Recommendation: Return to Authors for Major Revisions

Overview

The manuscript by Yuya Baba (2026), “Transfer learning-based hybrid machine learning in single-column model of AFES v4”, presents the development of a hybrid machine learning (ML) framework implemented within a single-column configuration of an atmospheric general circulation model. The ML component of the hybrid system appears to reduce model biases in temperature and humidity and therefore shows potential for improving numerical weather prediction skill. The topic is timely and relevant to the advancement of weather and climate modeling.

However, from the perspective of an atmospheric scientist with limited background in machine learning, the presentation is difficult to follow in several key sections. Important methodological details and diagnostics are either insufficiently explained or presented primarily in ML-specific terminology, which may limit accessibility for readers in the atmospheric science community. I therefore recommend a thorough revision to improve clarity, completeness, and accessibility. Major revision is necessary before the manuscript can be considered for publication.

Below are three major comments.

Major Comments

1. Quantification of model biases in physical units

To improve accessibility for atmospheric science readers, I recommend presenting temperature and humidity biases in commonly used physical units (e.g., K and g kg^{-1}), in addition to any normalized error metrics. For example, Figure 2 could show vertical profiles of temperature and specific humidity, along with their corresponding absolute biases, so that the magnitude and vertical structure of improvements are clearly visible.

The use of the “normalized L2 norm” is not immediately intuitive for readers outside the ML community. Before presenting advanced error metrics, it would be helpful to first quantify model improvements in terms of absolute changes in physically meaningful units. This would provide a clearer physical interpretation of the hybrid model’s impact.

2. Clarification of experimental setup and physical consistency
The experimental design requires further clarification. In particular, the manuscript should provide:

- A more detailed description of the physical parameterization schemes used in the SCM;
- The model output frequency;
- The training and evaluation data periods and spin-up time period in SCM simulation (if any);
- Any relevant boundary conditions or forcing specifications.

In addition, it should be explicitly stated whether the ML component is coupled online with the SCM during integration. My understanding is that the ML model is indeed coupled online. If this is the case, modifications to temperature and humidity tendencies could influence cloud formation and radiative fluxes. Given this coupling, it is important to evaluate not only temperature and humidity, but also macrophysical and microphysical cloud properties, as well as radiative quantities if available. Clouds are generally more challenging to simulate than temperature and humidity fields, and improvements in thermodynamic variables may have important implications for cloud processes. An assessment of cloud-related diagnostics would therefore strengthen the physical credibility of the hybrid approach.

3. Clarification of ML inputs and outputs

For the hybrid model training and testing flowcharts (Figure 1), the manuscript would benefit from a clearer indication of:

- Which variables are provided as inputs to the ML model;
- Which quantities are predicted by the ML model;
- How these predicted quantities are reintegrated into the SCM.

A schematic diagram explicitly showing the data flow between the dynamical core, physical parameterizations, and the ML component would greatly improve clarity. For example, a figure similar in structure to Figure 1 of Gettelman et al. (2021) would be helpful in illustrating the hybrid architecture and its coupling strategy.

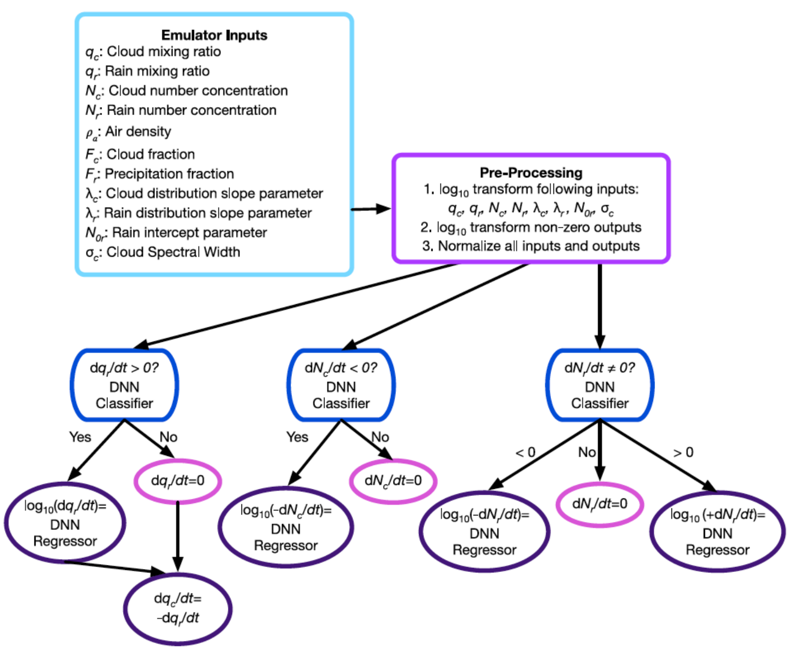

Figure 1. Flow diagram describing the preprocessing and machine learning pipeline for predicting each microphysical tendency.