Response to reviewers
ACP-2016-653
Original reviewer comments in normal typeface. **Responses in bold.**

Reviewer #1: Daniel Kirk-Davidoff

In this paper, Kravitz and co-authors introduce the System Identification (SI) technique to probe the linear dynamical response of climate models to localized perturbations. They demonstrate that by perturbing the temperature of adjacent regions of the ocean with orthogonal noise and then filtering the global response by the time series for each region, they can approximately reproduce the effect of a step change in temperature in one of those regions on the low cloud and surface latent heat flux fields.

We thank the reviewer for the careful attention to our manuscript.

While I found the approach intriguing, I feel that there are three key analysis steps missing from a publishable paper. First, since the study reported here involved three simulations using the SI procedure, it seems odd that only ensemble means, but no ensemble variability was shown. The results were said to be robust across the three runs, but why not show the readers the variability? Why not perform a few more runs, so that some measure of statistical significance of the results could be shown, for example by stippling, in the figures?

The reviewer brings up a couple of valuable points. We have removed these lines and added a new section in which we discuss this more quantitatively.

First, we conducted two additional system identification simulations; all results now show the ensemble average of five runs. The addition of these runs did not change broad features of the sensitivity maps shown in the manuscript. Moving from three to five ensemble members reduces the noise by a factor of sqrt(3/5)=0.77, so the plots should not be visually different. One might expect differences in estimates of the inter-ensemble standard deviation, which we have now included in the manuscript as a figure.

To address the problem of statistical significance, we have now included a figure showing the ensemble-mean sensitivity, masked out where the inter-ensemble variability failed a Student's *t* test at the 95% confidence level. We performed a different method of calculating statistical significance by creating 1000 random sequences with the same frequency content as the other sequences used in the paper (although this time they are not mutually uncorrelated), projected the control simulation onto each sequence, and calculated the standard deviation across all 1000 projections. This gives us a means of estimating natural variability in the sensitivity fields. All values in the ensemble mean sensitivity field were over 2 standard deviations of the natural variability.

We have included a discussion of all of these new results in the revised manuscript.

Second, there's little discussion of the character of the differences between the step-change runs and the SI results. What's going on with the synoptical scale variability in latent heat flux found in the mid-latitude storm tracks in SI runs (figure 4, middle row, right column)? Are these natural modes of variability whose frequency happens to correspond to the frequencies excited over the Indian Ocean? Are these patterns also found in the response for many other regions. In this regard, it would be helpful to see the response patterns to the perturbation introduced in at least one other region besides the Indian Ocean, even without a corresponding step-change experiment. Are the remote responses corresponding to the perturbations in these regions similar to the ones from the Northwest Indian Ocean? If so, is there a filtering procedure that could potentially be used to screen them out?

This is a really useful comment. We note that these values are not statistically significant using the metric presented in one of the new figures, so we are reluctant to discuss those features without strong evidence for a physical mechanism underpinning those results. Moreover, because our intention with this manuscript is that the scope should remain within that of a Technical Note, we have eschewed detailed explorations of particular features.

Nevertheless, the reviewer brings up an excellent point, in that we have not fully described how the advantages provided by system identification could be used to gain deeper insight into teleconnections and their underlying physical mechanisms. We have added an additional paragraph to the discussion section describing these procedures, using the suggested mid-latitude storm tracks as an example.

Finally I believe it would be helpful if the authors could clarify the usefulness of the detection of purely linear responses to climate perturbations (subject to some accidental contamination with non-linear responses). Since the response to any observable forcing will include non-linear features, the authors should say a little more about the benefits and hazards of isolating the linear response using SI.

Agreed. We have combined this with a suggestion from the other reviewer to include a more thorough discussion of other linear methods. We have added an additional paragraph to the manuscript that addresses these issues.