Geosci. Model Dev. Discuss.,
https://doi.org/10.5194/gmd-2018-198-AC2, 2019

Geoscientific Model Development Discussions

Interactive comment on "Improving climate model accuracy by exploring parameter space with an O(10^5) member ensemble and emulator" *by* Sihan Li et al.

Sihan Li et al.

sihan.li@ouce.ox.ac.uk

Received and published: 7 March 2019

General response: Thank you very much for these comments. We feel very encouraged from this review and the similar encouraging comments from the other reviewer. We have made the best attempt to respond to these constructive comments.

Response to item 1: Thank you for this suggestion. The title has now been changed to "Reducing climate model biases by exploring parameter space with large ensembles of climate model simulations and statistical emulation".

Response to item 2: Thank you for these suggestion. Lines 53-54 now reads 'reduce

the reliability of' as suggested. In the revised manuscript we have moved the previous 5th paragraph which introduces parameter perturbation to earlier in the paper (2nd paragraph now) as suggested.

Response to item 3: Thank you for this suggestion. In the revised manuscript, we have added a few sentences about using PPEs to estimate model PDFs and uncertainty at the end of paragraph 7 "Besides parameter refinement, PPEs have also been used in many studies to estimate probability distribution functions (PDFs) of equilibrium climate sensitivity (e.g., Murphy et al., 2004) and transient regional climate change (e.g., Sexton et al., 2012), permitting probabilistic projection of climate change (Murphy et al., 2007, 2009; Harris et al., 2013). PPEs are becoming common as a means to assess the range of uncertainty in climate model projections (Murphy et al., 2004; Stainforth et al., 2005; Collin et al., 2006; Sanderson, 2011; Sexton et al., 2012; Shiogama et al., 2012)".

Response to item 4: Thank you for this suggestion. The wording on line 146 has been changed to "varied systematically or randomly" as suggested.

Response to item 5: Thank you very much for this suggestion. We have added a few sentences in the main text to point this out at the end of paragraph 10: "It is worth pointing out that the second and the third categories may not be different from each other if a sufficient number of model simulations are used to train a statistical emulator over the full parameter space. With a good emulator, it is possible to rule out parameter space and optimize parameter values, in which case categories two and three are post-processing steps".

Response to item 6: Thank you for pointing this out. In the revised manuscript, we have included a few references on Bayesian climate model calibration and MCMC, as well as optimization over multiple objectives in the introduction as suggested.

Response to item 7: Thank you for this suggestion. We have added a few references of prior work using PPEs for parameter refinement to improve regional climate over

Europe and North America in paragraph 11 (line 207 in the original manuscript), which now reads "However, very little (Bellprat et al., 2012; 2016) has been published on using PPEs for parameter refinement with the aim of improving regional climate models (RCMs)."

Response to item 8: Thank you for this suggestion. We have added a few sentences commenting on the potential benefits of using posterior parameter PDFs toward the end of the introduction.

Response to item 9: If we were to swap the order of phases 1 and 2, we could possibly get rid of the regional temperature biases, seeing the reduction in temperature biases is accompanied by increased TOA reflected SW radiation (Fig. 4), implying there are more clouds in those PPE simulations with reduced JJA temperature biases. Then we could end up in a corner of the parameter space where there is minimal JJA temperature biases, but out of balance TOA energy fluxes. Our premises was that TOA radiation balance is an emergent property in GCMs (Solomon et al., 2007), so we chose to carry out the parameter refining process following phases 1 and 2, preserving energy balance first, then reducing biases in NWUS.

Response to item 10: Thank you for these comments. Some model parameters are different between the global and regional model (adjusted for scale) but these are not among the parameters that were perturbed in this study. We are mindful of the possibility that among the parameters that were perturbed here, for some parameters the parameter values may be resolution-dependent, especially in a topographically complex region such as the NWUS. This is an important issue that needs to be further explored. However, currently there is no clear guidance on how the parameters should scale with changes in resolution between the global and regional model. Without information on how these should be adjusted or performing a further nested parameter sweep (which is beyond the scope of this study), it would be hard to know which parameters to adjust, to what extent are they resolution dependent, and how to adjust them. Therefore, the same values are applied in HadAM3P and HadRM3P as a first

estimation, without adjustment to account for differences in scale. But it would be an interesting and useful follow-on study to look at setting parameters differently in the global and regional model. As part of our ongoing work, we have performed additional PPEs, where the parameter values are set to be different in HadAM3P and HadRM3P, and we will attempt to address the resolution-dependency of parameter values in a following paper using those PPEs.

