

This study uses the machine learning (ML) approach to speed up the radiative transfer computation in the Global Forecasting System of China Meteorological Administration (CMA-GFS). Offline simulations of the Rapid Radiative Transfer Model for general circulation model applications (RRTMG) adopted as the CMA-GFS radiation scheme were performed to generate training and validation output datasets. The authors generated the input datasets by consecutively running 5-day CMA-GFS forecasts with 25 hours increments. The constructed residue convolutional neu network (ResCNN) radiative flux prediction model shows good accuracy. The authors subsequently linked the radiative flux ML emulator to their operational forecast model via the Fortran Torch adapter, resulting in a computational speed 7 times as fast as that with the original RRTMG scheme. They showed that it is unnecessary to use the ML emulator to predict the atmospheric radiative heating, which is of larger uncertainties than if it is calculated using the predicted radiative fluxes at the layer bounds. In addition, to avoid the forecast error blowup due to encountering extreme conditions—such as hurricanes, heat waves, cold surges, etc.—which were not sufficiently represented in the training dataset, the authors further restricted the values of predicted variables to physically reasonable ranges and retrained the ML emulator with more extreme conditions included. The performance of the CMA-GFS model with the ML emulator enabled is comparable to that of the operational CMA-GFS model. The manuscript also includes a clear and informative introduction of the AI applications in weather and climate predictions, which I enjoyed reading. However, the introduction of the radiative transfer models and the detailed treatments therein may be improved. I suggest this study be accepted for publication after a major revision. Below please find my suggestions.

Major Suggestions

Lines 173-175. The statement is hard for me to understand. What does “multi-channel one-dimension vector” mean? As documented in Table 1, input cloud parameters could have dimensions of (ngptlw, ncol, nlay), (ncol, nlay), and (ngptsw, ncol, nlay). What did the authors specifically do to merge the cloud parameters at different g-points (i.e., quadrature points)? What does “6-channel one-dimensional vector” mean? It would be better to be specific.

Line 332. What do you mean by “converging to highly fragile local minima”? Would it be helpful to explain more using mathematical language?

Line 340. Figure 5. Would it be better to add an introduction of what the curves and bars represent in the caption of this figure? It appears that the curves in each panel above $y = 0$ show anomaly correlation, but what do the bars in the lower portion of each panel show?

Line 360. Figure 6. Are the threat scores shown in the figure global or regional results? According to what shown in Figure 8, it seems the scores are computed based on the precipitation forecasts over China? I am not a weather forecaster. This may be commonly known in the weather forecast community, but I am curious why boreal summer (July) precipitation forecasts have higher threat scores than boreal winter (January) precipitation forecasts, particularly for heavy and torrential rain forecasts?

Lines 360-365. Figure 7. Do the biases shown in this figure have units of mm? It appears the thick horizontal red line in each panel marks $y = 1$ (mm?). Would it be better to adjust the y limits of the relevant panels to avoid the overlapping of the red lines with x labels or tick-labels?

Lines 353-354. I assume the authors must have verified the statement about the forecast of precipitation extremes using some metric not presented here. It seems the authors used the purple curves to show the major rivers and lakes in China in the upper panels of Figures 8 and 9. However, according to the colormaps of the two figures, purple also means the regions with accumulated precipitation > 100 mm, making it hard to see that the hybrid model shows a slightly increased consistency with the observation.

Lines 390-395. Figures 10 and 11. What do the two contractions “MAE” and “ME” in the two figures stand for? Are the results shown in the two figures also the forecasts over China?

Lines 406-408. I assume the authors must have verified the statement about the forecast of the hurricane precipitation using some metric not presented here. By eyeballing Figure 14a, it looks that the hybrid model succeeded in forecasting the outer rain belt of the hurricane while the operational model did not?

Lines 415-416 and 430-431. The statements are vague. What and how can satellite observation data be used to improve the performance of the hybrid model in forecasting synoptic fields and 10-meter wind speed? How can LBLRTM be used to achieve this goal? Depending upon how fine the selected wavenumber interval is, LBLRTM could be several orders of magnitude more computational expensive than RRTMG.

