To begin, the authors would like to thank the reviewer for their time, attention to detail, and thoughtful insights on the paper and research. Each comment will be addressed point by point.

5

Specific comments:

Page 3, lines 23-24: Upon examination of Ciesielski et al. (2014), which documents the sounding types used at the sites in question during the DYNAMO experiment (October 2011 to March 2012), they note that the Singapore site changed sonde types during DYNAMO. Did the other sites use the same sonde type for the entire period examined (2008-2016)?

10

According to the World Meteorological Organization Catalogue of Radiosondes and Upper-air Wind Systems, there is no record of the Ranai (Modem M2K2DC sonde) or Puerto Princesa (Sippican Mark II sonde) sites switching sonde types during the study period (2008-2016).

15   Page 3, line 24: In what fields are the constant bias corrections available (e.g., humidity, temperature)?

The reporting biases for temperature and wind are small across sonde type and are generally accepted as is, barring external influences at the site. Bias corrections therefore exist mainly for humidity, which is the most uncertain measurement in radiosonde observations. For the VRS92 sondes (Singapore post 21-Dec-2011), the NCAR radiation bias correction (Wang et
20   *al. 2013) can be applied to correct for a solar radiation dry bias found in daytime soundings and is a form of mean bias correction. The Modem sonde used at the Ranai site can be corrected with the constant bias technique outlined in Nuret et al. 2008. The Sippican sonde used at the Puerto Princesa site was found to have relatively small RH errors (Wang and Zhang, 2008), along with the Graw sonde (Singapore pre-21-Dec-2011). The biases in the Sippican and Graw sondes were deemed small enough that correction was unnecessary. (Ciesielski et al., 2014). To err on the side of caution, during quality control the*
25   *inaccurate low-level soundings of the Graw sonde, specifically at the Singapore site (Ciesielski et al. 2014), were removed rather than corrected. For the three sites in this study, the sonde type was either reported as precise enough for scientific research, or a form of mean / constant bias correction was available to improve its accuracy.*
      However, rather than implement the corrections, a mean bias test was performed (Supplementary Figure 1), whereby a positive and negative humidity mean bias were introduced to the entire Ranai dataset from xx hPa to xx hPa. The
30   *choice of both a positive and negative bias was deliberate to ensure that an inflection point was represented. The structure functions from the original dataset were compared to the perturbed-mean dataset and were found to be identical. Because the current correction techniques seek to modify the measurement mean, the amendments themselves have no impact on system variability. Principal component analysis focuses on quantifying system variability and is indifferent to mean quantities or biases therein. Therefore, the application of humidity bias corrections was unnecessary for this study but can be utilized in*
35   *instances where precise mean values are required.*

Do biases remain constant when the sonde type changes such as occurred at Singapore in late December 2011. Is there any evidence that the characteristics of the RCs changed at Singapore when they switched sonde types?

40   *The Singapore site was also tested to determine if the change in sonde type affected system variability. The Singapore dataset was split in two at the date of sonde replacement (21-Dec-2011) and the principal component analysis was performed three times: once on the full dataset (2008-2016), once on the Graw sonde period (2008-2011), and once on the VRS92 sonde period (2011-2016). The resulting structure functions are almost identical (Supplementary Figure 2), with only a few variations at levels above 200-hPa and near 700-hPa. While slight variations between the components exist, it is not possible*
45   *to attribute this simply to the change in sonde type; it could also be an artifact of natural system variability between the two time periods, rather than instrument discrepancies. To further complicate matters, it could very well be a combination of both the natural climate variability and the change in instrument.*

To test this further, the Koror release site was split in two at the same date (21-Dec-2011) and the analysis was repeated.
50   *Koror was chosen instead of Ranai or Puerto Princesa because it has the most reliable release record before 21-Dec-2011 and after. Both the Ranai and Puerto Princesa site are skewed in that they have more observations after 21-Dec-2011 than before, which could distort the analysis. Looking at the right panel of Supplementary Figure 2, the Koror site also has upper level differences in the RH RCs, it but is missing the mid-level differences at 700-hPa. Because the Koror site did not switch sonde type, this is most likely natural variability. Owing to the fact that the Koror site also has differences when the data set is*
55   *split in two, researchers are hesitant to assign the RC fluctuations only to the change in sonde type at the Singapore release site.*

Page 5, line 14: I'm assuming you mean "temporal range"? Please clarify.

Unfortunately, PCA does not possess the capability of quantifying the time scales associated with variability. Time scales must be attributed by researchers once the physical interpretation is complete. In this study we can only seek to quantify magnitudes of variability, and here the use of the word "range" comes from the range of magnitudes (Figure 2). To clarify, temperature values near the surface deviate from the mean only by only +/- 2°C, whereas at 100-hPa the range of values is much larger at approximately +/- 7°C from the mean. By comparing the different limits, it is obvious that temperatures in the upper troposphere and lower stratosphere are more variable than the well-mixed layers of the troposphere. The wording in the paper has been updated to clarify the use of the word "range" to indicate that it refers to magnitudes and not time.

Page 8, line 6 and page 9, line 20: Past studies (e.g., Houze et al. 1981 MWR, Ciesielski et al. 2006 MWR) would suggest that during certain seasons, the Ranai site is not entirely free of the diurnal cycle or the influence of land convection.

A statement was added to address that the Ranai site can experience a diurnal convective cycle in December due to the northerly monsoon winds and the sea / land breeze circulations off of Borneo (Houze et al., 1981) and that even in non-monsoon months there is some evidence of a diurnal cycle (Ciesielski and Johnson, 2006). Incontestably, no site, maritime, island, or continental, would be completely free of diurnal effects. However, compared to other land masses in the MC, Ranai experiences relatively fewer. For example, in Figure 18 in Ciesielski and Johnson (2006) the summertime months only show a diurnal difference in precipitation on the order of 2 mm/day at Ranai (108°E), compared to Borneo (110°E – 117.5°E) where the difference can be over 18 mm/day. Despite these impacts, compared to the Singapore and Puerto Princesa sites, Ranai will be relatively more maritime.

Page 8, line 24: are the wind reversals noted here a reflection of the QBO?

The variability above 100-hPa in the meridional wind is most likely the QBO (RC5). The pressure level where this variability is detected correlates with the time series in Figure 9 where the QBO wind reversal is evident. Also, see Supplementary Figure 3, where the QBO cycle in the meridional wind is clear in RC5.

Page 9, line 7-8: Please clarify what you mean that the active and suppressed phases are unique to Ranai and not representative of the entire MC.

Because the MJO propagates eastward, its effects have a spatial and temporal nature. The categorization of the MJO into 8 phases stems from this cyclic propagation. Depending on location, the convectively enhanced phases of the MJO, or active phases, and the convectively suppressed phases will be different. For example, in Peatman et al. (2015) Figure 3, Ranai has positive daily mean precipitation anomalies in phase 3 and is convectively enhanced, while in the same phase Paupa New Guinea presents negative anomalies and is convectively suppressed. Essentially, specific MJO disturbances will be observed at different times at different sites and this must be accounted for when assigning phase numbers as either active or suppressed.

