# Reply to reviewers’ comments

We thank the reviewer for the time spent on reviewing this manuscript and for providing helpful comments and suggestions.

# Reviewer #2

Thank you to the authors for thoroughly addressing the comments. However, further emphasis on the study's contribution to the community would strengthen the manuscript.

1. The authors acknowledge the comment regarding the ultimate goal of developing an ML parameterization trained from superparameterization or cloud-resolved models. An ML parameterization offers two potential advantages: first, it can help reduce uncertainty. Second, it can reduce the computational cost compared to running superparameterization or high-resolution models directly. However, this study presents an ML model as a surrogate for an existing microphysics scheme rather than training it on high-fidelity data, which cannot reduce uncertainty. In order to highlight the contribution, the authors could present the computational efficiency. To better evaluate the ability to reduce costs, the authors could estimate the computational requirements of superparameterization or high-resolution runs to directly compare against the ML model performance. This would strengthen the case for ML as a lower-cost alternative to traditional parameterization approaches.

**Response: As suggested by the reviewer, we added the following sentence in the appendix.**

**Khairoutdinov et al. (2009) employed Large-Eddy Simulation (LES) to model deep tropical convection over an area of approximately 205 km x 205 km, focusing particularly on maritime regions. They conducted a benchmark simulation spanning 24 hours, with a spatial resolution of 100m and 256 vertical levels. This benchmark simulation utilized 2048 processors and took approximately 6 days of wall-clock time to complete. Additionally, we attempted a cloud-resolving simulation, covering a domain of 600 km x 500 km domain with a grid spacing of 500 m (resulting in a grid of 1200 x 1000 points) and employing 45 vertical levels. The wall-clock time for this simulation was approximately 40 times the forecast time (dt = 2 seconds). For a single 36-hour simulation, the computational time is around 60 days, which far exceeds our current computational resources. Therefore, implementing machine learning-based parameterization would offer a significant advantage in reducing computational costs when replacing the super-parameterization scheme or cloud-resolving model.**

2. This study presents a novel multi-task ML approach for both trigger function classification and tendency regression. This multi-faceted application of ML could be emphasized as another key contribution of the work. Previous studies have applied ML either to trigger function classification alone (Zhang et al., 2021) or tendencies regression independently (Brenowitz & Bretherton, 2019; Rasp et al., 2018; Wang et al., 2022). As the authors point out that "models based only on regression can yield inconsistent tendencies, resulting in conflicting indications for convection triggering at specific grid points. In contrast, models that rely exclusively on classification are also deficient, as they do not generate the necessary tendencies for the CP scheme". To further validate this assertion, ablation experiments removing each individual task (i.e. classification-only vs regression-only models) could demonstrate the benefits of the proposed multi-task framework. Such an analysis would help substantiate the value of the multifaceted ML approach over single-task baselines, strengthening the novel aspects highlighted in this work.

**Response: As suggested by the author, we added the following sentences to strengthen the novel aspects of multi-output ML model over single-task baselines.**

**First, we added the following sentence in the abstract:**

**"This multi-output Bi-LSTM model is capable of simultaneously predicting the convection trigger while also modeling the associated convective tendencies and precipitation rates with high performance."**

**Secondly, we added the following sentence in the subsection "ML model structure":**

**"Previous studies have applied ML models to address these objectives, with some dedicated solely to the classification task of convection trigger (Zhang et al., 2021a), while others have independently pursued the regression of convective tendencies (Rasp et al., 2018; Brenowitz and Bretherton, 2019; Wang et al., 2022)."**

**More importantly, we followed the review's suggestion and conducted the ablation experiments by removing each individual task (i.e. classification-only vs regression-only models), and demonstrated the benefits of the proposed multi-task framework. We added the following sentences and figures in the appendix.**

**"Two separate Bi-LSTM models were trained with slight modifications to the multi-output Bi-LSTM model illustrated in Figure 3. The first model aimed to predict convection triggers alone, termed Bi-LSTM-trigger, while the second model aimed to predict convective tendencies, termed Bi-LSTM-tendency. In predicting convection trigger, both the Bi-LSTM-trigger model and the multi-**

output Bi-LSTM model demonstrated comparable accuracy, as observed in Figures A1 and A2. However, while the convection trigger predicted by the Bi-LSTM-trigger model were indistinguishable from those of the multi-output Bi-LSTM model, the former failed to accurately predict corresponding convective tendencies. Consequently, it cannot serve as a replacement for convection schemes within NWP models.

