Supporting Information for "Probing the Skill of Random Forest Emulators for Physical Parameterizations via a Hierarchy of Simple CAM6 Configurations"

Garrett C. Limon[1], Christiane Jablonowski[1]

[1]Department of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI

Contents of this file

Text S1. Aquaplanet Details

The aquaplanet configuration was used to inform parameter choices for the BM convection scheme discussed in section 2.1. An aquaplanet is an ocean-covered model with prescribed sea surface temperatures (SST) in which the exchange of heat and moisture between the ocean and the atmosphere provides additional quasi-realistic atmospheric fluid flow. It is a widely used configuration for simplified physics studies of GCMs. We used the aquaplanet configuration with the older CAM4 physics package with the CONTROL SST profile configuration described in Neale and Hoskins (2000) to guide our choice of $\mathrm{RH}_{\mathrm{BM}}$ and τ in the BM scheme (Neale et al., 2010). Zonal-mean, time-mean fields for various model output fields comparing the aquaplanet and the convection scheme are shown in Figs. (S1) and (S2) and were used to inform our decision for the chosen parameters.

While we acknowledge that these two cases are not intended to appear identical when compared, there are many fields in which the results are quite similar. In particular, the temperature, specific humidity, relative humidity, zonal wind, and precipitation rates share many similarities in their averaged profiles. The tendencies in Figs. (S1 d & e) and (S2 d & e) are where the stark differences arise, however this is expected as they are significantly different physics schemes. These cases are run at the same spatial and temporal resolution as all other CAM model runs in this work. Since the CONTROL case for the aquaplanet setup in CAM4 is not the default setup, we note here that the compset 'long name' format is "2000_CAM40_SLND_SICE_DOCN%AQP1_SROF_SGLC_SWAV" and is needed to reproduce Fig. (S1).

Text S2. Machine Learning Hyperparameter Tuning

Parameters like the number of trees in a random forest, the number of training samples, as well as the maximum leaves on a branch are all examples of a ML hyperparameter. These can impact the effectiveness of our random forest as emulators. The majority of these were chosen via the SHERPA hyperparameter optimization library. Tables S1 to S8 show the hyperparameter choices for the various random forest emulators. The symbols RELHUM, LHFLX, and SHFLX stand for the relative humidity, surface latent heat flux, and surface sensible heat flux, respectively. For further details on these parameters and how they work to impact the overall model, we direct the reader to the SciKit-Learn documentation (Pedregosa et al., 2011).

Figures

Tables

References

Neale, R. B., & Hoskins, B. J. (2000). A standard test for AGCMs including their physical parameterizations: I: The proposal. *Atmos. Sci. Lett.*, *1*, 101–107.

Neale, R. B., Richter, J. H., Conley, A. J., Park, S., Lauritzen, P. H., Gettelman, A., … Lin, S.-J. (2010, April). *Description of the NCAR Community Atmosphere Model (CAM 4.0)* (NCAR Technical Note Nos. NCAR/TN-485+STR). Boulder, Colorado: National Center for Atmospheric Research. (212 pp., available from http://www.cesm.ucar.edu/models/cesm1.0/cam/)

Pedregosa, F., Varoquaux, G., Gramfort, A., & Michel, V. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*(Oct), 2825–2830.

