Review of "ClimKern v1.1.2: a new Python package and kernel repository for calculating radiative feedbacks"
by Janoski et al
MS No.: egusphere-2024-2561

Summary

In this paper the authors describe a new python package for computing radiative feedbacks using radiative kernels and a corresponding repository of 11 radiative kernels that have been developed by various groups since the technique was introduced in 2008. The authors have brought these kernels together, placed them on consistent grids, given them consistent sign and variable naming conventions, and done additional curation in an effort to better facilitate community usage. At this time, only a subset of the most commonly used kernels (non-cloud kernels for top-of-atmosphere radiation) are part of the repository, with future plans to incorporate other kernels that are used in the community (e.g., cloud radiative kernels and kernels for surface radiation). The python package that the authors have developed for using the kernels to compute radiative flux anomalies is a major advance, as authors wishing to compute radiative feedbacks have generally either had to write code from scratch or follow someone else's code that is generally not well documented, commented, etc. It has basically been the wild west on this front for ~15 years. Given the number of methodological choices that need to be made in computing radiative feedbacks with kernels – choices that can have sizable impacts on the resulting feedback values – it is not ideal for the community of practice to be reinventing the wheel for these calculations. Having a dedicated package to perform these calculations and to quickly assess sensitivity to kernel and some methodological choices is very much welcome. I found the paper to be well written and illustrated, and I recommend acceptance of this manuscript after the revisions detailed below.

Major Comments

- **Role of effective radiative forcing (ERF) in the calculations.** First, on L134, I suggest providing a little more detail here regarding how ERF is computed. This is an input for the adjusted CRE calculation, so if users wanting to compute feedbacks outside of the tutorial dataset will need to know how to compute ERF. (Side note: is it worth at some point incorporating an ERF calculation capability into ClimKern?) Second, and more importantly, what if the end-user does not have ERF or chooses not to provide it? Can the calc_cloud_LW and calc_cloud_SW functions still be used if ERF is not provided? In the case of abrupt-4xCO2 simulations, the forcing is not changing through the course of the run, so if one is computing feedbacks via regression of the TOA anomalies on global mean surface air temperature (Gregory et al. 2004), the ERF term in Eq 4 and the ERF masking term in equation 5 should be zero (I think – correct me if wrong). Alternatively, when computing feedbacks from idealized atmosphere-only warming experiments (e.g., amip-p4K minus amip), there is no radiative forcing, so this term is zero by definition. I suppose the end user could provide a DataArray of zeros for the ERF term, but this is sort of klunky relative to the code allowing for this to be an optional input field.
- **Computing water vapor anomalies (L220-234).** Feedback junkies like myself have been down this dark and lonesome road, but the average reader is likely to get rather lost in this section. I think providing the relevant equations would help the reader to understand that there is some ambiguity in the right way to compute humidity anomalies and to better rationalize the four choices. A follow up question is have you assessed whether one these four choices is clearly superior and/or whether one or more are clearly inferior? Surely they can't all be equally useful, right? I think you may be in a unique position to weigh in on this, or at least report on a null

result. In my own experimentation, I seem to recall these things tending to be equivocal – some methods work better for some models and some work better for others; do you find the same?

- **Clear-sky linearity tests to evaluate kernels.** Related to the previous comment, I was surprised that you did not present clear-sky linearity tests (Shell et al. 2008), which would allow for an evaluation of which kernels best close the TOA energy budget. There seems to be a desire not to evaluate whether certain kernels are better or worse, but this would be a very useful thing for the community to know. I suppose one issue is that you have only applied the kernels to a single model, one that happens to have a corresponding kernel, which could give it an advantage in this test. So I understand the choice not to weigh in on this. However, I can't understand the statement in the conclusions that using a mean kernel would be advantageous in computing feedbacks. If one kernel is superior, then averaging it with inferior kernels should not improve things. I would expect, for example, that kernels built in late-2000s era climate models (Soden et al. 2008) would be inferior to those built from more modern GCMs or reanalyses with vastly better mean-state cloud properties and improved representation of gas optics in the radiative transfer schemes (Huang and Huang 2023). I recommend deleting these statements in the conclusions.

