

Interactive comment on “Multivariable Integrated Evaluation of Model Performance with the Vector Field Evaluation Diagram” by Zhongfeng Xu et al.

Anonymous Referee #1

Received and published: 14 June 2017

Multivariable Integrated Evaluation of Model Performance with the Vector Field Evaluation Diagram by Z. Xu, Y. Han, and C. Fu

The authors describe a method to assess the performance of climate models in simulating an arbitrary number of equally important variables based on the concept of the vector field evaluation (VFE) introduced by Xu et al. (2016). In addition, the authors describe a method to collapse the three different metrics root mean square length (RMSL), vector field similarity coefficient, and root mean square vector deviation (RMSVD) into a single index that can be used to rank the models by overall performance for the set of given variables.

C1

The manuscript is generally well written and I suggest minor revisions to the manuscript before publication in Geoscientific Model Development addressing the points given below.

General comments

- As an example application of the model evaluation method presented in this paper, three different temperature / precipitation datasets are averaged as reference dataset. Datasets such as the CRU 2m temperature data typically contain missing values (also over land). How are missing values being treated in this study? Also, are the grid cells weighted by their surface area? This would be needed to make sure that the skill scores are representative for the global mean values of the respective quantities. If not, the calculated average metrics would be very hard to interpret as e.g. grid cells in polar region would receive more relative weight than grid cells in low latitudes. As the method presented here suggests evaluation of global averages, I would like to see one or two sentences on this issue. Also, please be more specific on the processing of the observational data regarding e.g. missing values.
- Some parts of the manuscript are somewhat repetitive and could be shortened. For example on p.8, l. 9-10 it reads "Thus, three statistical quantities can be indicated by a single point on the VFE diagram", which is almost identical to p. 10, l. 10-12: "[...] the three quantities [...] can be represented by a single point for each model on the VFE diagram." or to p. 10, l. 13-15: "Thus, each point on the VFE diagram can represent the overall performance of an individual model in terms of 3 statistical quantities [...]" I suggest to go through the text and remove

C2

such repetitions where possible.

- The authors propose to include the standard deviation of the RMS values of multiple scalar fields into the VFE diagram as an additional performance measure (p. 11, l. 12-14). It remains unclear to me whether the length of the proposed additional line segments in figure 3 are σ_{RMS} or actually $\pm\sigma_{RMS}$, i.e. $2\sigma_{RMS}$. Or did the authors mean variance of the RMS values (equation 23)? Please clarify.
- A proper evaluation of the performance of climate models usually requires to take into account observational uncertainties. Differences between models and observations can only be interpreted as model errors or lack of model skill if the differences are larger than the observational uncertainty. This is particularly the case for variables with a large uncertainty such as, for instance, ice water path, but also important when ranking models by performance. More and more observational datasets provide estimates of the observational uncertainty. What are the authors' thoughts about including such additional information into their calculations, in particular when calculating skill scores such as the presented multivariable integrated evaluation index (MIEI) that is then used to rank models according to their average performance skill?

Specific comments

- p. 1, l. 21: "[...] evaluation *of* model performance."
- p. 10, l. 5: "what is meant by "summer SAT and precipitation"? Is this an average over the months June, July, August? Please be more specific.
- p. 12, l. 20, "*In comparison with* the RMSVD, [...]" : did you mean "*In contrast to* [...]"?

C3

- p. 13, l. 2, "Index" → "index"
- p. 13, l. 5, "[...] but a larger MIEI *relative to* [...]" : did you mean "compared to"?
- p. 26, l. 3, "CMIP5 model" → "CMIP5 models"
- p. 26, table 1: it would be interesting to add the performance of the individual observational datasets to the table as a "rough estimate of the observational uncertainties" as stated on p. 9, l. 22-23.
- p. 27, caption of figure 1, l. 4: delete "apart"
- figure 2 is fully included in figure 3 and could be deleted
- p. 31, figure 5: the second level of metrics includes σ_{RMS} while the caption and the referenced equation 23 specify the variance of RMS values (σ_{RMS}^2). Which one is correct? Is there a "²" missing?

Interactive comment on Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2017-95>, 2017.

C4