I would like to thank the authors for the revisions they have provided. They clearly put time and effort in addressing the comments and suggestions we made. The manuscript reads a lot better now, and with the information added for example on the uncertainty, it is easier to appreciate the results to the fullest. I still have a list of rather minor comments which I would like to be addressed before publication, but I think this version of the manuscript is close to final.

- l. 27 and after: “CERES-based kernel” reads more appropriate than “CERES kernel”, as CACK is derived from CERES data but not included in the same dataset and not produced by the same researchers.

- l.89: can you make the acronym more explicit?

- L.268: “Suggested” or “proposed” is more appropriate than “novel”, which may be understood as “recent”

- L. 274: I think I understand what the authors mean, but were the shortwave boundary fluxes really directly compared with the GCM kernel? To me it reads rather like a shortened description of the actual methodological step the authors undertook, in which case I think a more exact description would be necessary. It is still not crystal-clear which CERES variables were considered as potential predictors for the tested models (on l. 271 the authors only refer to “GCM boundary fluxes”, which is less restrictive than “shortwave boundary fluxes”).

- L. 275: where do the ~200,000, 50%, 97% and 32% numbers come from? How many years were taken into account for each GCM?

- L. 295: I believe it is now Section 2a

- L. 296: consider adding “introduced from Section 3b to 3d” after “six simple model candidates” to clarify the procedure. There could be some confusion with the model candidates examined by the machine learning algorithm.

- L. 321: actually if I understand correctly only two candidate models were used for the emulation

- L. 380: are the subscripts of the numerators of the right-hand term correct? Isn’t it supposed to be alpha_CRO,m and alpha_EBF,m? In which case the covariance term between these two variables should also be included, I believe?

- L. 394: “interannal”

- L. 434: “RMSE” instead of “RMSD”

- L. 476: “in all months.” Specify that this is on a global average.

- L. 494: There is an interesting pattern on Fig. 5D, where one can observe, in each hemisphere, a thin band located between 40° and 60° where the relative error is

higher. Can the authors advance reasons for this pattern? And do they know what happens over Eastern China?

- L. 498: if I understand well the results from Figure 6. should be understood as “what happens if these pixels were initially completely covered by evergreen broadleaved forests which would then be replaced by grasslands”. This is different than “what happens if all evergreen broadleaved forests in these regions were to be replaced by grasslands”, but I think the authors should provide more explanation to avoid that these wrong conclusions are drawn by readers.

- L. 508: To be exact, the effect of an increasing albedo trend also emerges, right?

- L. 566: It reads peculiar to have just one subsection.

- L. 8564: the authors could clarify that the “CACK model candidates” are not those of the selection phase by the machine learning algorithm

- L. 896: “mean local” over which domain?