**Figures A3 and A4 present snapshots of rthcuten and rqvcuten predicted by the Bi-LSTM-tendency model. These figures reveal that the Bi-LSTM-tendency model predicts non-zero values across nearly the entire domain. Since the Bi-LSTM-tendency model exclusively focuses on predicting convective tendencies, convection trigger are derived using certain threshold values.The spatial distribution of these triggers is notably influenced by the choice of threshold values, and the patterns of convection trigger derived from rthcuten and rqvcuten exhibit considerable discrepancies. This confirms that models based solely on regression yield inconsistent tendencies. In contrast, the multi-output Bi-LSTM model does not encounter the aforementioned issues of the Bi-LSTM-tendency model and generates a more consistent spatial pattern of rthcuten and rqvcuten (see Figures A5 and A6).”**

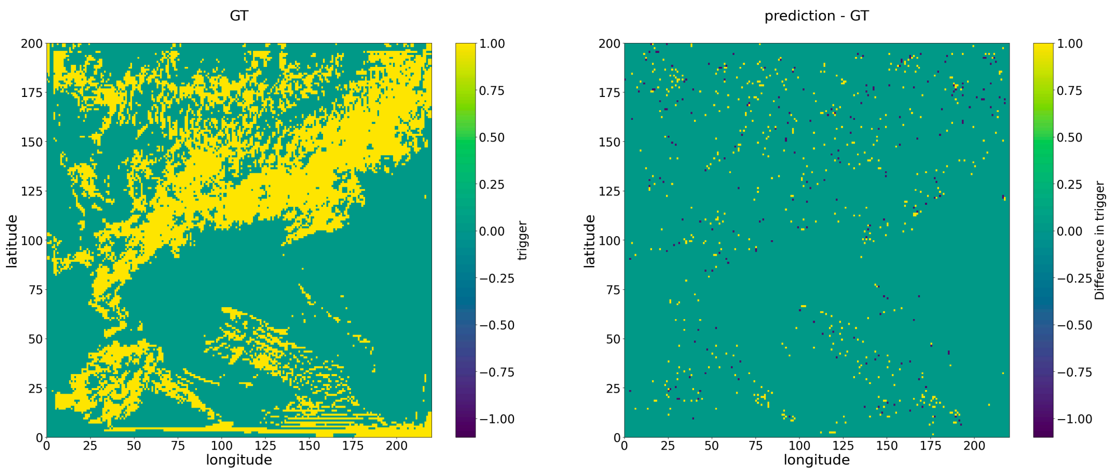


**Figure A1.** Snapshot example of convection trigger, with the left column showing the ground truth (GT), and the right column showing the difference between convection trigger as predicted by the Bi-LSTM-trigger model and ground truth values, for the 25-hour WRF simulation initialized at 12UTC on May 20th, 2021.

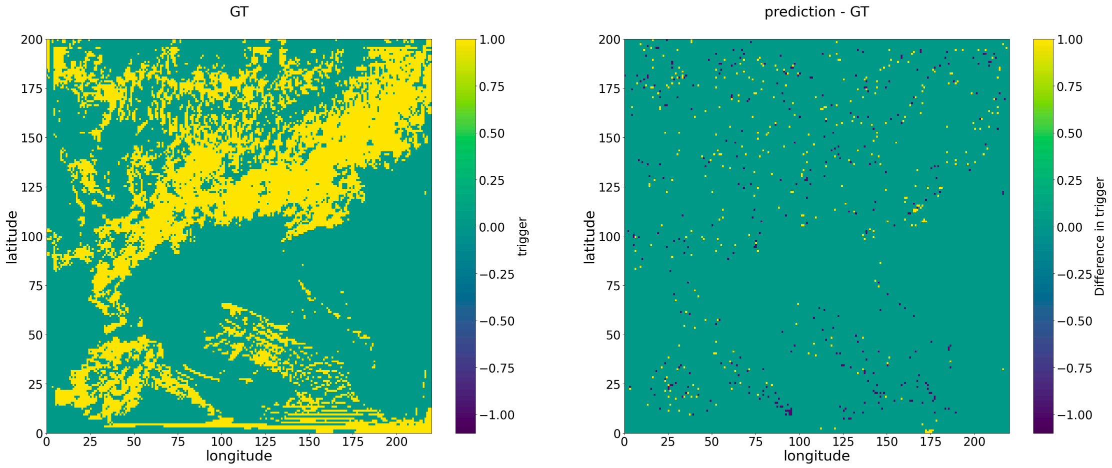


**Figure A2.** Snapshot example of convection trigger, with the left column showing the ground truth (GT), and the right column showing the difference between convection trigger as predicted by the multi-output Bi-LSTM model and ground truth values, for the 25-hour WRF simulation initialized at 12UTC on May 20th, 2021.

