Supporting Information for "Probing the Skill of RF 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 Figures S1 and S2 and were used to inform our decision for the chosen parameters.

While we acknowledge that these two cases are not identical, there are many fields with similar flow characteristics. In particular, the temperature, specific humidity, relative humidity, zonal wind, and precipitation rates share many similarities in their averaged profiles. The physical tendencies in Figures S1d,e and S2d,e display greater differences. However, this is expected as the complexity of the physical parameterizations differs. All cases are run at the same 1.9 $\times$ 2.5 degree spatial resolution with 30 model levels. 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".

This is needed to reproduce Figure S1.

Text S2. Machine Learning Hyperparameter Tuning

Parameters like the number of trees in an RF, the number of training samples, as well as the choice of activation functions in a neural network are examples of hyperparameters. These impact the effectiveness of the emulators. The majority of the RF parameters for this study were chosen via the SHERPA hyperparameter optimization library. Tables S1 to S8 show the hyperparameter choices for the various RF emulators. For further details on the RF parameters and how they work to impact the overall model, we direct the reader to the SciKit-Learn documentation (Pedregosa et al., 2011). We also show choices for the neural network setups in able S9, all of which were informed by Beucler et al. (2021). Each field uses an identical setup, however precipitation rates use a *sigmoid* activation (rather than *tanh*) on the final layer in order to enforce positive-definite solutions. Our NNs also use Keras' Normalization layer for our features in order to transform the input to be unitarily invariant, see Keras documentation for further information on this normalization process (Chollet, 2017). The symbols RELHUM, LHFLX, and SHFLX stand for the relative humidity, surface latent heat flux, and surface sensible heat flux, respectively. We note that upon review we found that reducing the number of trees in our RFs from the SHERPA suggestion down to 50 trees across each configuration did not noticeably impact our results. Therefore, we kept the number of trees consistent across all RF models at 50 trees.

Figures

Tables

References

Beucler, T., Pritchard, M., Rasp, S., Ott, J., Baldi, P., & Gentine, P. (2021). Enforcing analytic constraints in neural-networks emulating physical systems. *Phys. Rev. Lett.*, *126*, 098302. doi: 10.1103/PhysRevLett.126.098302

Chollet, F. (2017). *Deep Learning with Python* (1st ed.). USA: Manning Publications Co., 384 pages.

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.

