Review of "Radiative sensitivity quantified by a new set of radiation flux kernels based on the ERA5 reanalysis"
By Huang and Huang
essd-2022-474

Summary
Radiative kernels, which quantify the impact of unit changes in individual fields on radiative fluxes, have become a key tool in diagnosing radiative feedbacks both in climate models and in observations. In this study the authors develop a new set of radiative kernels using atmospheric and surface fields from the ERA5 reanalysis as inputs to the RRTMG radiation code. Unlike many previous kernels, they generate kernels for both the top-of-atmosphere (TOA) and the surface (SFC) such that impacts of changes in temperature, humidity, clouds, and surface albedo on surface radiation can be diagnosed. The ERA5 kernels are compared with previously-generated kernels, and inter-kernel differences are illuminated. The authors also explore the degree to which the derived kernels depend on the state of the climate, with input data from years impacted by El Nino events or with anomalous sea ice concentration resulting in kernels of different strength.

Overall I find the analysis to be solid and the presentation to be mostly clear. I have some suggestions for improving the readability of the paper and for presenting the relative importance of inter-kernel versus inter-model feedback differences. I also would like the authors to provide more evidence of the quality of this new kernel versus existing kernels. I recommend acceptance pending minor revision, as detailed below.

Mark Zelinka

Major Comments

- Abstract: Since the goals of this data journal are to publish work that documents useful datasets, with the scientific results being secondary, I felt that the abstract spent too much time on the inter-kernel comparison and not enough on the evaluation of the specific ERA5 kernels developed here. For example, it would be good to know in the abstract whether the new kernels have smaller residuals in the global mean or regionally than previous kernels. The bulk of the abstract describes results from all kernels collectively rather than focusing on the ERA5 kernels.
- The paper discusses TOA and SFC kernels but does not discuss the implied atmospheric kernels, derived via differencing the TOA and SFC kernels. Perhaps this would make the paper too long, but the authors might consider adding something on ATM kernels.
- Organization of the figures: I found it to be really taxing and distracting to have to jump between eight large figures on separate pages during Sections 3.1 to 3.2.
    - Section 3.1 discusses the ERA5 kernels in isolation. I think it would be more logical to have the first figure or two just show all the ERA5 kernels. This would include the first column of Figs 1-8, which is 32 panels. Perhaps you could have 2 figures with 4 rows and 4 columns each. This way a reader can see all of the new kernels just by looking at 2 figures, and can more easily match the discussion in Section 3.1 with the individual figure panels being discussed without flipping between 8 pages. If you do

this, I suggest re-labeling so it is obvious above each panel what one is looking at (i.e., “All-sky SFC Air Temperature Kernel”, “Clear-sky TOA LW Water Vapor Kernel”, etc.)

  - Section 3.2 discusses the inter-kernel comparison, which refers solely to the two right columns of Figs 1-8. I would suggest making these their own figures. Perhaps some of this material could go in supporting information or the appendix, if you don’t spend much time discussing it. Given the choice of journal, the focus of this manuscript should be to present and evaluate the new dataset, so this intercomparison is somewhat superfluous as it currently stands. It might be worth doing a more rigorous evaluation of ERA5 against other datasets rather than this discussion of the kernel differences collectively.