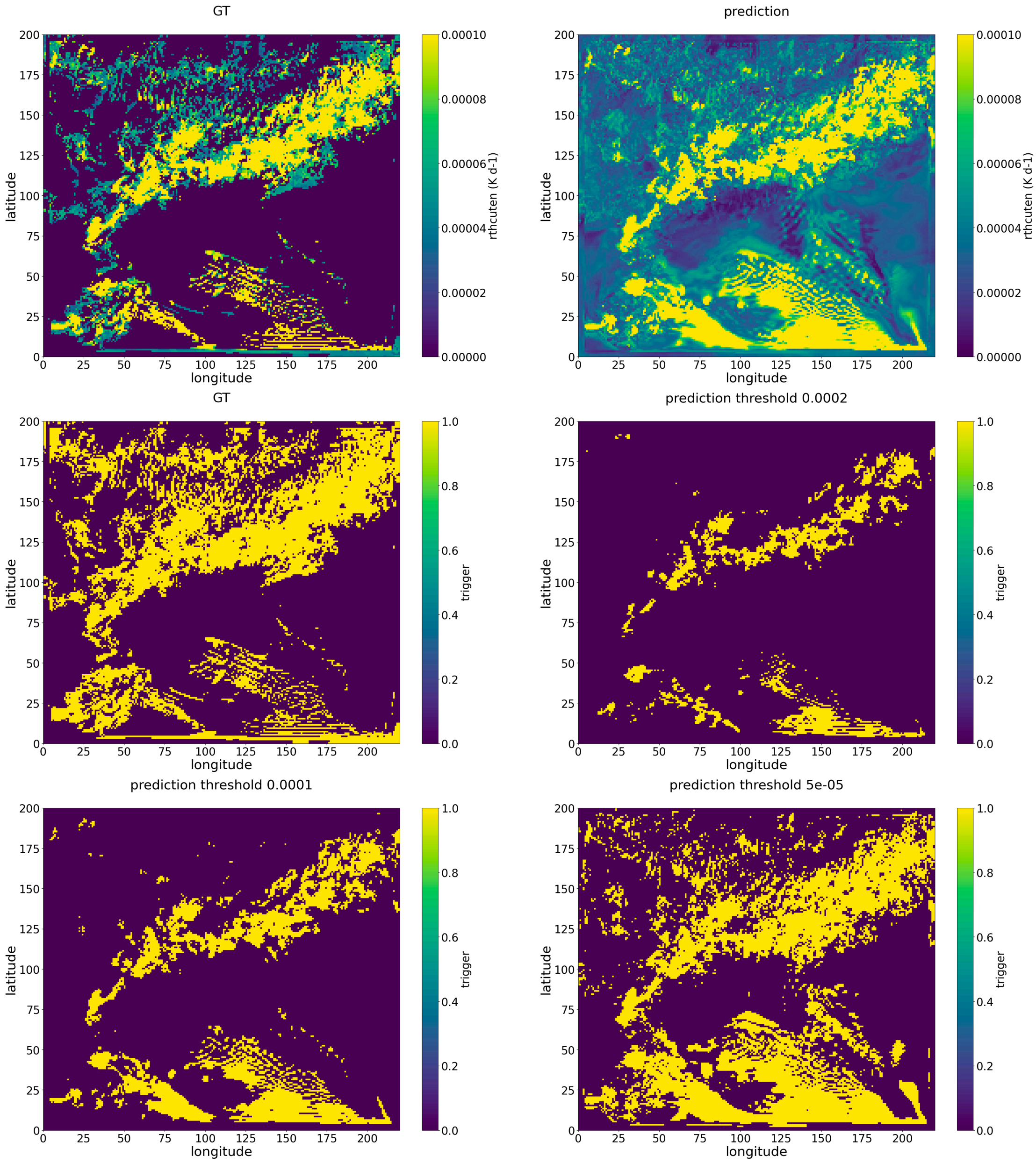


**Figure A3.** Snapshot examples of rthcuten summed along the vertical direction, with the top left panel showing the GT values and the top right panel showing the rthcuten predicted by the Bi-LSTM-tendency model, for the 25-hour WRF simulation initialized at 12UTC on May 20th, 2021. Similarly, snapshot examples of trigger, with the GT shown in the middle left panel, and the predictions from the Bi-LSTM-tendency model using varying threshold values of rthcuten shown in the middle right column, bottom left panel, and bottom right panel, respectively.