Minor Comments

- Title: should "v1.1.2" be in the title? Most of what is described is applicable beyond this specific version, I would presume.
- L62-66: It may be worth noting that Zelinka et al. (2020) assessed sensitivity of results to kernel choice as well (their Figure S2).
- L135: Why is the IRF provided?
- L190: suggest clarifying that the tropopause height input is optional
- L243: Somewhere in here I think you need to mention that the package computes all the previously described feedbacks for clear-sky conditions as well, using the respective clear-sky radiative kernels. Otherwise when you get to the cloud feedback calculations, it is unclear where the clear-sky feedbacks come from.
- L245: should "most" be "all"?
- Eq 5: I think there should be parentheses around the two ΔR_i terms that follow the summation. Also I think the nomenclature could be confusing, since the subscript "all-sky" appears in some equations but not in others.
- L304-305: it is stated that each kernel exhibits differences in the standard deviations; I think you mean "as evidenced in the standard deviations" or something like that? Also, each kernel exhibits differences between the all- and clear-sky versions. That doesn't seem surprising to me. Or are you talking about the interkernel differences in how different the all- and clear-sky kernels are? I think this sentence needs to be re-written for clarity, since the first part deals with inter-kernel spread while the latter deals with all- vs clear-sky differences within a given kernel (I think).
- L319: I would have thought solar path length through the atmosphere would be highly relevant too.
- L330-334: I think you should provide more explicit detail about how you did your feedback calculations here. Which WV feedback option was used? Did you integrate up to the default tropopause, or did you compute the tropopause explicitly? (Side note: is it worth at some point incorporating a tropopause calculation function into ClimKern, something like PyTropD?). To clarify: are you differencing a climatology from the last 30 years of abrupt simulation and a climatology from the last 30 years of the piControl simulation, or are you using a climatology

from the full 150-year abrupt simulation? In either case, I suggest mentioning that this difference of perturbed and control climatologies is not ideal for computing feedbacks in abrupt 2x or 4x CO2 runs because rapid adjustments are aliased into the feedback when computed this way. Computing the TOA anomalies throughout the duration of the 150-year abrupt experiment and regressing them on coincident global mean surface air temperature anomalies is preferred. Related to this, does the code require that both the perturbed and control data inputs have no more than 12 months? Can one input perturbed climate fields that are length N*12 months (where N is the number of years) and have the code difference them with the 12-month long piControl climate, yielding N*12 month TOA anomalies?

- Figure 2: Why is the standard deviation multiplied by 2? I don't love how the colorbar scales change among the figures, especially for the right column. Could the standard deviation colorbars be objectively related to the means (e.g., from 0 to some percentage of the range of mean magnitudes)? Currently the tiny interkernel WV and Planck feedback spreads are over-emphasized relative to, say, the cloud feedback spread.
- Section 4.1.1: Suggest reiterating somewhere in here (or at multiple places) that these are just results from a single model (CESM1-LE). Also, I may have missed it, but are you using just one ensemble member? Are the other members of the LE just used for diagnosing ERF?
- L363: I don't really see this (much of the ocean has a positive cloud feedback and much of the land has a negative cloud feedback), so I don't think it should be the primary feature to highlight.
- Figure 3: The fact that the y-axis ranges are so different (some only span 3 W/m2/K while others span 12 W/m2/K) tends to mislead regarding interkernel spread. I think these should either be put on equal footing or this plotting choice should be pointed out more explicitly.
- L400: I don't see this. Table 2 shows that the interkernel standard deviation of Planck is larger than for 3 other feedbacks (WV, surface albedo, and LW cloud). Are you referring to the sensitivity as a fraction of the mean?

References

Gregory, J. M., and Coauthors, 2004: A new method for diagnosing radiative forcing and climate sensitivity. *Geophys. Res. Lett.*, **31**, https://doi.org/10.1029/2003GL018747.

Huang, H., and Y. Huang, 2023: Radiative sensitivity quantified by a new set of radiation flux kernels based on the ECMWF Reanalysis v5 (ERA5). *Earth Syst. Sci. Data*, **15**, 3001–3021, https://doi.org/10.5194/essd-15-3001-2023.

Shell, K. M., J. T. Kiehl, and C. A. Shields, 2008: Using the Radiative Kernel Technique to Calculate Climate Feedbacks in NCAR's Community Atmospheric Model. *J Clim.*, **21**, 2269–2282, https://doi.org/10.1175/2007JCLI2044.1.

Soden, B. J., I. M. Held, R. Colman, K. M. Shell, J. T. Kiehl, and C. A. Shields, 2008: Quantifying Climate Feedbacks Using Radiative Kernels. *J Clim.*, **21**, 3504–3520, https://doi.org/10.1175/2007JCLI2110.1.

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