah, W. M., & Leung, L. R. (2022). Convective momentum transport and its impact on the madden-julian oscillation in e3sm-mmf. *Journal of Advances in Modeling Earth Systems*, 14(11), e2022MS003206.

Yoshida, K., Mizuta, R., & Arakawa, O. (2018). Intermodel differences in upwelling in the tropical tropopause layer among cmip5 models. *J. Geophys. Res. Atmos.*, 123, 13 658–13 675.

Yu, S., Hannah, W., Peng, L., Lin, J., Bhouri, M. A., Gupta, R., Lütjens, B., Will, J. C., Behrens, G., Busecke, J., et al. (2024a). Climsim: A large multi-scale dataset for hybrid physics-ml climate emulation. *Advances in Neural Information Processing Systems*, 36.

Yu, S., Hu, Z., Subramaniam, A., Hannah, W., Peng, L., Lin, J., Bhouri, M. A., Gupta, R., Lütjens, B., Will, J. C., Behrens, G., Busecke, J. J. M., Loose, N., Stern, C. I., Beucler, T., Harrop, B., Heuer, H., Hillman, B. R., Jenney, A., Liu, N., White, A., Zheng, T., Kuang, Z., Ahmed, F., Barnes, E., Brenowitz, N. D., Bretherton, C., Eyring, V., Ferretti, S., Lutsko, N., Gentine, P., Mandt, S., Neelin, J. D., Yu, R., Zanna, L., Urban, N., Yuval, J., Abernathey, R., Baldi, P., Chuang, W., Huang, Y., Iglesias-Suarez, F., Jantre, S., Ma, P.-L., Shamekh, S., Zhang, G., & Pritchard, M. (2024b). Climsim-online: A large multi-scale dataset and framework for hybrid ml-physics climate emulation.

Yulaeva, E., Holton, J. R., & Wallace, J. M. (1994). On the cause of the annual cycle in tropical lower-stratospheric temperatures. *J. Atmos. Sci.*, 51, 169–174.

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.

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), 3295.

Zelinka, M. D., Myers, T. A., McCoy, D. T., Po-Chedley, S., Caldwell, P. M., Ceppi, P., Klein, S. A., & Taylor, K. E. (2020). Causes of higher climate sensitivity in cmip6 models. *Geophysical Research Letters*, 47(1), e2019GL085782.

Zhao, M., Golaz, J.-C., Held, I. M., Ramaswamy, V., Lin, S.-J., Ming, Y., Ginoux, P., Wyman, B., Donner, L., Paynter, D., et al. (2016). Uncertainty in model climate sensitivity traced to representations of cumulus precipitation microphysics. *Journal of Climate*, 29(2), 543–560.

Zhou, W. & Xie, S. P. (2019). A conceptual spectral plume model for understanding tropical temperature profile and convective updraft velocities. *Journal of the Atmospheric Sciences*, 76(9), 2801–2814.