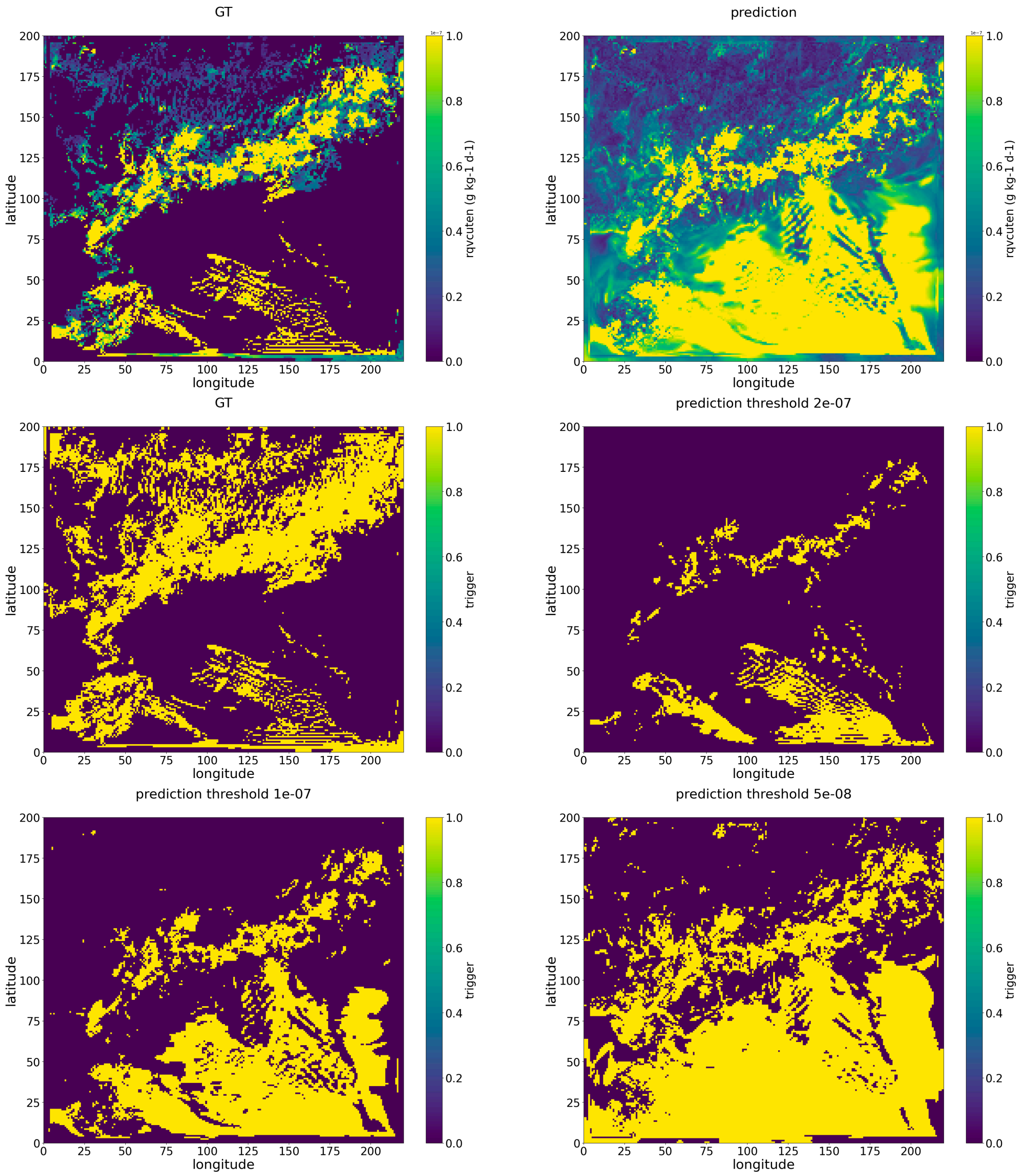


**Figure A4.** Snapshot examples of rqvcuten summed along the vertical direction, with the top left panel showing the GT values and the top right panel showing the rqvcuten predicted by the Bi-LSTM-tendency model, for the 25-hour WRF simulation initialized at 12UTC on May 20th, 2021. Similarly, snapshot examples of trigger, with the GT shown in the middle left panel, and the predictions from the Bi-LSTM-tendency model using varying threshold values of rqvcuten shown in the middle right column, bottom left panel, and bottom right panel, respectively.