Response to item 11: We are not quite sure what the reviewer means by 'upscaling variability'. Assuming that means variability in upscaling methods in regridding PRISM to HadRM3P, to clarify, we chose the method where for each HadRM3P grid point, an average was taken over all the PRISM points that fall in the bounds of that HadRM3P grid point, and the averaged value was assigned to that HadRM3P grid point.

Response to item 12: Thank you for these comments and suggestions. Yes, emulators are used for sensitivity analysis. Apologies for not being clear about this. We have added a sentence in the sensitivity analysis section to clarify this "Emulators are used for the sensitivity analysis". As suggested, we have included a brief summary of the emulators (the new section 2.5. Emulators) before the sensitivity analysis (the previous section 2.5, which is now section 2.6).

FAST is also a variance-based quantitative sensitivity analysis method. The fraction of the variance due to an input parameter (main effect) is calculated as the sum of the Fourier coefficients for the frequency assigned to the input parameter and it's harmonics. The total contribution of each parameter, xi, to the output variance includes main effects and interactions with all other parameters and can be calculated by summing Fourier coefficients for the set of frequencies complementary to the frequency assigned to input the parameter xi. The residual variance, not accounted for by the main effect, is therefore attributable to interactions between the parameter xi and any of the other parameters. We agree with the reviewer that Sobol analysis could be used, but FAST is a quantitative method that can be used to identify the dominant parameters as well.

Response to item 13:

Thank you for these comments and suggestions. We have added a few sentences to summarize the quality of the emulators earlier in the manuscript (section 3.1) as suggested. Regarding the relationship between the LW and SW, the bias reduction is accompanied by reduced TOA reflected SW, suggesting changes in clouds, but given different cloud types have different radiative effects on SW and LW (Zelinka et al., 2012, Fig. 8), we do not expect a clear positive or negative correlation between SW and LW by increased clouds, without knowing if the increased clouds are high, medium, low, thin, medium, thick, or any combination of these cloud types. We suspect the reason why there are no simulations in the blue ellipse with high LW is because the net effects of changes in clouds in these PP experiments are so that they do not lead to too much infrared radiation emitting to space. To answer this in further detail, we would need to run additional experiments where detailed could covers (low, medium and high) and cloud properties (e.g., optical depth) are saved as outputs, which is beyond the scope of this study.

Response to item 14: Thank you for these comments and suggestions. We have regenerated the OAAT plot using SP values as suggested (please see Figure 1). The results are very similar to those shown in the main Figure 2 holding other inputs at their mean values.

Figure 1. One-at-a-time sensitivity analysis of magnitude of MAC-T over Northwest to each input parameter in turn, with all other parameter held at SP values. Heavy lines represent the emulator mean, and shaded areas represent the estimate of emulator uncertainty, at the ± 1 SD level.

Thank you for suggesting to use partial dependence plots. We have attempted to compute and plot partial dependence plots (using scikit-learn in python https://scikit-learn.org/stable/). Figure 2 shows the results for the four most important parameters - cw_land, vf1, entcoef, and v_crit_alpha. The results are very similar to those shown in

the OAAT plots, and the results for the other parameters are very similar as well (not shown here). Since the interaction terms are quite small (from the sensitivity analysis shown in Figure 3 in the main text), it is not that surprising that the OAAT plots and the partial dependence plots are similar.

Figure 2. Partial dependence plots of magnitude of MAC-T over Northwest to CW_LAND, VF1, ENTCOEF, and V_CRIT_ALPHA.

Response to item 15: We agree with the reviewer that non-linear models can have coefficients that sum to greater than one. However, in the extended Fourier Amplitude sensitivity analysis (FAST; Saltelli et al., 1999), the fraction of the variance due to an input parameter (main effect) is calculated as the sum of the Fourier coefficients for the frequency assigned to the input parameter and it's harmonics. The total contribution of each parameter, xi, to the output variance includes main effects and interactions with all other parameters and can be calculated by summing Fourier coefficients for the set of frequencies complementary to the frequency assigned to input the parameter xi. The residual variance, not accounted for by the main effect, is therefore attributable to interactions between the parameter xi and any of the other parameters. The fraction of the total variance due to interactions is not resolved as the sum of individual effects and will never sum to a value other than 1. For further information please refer to Saltelli et al. (1999).