Page 9, line 9-10: In addition to presenting the time-height plots, which are informative, showing the temporal behavior of the RCs might help clarify what physical phenomena the different RCs are associated with. For example, does RC1 for temperature vary seasonally (page 8, line 20), is RC5 for zonal wind associated (page 8, line 25) with the QBO time-scale changes, do variations in RC1 and RC2 for meridional wind vary on a monsoonal time scale as suggested, and so on?

Generally, the temporal signature of variability is lost during PCA. However, some structure of this remains in the RC-weight time series (Supplementary Figure 3). Looking at the figure, for almost all of the RCs, the seasonal monsoon is the dominant cyclical signal. These plots can help attribute meaning to some of the RCs, but we cannot overlook the fact that the monsoon cycle will modify the thermodynamic signature.

Page 9, lines 11-12: Are the moistening and the mid-tropospheric temperature changes noted here associated with an MJO?

In Figure 9 it looks like the moistening and temperature changes are mostly associated with the monsoon cycle. However, a MJO event is embedded within the monsoon and would act to further modify the thermodynamics.

Page 9, line 25: How high is the topography on this island? How does this compare with other sites in Indonesia that might be impacted by steep topography? This relates back to my earlier comments on the representativeness of this site.

Ranai island is flat overall and the airport release site sits at 2 meters above mean sea level (AMSL). The Singapore city-state is located in the island lowlands at under 200 meters AMSL with the release site at 16 meters AMSL. Nevertheless, the topography north of Singapore on the Malay Peninsula encompasses the Titiwangsa Mountains, which peak at over 2,000 meters AMSL at Mount Korbu. Like Singapore, the release site at Puerto Princesa is also at 16 meters AMSL and is surrounded by mountainous terrain on the island of Palawan. Here the elevation also surpasses 2,000 meters AMSL at Mount Mantalingajan. While both Singapore and Puerto Princesa reside in the island lowlands, the sites are still influenced by the irregular mountainous and jagged coastal terrain. Ranai will still be influenced by the larger islands in the area, but is further removed from high topography.

Page 9, line 26: What is meant by a "portion region of the southwest monsoon"?

This is a typo. It has been edited to read "a portion of the southwest monsoon."

Fig. 9. Is perturbation temperature being plotted here? If so, please specify in figure caption. What are the differences between the black and white spaces? Would it be possible to linearly interpolate in time the missing data, at least for reasonably short time gaps, shown in the right-hand panels? This would make the structures referred to in the text refers to easier to identify. In the last sentence of the caption are colors red and blue reversed.

Perturbation temperature is being plotted and the caption has been updated, along with the color description for the MJO. The figure has also been updated to remove black spaces and has been interpolated for short time scales (< 5 days of missing data in a row).

Is any ENSO time-scale signal detected in the RCs?

Currently no ENSO signal is apparent in the RCs, although a future bin-analysis may reveal one. The ENSO signal may also lie in a higher order RC that the researchers did not feel comfortable applying physical meaning to. Lastly, the signal may be embedded in variability already captured by the current RCs.

Page 10, lines 9-10: Does RC1 of temperature shown in Fig. 10 reflect differences in mean tropopause height of these sites with Singapore have the highest mean height and Ranai the lowest? If so, would it be possible to list the mean tropopause heights of these three sites?

Both RC1 and RC5 represent tropopause height. Looking at only the most recent full year of data (2016) the mean tropopause height reported for the soundings are Ranai: 91.83 hPa; Puerto Princessa; 91.38 hPa, and Singapore: 90.84 hPa. This pressure level is calculated via the World Meteorological Organization's definition of the tropopause: "The lowest level at which the lapse rate decreases to 2 °C/km or less, provided that the average lapse rate between this level and all higher levels within 2 km does not exceed 2 °C/km." While the results follow the relative order of the RCs, the total range between the sites is only 1°C and may not be significant. Small changes in quality control procedure or number of data points could easily alter the mean values.

Furthermore, the calculation of tropopause height is somewhat arbitrary and difficult to interpret. For the purpose of this study, it's appropriate to envision the tropopause as a transition layer, rather than an individual point. Because the PCA algorithm was applied to the vertical column, it captures the variability due to this transition from the troposphere into the stratosphere.

In Fig. 1 instead of using some random MTSAT image it would be more instructive if the long-term mean rainfall map for this region were shown. For example, this map could be based on MERRA analyses or preferably TRMM or GPCP observations which are easily accessible. In this manner the reader could see how the sites chosen were located with respect to regions of major convection.

A TRMM seasonal climatology plot has replaced the MTSAT image as Figure 1 (Supplementary Figure 4)

Page 3, line 27: Please state when RH is first mention in the paper whether it was computed with respect to ice or water at temperatures less than 0C.

RH was calculated with respect to water for the entire vertical column, including at temperatures less than 0 °C. Switching the calculation from water to ice would produce a discontinuity, which the PCA algorithm does not handle well.

Technical corrections:
Page 1, line 9: continent misspelled.
Page 1, line 13: I suggest changing "sounding release sites" to "upper-air sounding sites" here and other places where this phrase is used.
Page 5, line 29: decomposition misspelled.
Page 4, line 9: Suggest rewording, "... the database was subjectively analyzed in the following fashion."
In Table 1 title and caption change "Control Flags" to "Checks". Also suggest changing "checkpoint in" to "pass at". Finally suggest changing last sentence in caption to: "This is not a comprehensive list as additional subjective QC was applied."
In a few of the figures (Figs. 6, 9 and 10) where lines overlap, it's difficult to distinguish between black and dark blue lines. You might try using cyan instead of dark blue to make the lines more distinguishable.

The above technical suggestions and corrections have all been addressed and updated in the paper.

General comments:
The papers examines temporal and spatial modes of atmospheric variability over the Maritime Continent (MC) by applying principal component analysis to upper-air data from three "representative" sounding sites. With this approach they conclude that the fundamental modes of spatial and temporal variability of this region can be captured with small set of coherent structures. The manuscript is reasonable well written and organized but could you use some additional clarification in several places as noted in comments below.

In addition, I was left wondering if the three sites examined, all of which border or are within the South China Sea, are truly representative of the entire the MC region which spans over 60 degrees of longitude and contains over 17,000 islands. It's unclear to what degree the RC analyses of these three sites capture the major modes of variability in this large MC region? What measures did the authors take to conclusively demonstrate this point? Were other sites examined, say sites bordering the Java Sea or coastal sites in Borneo or Papau New Guinea where topographic effects and the diurnal cycle dominate substantially more than the three sites examined here (see Peatman et al. 2014, QJRMS) or ENSO effects might be more prominent? Short of such efforts, the authors may need to temper their conclusions somewhat to reflect the more limited scope of their analyses.

