Review of "Cloud Regimes Over the Amazon Basin: Perspectives From the GoAmazon2014/5 Campaign" by S. Giangrande, D. Wang and D. Mechem

Yizhou Zhuang (zhuangyz@atmos.ucla.edu)

The authors present a k-means clustering analysis of thermodynamic conditions in the Central Amazon as represented by the radiosonde launches during the GOAmazon 2014/5 field campaign. They identified five regimes related to wet, transitional, and three dry types, respectively. Composite cloud and precipitation properties, convection statistics, large-scale circulation, and moisture advection related to these regimes are further contrasted. Finally, the authors discussed how these thermodynamic regimes can be linked to occurrence of different convection types. This manuscript is well written and it's interesting to see the clustering technique being applied to segregate local thermodynamic controls as compared to simple seasonal composite analysis in most previous studies. Overall, most of their conclusions are consistent with previous research efforts, but they also provide another angle to understand the relationship between the complex convection characteristic over the Amazon Basin and various types of seasonal thermodynamic controls. However, I do have a few relatively minor comments, which are mostly related to clarification of some points and improvement of figures. After addressing these I think this manuscript should be ready for publication in ACP.

## Major comments

### Selection of Radiosonde in Clear Conditions – section 2.1

In Line 99-100, the authors state that they only use radiosondes that launched in clear conditions. This is a very good practice for capturing pre-convection condition and studying shallow-to-deep convection transition. In Line 301-306, the authors also state that the enhanced moisture advection in regime 4 and 5 are not influenced by precipitation constrains since 1200UTC is prior to significant precipitation. However, I think the one-hour constraint for clear condition is probably too short to reduce the influence from early morning convection on the 1200 UTC sounding, especially during the wet and transition seasons. I would suggest use clear condition for at least 3-6 hours prior to sounding time and 1 hour after that, or at least discuss how the nocturnal convection can influence the 1200 UTC sounding and your results, especially those related to moisture.

### K-means Clustering Method – section 2.2

In Line 141-142, the authors described the clustering process as "radiosonde temperature and wind information is input at …" following Pope et al. 2009b. I think it can be made clearer that how many variables go into the clustering process (temperature, eastward and northward wind speed?)? Also, I'm concerned about why humidity information is not included as input since humidity is also a very important aspect in thermodynamics, and it can be very different during different seasons in Amazon and show significantly influence on buoyancy profile (e.g. Zhuang et al. 2017;

2018). I think justify this point will help readers better understand your basis for clustering. Also, I did not find out if the author preprocess the data before inputting them to k-means clustering.

In Line 145-147, "Although the authors prefer the solution that does not use normalized inputs … select consequences are discussed when these inputs result in divergent solutions". I'm still confused here, what kind of input is finally used to produce the final clustering results shown in the manuscript. I didn't find discussion about how this choice of input type would affect your results either. Perhaps it's better to move/add related discussion here or in the summary and discussion section. In addition, if you are using original or anomaly profile as input, did you assign weights for different variable? Would this affect your results? I'm asking this because these variables are in different units and the weighting can still have some influences even if the units are the same.

Line 141-142, "input at 20 equally-spaced levels from the 1000 hPa to 200 hPa, …". I assume that this means equal weighting for different vertical layers. However, it seems to me the middle and upper level thermodynamics is much less important for convection than the lower troposphere. I'm wondering if some upper level thermodynamic disturbances could mask the lower level information and thus affect the clustering results. Maybe the authors can briefly comment this point. Also, the authors only show median profile in Figure 3, but I think a figure (either in the manuscript or the supplementals) showing both the mean/median and one standard deviation range of the input profiles in each regime would help address this point and show how well these five regimes represent the data.

**Large-Scale Synoptic Conditions – section 3.2**

Please justify the use of 1000-hPa geopotential to represent large-scale circulation. For me, 1000-hPa is not a commonly used level for this kind of analysis, and I would prefer mean sea level pressure for surface system, 500-hPa or 200-hPa streamline for mid- to upper level circulation, 850-hPa wind for moisture advection analysis (consistent with many studies that low level moisture is more important for convection development and also your later results in Figure 7, 8 & 14).

**MCS in Regime 4 – section 4.3**

Many studies (e.g. Williams et al. 2002, Zhuang et al. 2017) has shown that the transition season has a more unstable environment possibly contributed to its more intense convection than the wet season. It's also very interesting here (Line 441) to see that nearly half of the locally formed MCS are observed during the transitional regime. The authors have compared the thermodynamics between the nonMCS and MCS cases in regime 4 (Line 454-461), but I'm more interested about why regime 4 can produce about twice MCS cases as many as those in regime 5. Can the results from early sections be used to explain this? Perhaps some of the discussions from Line 238-247 can be moved here.