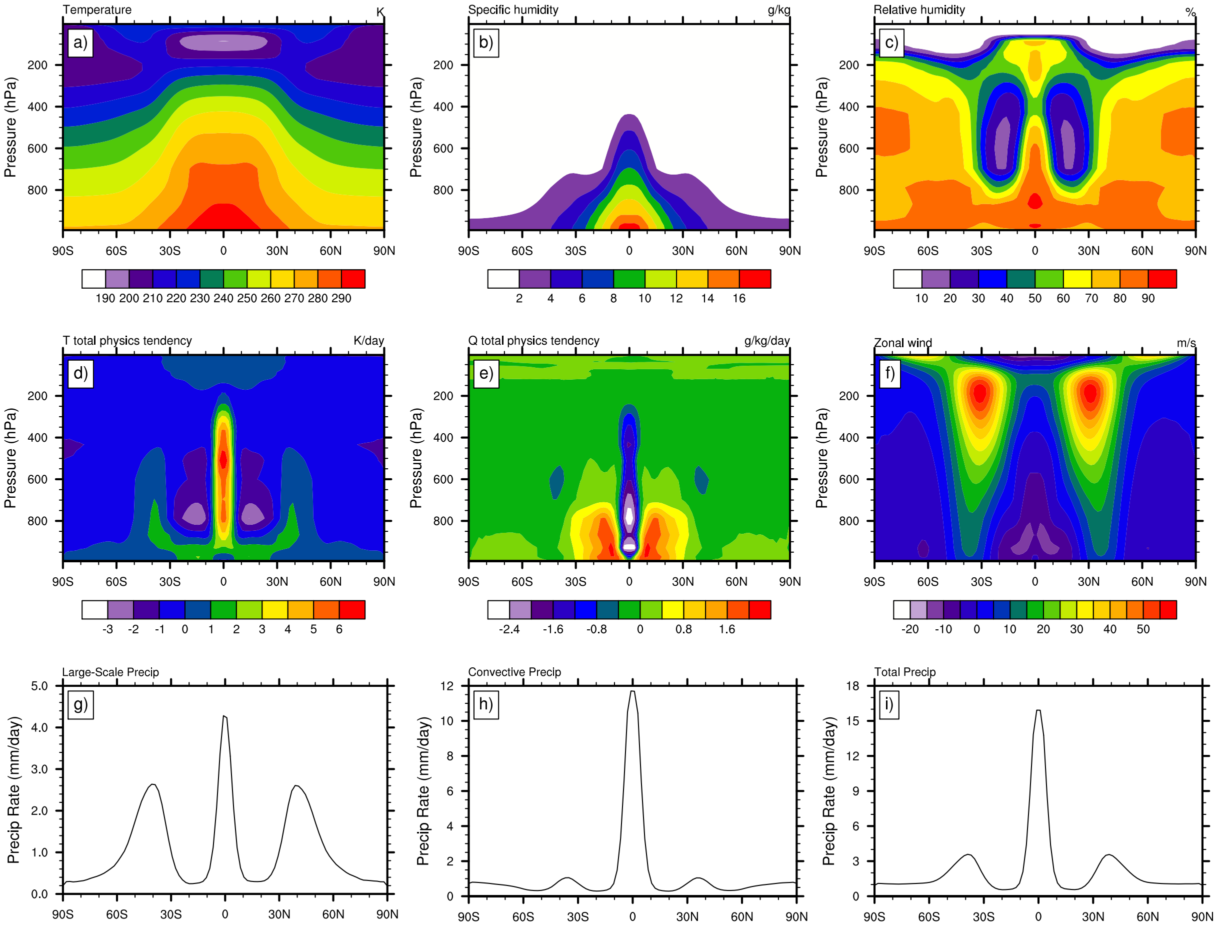

Figure S1. Zonal-mean time-mean panel of (a) temperature, (b) specific humidity, (c) relative humidity, (d) temperature tendency, (e) moisture tendency, (f) zonal wind, (g) large-scale precipitation, (h) convective precipitation, (i) total precipitation rate for the CAM4 aquaplanet setup with the CONTROL SST profile.