The limiting factor in expanding this work across other MC sites is the need for a reliable radiosonde release record. While many MC sites have a decent record of radiosonde observations, the availability of moisture observations above 250 hPa is scarce. The transition from the troposphere into the stratosphere is an important source of variability and should be included. Ranai, Singapore, and Puerto Princesa all report humidity beyond 250 hPa for a significant portion of their release history. Only a handful of other MC sites can be included in the analysis with this requirement, and generally the recording history of sites that report above 250 hPa is short, or the location is close to one of the three original sounding sites in the South China Sea. A statement has been added that most of the focus of this study will be the South China Sea region, but that the methodology can be applied elsewhere.

Nevertheless, the authors tested three additional sites at Koror, Palau (2008-2016), Sorong, Indonesia (2014 through 2016), and Cilicap, Indonesia (2014 through 2016). The number of observations for the additional sites can be found in Supplementary Table 1. Many of the structure functions between the sites are identical (Supplementary Figure 5), even at Cilicap and Sorong, which have much shorter release histories. Koror has the most robust release history and is similar to Ranai throughout. It is likely that the major differences between Ranai, Sorong, and Cilicap exist because the full spectrum of of short-term climate variability is not being captured in the sparse record history. Although, the fact that the PCA algorithm can identify signals at Sorong and Cilicap speaks to its robustness as a statistical method. Furthermore, it also demonstrates that a majority of the variability in the thermodynamic system is the result of predominant tropical meteorology and climate, specifically tropical waves and the monsoon.

While it might be beyond the scope of this study, it would be helpful to put the variability of these MC sites in context by showing how they differ from sites in the Indian Ocean and West Pacific (e.g, Gan and Manus) where MJOs often typically

initiate and dissipate and even a midlatitude continental site which should show dramatically different structures. Another natural extension of this work would be a PC analysis of rainfall at each site to better understand the relationship of the RCs presented in this paper to convection.

5 *While outside the scope of this project, the researchers have started looking at sites outside the Maritime Continent, both in the tropics and the mid-latitudes. This will be the subject of future work and will include a bin analysis, where observations will be placed in categories based on the background environment such as MJO phase, monsoon onset, and ENSO index.*

Additional Citations

10

*Ciesielski, P. E., and R. H. Johnson, 2006: Contrasting Characteristics of Convection over the Northern and Southern South China Sea during SCSMEX. Mon. Wea. Rev., **134**, 1041-1062, doi: http://dx.doi.org/10.1175/MWR3113.1.*

Houze, R. A., S. G. Geotis, F. D. Marks, and A. K. West, 1981: Winter Monsoon Convection in the Vicinity of North Borneo.
15 *Part I: Structure and Time Variation of the Clouds and Precipitation. Mon. Wea. Rev., **109**, 1595-1614, doi: http://dx.doi.org/10.1175/1520-0493(1981)109<1595:WMCITV>2.0.CO;2*

*Nuret, M., J.-P. Lafore, F. Guichard, J.-L. Redelsperger, O. Bock, A. Agusti-Panareda, and J.-B. N'Gamini, 2008: Correction of humidity bias for Vaisala RS80-A sondes during the AMMA 2006 observing period. J. Atmos. Oceanic Technol., **25**, 2152–*
20 *2158, doi:10.1175/2008JTECHA1103.1.*

*Wang, J., and L. Zhang, 2008: Systematic errors in global radiosonde precipitable water data from comparisons with ground-based GPS measurements. J. Climate, **21**, 2218–2238, doi:10.1175/2007JCLI1944.1.*

25 *Wang, J., L. Zhang, A. Dai, F. Immler, M. Sommer, and H. Vömel, 2013: Radiation dry bias correction of Vaisala RS92 humidity data and its impact on historical radiosonde data. J. Atmos. Oceanic Technol., **30**, 197–214, doi:10.1175/JTECH-D-12-00113.1.*

30

35

40

Supplementary Figures
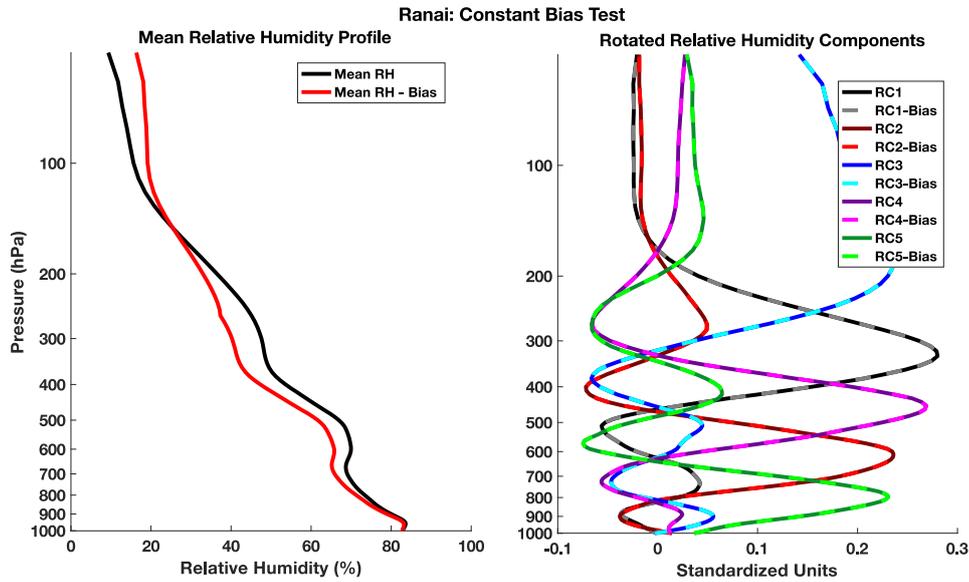
Supplementary Figure 1) Mean bias test performed on the rotated relative humidity components. An arbitrary mean humidity bias was introduced to the entire Ranai data set (left) and run through the PCA algorithm. On the right, the rotated RH components from the original Ranai data set (dark colors) plotted together with the rotated RH components from the biased data set (dashed light colors). There are no distinguishable differences between the original data and the data with an introduced mean bias.