- Multi-kernel dataset: Have you considered doing the community service of placing the common-gridded multi-kernel dataset discussed on lines 294-296 on a public website?
- Throughout: The inter-kernel differences are referred to as “biases”. Perhaps the authors are referring to the fact that all model-based kernels have a biased mean-state with respect to observations, but I think this verbiage is misleading. Also, the definition in L306 quantifies the bias with respect to the multi-kernel average, implying that the multi-kernel average is truth. The inter-kernel differences are a mix of model differences (in mean-state, radiation codes, etc.) and possibly the influence of actual biases (like the issues identified here in the HadGEM and Oslo kernels). If a kernel were to be built from a preindustrial control state, it may be less biased for computing feedbacks with respect to that state than the ERA5 kernels developed here; it depends on the context whether a given kernel is biased. I suggest changing all instances of “bias” to “differences” unless it can be shown to be a true bias with respect to a correct value.
- Tables 3-6: Could these results be presented more effectively? I’m not sure how insightful it is to present all the individual model results in four big tables. The message you are trying to convey is the relative importance of inter-kernel differences versus inter-model differences in SFC and TOA feedbacks, either broken down into LW, SW, and net, or into individual feedback components. I wonder if something analogous to Figure 1 of Chao and Dessler (2021) might be more effective. In this case, you would show the spread in each feedback from inter-kernel vs inter-model differences. Or would simply showing a figure comparing inter-kernel and inter-model standard deviations (ignoring the multi-model mean values) be more effective? Deciding on the most important points and then creating a figure that supports those points clearly would be worthwhile. Right now it is a bit hard for the reader to wade through these four big tables and extract the messages.
- In the end it is still a little unclear to me whether the new ERA5 kernel has a smaller residual than the other kernels. Can you make a stronger case for why we need this new kernel, and whether it is more accurate? Figures 11 and 12 suggest to me that the residuals are comparable to previous kernels; but this should be noted explicitly. If it is not more accurate, why should I use it over previous kernels? If it is more accurate, do you advocate that the community use this instead of the others? I think it is well established here and elsewhere that the inter-kernel differences are small relative to inter-model spread; why are we regularly making new kernels in this case?

Minor Comments

- Verbiage: Throughout the paper, I found some of the verbiage to be unnecessarily long-winded. Could "kernel of the surface flux" be the "surface kernel", for example?
- L22: "in" should be "for"
- L32: I don't understand what is meant by "inter-kernel bias-induced uncertainty", which appears in slightly modified phrasing in other places as well (L557). Is this just "inter-kernel differences"?
- L38: delete "on the other hand"
- L60: suggest also citing the recent work of Chao and Dessler (2021)
- L75: Suggest citing some additional work, some of which includes surface and atmosphere cloud radiative kernels (Zhang et al., 2021; Zhou et al., 2022, 2013)
- L81: suggest specifying "largely insensitive"
- L83: "are" should be "is"
- L107: suggest simplifying to "we intercompare"
- L109-111: suggest rephrasing this sentence, which I found hard to parse. Also, you probably want to specify that you are comparing across-model vs across-kernel differences in this sentence (I think)
- L150-152: I'm confused by how you describe the analysis. I thought kernels were constructed using one experiment, performing many calls to the radiative transfer code, each time with a single field / level / location perturbed. This is not how the procedure is described here.
- L168-169: Probably want to remind the reader why the factors of 4 and 8 are present in these expressions. It is because the radiation calculations are done 4- or 8-times daily, I think.
- L168: "kernels" should be singular
- L190: suggest "upwelling" instead of "outgoing". Also, suggest simplifying to "the kernel is negative"
- L206: should be "(f,l)" rather than "(g,l)"
- L253: "reduce" should be plural
- L257: I think you should specify that you are talking about the clear-sky TOA kernel here.
- L336: "by the inconsistency in" should be "by inconsistencies in"
- L343-344: could this be simplified to "state-dependency in the kernels"?
- L354-355: "the" before "interannual" and "cloudiness" is not needed
- L359: what is meant be "seasonal SST anomalies" Previously, it is stated that you are examining annual means.
- L363: "since" should be "in the"
- L364: "exemplify" should be "illustrate" or "highlight"
- L365: All sky what? Kernels?
- L370: I think some explanation of this result is warranted. Why does Figure 9e have that structure, wherein some regions that are moister and cloudier have a larger SW WV kernel but some do not (NE Pacific). Also, the panel titles in Figure 9 are a little ambiguous; suggest explicitly stating what is shown in each.

