I have a couple of comments. Did you consider using the CERES dataset instead of ERBE as it is considered more reliable and accurate. Also a measure of a good climate model has to include energy balance at the top of atmosphere and even though the ERBE LW/SW components may match well they combine to have a 6W energy imbalance I believe. Have the authors considered this as a metric for tuning algorithm?

Reply: Thank you for your comments. We agree with the reviewer that more recent observations (such as CERES-EBAF, having a better TOA radiation balance), can be used as the reference metrics. With the proposed method, we may get better model performance using these more reliable and accurate observations. The focus of the study is the description of the proposed optimization method, which targets at the cost-expensive optimization problem arising from parameters calibration of climate system models.

In principle, different reference metrics can be used and the final tuned parameters will be dependent on the metrics used. As pointed out by the reviewer, ERBE has a TOA imbalance of approximately 5~6 W/m^2. However, the tuned simulation in this work has a TOA radiation imbalance less than 0.1 W/m^2, which is even better than that of control run (0.8 W/m^2). See the figure below for more details. In this study, we transform 16 output variables into a scalar value (Eq. 3). The optimal metrics value embodies the comprehensive optimal performance of all the output variables, not only part of them. In the following studies, we are going to use more reliable and accurate observations, and construct a much more comprehensive and reliable metrics. Accordingly, we have included the following to the manuscript.
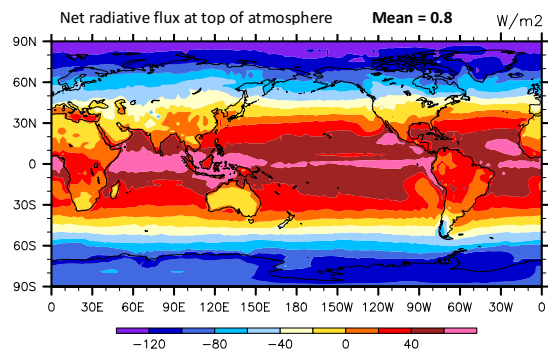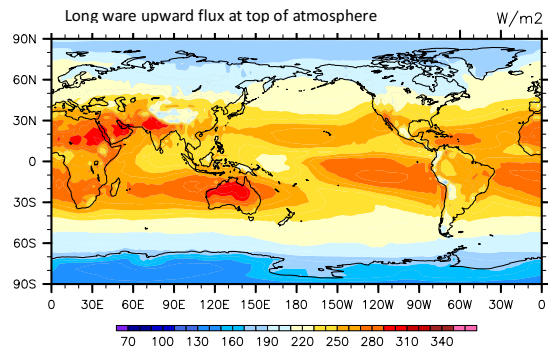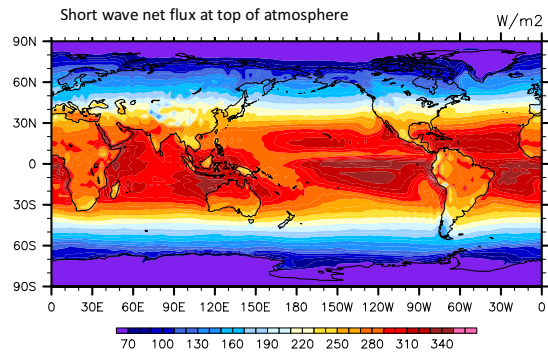
1. The end of the first paragraph on page 16.
   "Overall, the tuned simulation has a TOA radiation imbalance of 0.08 W/m^2, which is better than that of the control run (0.8 W/m^2)."
2. The end of the last paragraph on page 16.
   "The choosing of appropriate reference metrics and related observations are very important for the final tuned model performance. In future studies, we are going to use the more reliable and accurate observations, and add some constraint conditions for parameters tuning to construct a more comprehensive and reliable metrics."

Default: Optimal:

Short wave net flux at top of atmosphere   W/m2   Short wave net flux at top of atmosphere   W/m2

70 100 130 160 190 220 250 280 310 340   70 100 130 160 190 220 250 280 310 340

Long ware upward flux at top of atmosphere   W/m2   Long ware upward flux at top of atmosphere   W/m2

70 100 130 160 190 220 250 280 310 340   70 100 130 160 190 220 250 280 310 340

Net radiative flux at top of atmosphere   **Mean = 0.8**   W/m2   Net radiative flux at top of atmosphere   **Mean = 0.08**   W/m2

−120   −80   −40   0   40   −120   −80   −40   0   40