Rotated Relative Humidity Components - Sonde Switch Test
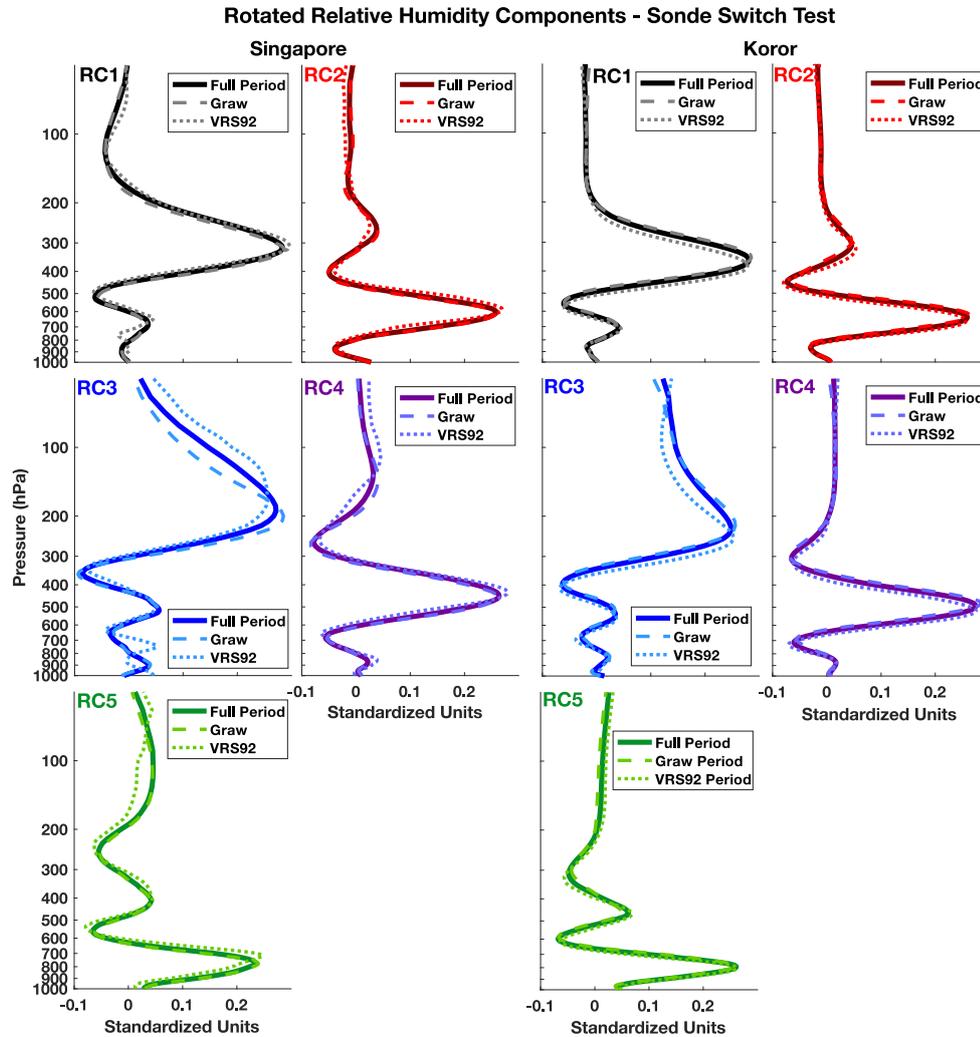
Supplementary Figure 2) Comparison of rotated relative humidity components at the Singapore site (left) before and after the sonde type was switched on 21-Dec-2011. The data set was split in two on the day the sonde type was changed from the Graw sonde (light dashed lines) to the VRS92 sonde (light dotted lines). The PCA algorithm was then run on the two split data sets and the full period (dark continuous lines). There are slight variations at levels above 200-hPa and near the 700-hPa

5

10

Supplementary Figure 4) TRMM 1998-2016 average seasonal rainfall climatology with the release site locations listed for reference.

5

Number of Soundings (2008-2016)				
Variable	**Ranai**	**Koror**	**Sorong**	**Cilicap**
Temperature	3429	5304	624	566
Relative Humidity	1401	5259	623	561
U and V Wind	3380	5237	625	548

Supplementary Table 1) Total number of quality controlled radiosonde observations for each of the additional upper-air sounding sites.

Rotated Structure Functions: Site Comparison
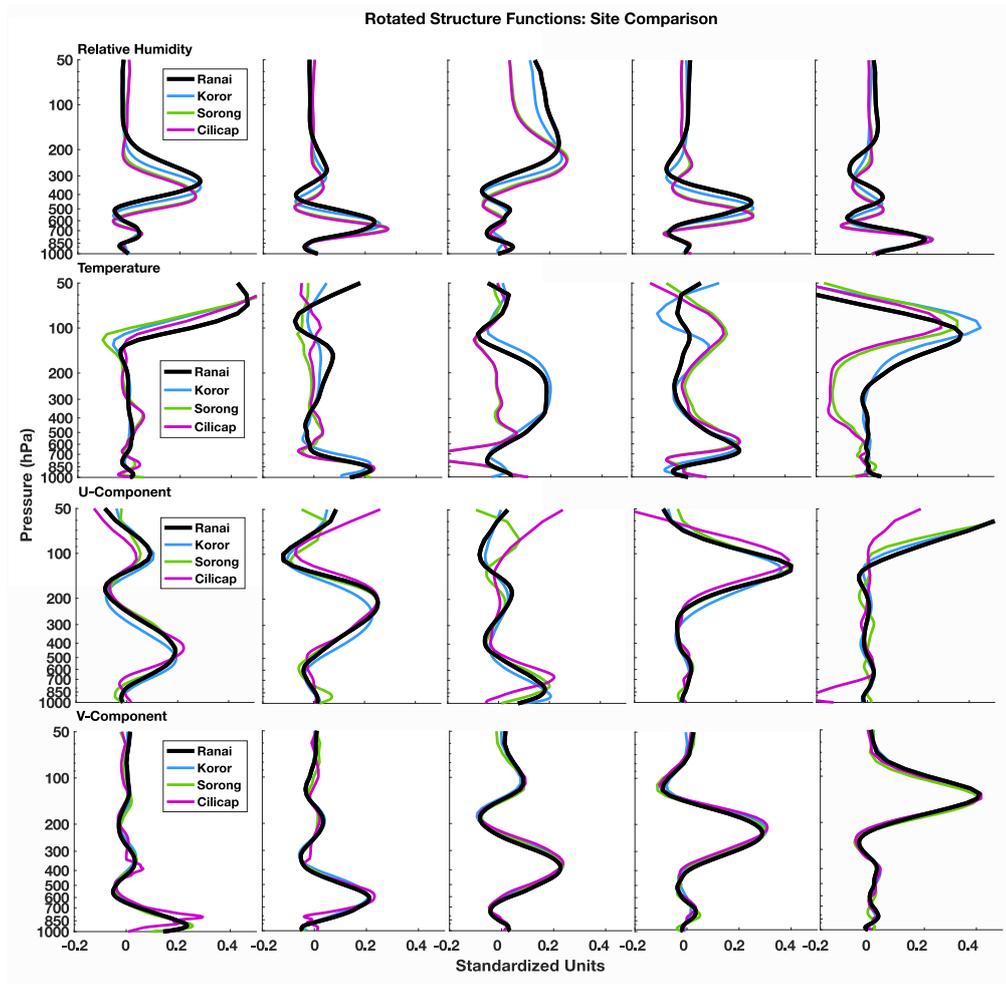
Supplementary Figure 5) Additional release site comparison including Koror, Sorong, and Cilicap. Koror has the most robust release history and the RCs closely match the Ranai RCs. Sorong and Cilicap have much shorter release records and deviate more from the Ranai RCs.