**Summary and comparison of the results to broader literature – section 5**

The summary section in the manuscript only lists some major findings throughout the previous results section. This section should include a more detailed discussion about how the results relate to and differ from previous studies. The bullet points for major finds should also be shortened to be simpler and more precise. It is also worth mentioning the advantage of applying this clustering technique to study thermodynamic controls of Amazon convection compared to regular seasonal composite analysis.

## Minor comments

**Abstract**

Line 11: "three dry-season clusters". There are many places in the manuscript that use "dry season regime/cluster" or "drier season regime/cluster". I would suggest drop the "season" and simply use something like "dry regime/cluster" since these dry regime samples are also observed during the commonly defined wet or transition season. Also make sure the terminology is consistent throughout the manuscript.

Line 12: "… for each regime for characteristic cloud frequency …" looks confusing. Please rephrase.

Line 15: Again, what is "driest regimes". Is it just regime 1 or regime 1-3? Simply use "three dry regimes" if you were referring to regime 1-3?

Line 15: What is "those" refer to?

Line 15: "convective inhibition CIN". No need to write down abbreviation here.

**Section 2**

Line 138-139: Please provide references for these commonly defined seasons.

Line 141: "is input" → "are input"

Line 142: "over North Australia" → "over the North Australia"

Line 165-166: This sentence looks weird and hard to follow. Do you mean in their studies, rainfall trends and onset measures indicate 2014-2015 wet season onset occurred later? How can rainfall trend relate to onset time? Please rewrite and make it clearer.

**Section 3**

Line 206: This information should be also included in the caption of Figure 4.

Line 212: Is it 4-6 m/s in the dry season versus 2-4 m/s in the wet season? The dry regime spread looks wider than wet regime in Figure 4f.

Line 267: "composite westerly wind components over the MAO T3 site"? Where does this information come from? Figure 5e? I don't think the wind above the green star is significant westerlies.

Line 267: Be consistent with site name. You used MAO site many times and MAO T3 a few times throughout the manuscript.

Line 268: Same as the previous comment, I don't find the wind field above MAO in regime 4 much different from regime 5. Also, as I pointed out in the major comment, I would suggest use 850 hPa if you want to use wind to indicate moisture transport. This is more consistent with previous studies and your results in Figure 7-8.

**Section 4**

Line 312: How is Figure 9 correspond to Figure 7? If there is no specific link, I think you can simply drop "that correspond to Figure 7".

Line 353: "moister" → "wetter".

Line 381: In Figure 11, why is the overall average rain rate higher than that of any regime during 03-12 UTC? Also, why is there no nocturnal precipitation here while there are significant clouds during late night and early morning in regime 3-5 in Figure 9. What is "regime-events having measurable precipitation"? Did you explain this before?

Line 385-386: You mentioned the uncertainty of radar estimated precipitation here. Can you also briefly introduce in the method section how the precipitation is derived from radar reflectivity (Z-R relation)? As I can recall, they only use the wet season Z-R relationship from the disdrometer to calculate all precipitation data. This information can be found in the ARM-MAO PI dataset.

Line 394: "the most frequent clouds we observe are" → "the time with most frequent clouds are"

Line 399: "lower-relative domain rainfall rate"? Do you mean lower domain rain rate?

Line 426: "e.g., defined by a minimum area of Z>20dBZ of <200km2". If this is the definition you used for non-precipitating event, remove "e.g.,". Also, perhaps "minimum area of Z>20dBZ is less than 200 km$^2$" is better. minimal area of Z>20dBZ of 200 km2? What is definition of isolated, and widespread precipitation event.

Line 431: "km2". Use superscript for square. Also check elsewhere in the manuscript.

**Figures**

Figure 1. Texts and numbers in this figure are too small. Consider increase the font size (also apply to some other figures), and use a legend like Figure 2 instead of listing R1, R2, … for all of the pie charts.

Figure 2. Add a legend for different colors of dot in Figure 2a. Increase font size of R1, R2, … in Figure 2b. Also, to match the definition of seasons and make it easier for readers to understand the result in Figure 2b, please consider only use four main color tones to represent wet, dry, and two transitional seasons. For different month in one season, just use different levels of darkness of the same color.

Figure 4. Explain in the caption what's the thick black line in the middle of the density plot. Make the white number in bold font (also apply to Figure 12).

Figure 7. “600-hPa/700-hPa” → “600-hPa (f-j) / 700-hPa (a-e)”

Figure 8. What you plot is dash-dotted line not dashed line.

Figure 9. Add unit to the colorbar. Why is tick numbers not aligned with the color?

Figure 11. The shading areas look very narrow for standard deviation. Is it one standard deviation or standard error?