Lines 416-418. Yes, you did a great job. I have a suggestion that I do not expect the authors to address in this study. This study already has abundant results. Now, the authors have successfully speeded up the radiation computation in the weather forecast model by a factor of 7. What if the authors increase the frequency of calling the radiation scheme in running the hybrid model? Will weather forecast improve if the radiation scheme is called more frequently? Experiments for answering the questions may be a future study.

Minor Suggestions

Lines 62-64. The SPeedy Algorithm for Radiative TrAnsfer through CloUd Sides (SPARTACUS) solver in *ecRad* is a 3D solver that incorporates 3D cloud radiative effects on grid-box-mean radiative fluxes. The statement here as written reads like the *ecRad* had not have a 3D radiative transfer computational capability until Alves et al. (2023) added one. It would be better to make it clearer.

Line 67. What do you mean by “non-cooling effect of cloud reflection”? Cloud sunlight reflection results in a heating of the cloud top, i.e., a heating effect locally. For example, see the offline radiation simulation study of Ren et al. (2021).

Line 100. Clouds and aerosols in the atmosphere are particles, not gases, but how electromagnetic waves interact with them is important and handled in the atmospheric radiative

transfer computation. Hence, would it be more accurate to say “an absorption and scattering medium” rather than “a gas medium”?

Lines 115 and 120-121. Can the authors be more specific? What is “subgrid-scale cloud variability” is considered in RRTMG but not in RRTM? Correct me if I am wrong, the impact of cloud mass horizontal variability in a cloud-containing model layer on grid-box-mean radiative fluxes is not handled in either RRTM or RRTMG. In other words, cloud is assumed to be homogenous in each cloud-containing model layer in both RRTM and RRTMG. In addition, correct me if I am wrong, random and maximum-random cloud overlapping treatments are available in both RRTM and RRTMG?

Lines 121-122. The McICA method is developed for computational efficiency. What aspects of cloud-radiation interactions does this method improve?

Lines 123-124. Can the authors provide references for this statement? The largest longwave wavelength band of RRTM/RRTMG has a wavenumber range of $10\text{--}350\text{ cm}^{-1}$ or a wavelength range of $28.57\text{--}1000\text{ }\mu\text{m}$. What and how have measured sub-millimeter radiance spectra been used to validate RRTMG?

Lines 165-170. Table 1. I can guess out the meanings of all the dimension name abbreviations except for this one *naerec*. Does it refer to the number of aerosol species? Would it be better to spell out the full names of all the abbreviations somewhere? The two variables *reicmcl* and *relqmcl* in the Inputs table each shows up two times. Would it be better to combine the two rows of the same variable?

Line 175. It is hard to understand “full model layers” and “half layers”. Would it be clearer if 88 model layers and 89 model levels were used instead?

Line 176. “RseCNN” or “ResCNN”?

Line 194. The phrase “machine learning” appears in the main text for the first time in Line 21. Would it be better if the contraction “ML” were introduced at the very beginning?

Line 195. “panel compares” or “panels compare”? In the left panel of each of the two figures, I find only 3 variables, not 6 variables.

Line 313. “becomes” or “become”?

Line 320. Would it be better to use “in” or “of” rather than “at” right be “layer *i*”? The atmospheric heating rate of a layer is computed with the variables at its upper and lower levels (or interfaces) next to the adjacent layers.

Line 373. There is an unwanted backtick symbol “`” before the word “model’s”.

Line 402. “on the afternoon” or “in the afternoon”?

Reference

Ren, T., Li, D., Muller, J., & Yang, P. (2021). Sensitivity of radiative flux simulations to ice cloud parameterization over the equatorial western Pacific Ocean region. *Journal of the Atmospheric Sciences*, 78(8), 2549-2571. <https://doi.org/10.1175/JAS-D-21-0017.1>