Response to Reviewer 2 Comments

To begin, the authors would like to thank the reviewer for their time, attention to detail, and thoughtful insights on the paper and research. Each comment will be addressed point by point.

5
General Comments

However, I am not really sure what new it brings to the table in addition to the usual procedure of generating T/RH composites about high rain events, and I think the paper oversells how useful this technique is likely to be in the future. For example, from work by Kiladis and others, we already know a lot about the temperature and wind anomalies associated with different types of convectively coupled waves in the tropics. This type of procedure, in which we project anomalies on to the types of physically known propagating

10
3D convectively coupled waves seems more insightful than the 1D PCA done here, in which there is no attempt to physically separate any of the myriad of influences on a particular profile.

Looking at a 1D-representation is useful because interpretation is uncluttered, which can illuminate confounding signals and relationships. Attempting to model 3D meteorological processes is extremely difficult and the results are not always straightforward

15
due to model error and fundamental gaps in our process-level knowledge. This project is an attempt to take a step backward and ask / answer fundamental questions about the variability of thermodynamics in the MC region: Do we know which layers of the atmosphere are the most variable? How can we quantify this variability? How do perturbations to one of these layers affect adjacent and non-adjacent layers? Which thermodynamic states are most often observed? Analysis of the response of various modes of tropical convection to changes in the environment is only possible after we sufficiently understand the answers to these elementary

20
questions (or represent them quantitatively).

This relatively simple approach uses PCA, which is not to be confused with a factor analysis. The two methods are not synonymous techniques and have unique motivations for their respective applications. PCA seeks to identify where variability occurs and remove data redundancies to reduce dimensionality. PCA has no prior assumptions about the underlying causes of the identified variability,

25
which can be both a strength and a weakness. Conversely, factor analysis aims to attribute the known system variability to a set of predefined processes and influences. A factor analysis performed on variable layers identified from the PCA would be an interesting project, but it is outside the scope of this study.

We employed the PCA approach because it is easily implemented, the outcome is straightforward and intuitive, and it is based on

30
observations rather than modeled output. The results can be used in several ways from both a data analysis point of view, such as expanding into a factor analysis, or from a modeling perspective. The observational RCs can be used to test whether models have accurately represented variability in the system or as initialization / boundary conditions in ensemble-based modeling techniques.

Specific Comments

35
(1) Section 4.1. The justification for the physical interpretations to which the various RC's are assigned is often unclear. For example, "shallow convective heating is represented in RC4".

Interpretation of the PCs / RCs is admittedly subjective. When the physical attribution was not obvious, we relied on Figure 9 as a starting point. A motivation behind this project was to rely on a statistical framework to identify variable layers, rather than

40
preconceived ideas. Fortunately, by comparing the RCs to the time series in Figure 9 we see that oftentimes our intuition about which layers might be variable (e.g. that the melting layer, top of boundary layer, and the tropopause transition would all be good candidates for regions with significant thermodynamic variability) are also identified by the PCA. While to first order, our eyes could distinguish these layers from Figure 9, we now have a robust mathematical representation of their location and magnitude and an understanding of how variability in one layer correlates with variability in surrounding levels. The last point is the most problematic

45
when approaching this problem from a non-statistical framework. We could infer the relationships or attempt to model them, but the signal is already present in existing observations and can be captured through the PCA algorithm.

Depending on the application of the PCs / RCs, precise physical interpretation may or may not be necessary. From a modeling perspective, whether the interpretation exists or not the information on variability is retained and can be used to initialize or compare

50
simulations. Conversely, from a data analysis perspective, it is important that the interpretation is correct. Overall, the physical attribution has to answer two questions "where would variability in this layer originate from on a basic level, and on what time scales is it acting or being modified?"

55
But do we really know what kinds of temperature anomalies are likely to be associated with shallow convective heating? For example, if shallow convective clouds occur more frequently (e.g. are triggered) by the moistening and cooling associated with low level upward motion (likely, especially in the vicinity of deep convection), then perhaps shallow clouds are correlated with low level cold anomalies, and any positive correlation between shallow convection with positive RH should not be interpreted as a consequence of detrainment moistening, but some external dynamically imposed influence.

11

The variability associated with what we infer to be the action of shallow clouds (RH5 and T4) is captured exactly as described here. The shallow convective heating peaks at 700 hPa and has an associated cooling below 850 hPa. As stated, we can infer that the shallow clouds are correlated with low-level cold anomalies. This is also apparent in Figure 9, where warm temperature anomalies near 700 hPa sit atop cold anomalies, and vice versa. This could be the result of cooling via ascent, evaporative cooling, and / or radiative effects.

For example, even the net effect of precipitating shallow convection on the RH of a particular level is unclear. It is a residual of the drying associated with induced descent, moistening from detrainment and evaporative moistening, and then a slower dynamical response driven by the geopotential anomalies associated with the convective heating.

In most cases, this method does not allow ascription of the individual forcing or combination of environmental factors that go on to produce the observed variability. This sort of attribution would require an in-depth synoptic and mesoscale analysis for every observation in the dataset. While the source may be dynamic, the result is still going to manifest in the thermodynamic signature in predictable ways. Naturally, many different dynamical and / or microphysical processes could go on to produce the same thermodynamic signature, but this sort of factor analysis will have to be part of a future study.

Furthermore, in this more basic framework we can only focus on correlations between layers for one variable at a time. At this point the method is limited in that we cannot relate the thermodynamic variables to each other. Correlations between the thermodynamic variables will also have to be explored in future work.

More generally, causality between T, RH, u, v anomalies in the background atmosphere and convective clouds always goes both ways. There can't be a simple one to one relationships between certain types of T/RH anomalies and certain cloud types or heating profiles, as implied here. (Otherwise it seems to me that convectively coupled waves in the tropics could not exist.)

The tropical wave cycle is ideal for a PC analysis. The PCA results in both a positive and a negative mode for each PC (Figure 7), which would be required to describe convectively coupled waves. For simplicity, only the positive mode was displayed in Figures 8 and 10, although the opposite mode is also valid. Thus, for a convective disturbance (say a positive RH mode), there would be an associated negative RH mode preceding and following the wave.

(2) Similarly, sometimes the RC's for U and V are assigned physical interpretations and again the justification is unclear. E.g. "The overwhelmingly dominant signal in the V-component of wind is the seasonal monsoon. The MC monsoon is characterized by a complete reversal ...". I guess it is not clear to me here what exactly is meant by "monsoon" in a region of such complicated topography, or why it must have these impacts on U and V. For example, the three radiosonde locations are at quite different locations in the Marine Continent, so the dynamical signature of the monsoon must vary between locations, but the RC's of the three locations are the same almost (except for ordering).