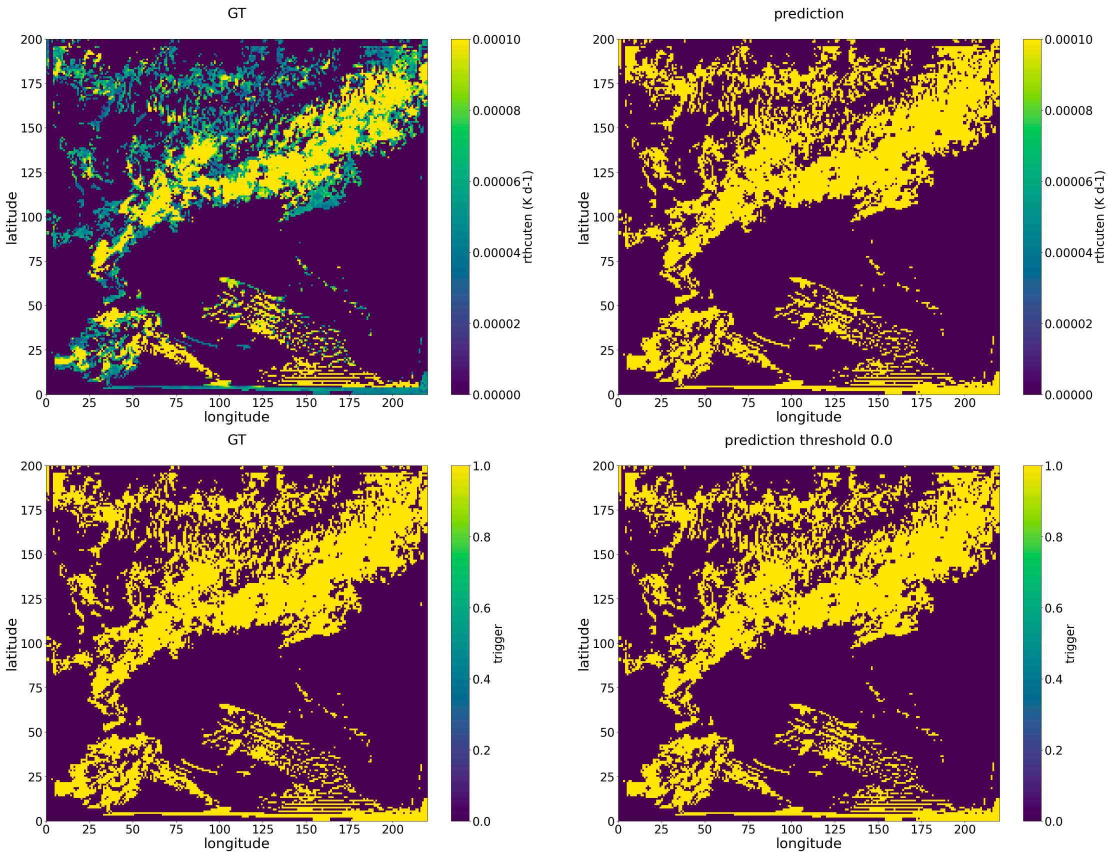


**Figure A5.** Snapshot examples of rthcuten summed along the vertical direction, with the top left panel showing the GT values and the top right panel showing the rthcuten predicted by the multi-output Bi-LSTM model, for the 25-hour WRF simulation initialized at 12UTC on May 20th, 2021. Similarly, snapshot examples of trigger, with the GT shown in the bottom left panel, and the predictions from the multi-output Bi-LSTM model using a threshold value of 0 shown in the bottom right panel.

