

We thank the reviewer for their comments and have elaborated in the Discussion section on the points raised.

*After reading the manuscript, several questions came immediately on my mind:*

*(1) Why are the improvements by the categorical substantial less convincing over land than over ocean?*

There are likely additional factors influencing retrieval quality and averaging kernel structure over land that we have not considered. For  $p_D$  in particular, the observational controls over land were weaker (Table 1), making their approximation in the categorical operator more difficult. As the categorical operator evolves, we will start by testing topography, land cover type, and, related to both, thermal contrast between ground and air, which will be greater over land than ocean. In the latter case, the poorer performance over land could be amplified by the fact that we considered daytime retrievals only. Also, the land-based categories will be based on fewer measurements, making them less robust than over the ocean. We have expanded on this on P26L22 of the revised manuscript.

*(2) Why does the categorical approach yield rather poor results if applied for extra-tropical regions?*

We suspect this is because the averaging kernels were aggregated over too broad a range of conditions. The first change that could be considered would be to compute categorical retrieval quality and averaging kernels over several different latitude bands, although it is less than ideal. We note too that the poor performance shown in Fig. 17 is not specific to the extratropics. Computing the categories over such a wide latitude range actually worsens the performance over the tropics also.

Over land, the deficiencies brought up in the first point will be exacerbated as the latitude range and variety in land-cover characteristics increases (especially in the northern hemisphere). Over the oceans, assuming that vertical moisture gradients are in fact the dominant factor outside of the tropics, a strong possibility is that the  $PW_B$  and  $PW_F$  bins are too coarse. At low latitudes, we used as few bins as possible to capture the difference between the tropical rain bands and the subtropical anticyclones. More bins at the low end of the moisture ranges will probably be needed in the extratropics.

We have expanded upon this at P27L7 of the revised manuscript.

*(3) How strongly will the GISS results with the categorical approach change if model results from a free (non-nudged) simulation are used?*

We would expect that the basic relationships between retrieval quality,  $p_D$  and the control variables would be similar (i.e. in Figures 11 and 13). The modeled retrieval quality and  $p_D$  fields (i.e. in Figures 10 and 12) will change to the extent that the underlying control fields change, and more specifically, the covariation between the control variables changes. One potential weakness is that a new model configuration will have an increase in conditions corresponding to categories that were not well populated and for which the retrieval quality and mean averaging kernels are less robust (although the opposite could also be true).

*(4) How well will the categorical approach work for other isotope-enabled AGCMs?*

We would expect that it would work in other models if a similar approach to subgrid cloud characteristics is taken. Ideally, the performance of the categorical approach should be model-independent. It will more strongly depend on the categorization used, from the different variables and coarseness of their binning. We hope that this can be examined by other groups working with isotope-enabled AGCMs. The simulator code is available as part of the ModelE distribution:

<http://www.giss.nasa.gov/tools/modelE/>.

We have expanded upon points 3&4 P28L5 in the discussion.

*Minor suggestions / corrections: none*

- page 13845, line 6: delete first “in” after “regions” Done – thanks.

- page 13846, first paragraph: *The results for  $T_s$  (Fig. 11g-j) should be explained here, too*  
We have changed the description at P19L25 in the revised manuscript to explain the results over land.

- *Fig. 11 and Fig. 13 are too small*

We have removed the ‘CPW’ panels from both (along with those in Fig 10, Fig 12 and Fig. 15), and increased the size of the text in the figures. We have removed discussion of the CPW categorization from the text, with, we hope, little ill effect.