This is noted in point 4 of the conclusion. Should the "monsoon" have the same dynamical signature in all three locations? Perhaps give some explanation of what is really meant by "monsoon". It seems that the authors have simply defined a particular RC as a monsoon signature, and then remarked that this RC is the same at all three locations, and then say the monsoonal signature is the same at all three stations. Everything proceeds from the initial categorization. But is this really more than a semantic game? Do you really know for certain what types of large scale dynamical motions are associated with a particular RC? How would you prove this? I realize there is some discussion of this in lines 13-14 of Section 4.1, but this wasn't fully convincing to me.

The temporal signature of variability is generally lost during PCA. However, some structure of this remains in the RC-weight time series (Supplementary Figure 1). Looking at the figure, for almost all of the RCs, the seasonal monsoon is the dominant cyclical signal. The three sites in the study are all in the South China Sea, so it is possible that at different sites the monsoon signal would be stronger, weaker, or non-existent. The sites may feel the influence of the monsoon at different times, but because the RCs in themselves are non-temporal, the physical wind reversal is all that can be captured.

Because this analysis is on a short-climatology timescale, individual dynamical motions are not the desired result. The much higher resolution time scales would be a perfect project for modelers, or would require a much more in-depth data and meteorological examination than is possible here. This is only a first step in a much larger and more complicated problem. But looking at variability in a lower dimensional subspace allows researchers to ignore system noise (the individual dynamic motions) and categorize specific modes of variability.

(3) Figure 9. I found this hard to interpret. Especially there was so much variability in the top 4 panels, that the features discussed in the text were not clear to me.

Supplementary Figures

5
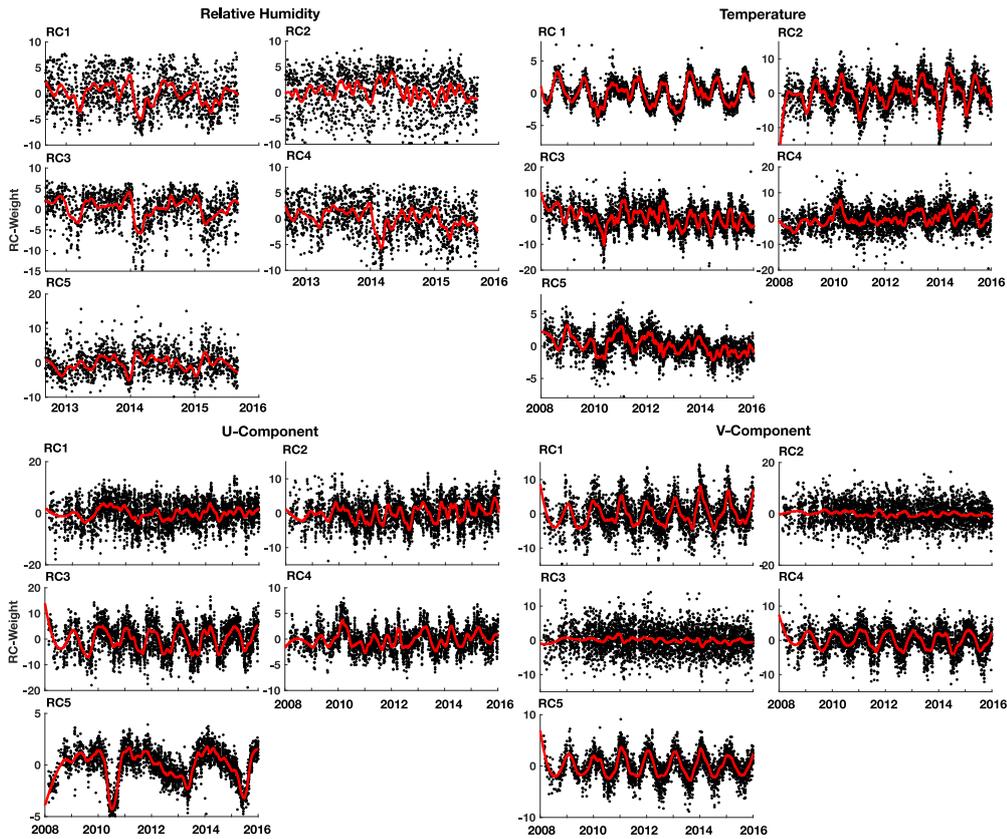
Supplementary Figure 1) Rotated component weight time series. RC-weights (black dots) are accompanied by a best fit line (red) to highlight their cyclic nature. The best fit line was calculated with a weighted linear least squares method combined with a 2nd degree polynomial local regression with a span no larger than 5%. In almost all of the RCs, the dominant signal is the seasonal / monsoon cycle.

10

Revised manuscript with changes tracked follows

Modes of Vertical Thermodynamic and Wind Variability over the Maritime Continent

Jennie Bukowski[1], Derek J. Posselt[1], Jeffrey S. Reid[2], Samuel A. Atwood[3]

[1]University of Michigan, Ann Arbor, MI, USA
5 [2]Marine Meteorology Division, Naval Research Laboratory, Monterey, CA, USA
[3]Department of Atmospheric Science, Colorado State University, Fort Collins, CO, USA

Correspondence to: Derek J. Posselt (Derek.Posselt@jpl.nasa.gov)

Abstract. The Maritime Continent~~Content~~ (MC) is an exceedingly complex region, both from the perspective of its
10 meteorology and its aerosol characteristics. Convection in the MC is ubiquitous, and assumes a wide variety of forms under
the influence of an evolving large- scale dynamic and thermodynamic context. Understanding the interaction between
convective systems and their environment, both individually and in the aggregate, ~~requires~~ entails knowledge of the
dominant patterns of spatial and temporal variability. Ongoing cloud model ensemble studies require realistic perturbations
to the atmospheric thermodynamic state to devise system sensitivities. Apart from modeling studies, evanescent signals in
15 the tropical system are obscured by the underlying broad-scale meteorological variability, which if constrained could
illuminate fine-scale physical processes.

To this end, radiosonde observations from 2008-2016 are examined from three upper-air sounding sites, ~~sounding release
sites~~ within the MC for the purpose of exploring the dominant vertical temperature, humidity, and wind structures in the
region. Principal CComponent AAnalysis is applied to the vertical atmospheric column to transform patterns present in
20 radiosonde data into canonical thermodynamic and wind profiles for the MC. Both rotated and non-rotated principal
components are considered, and the emerging structure functions reflect the fundamental vertical modes of short-term
tropical variability. The results indicate that while there is tremendous spatial and temporal variability across the MC, the
primary modes of vertical thermodynamic and wind variability in the region can be represented in a lower-dimensional
subspace. In addition, the vertical structures are very similar among different sites around the region, though different
25 structures may manifest more strongly at one location than another. The results indicate that, while ~~very~~ different
meteorology may be found in various parts of the MC at any given time, the processes themselves are remarkably consistent.
The ability to represent this variability using a limited number of structure functions facilitates analysis of co-variability
between atmospheric structure and convective systems, and also enables future systematic model-based ensemble analysis of
cloud development, convection, and precipitation over the MC.