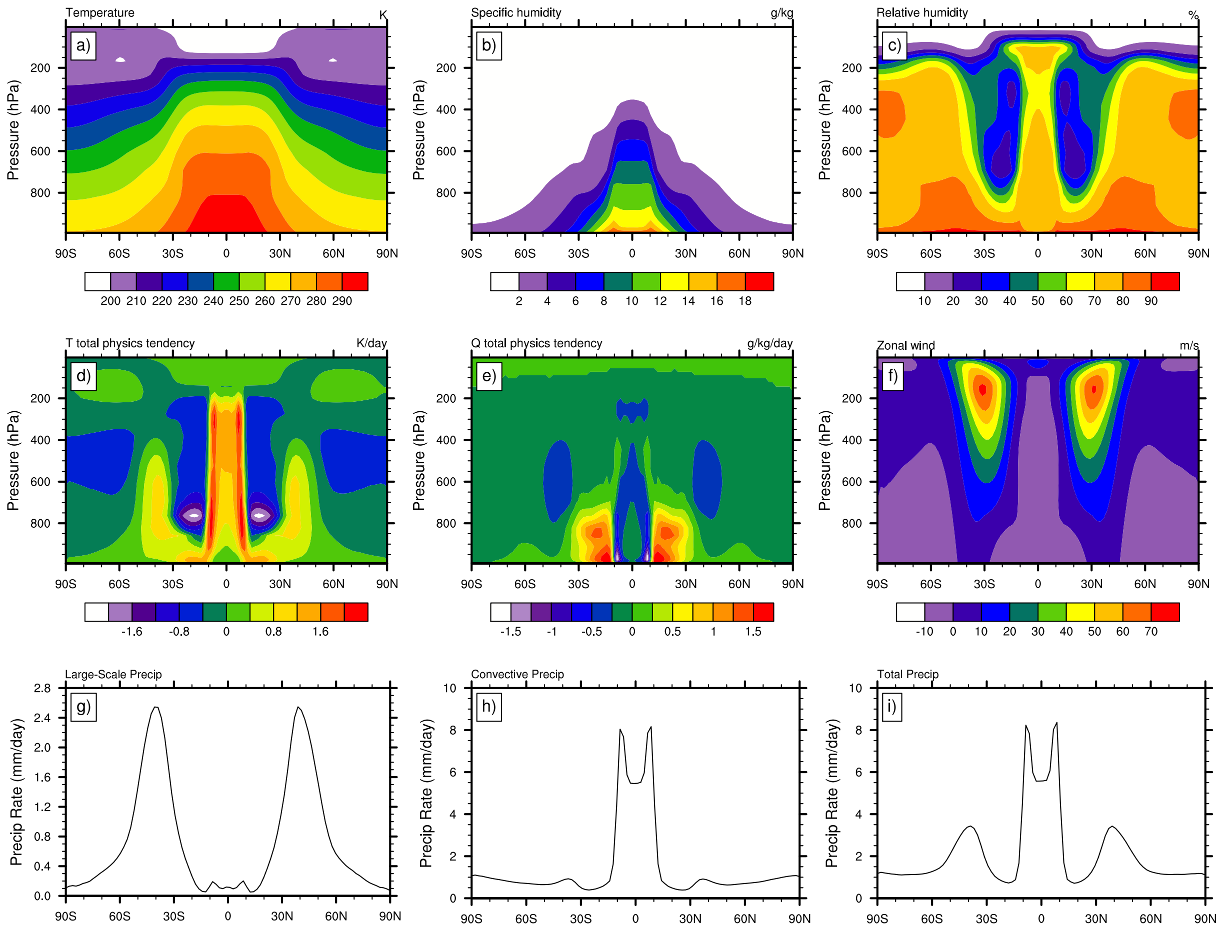

Figure S2. Zonal-mean time-mean panel of (a) temperature, (b) specific humidity, (c) relative humidity, (d) temperature tendency, (e) moisture tendency, (f) zonal wind, (g) large-scale precipitation, (h) convective precipitation, (i) total precipitation rate for the TJ16 configuration in CAM6 coupled with the BM convection scheme with $\tau = 4$ hr and $\mathrm{RH_{BM}} = 0.7$.

Table S1. Dry dT/dt Hyperparameters

ML Option	Choice
Input Variables	T, p, ϕ
Number of Samples	23 Million
Number of Trees	157
Max Depth	39
Min Samples Split	17
Min Samples Leaf	6

Table S2. Moist dT/dt Hyperparameters

ML Option	Choice
Input Variables	T, p, q, RELHUM, LHFLX, SHFLX
Number of Samples	12 Million
Number of Trees	250
Max Depth	30
Min Samples Split	20
Min Samples Leaf	15

Table S3. Convection dT/dt Hyperparameters

ML Option	Choice
Input Variables	T, p, q, RELHUM, LHFLX, SHFLX
Number of Samples	15 Million
Number of Trees	269
Max Depth	22
Min Samples Split	23
Min Samples Leaf	18

Table S4. Moist dq/dt Hyperparameters

ML Option	Choice
Input Variables	T, p, q, RELHUM, LHFLX, SHFLX
Number of Samples	20 Million
Number of Trees	83
Max Depth	32
Min Samples Split	45
Min Samples Leaf	13

Table S5. Convection dq/dt Hyperparameters

ML Option	Choice
Input Variables	T, p, q, RELHUM, LHFLX, SHFLX
Number of Samples	20 Million
Number of Trees	98
Max Depth	32
Min Samples Split	19
Min Samples Leaf	17

Table S6. Moist Large-Scale Precip. Hyperparameters

ML Option	Choice
Input Variables	T, p, q, RELHUM, LHFLX, SHFLX
Number of Samples	20 Million
Number of Trees	395
Max Depth	30
Min Samples Split	28
Min Samples Leaf	6

Table S7. Convection Large-Scale Precip. Hyperparameters

ML Option	Choice
Input Variables	T, p, q, RELHUM, LHFLX, SHFLX
Number of Samples	20 Million
Number of Trees	395
Max Depth	30
Min Samples Split	28
Min Samples Leaf	6

Table S8. Convection Convective Precip. Hyperparameters

ML Option	Choice
Input Variables	T, p, q, RELHUM, LHFLX, SHFLX
Number of Samples	20 Million
Number of Trees	137
Max Depth	37
Min Samples Split	2
Min Samples Leaf	11