We have added a few sentences in 'Sensitivity Analysis' section to clarify this "In the FAST method, the fraction of the total variance due to the interactions is not resolved as the sum of individual interactions, but is computed from the parameter contribution to the residual variance , i.e., variance not accounted for by the main effects."

Regarding the emulator uncertainty, Figure 2, which shows the emulator uncertainty as the shaded area in each panel, illustrates that the contribution of the emulator uncertainty to the variance of the emulated output is small.

Response to item 16: Thank you for these comments. We agree with the reviewer that

Figure 4 contains a lot of information. In the original figure, we added the horizontal and vertical red lines to mark the transition from parameter inputs to model output metrics, hoping that would make digesting the figure easier. In the updated Fig 4 (which is shown in Figure 3 below), we have added additional labels in the figure to mark the three quadrants of the figure as a) input-input, b) input-output, and c) output-output.

Response to item 17: Thank you for this suggestion. We have plotted the temperature surfaces (MAC-T, JJA-T, and DJF-T) as a function of each pair of parameter inputs using the emulators (please see Figures 4-6).

Figure 4. MAC-T biases projected into the two-dimensional spaces of each pair of input parameters using the emulator.

Figure 5. JJA-T biases projected into the two-dimensional spaces of each pair of input parameters using the emulator.

Figure 6. DJF-T biases projected into the two-dimensional spaces of each pair of input parameters using the emulator.

Response to item 18: Thank you for point this out. This seems like an artifact from the original plotting. We will update the spatial maps to in the revised manuscript.

Response to item 19: Thank you for these comments. We have recomputed and replotted the sensitivity indices for the refined parameter space in phase 3 as suggested (please Figure 7, and this figure has been added to the supplementary information of the revised manuscript). The dominant parameters for SW, LW, and DJF-Pr are still the same, whereas the dominant parameters for the other output metrics are different. V_CRIT_ALPHA becomes the most important parameter for MAC-T, JJA-T and JJA-Pr, and ASYM_LAMBDA becomes the most important parameter for DJF-T. The dominant parameter is partially a function of the plausible range set for that parameter. We agree with the reviewer that the parameters that were well constrained after phase 2 may not be dominant in phase 3 simply because the perturbation range was reduced

after phase 2.

References: Bellprat, O., Kotlarski, S., LuÌĹthi, D., and SchaÌĹr, C.: Objective calibration of regional climate models, Journal of Geophysical Research: Atmospheres, 117(D23), https://doi.org/10.1029/2012JD018262, 2012. Bellprat, O., Kotlarski, S., LuÌĹthi, D., De EliÌĄa, R., Frigon, A., Laprise, R., and SchaÌĹr, C.: Objective calibration of regional climate models: application over Europe and North America, Journal of Climate, 29(2), 819-838, https://doi.org/10.1175/JCLI-D-15-0302.1, 2016. Collins, M., Booth, B.B., Harris, G.R., Murphy, J.M., Sexton, D.M. and Webb, M.J.: Towards quantifying uncertainty in transient climate change. Climate Dynamics, 27(2-3), pp.127-147, 2006. Harris, G.R., Sexton, D.M., Booth, B.B., Collins, M. and Murphy, J.M.: Probabilistic projections of transient climate change. Climate dynamics, 40(11-12), pp.2937-2972, 2013. Murphy, J.M., Sexton, D.M., Barnett, D.N., Jones, G.S., Webb, M.J., Collins, M. and Stainforth, D.A.: Quantification of modelling uncertainties in a large ensemble of climate change simulations. Nature, 430(7001), p.768, 2004. Murphy, J.M., Booth, B.B., Collins, M., Harris, G.R., Sexton, D.M. and Webb, M.J.: A methodology for probabilistic predictions of regional climate change from perturbed physics ensembles. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 365(1857), pp.1993-2028, 2007. Murphy, J.M., Sexton, D.M., Jenkins, G.J., Booth, B.B., Brown, C.C., Clark, R.T., Collins, M., Harris, G.R., Kendon, E.J., Betts, R.A. and Brown, S.J.: UK climate projections science report: climate change projections, 2009. Sanderson, B.M.: A multimodel study of parametric uncertainty in predictions of climate response to rising greenhouse gas concentrations. Journal of Climate, 24(5), pp.1362-1377, 2011. Sexton, D.M. and Murphy, J.M.: Multivariate probabilistic projections using imperfect climate models. Part II: robustness of methodological choices and consequences for climate sensitivity. Climate Dynamics, 38(11-12), pp.2543-2558, 2012. Shiogama, H., Watanabe, M., Yoshimori, M., Yokohata, T., Ogura, T., Annan, J.D., Hargreaves, J.C., Abe, M., Kamae, Y., O'ishi, R. and Nobui, R.: Perturbed physics ensemble using the MIROC5 coupled atmosphere–ocean GCM without flux corrections: experimental design and results. Climate dynamics, 39(12), pp.3041-3056, 2012. Solomon, S., Qin, D., Manning, M., Averyt, K. and Marquis, M. eds.: Climate change 2007-the physical science basis: Working group I contribution to the fourth assessment report of the IPCC (Vol. 4). Cambridge university press, 2007. Stainforth, D.A., Aina, T., Christensen, C., Collins, M., Faull, N., Frame, D.J., Kettleborough, J.A., Knight, S., Martin, A., Murphy, J.M. and Piani, C.:. Uncertainty in predictions of the climate response to rising levels of greenhouse gases. Nature, 433(7024), p.403, 2005. Zelinka, M.D., Klein, S.A. and Hartmann, D.L.: Computing and partitioning cloud feedbacks using cloud property histograms. Part I: Cloud radiative kernels. Journal of Climate, 25(11), pp.3715-3735, 2012.