1 Introduction

The Maritime Continent (MC) constitutes a region in the Western Pacific critical to the formation and/or evolution of many of the key drivers of tropical atmospheric variability (Neale and Slingo, 2003). It plays host to a broad spectrum of convective systems that interact with the environment and each other in ways that are difficult both to observe and to model. It is also the site of one of the most complex aerosol environments on Earth (Reid et al., 2012; Atwood et al. 2013; Reid et al., 2015; 2016a; 2016b; Atwood et al. 2016). Despite its significance to the global climate system, the MC is an area in which models struggle to reproduce the full range of observed atmospheric fluctuations (Jourdain et al., 2013). At the heart of the challenge is the fact that short- to medium-term variability in the region is affected by processes that span a tremendous range of temporal and spatial scales (Zhang, 2014). Among these are the diurnal cycle of convection, the Madden-Julian Oscillation (MJO; Zhang, 2005), equatorial waves, seasonal monsoons (Misra and Li, 2014) progression of the Intertropical Convergence Zone (ITCZ), the El Niño Southern Oscillation (ENSO), as well as a host of mesoscale and local circulations. Models have difficulty realistically simulating the propagation of tropical waves through the region in part due to the interactions across scales (Peatman et al. 2014; 2015). They have also proven largely unable to reproduce the diurnal cycle of convection, which exhibits a nocturnal maximum over the ocean (Chen and Houze, 1997; Sui et al., 1997; Yang and Slingo, 2001; Dai, 2001). Taken together, the spectrum of convective features observed in the MC are non-independent, their interactions are non-linear, and many are difficult to capture in regional and global models. All of them are sensitive to, and feedback on, the tropical vertical thermodynamic and wind structure, which serve as both a driver and consequence of the state of ~~discernable~~discernible weather ~~and climate~~and downstream applications such as climate and atmospheric composition.

Several previous studies have investigated the dominant modes of tropical, and specifically, Maritime Continent climate variability, in terms of rainfall (Teo et al., 2011; Yanto et al., 2016; Wang et al., 2014) and weather types (Moron, et al., 2015). Thermodynamic variability in the tropics has also been approached from a spatial framework (Serra et al. 2014), and a number of studies have used sounding datasets to examine tropical vertical thermodynamic and wind signatures (Brown and Zhang, 1996; Folkins et al. 2008) and diabatic heating profiles (e.g., Zhang and Hagos, 2009). While these studies have provided insight into the mean thermodynamic structure of the tropics and the range of variability, none have yet focused on the MC specifically, nor have there been attempts to represent tropical troposphere variability in terms of vertical patterns that summarize the dominant modes of variability in the system. These vertical modes contain information on the system itself, and can also be used in numerical model-based analysis of tropical systems.

Specifically, consider a model-based analysis of the response of convective systems to changes in the environment, which would require systematic perturbation of the various factors that control the outcomes of convection (e.g., precipitation, latent heating, radiative fluxes). The input factors should include at minimum the details of the vertical thermodynamic structure (e.g., temperature and water vapor content) that determine buoyancy (and hence convective vertical motion and detrainment) as well as the organizing influence of wind shear (changes in the u and v components of the wind with height).

As a thought experiment, let us assume that a particular environment can be characterized by a representative sounding consisting of temperature, relative humidity, u, and v; and let us further assume that this sounding is resolved in 25 hPa intervals from the surface (1000 hPa) to the lower stratosphere (50 hPa). If one were to examine the sensitivity of convection to changes in this sounding via brute force, one would need to perturb each variable of interest at each level independently. If we consider a simple experiment in which each variable at each level is perturbed just once, independent of the others, this equates to (4 variables x 39 layers)^2 = 24,336 simulations. If we consider a broader range of variability so that we require more than two values of each thermodynamic and wind variable at each level, this becomes (4 variables x 39 layers)^N, where N is the number of discrete values of each variable to be examined at each level. This is clearly computationally unfeasible, and probably non-sensical since it is clear from observations that there is co-variability between different layers, and that there are large-scale patterns of variability that should allow reduction in the number of effective degrees of freedom in the system. The fundamental questions, therefore, are:

1. Can the modes of vertical and temporal variability in MC soundings be reduced to a smaller set of coherent structures?

2. How do these modes of variability change among different sub-regions in the MC, including those that are located in the remote ocean, those that are near significant topography, and those that are proximal to larger land masses?

This study answers these questions by utilizing soundings from three sites in the Maritime Continent region spanning a period of eight years, in concert with an Empirical Orthogonal Function/Principal Component component analysis, to explore the dominant modes of vertical variability in the MC. Each site is representative of a different MC sub-region in the South China Sea, and the results are examined to explore the differences in modes obtained across the region. The results are useful not only for understanding the local meteorological variability, but also for informing future model-based sensitivity experiments. The remainder of this manuscript is organized as follows. Data sources are described in Section 2, while the analysis methodology is presented in Section 3. Results of the EOF/PC analysis are presented in Section 4, while a summary and conclusions are offered in Section 5.

2 Data Sources

Radiosonde observations from 2008-2016 were collected from three upper-air sounding release sites, including Ranai, Indonesia (Riau Island), Puerto Princesa, Philippines, and Singapore (Fig. 1) from 0Z and 12Z soundings. These three sites were selected from only a handful of possible MC locations which have a to represent the spatial extent of the MC and because of their reliableconsistent radiosonde release record and report upper-level humidity measurements. The Ranai, Singapore, and Puerto Princesa sites all have sufficient observations for the analysis, and coincidentally lie along the South China Sea. While the approach presented in this paper can be applied to other MC sites, the focus will be specifically on the South China Sea due to data constraints. Since the Ranai sounding site has a unique location more representative of

16

maritime MC meteorology, it will be featured as a case study in Section 4a and then compared with the other two sites in Section 4b.

