Authors Response to Reviewer #2

In this study, the authors presented a generative AI model for the region climate simulation. This model generates the large ensembles and can well capture the intrinsic climate variability. The topic is very interesting. But there still are several questions that need to be addressed.

We are grateful to Reviewer #2 for the clear summary of our study and for the insightful comments, which have been very helpful in revising and improving the manuscript.

1. How was the MESMER-RCM model trained? Please provide more details about the training of the model. For example, how to divide the training and testing sets? How to set the model parameters?

To improve clarity, we have restructured the manuscript: Section 3 now provides the methodological framework, while Sections 4.1 and 4.2 give detailed accounts of the data preparation and training procedure, including the division of training and testing data sets as well as the calibration of the model parameters. Furthermore, we have added a new Figure 1 to present a comprehensive schematic of MESMER-RCM. We believe these additions can improve the transparency and accessibility of the MESMER-RCM methodology.

2. There are some parameters in the model. Are the results sensitive to the choices of the parameters? The detailed tests should be done.

We added additional sensitivity experiments to clarify this. Regarding the deterministic response module, we examined the sensitivity to the number of nearest GCM grid points used as predictors, k. As shown in the new Figure 5a–c, we compared $k = 1$ and $k = 9$. Using $k = 9$ provides a good balance between interpretability and emulation quality: it captures the local GCM-to-RCM temperature response effectively, while avoiding the blocky artifacts that typically arise from the simple nearest-neighbor regression ($k = 1$). For the residual variability module, we tested the sensitivity to the number of residual samples used for prior P construction, k_r (Figure 4). Due to limited data availability, the choice of k_r reflects a trade-off between physical consistency and numerical stability: while increasing k_r enhances physical consistency and improves emulation performance (as indicated by the rank histogram), it compromises the numerical stability of the prior and may introduce distortions due to ill-sampling.

3. The figure 1 shows the 2-m temperature in a region. Why there are some blank areas? Additionally, could you add the latitude and longitude in the figure? Because the readers may be not familiar with that region.

We thank reviewer #2 for this helpful comment. MESMER focuses on emulating regional land warming within the EURO-CORDEX domain (−25°E to 45°E, 26°N to 72°N), which is consistently applied throughout this study. The blank areas in Figure 1 correspond to ocean regions outside the land domain. We have added latitude and longitude information in Section 2 (Data description) for clarity.

4. In this study, only a simple example was displayed. To show the advantage of the model, more examples in different areas should be presented. Whether can this model be extended to other regions?

We thank reviewer #2 for this valuable question. At present, EURO-CORDEX is the only region that provides a sufficiently large ensemble of simulations to meet the requirements for emulator training. MESMER-RCM is designed to be applicable, in principle, to any region, provided that a sufficiently large set of GCM–RCM model-chain simulations is available to construct the prior and thereby ensure physical consistency and numerical stability. Due to current data limitations, extending the emulator beyond Europe is not yet feasible. However, the upcoming CMIP6-CORDEX initiative is expected to provide a broader set of simulations, which would enable robust applications of MESMER-RCM in other regions, such as East Asia and North America. We highlight this as a promising avenue for future work in the conclusions.