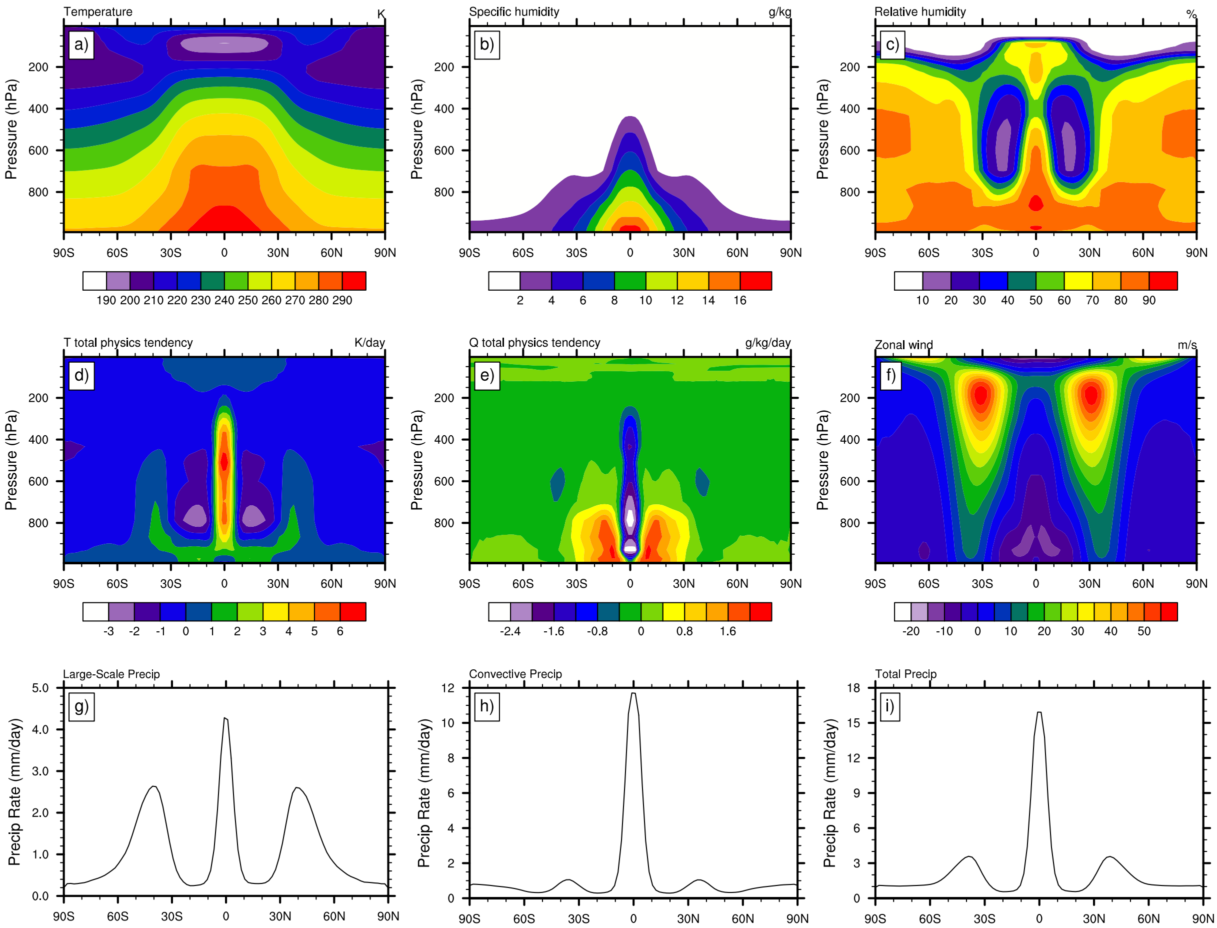

Figure S1. Zonal-mean time-mean panel of (a) temperature, (b) specific humidity, (c) relative humidity, (d) temperature tendency, (e) moisture tendency, (f) zonal wind, (g) large-scale precipitation, (h) convective precipitation, (i) total precipitation rate for the CAM4 aquaplanet setup with the CONTROL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