Interactive comment on Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2018-198, 2018.

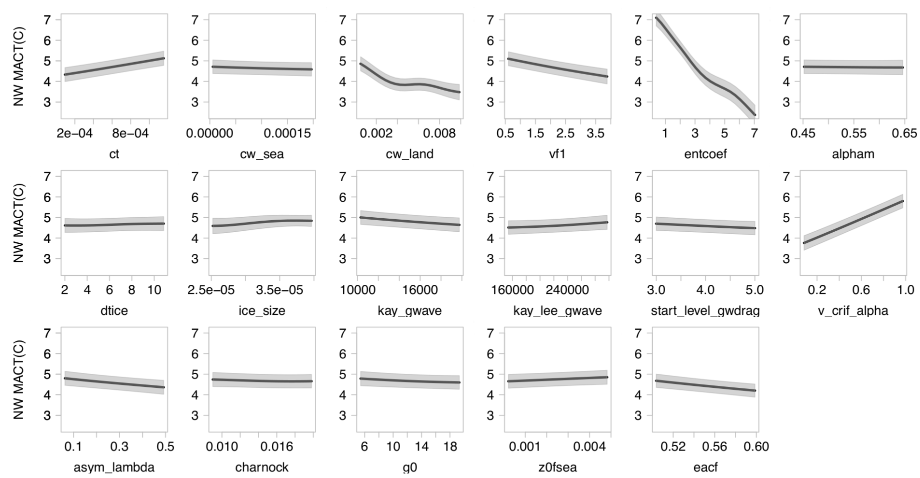

Fig. 1. One-at-a-time sensitivity analysis of magnitude of MAC-T over Northwest to each input parameter in turn, with all other parameter held at SP values. Heavy lines represent the emulator mean, and shaded

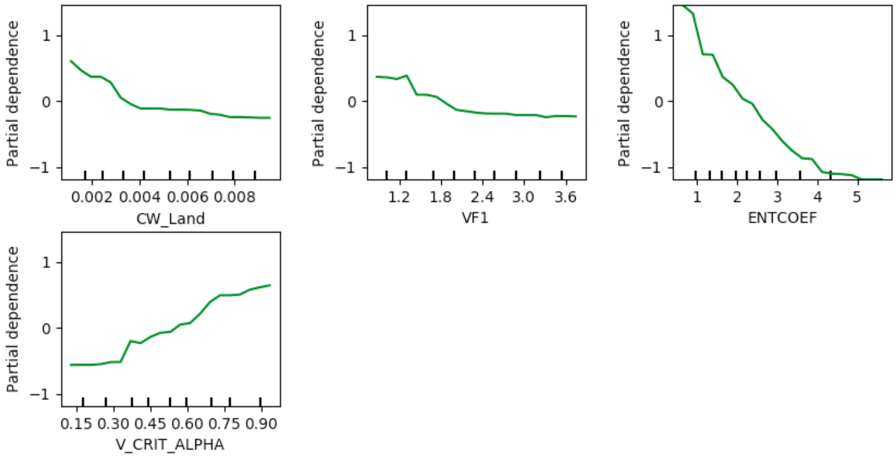

Fig. 2. Partial dependence plots of magnitude of MAC-T over Northwest to CW_LAND, VF1, ENTCOEF, and V_CRIT_ALPHA.

