Review of Duran, et al. – "A new method for diagnosing effective radiative forcing from aerosol-cloud interactions in climate models"

This paper describes a technique to calculate aerosol ERF using monthly mean LWP-reff histograms from COSP and radiative kernels. The ERF can also be broken down into contributions from droplet number, LWP and cloud fraction changes. The use of monthly mean histograms means that high frequency bespoke output is not required, hopefully making the technique applicable to more models. The results compare well to other techniques where it is possible to compare them. ERFs and its components are compared between a few different climate models. Evaluation of the histograms from the model those from observations are performed, which is a useful way to assess model performance. The MIRCO6 models perform poorly, which may be due to issues with the cloud phase identification from COSP. An emergent constraint is attempted based on the relationship between the Twomey/LWP adjustment and liquid cloud fraction (constrained by observed cloud fraction).

The paper is very well written and the analysis is sound. I have very few criticisms to make, just a few small points below. The biggest concern is the situation with the MIROC model - it's use in the emergent constraint is dubious given that the results from that model might be being caused by issues identifying the cloud phase from COSP – either that or the model cloud field is very unrealistic. It would be good to determine which is the case.

Issues

L32 – "The Twomey effect, or instantaneous radiative forcing (IRFaci)"

- IRF refers to the initial perturbation to the Earth's radiation budget without any tropospheric, land surface or stratospheric adjustments (e.g., Smith, 2020). So the Twomey effect could be an IRF, RF or ERF depending on the climate model set up for how it was quantified. Perhaps you could just use ERF for consistency with the other terms? Or, if the Twomey component is different to the other components (no adjustments vs adjustments) in your breakdown calculation, it would be good to explain why the Twomey component is an IRF and the difference between IRF and ERF. But since this is the introduction and you haven't wrote about what simulations you use yet it might be better to stick to ERF? If so you'd have to correct it throughout the paper.

L62 – It would also be good to mention the technique of Grosvenor (2020) that used a simplified radiative calculation for low liquid clouds to estimate the breakdown of ERFs.

MIROC6 results – the cloud phase identification issue is very important for the evaluation of MIRCO6. It might be good to discuss why there is this issue with MIROC some more. Presumably the fraction of undetermined phase clouds is higher than observed by MODIS, which may tell us something about whether the undetermined categorization is realistic. Is it

possible to evaluate MIROC in stratocumulus regions where it is know that there is little ice – then you could assume all of the model clouds were liquid and evaluate the MIROC cloud fraction again. Maybe this is what you allude to at line 498, but it's not clear whether you are still restricting to only liquid clouds as identified by COSP there. I.e., does MIROC really have such unrealistic cloud fractions, or is it due to the phase issue?

Since the results for this model are based on a very limited spatial sample (due to the cloud phase identification issue) it is perhaps not ideal to use it in the emergent constraint.

Typos/grammar

P8, L231 – "The two versions of MIROC6 have the weakest SW ERFaci, because they have less liquid cloud in the control state, so fewer clouds susceptible to aerosol perturbations and weaker aerosol forcing."

This would be better as "The two versions of MIROC6 have the weakest SW ERFaci because they have less liquid cloud in the control state and so fewer clouds susceptible to aerosol perturbations."

P14, L414 – "The relationship between mean state biases and the historical SW ERFaci identified in the previous section suggest" – "suggest" would be better as "suggests".

References

Smith, 2020 - https://acp.copernicus.org/articles/20/9591/2020/

Grosvenor, 2020 - https://doi.org/10.5194/acp-20-15681-2020