Review of the preprint entitled "Global Climate Modeling with Improved Precipitation Characteristics by Learning Physics (GRIST-MPS v1.0) from Global Storm-Resolving Modeling" by Y. Wang et al.

General comments

This preprint describes the improved long-term global climate simulations using a general circulation dynamic core coupled with machine-learning (ML)-based physics instead of the conventional parameterization schemes. While economic ML large models have been long doubted for lacking of physical constraints, the dynamical models suffer from expensive computational cost with conventional parameterization or unstable computation with online ML physics schemes (MPS). The present study shows a practical way of coupling the ML-physical processes for stable model integration by selecting balanced spatiotemporal resolution of the training data, besides the evaluation of three neural network architectures. The proposed method effectively ameliorates the precipitation simulation and structure of thermodynamic quantities (T and q) via the improved convective-diabatic interaction, even though no physical feedback to momentum is considered. Descriptions of the methodology and corresponding consummate skill in the MPS implementation, which must be interested and also helpful to researchers in this field, are provided in detail for reference.

The use of high-resolution model output as the training data, being of a result of multi-scale dynamical interaction, helps the neural network to learn detailed spatial structure and temporal variation of the corresponding variables from the fine-modelled data. The induced MPS is then capable of reproducing the high-resolution-model pattern of precipitation and prognostic variables with coarse-resolution GCM, just as show in this preprint.

Beside the quality improvement, the computational cost is also an issue of concern. It is better to add a short note on the cost of MPS vs CPS. In this study, the hybrid coupling of CPS and MPS to dynamical core, in addition to retaining of radiation scheme for the land surface model, seems introduce excessive computation. How about the cost reduction with the MPS?

Another concern issue is the diagnosis of the Q1 and Q2 on pressure levels. The interpolation between the model levels and the pressure levels introduces errors inevitably. Why not diagnose directly on the model levels?

The preprint is generally well written and organized. The technical details must be highly evaluated for the online stable integration. I recommend accept for publication after necessary revisions.

Specific comments

(1) The GRIST model should be mentioned in the abstract.
(2) Please confirm the "precipitation frequency-intensity spectra" in Line 23, may be "precipitation frequency" because no frequency-intensity spectra is shown in the preprint.
(3) Please rephrase the "explicitly resolved global storm-resolving models" in Line

32.

(4) Line 90, please note the abbreviation “PhysW” here.

(5) (Line 120) In case of a different resolution, the use of MPS needs additional training or not? For the specific resolution, the trained MPS may resolve interactions between systems above the doubled grid scale. For finer resolution, however, the trained MPS may miss the processes smaller than training data resolution (0.25 degree). So, is it really scale-aware?

(6) Line 124, is the “GSRM” a typo of “GCM”?

(7) Why not directly output (U, V, T, q, P) data in a 20-minute interval in the GSRM simulation (Lines 142 & 179)? Only due to the limitation of data storage?

(8) Please check the formula of *Prec* in Line 216. What unit is used here?

(9) Please define X, μ and σ in Line 269.

(10) The word “extreme event” is better replaced with “heavy-precipitation event” for easy understanding. The extreme event, out of the training data, is difficult for MPS to resolve.

(11) What is the “weight file” in Line 292.

(12) The sentence on Line 322 needs to be revised. Do you mean: CNN nearly doubled the frequency of light precipitation occurrence in conventional GCMs?

(13) Can the GRIST-MPS stably run more than 6 years?

(14) It is helpful to label the averaged pressure in right side of panels in Fig.2

(15) In Fig.3, temperature difference should be the shaded (being wrongly noted in figure caption). Why different terrain data (blanking or not) in bottom-left corner of the panels are shown?

(16) Fig.8a is slightly different from Fig.3b, why?

Technical corrections

1. Using “from” instead of the “of” in “temperature deviations (shading) within ±5 K of ERA5 reanalysis” may be better, in Line 235.
2. “evolution” in Line 270, and “East Asian Monsoon Index ()” in Line 271 can be removal with the EAMI defined previously.
3. “maintaining comparable spatial RMSE to observations” (in Lines 306-307) ? Is it “maintaining comparable spatial RMSE to CNN”?