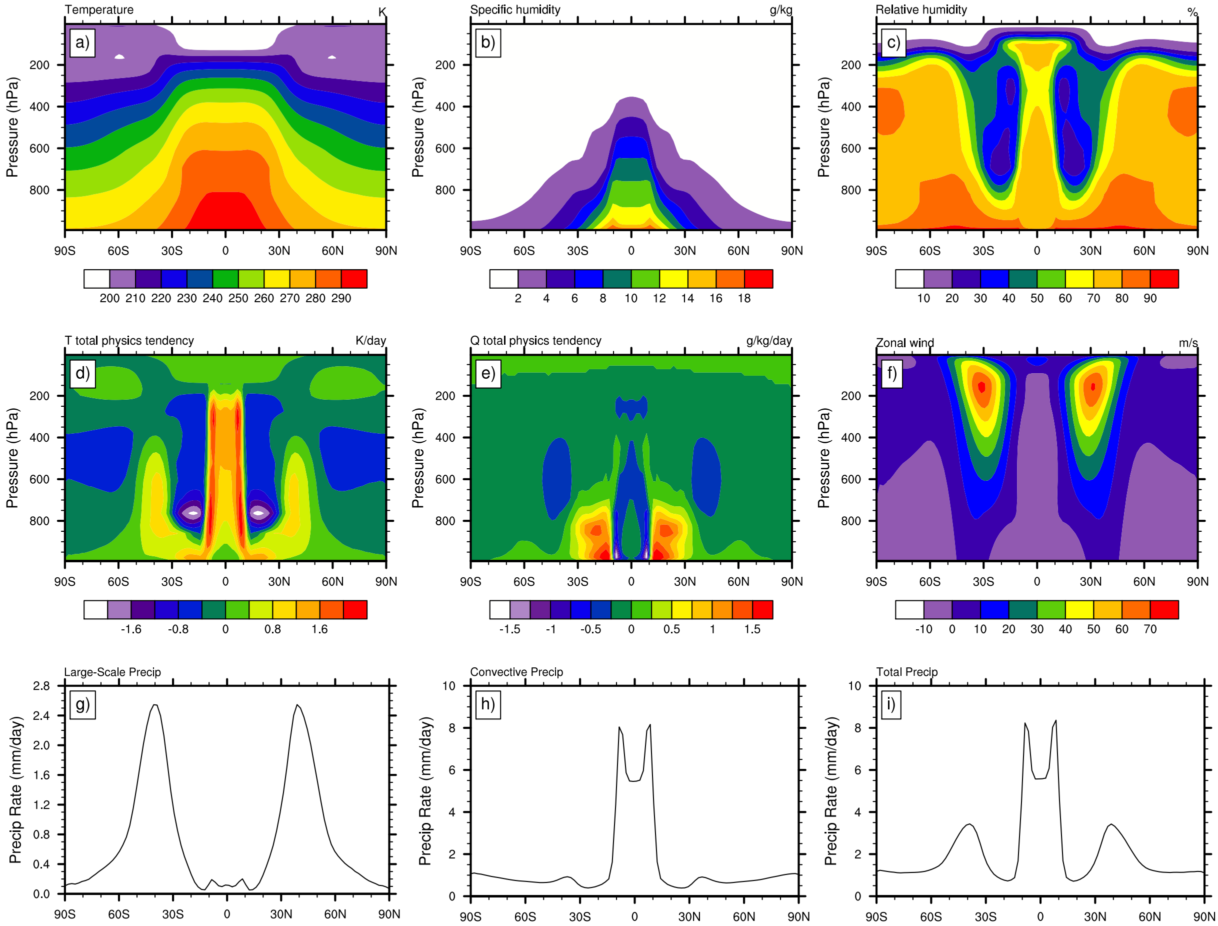

Figure S2. Zonal-mean time-mean panel of (a) temperature, (b) specific humidity, (c) relative humidity, (d) temperature tendency, (e) moisture tendency, (f) zonal wind, (g) large-scale precipitation, (h) convective precipitation, (i) total precipitation rate for the TJ16 configuration in CAM6 coupled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

Table S1. Dry dT/dt Hyperparameters

RF Option	Choice
Input Variables	T, p, ϕ
Number of Samples	20 Million
Number of Trees	50
Max Depth	39
Min Samples Split	17
Min Samples Leaf	6

Table S2. Moist dT/dt Hyperparameters

RF Option	Choice
Input Variables	T, p, q, RELHUM, LHFLX, SHFLX
Number of Samples	15 Million
Number of Trees	50
Max Depth	30
Min Samples Split	20
Min Samples Leaf	15

Table S3. Convection dT/dt Hyperparameters

RF Option	Choice
Input Variables	T, p, q, RELHUM, LHFLX, SHFLX
Number of Samples	15 Million
Number of Trees	50
Max Depth	22
Min Samples Split	23
Min Samples Leaf	18

Table S4. Moist dq/dt Hyperparameters

RF Option	Choice
Input Variables	T, p, q, RELHUM, LHFLX, SHFLX
Number of Samples	20 Million
Number of Trees	50
Max Depth	30
Min Samples Split	45
Min Samples Leaf	15

Table S5. Convection dq/dt Hyperparameters

RF Option	Choice
Input Variables	T, p, q, RELHUM, LHFLX, SHFLX
Number of Samples	20 Million
Number of Trees	50
Max Depth	32
Min Samples Split	19
Min Samples Leaf	17

Table S6. Moist Large-Scale Precipitation Hyperparameters

RF Option	Choice
Input Variables	T, p, q, RELHUM, LHFLX, SHFLX
Number of Samples	20 Million
Number of Trees	50
Max Depth	30
Min Samples Split	30
Min Samples Leaf	5

Table S7. Convection Large-Scale Precipitation Hyperparameters

RF Option	Choice
Input Variables	T, p, q, RELHUM, LHFLX, SHFLX
Number of Samples	20 Million
Number of Trees	50
Max Depth	30
Min Samples Split	30
Min Samples Leaf	5

Table S8. Convection Convective Precipitation Hyperparameters

RF Option	Choice
Input Variables	T, p, q, RELHUM, LHFLX, SHFLX
Number of Samples	20 Million
Number of Trees	50
Max Depth	37
Min Samples Split	2
Min Samples Leaf	11

Table S9. Neural Netork Setup/Hyperparameters

NN Option	Choice
Input Variables	T, p, q, RELHUM, LHFLX, SHFLX
Number of Samples	12.8 Million
Number of Layers	8
Nodes per Layer	512
Hidden Layer Activation	LeakyReLU ($\alpha = 0.25$)
Output Layer Activation	tanh (sigmoid for precip)
Dropout Rate	0.001
Loss Function	MSE
Batch Size	128
Epochs	15
Optimizer	Adam (learningRate= 0.00001)