Response Letter

Dear Editor:

We sincerely thank you for your comments on our manuscript entitled "An improved multivariable integrated evaluation method and NCL code for multimodel intercomparison (MVIETool version 1.0)" (gmd-2020-310). We have revised our manuscript based on your comments. The changes are marked in red in the revised manuscript. Our point-by-point responses to your comments are as follows:

I carefully evaluated your response to the reviewer comments and your revised manuscript. While your efforts in addressing the reviewer comments are very much appreciated, I think that your manuscript needs some further, minor revisions before publication:

Response:

Thanks. We have revised the manuscript and related codes.

First, I would like to encourage you to improve your manuscript in highlighting the novelty of your tool and the changes compared to previous versions. You provide a quite extensive discussion of the previous literature on performance metrics in the author's response, but your paper is rather short on this. I think the reviewer comment clearly shows that there is a need for clarification. So instead of addressing this issue only in the author's response, I suggest to extend the discussion of previous metrics and differences to your approach in the introduction.

Response:

Thanks for the comment. We have added discussion about previous metrics and differences to our approach in the introduction of the revised manuscript (Lines 44–54 on Page 2), which is also listed as follows for the convenience:

"Moreover, the vector field statistics employed in Xu et al., (2016; 2017) did not consider area weighting, which is a limitation especially for an evaluation of the global field. Although area weighting was considered in many previous statistical metrics, e.g., correlation coefficient and standard deviation, they were used to evaluate scalar fields rather than vector fields (e.g., Watterson, 1996; Boer and

Lambert, 2001; Masson and Knutti, 2011). The consideration of area weighting in the definition of vector field statistics is one of the novelty of our study relative to previous studies (Taylor, 2001; Boer and Lambert, 2001; Gleckler et al., 2008; Xu et al., 2016; 2017). In addition, we also improve MVIE method to allow a mixed evaluation of scalar and vector fields. Furthermore, based on MIEI, a multivariable integrated skill score (MISS) for a climate model is proposed, which allows us to adjust the relative importance of different aspects of model performance. Finally, we develop a Multivariable Integrated Evaluation Tool (MVIETool version 1.0) to facilitate multimodel intercomparison. These efforts are expected to improve the accuracy and flexibility of the VFE and MVIE methods."

Some technical corrections:

- L 44/45: consider rephrasing, spatial weight -> "area weighting"

Response:

We have replaced " *spatial weight* " to "area weighting" in Lines 44–45 on Page 2 of the revised manuscript.

- L65: This sentence still does not make sense and needs clarification: what is divided by what? Do you mean "... by dividing by the rms value of the corresponding observed variable"?

Response:

Thank you for the comment. We have revised the sentence as "We need to normalize each modeled variable using the rms value of the corresponding observed variable." in Lines 69–70 on Page 3 of the revised manuscript.

- L116: What does CG stand for?

Response:

In the definition of MIEI (Xu et al., 2017), CG represents a line segment when F is 2 (Fig. R1). The square root of the first term on the right side of Eq. 7a can be regarded as multivariable amplitudes' error, while the square root of the second term represents the pattern similarity error of multiple field. Xu et al. (2017) used them as the lengths of two perpendicular sides (i.e., BC and BG in Fig. R1) and constructed a

right triangle on VFE diagram. In this case, MIEI value can be regarded as the hypotenuse of the right triangle, which is line segment CG.

In the revised manuscript, the sentence has been reworded as "Considering that MIEI has a geometric meaning when *F* is 2, which represents the length of a line segment (referring to line CG in Figure 3 in Xu et al., 2017)." in Lines 119–120 on Page 4 of the revised manuscript.

- L 232/233: "No matter what kind of masks chosen, the missing points across all variables of one model are the same." This sentence needs clarification. Why would you apply the same mask for all variables of one model? Assume an observational temperature time series includes 10% missing data, while an observational ozone time series includes only 1% missing data: Would the tool discard valid data points from the ozone time series?

Response:

Thanks for the constructive comment from the editor. The previous MVIETool unified missing points across all variables of one model, which is valid for the evaluation of the spatial field. However, this processing to missing points is not suitable for all evaluations, such as the situation proposed by the editor. To deal with more common situations, we have added a new argument —'`Unif_VarMiss`' (Table 1 in the manuscript) in the updated MVIETool. With the help of this argument, users can choose whether unify missing points across all variables of one model or not. We have also modified the sentence " *No matter what kind of masks chosen, the missing points across all variables of one model are the same.*" as "Further, whether to unify missing points across all variables of one model can also be chosen with the help of the argument `Unify_VarMiss`." in Lines 237–238 on Page 8 of the revised manuscript.

- L276: "Lighter colors indicate that the model statistics is closer…."

Response:

We have revised the sentence as "Lighter colors indicate the model statistics are closer to observation and vice versa" in Lines 281–282 on Page 10 of the revised manuscript.

- L327: "… which is calculated using the climatologies…"

Response:

We have revised the sentence as " SD_{ij}^{obs} is the standard deviation of multiple observations, which is calculated using the climatologies of REA1 and REA2 (Fig. 7) " in Lines 331–332 on Page 11 of the revised manuscript.

- Caption Fig. 7: "There is a red horizontal bar centered at the REF point of which length is as a measure of the observational uncertainty." Sentence needs rephrasing, e.g., "The observational uncertainty is indicated by the red horizontal line centered at the REF point." Same for the sentence in lines 322/323. And replace "red bar" by "red line".

Response:

We have rephrased the sentence in the caption of Fig. 7 as " The observational uncertainty is indicated by the red horizontal line centered at the REF point.". In addition, we have replaced "*red bar* " with "red line" in Line 327/335 on Page 11 of the revised manuscript.

Reference

Xu, Z., Han, Y., and Fu, C.: Multivariable Integrated Evaluation of Model Performance with the Vector Field Evaluation Diagram, Geosci. Model Dev., 10, 3805–3820, 2017.

Figure

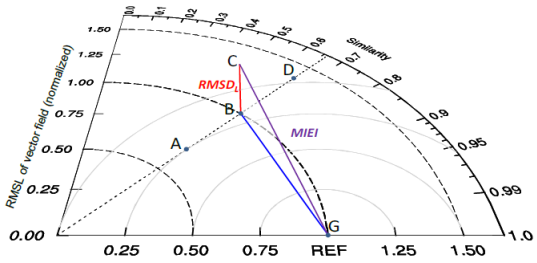

Figure 3. Schematic diagram displaying the relationship between the RMSVD, RMSD$_L$, and MIEI. The points A, B, and D represent different models. The RMSD$_L$ measures the overall difference between the modeled rms values and the observed ones. The line segment BC is vertical with respect to the VFE diagram. The length of line segment BG is determined based on the vector field similarity, which measures the overall pattern similarity of various scalar fields relative to the observed ones. Thus, the MIEI index takes both the pattern similarities and the rms values of various scalar fields into account.

Figure R1. Figure 3 in Xu et al. (2017)