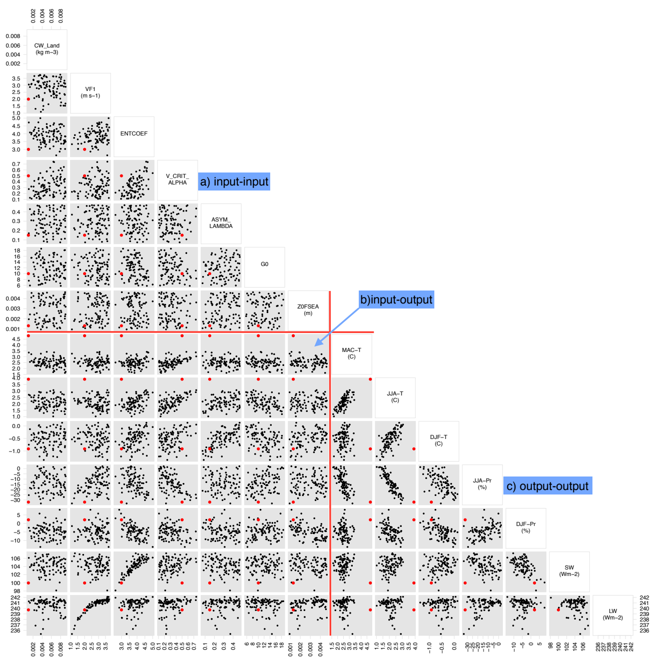

Fig. 3. Updated Figure 4 of the main text.

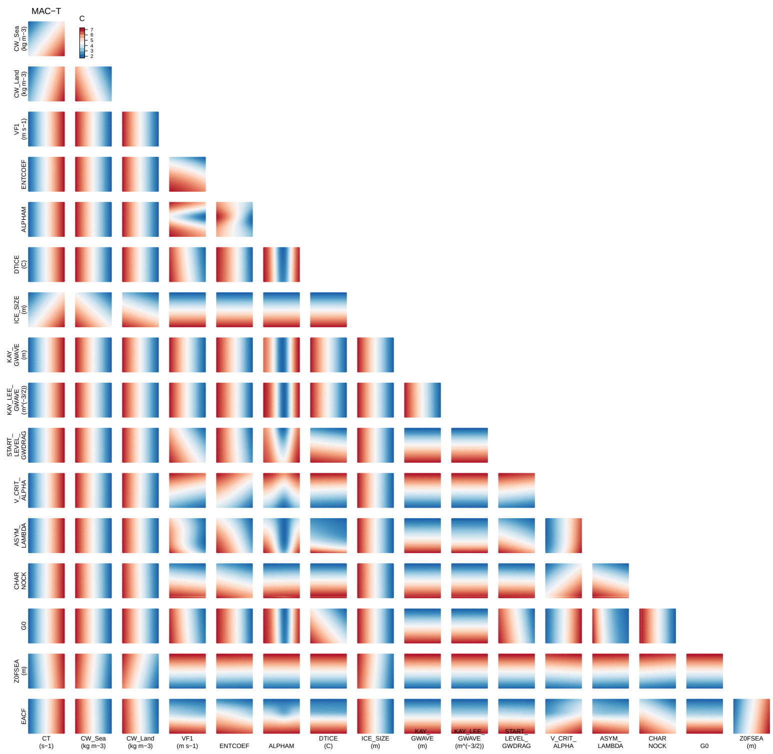

Fig. 4. MAC-T biases projected into the two-dimensional spaces of each pair of input parameters using the emulator.