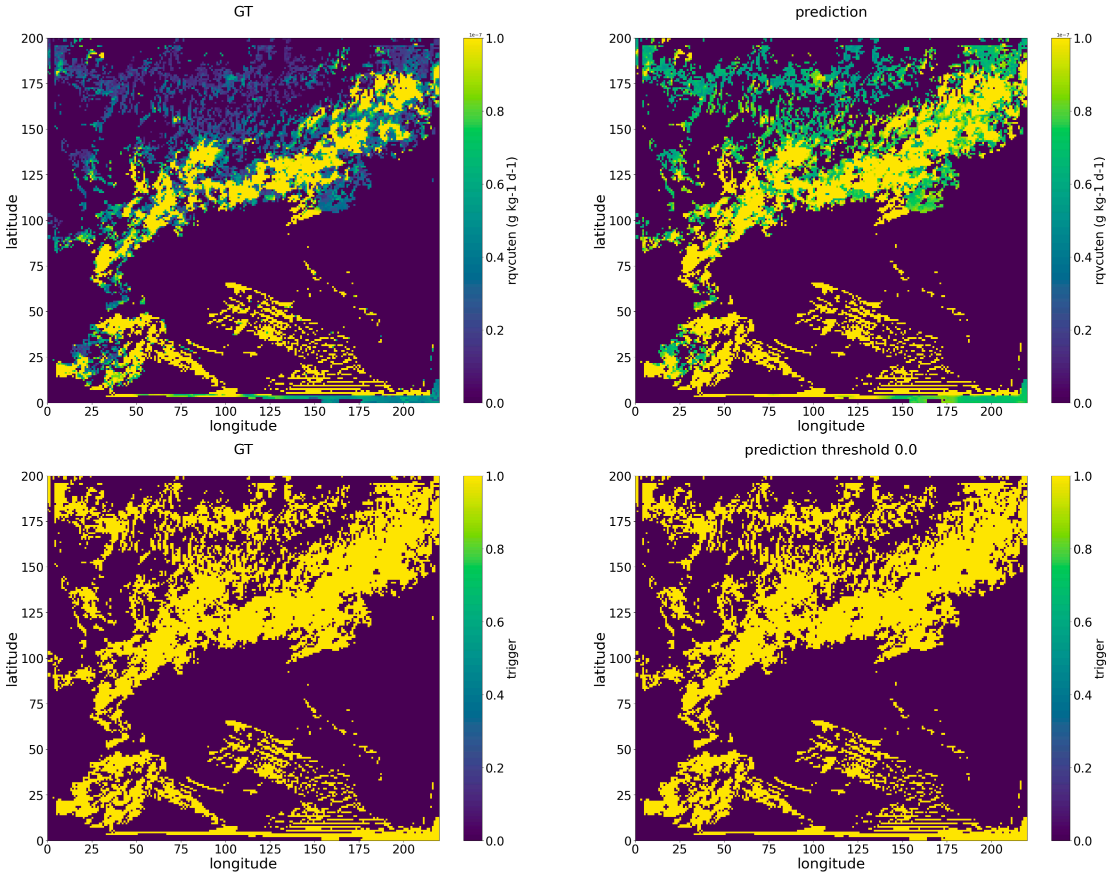


**Figure A6.** Snapshot examples of rqvcuten summed along the vertical direction, with the top left panel showing the GT values and the top right panel showing the rqvcuten predicted by the multi-output Bi-LSTM model, for the 25-hour WRF simulation initialized at 12UTC on May 20th, 2021. Similarly, snapshot examples of trigger, with the GT shown in the bottom left panel, and the predictions from the mult-ioutput Bi-LSTM model using a threshold value of 0 shown in the bottom right panel.

Ref:

Brenowitz, N. D., & Bretherton, C. S. (2019). Spatially extended tests of a neural network parametrization trained by coarse-graining. Journal of Advances in Modeling Earth Systems, 11(8), 2728–2744. https://doi.org/10.1029/2019ms001711

Rasp, S., Pritchard, M. S., & Gentine, P. (2018). Deep learning to represent subgrid processes in climate models. Proceedings of the National Academy of Sciences, 115(39), 9684–9689. https://doi.org/10.1073/pnas.1810286115

Wang, X., Han, Y., Xue, W., Yang, G., & Zhang, G. J. (2022). Stable climate simulations using a realistic general circulation model with neural network parameterizations for atmospheric moist physics and radiation processes. Geoscientific Model Development, 15(9), 3923–3940. https://doi.org/10.5194/gmd-15-3923-2022

Zhang, T., Lin, W., Vogelmann, A. M., Zhang, M., Xie, S., Qin, Y., & Golaz, J.-C. (2021). Improving Convection Trigger Functions in Deep Convective Parameterization Schemes Using Machine Learning. Journal of Advances in Modeling Earth Systems, 13(5), e2020MS002365. https://doi.org/10.1029/2020MS002365