- Figure 10: suggest deleting the longitude labels which clutter the figure and seem unnecessary given the provided coastlines.
- L385-394: More explanation of why you get these results is needed. Also, this is too long of a sentence.
- L390-394: Is one of the take-aways here that it may be necessary to average over multiple years when constructing kernels? Or at least that one has to be careful not to choose a year with an extreme Nino index or huge sea ice anomalies when constructing kernels? You might consider making this point explicitly.
- L403: missing space between "Table" and "2"
- Table 2: "model top level" is not an accurate description of what is reported in that column
- L412-419: I think more description and motivation for using these experiments is needed. The abrupt-4xCO2 experiment is a fully-coupled experiment whereas piClim-4xCO2 is an atmosphere-only experiment. You should also cite the relevant piClim-4xCO2 experiment description paper (Pincus et al., 2016). I've never seen these two experiments differenced in order to derive the temperature-mediated responses without the confounding effects of rapid adjustments; this is clever although it limits the number of models available to analyze. (Although more than just 6 models are available as far as I can tell.) I suggest explaining these choices a little better. I would also suggest mentioning this methodological difference when coming your values to those of Zelinka et al (2020) – that study used piControl simulations as the baseline and computed abrupt-4xCO2 anomalies and feedbacks differently. It is reassuring that the results of the two approaches agree as well as they do.
- L445-446: The end of this sentence is redundant with previous statements; suggest deleting.
- L460, L467: small relative to what?
- L477: Suggest stating the name of the row rather than making the reader count.
- L478-480: suggest citing some examples to explain how you arrive at these percentage numbers. Are you comparing inter-kernel standard deviations to inter-model standard deviations?
- L488: these numbers seem misleading, because most feedbacks have roughly the same absolute value of inter-kernel spread; they just vary in the central value. If all feedbacks had the same inter-kernel spread, but one feedback happened to be zero (e.g., if the SW cloud amount feedback perfectly compensated a SW cloud albedo feedback), the inter-kernel spread relative to this would be infinite, but that is not really meaningful.
- L541: Delete "First of all"
- L583: This sentence seems to run on and should probably be broken up for clarity.
- L585-591: this sentence is also way too long and should be broken up
- L594: "it is especially noticed that" can be deleted
- L599-600: suggest making this more concise by removing redundancy
- L601-602: How could inter-model spread come from inter-kernel spread? Please rephrase.
- L603: rephrase to "finding is consistent with previous"
- L762: specify whether this is an absolute or relative change. I'm pretty sure it is the former.
- L767: not sure what is meant by the last phrase
- L818: "trickiness" is probably too informal; suggest "challenge"

- L835: I don’t understand what is meant from “and accounting” onward
- Figure A2: are these SFC or TOA kernels? I assume SFC.
- L854: specify “in these cases”

References

Chao, L.-W., Dessler, A.E., 2021. An Assessment of Climate Feedbacks in Observations and Climate Models Using Different Energy Balance Frameworks. J. Clim. 34, 9763–9773. https://doi.org/10.1175/JCLI-D-21-0226.1

Pincus, R., Forster, P.M., Stevens, B., 2016. The Radiative Forcing Model Intercomparison Project (RFMIP): experimental protocol for CMIP6. Geosci. Model Dev. 9, 3447–3460. https://doi.org/10.5194/gmd-9-3447-2016

Zelinka, M.D., Myers, T.A., McCoy, D.T., Po-Chedley, S., Caldwell, P.M., Ceppi, P., Klein, S.A., Taylor, K.E., 2020. Causes of Higher Climate Sensitivity in CMIP6 Models. Geophys. Res. Lett. 47, e2019GL085782. https://doi.org/10.1029/2019GL085782

Zhang, Y., Jin, Z., Sikand, M., 2021. The Top-of-Atmosphere, Surface and Atmospheric Cloud Radiative Kernels Based on ISCCP-H Datasets: Method and Evaluation. J. Geophys. Res. Atmospheres 126, e2021JD035053. https://doi.org/10.1029/2021JD035053

Zhou, C., Liu, Y., Wang, Q., 2022. Calculating the Climatology and Anomalies of Surface Cloud Radiative Effect Using Cloud Property Histograms and Cloud Radiative Kernels. Adv. Atmospheric Sci. 39, 2124–2136. https://doi.org/10.1007/s00376-021-1166-z

Zhou, C., Zelinka, M.D., Dessler, A.E., Yang, P., 2013. An Analysis of the Short-Term Cloud Feedback Using MODIS Data. J. Clim. 26, 4803–4815. https://doi.org/10.1175/JCLI-D-12-00547.1