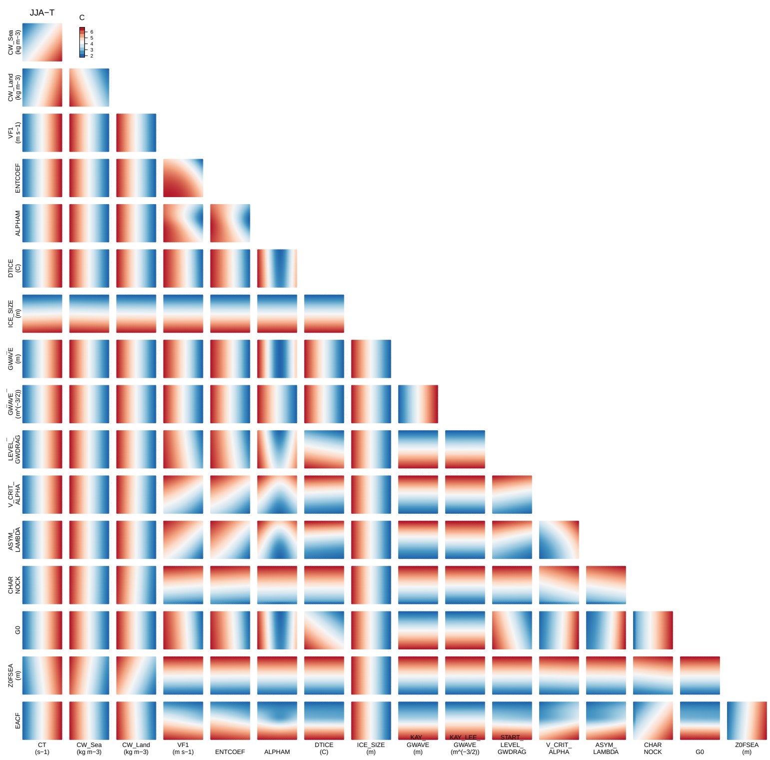

Fig. 5. JJA-T biases projected into the two-dimensional spaces of each pair of input parameters using the emulator.

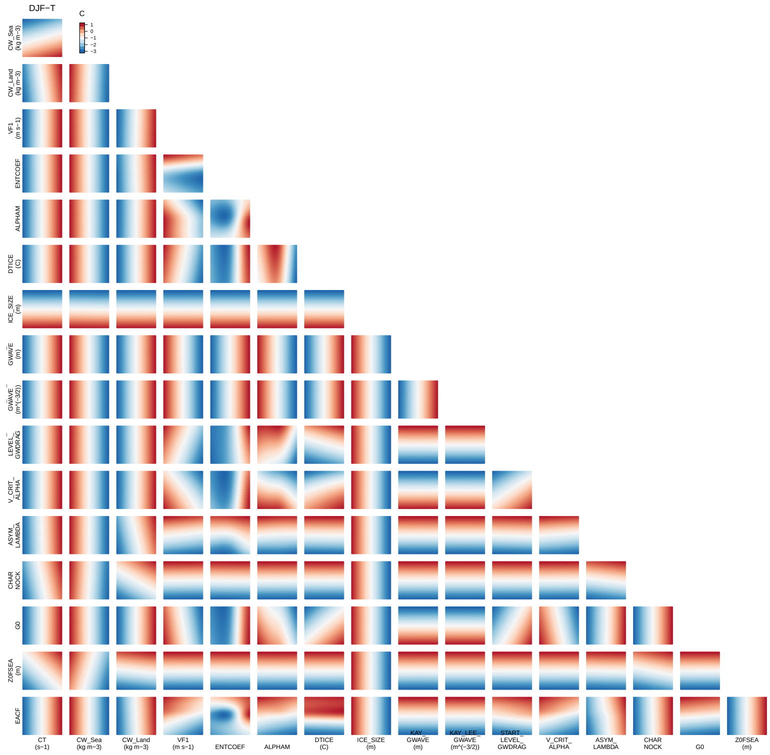

Fig. 6. DJF-T biases projected into the two-dimensional spaces of each pair of input parameters using the emulator.

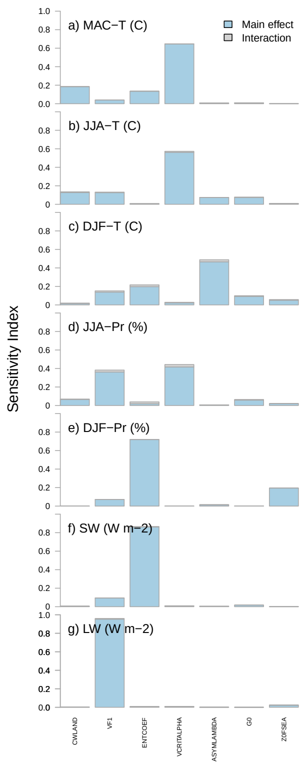

Fig. 7. The sensitivity indices for the refined parameter space in Phase 3.