

## Answer to the Report of Reviewer 1 of How biased are our models? – A Case Study of the Alpine Region

Thank you very much for your report of our paper. Your remarks and comments helped us to improve our paper. All changes are marked in red in the manuscript. Please, find below a detailed answer to the individual remarks.

This study demonstrated that focusing analyses purely on measurements could introduce bias. This research is performed based on a geothermal model, and Global Sensitivity Analysis and Reduced Order Modeling methods are used to illustrate the influence of data distribution. The authors used a weighting scheme to compensate for parts of this bias caused by data distribution. However, I think the more important issue in this study is the general rule of a reliable weighting scheme. In addition, some important details are not clear in this manuscript and should be addressed. Please see the specific comments below.

1. Line 140 Are the weights determined based on the number of data points? In addition, the author should describe how to use these weights briefly.

The weighting scheme is determined by both the number of data points and information about the data quality (as described in lines 136-142). We applied the weighting scheme directly to the quantity of interest in all analyses. An explanation and the corresponding formula has been added in line 143-149.

2. Line 145 The two models have different geological layers (31 and 34), and why they have the same model space?

The models have the same model space since the higher number of layers in the second model is obtained through a subdivision of several layers of the first model (as shown in Table A1). An explanation has been added to lines 157-158.

3. Line 178: what's the meaning of this threshold? the first- or total order indices?

The threshold is applied for the total-order indices. An explanation has been added to line 187.

4. Line 184: Before Figure 4, please explain all model parameters in a table.

The model parameters are all presented in Table A1. To highlight that this paper has a method focus and not a focus on this specific case study, we placed the table on purpose in the Appendix since that has been a misunderstanding in previous publications.

5. Line 304: Again, how to use these weights? Are these weights used for sensitivity analysis? Please give the formulas;

An explanation and the corresponding formula has been added in lines 143-149 (see point 1).

6. Line 316: It seems that the appropriate and reliable weighting scheme is crucial to compensate the data distribution problems. In addition, does the inappropriate weighting scheme may lead to bias to model output?

Yes, an inappropriate weighting scheme does induce a significant bias to the results. We added this for clarification in lines 343 and 344.

7. Line 325: The authors should describe the RB surrogate models, e.g., the accuracy, the cost.

We added Section 3.4 for a more detailed describe the RB surrogate models.

8. Line 340: Please give the results of model calibration, e.g., RMSE, R2.

We added the R2 values in lines 371-374.