General report on Edited version

This is an excellent combination of aircraft flux measurements, machine learning algorithms extrapolated to surface response functions. It also links to large scale modeling to allow contextual interpretation with respect to hydrological and energy budget response focused on an important climate sensitive region. It provides a significant advance in the area of airborne flux measurements and relevance to validation of surface response function dependent models. This is an important paper as this very thorough approach to airborne eddy covariance fluxes has really been missing from the scientific literature over the past 1-2 decades in general. The regional model comparison is a nice addition to capture the mesoscale variability and scales although the study is limited by the surface data availability. I do like the terminology used for model-aircraft comparison.

Updates
Relevant discussion on the design and implementation of the aircraft campaign is included and is sufficient for replication and addressing of issues and potential artifacts in such approaches.

The methodologies are very well described and relevant to the technique applied. These are appropriate to the conclusions arrived at with some limitations however in completing the full energy budget. These could have been discussed further with respect to the uncertainties but generally I don't think this could be improved on.

The relevant transform scales based on the flight track described appear consistent with the approach and is explained well and are also consistent with results previously published in the literature (although these were limited in terms of surface site comparisons). The relevant edge effects associated with the wavelet analyses are always an issue but I think these are within the uncertainties when scaling to the regional observations and looks quite reasonable. Whilst data quality control is critical for such wavelet analyses and could be quantified further I think we can assume this is good based on the results. There are other transform approaches that could have been compared but likely these would not have changed the results.

Page 8:
The addition of the distribution "rug" plot, Figure 5, is very useful.

Minor Questions and Formatting Issues
Item 1. Some brief comment on the appropriate optimization of relevant straight and level sampling altitudes for the flux measurements (discussed page 5 etc) with respect to heterogeneity scales within the flight track would be helpful but not essential here?

Item 2. Figure 4. Some of the arrows in the boosted regression tree figure overlap/obscure the text in the various nodes, e.g. a > 0.5, S↓ > 380 W m^{-2}, r> 7 k kg^{-1},

Item 3. Figure 7. It appears obvious that there are two clusters within the sensible and latent heat flux (predicted versus aircraft measured) domain with significantly different slopes with under-predictions at high values in each case. Can the authors comment on this? Is there a potential bias here?

Item 4. Figure 8. Legend: Has the standard error used in this figure been defined?

Item 5. As mentioned, the impact of enhanced convective conditions suggests potential under-sampling bias of all relevant scales in these conditions. It would be useful to mention the range therefore where such comparisons may break down, but this may require more detailed spectral analysis for another discussion. However I think this caveat/statement addresses the issue adequately for the work presented.

Item 6. Figure 7 is a brave plot (and we need more of them in the literature before relying overly on tower data). I think the discussion and literature references regarding the discrepancies with WRF are adequate but do highlight that there is still a great deal of work to do here.

Final comment: The authors are to be commended for delivering an excellent set of results.