The specific radiosonde equipment, accuracy, and observational biases from each MC site were documented in Ciesielski et al. (2014). The Ranai and Puerto Princesa sites used the same equipment throughout the study period, while the Singapore site switched sonde type in 2011. Nevertheless, the equipment substitution has little effect on the results of the analysis. Overall the reporting biases for temperature and wind are small across sonde type and are accepted as is, barring external influences at the site. Humidity measurements are the most uncertain, although several constant bias correction techniques are available for the equipment used at the three MC sites. For this study, the sonde type was either reported ~~as~~ to be precise enough for scientific research (Wang ~~&~~ and Zhang, 2008; Ciesielski et al., 2014), or a form of mean / constant bias correction is available to improve their accuracy (Wang et al., 2013; Nuret et al., 2008). Because the instrument error is consistent among the three sites (Ciesielski 2016, pers. comm.), any biases that exist in the radiosonde data will not propagate into the EOF/PC analysis, which quantifies system variability and is indifferent to mean quantities or biases therein. Since the current correction techniques only seek to modify the measurement mean, the amendments themselves have no impact on system variability. As such, the application of humidity bias corrections was unnecessary for this study and were not applied. ~~Because the instrument error is consistent among the three sites (Ciesielski 2016, pers. comm.), any biases that exist in the radiosonde data will not propagate into the EOF/PC analysis. Constant bias corrections are available for the three sounding sites but are not vital to the study and were not applied.~~

Raw vertical temperature (T), relative humidity (RH), and wind data (U and V) was obtained and decoded from the University of Wyoming atmospheric sounding database. Using linear-in-log-pressure interpolation, each variable was mapped from its original pressure levels to 10 hPa increments from 1000 hPa to 50 hPa for a total of 96 vertical levels. To be considered in the analysis, data points were required to report through the 40 hPa level, as no extrapolation techniques were employed. While the focus of this study is the troposphere, the 50 hPa ~~cutoff~~ cap above, rather than at the tropical tropopause was deliberate: the transition from the troposphere to the stratosphere is an important signal in the system and should be included. Additionally, the PCA algorithm can be skewed if the end points are located at a level where there is a strong signal. As such, levels above the tropopause were incorporated as a buffer so that the entire signal may be effectively captured. To increase the number of data points, each sounding variable was reviewed separately; retention of any one of T, RH, U, and V did not depend on all others being present through 40 hPa. However, because the U and V wind components were derived from the same values, if only one of the components failed the ~~QC~~ quality control tests they were both rejected. As a final step, a ~~five point~~ five-point boxcar moving average filter was applied once to the vertical profiles to smooth the sounding curves. Quality control ~~(QC)~~ was then performed on the set of sounding variables. Because PCA is highly sensitive to outliers and bad data points, the soundings were rigorously inspected for reporting, decoding, and interpolation errors. Various checks were applied to each variable based on the climatology of the region in order to remove overtly erroneous soundings.

In addition to the objective quality control criteria listed in Table 1, ~~the database was subjectively analyzed~~ in the following fashion~~the database was hand analyzed twice during the QC process~~. First, in order to remove poor logarithmic interpolations resulting from extended gaps in reporting levels, the maximum and minimum data points for each 10 hPa level were examined and vertical profiles with extreme values were discarded. Second, every Skew-T / Log-P diagram of the raw data sounding was visually inspected (Ciesielski et al., 2014) prior to dataset formation and questionable profiles not caught by the criteria in Table 1 were excluded. The final number of observations for temperature, relative humidity, and winds for the three upper-air sounding ~~release~~ sites is listed in Table 2. In the analysis that follows, data from each of the three ~~release~~ locations~~sites~~ was treated independently to highlight the structural similarities and differences between sites.

3 Methodology

3.1 Introduction to EOF / PC Analysis~~A~~

Empirical Orthogonal Function / Principal ~~Component~~ Component ~~Analysis~~ Analysis (EOF/PCA) is a statistical method and dimension reduction technique which transforms correlated variables into orthogonal, linearly uncorrelated principal components (PCs). The PCs are eigenvectors resulting from the eigenvalue decomposition of the correlation or covariance matrix of the original variables. Each individual observation in the time series will have a corresponding PC-weight. These weights take on both signs and represent the specific PC-space coordinate of a discrete data point. PCA is a ~~widely used~~widely-used technique in the atmospheric sciences and a rigorous treatment of the underlying mathematics and method utilizations can be found in the textbook *Statistical Methods in the Atmospheric Sciences* (Wilks 2011). While typically applied to horizontally distributed (and gridded) data, this study performs PCA calculations on the vertical atmospheric column from the time series of sounding profiles generated in Section 2. Execution of PCA calculations in this fashion has a precedent in the remote sensing community, whereby vertical PCA is used in infrared satellite retrievals as a noise reduction, data compression, and cloud filtering procedure (Huang and Antonelli, 2002; Smith and Taylor, 2003; Tobin et al., 2007). It has also been used to explore variability in aerosol vertical profiles (Chew et al., 2013; Reid et al., 2016c). Here, column PCA filters noise, but also removes spurious correlations between the artificially imposed 10 hPa atmospheric layers. The remaining underlying structures exhibit only the dominant signals of variability in the column and the relationships between these variable layers.

3.2 PCA Input Data

Temperature, relative humidity, and the U and V wind components were analyzed individually, with each 10 hPa layer treated as an input dimension. In PCA, either the correlation or covariance matrix can be used as input data. The correlation matrix approach assigns a unit variance, and therefore an equal weight to each dimension and is generally used on data with differing units of measure. Conversely, the covariance matrix emphasizes the maximization of variance by each PC. For this

analysis, the covariance matrix was selected over the correlation matrix, because each thermodynamic variable was run separately and contains only one unit of measure. Furthermore, the purpose of this study is to represent tropical variability, and the covariance matrix method accentuates that variability by attaching a larger weight to more variable dimensions.

Input data may also be standardized before computing the correlation or covariance matrix, and again the decision depends on the emphasis of the analysis. Leaving the data unstandardized inherently tags dimensions with a larger magnitude ranges as more variable. While this approach is intuitive, relatively smaller fluctuations and correlations between layers will be overshadowed by those with a larger spectrum of values (Fig. 2). For instance, the overall range in magnitude of tropical temperatures values is small from the surface to the tropopause at 150 hPa, beyond which temperatures fluctuate more. This relatively larger temperature range in temperature near the tropopause will be the first signal captured by the PCA algorithm (Fig. 7).

Conversely, standardizing the input data weights layers with diverse magnitude ranges equally and exposes more fine--scale system relationships and vacillations. Additionally, the non-standardized matrix retains skewness found in the original dataset, which is apparent in the distribution of PC weights. As an example, histograms of the RH PC-weight distributions are shown in Figure 3, along with the associated distribution fit, generated via the non-parametric kernel density estimation method. It can be seen that, as with the variable itself, the descriptive statistics for RH are also non-normal. The choice to standardize or not will depend entirely on the intended use of the structure functions and whether the signal itself or its amplitude is more important. In this studystudy, a non-standardized methodology was applied to underscore the relevance of magnitude in the representation of variability.

In summation, four 96 x 96 covariance matrices of non-standardized values of T, RH, and U and V winds were generated for each sounding site. The eigenvalue decomposition of the non-standardized covariance matrices was performed to transform and reproduce the radiosonde data in principal component space. The resulting orthogonal eigenvectors (PCs) are axes indicating a mean independent signal and represent a frame around which values (PC-weights) may fluctuate. Due to the nature of eigen-decompostiondecomposition, the exact sign of the PCs is irrelevant and can be inverted (Figure 6). Furthermore, because scalar multiples of eigenvectors are also solutions to the eigen-decomposition, the eigenvectors were scaled to have a unit length of one. This ensures that all PCs of the same variable have identical lengths for comparison and differ only in direction. In all, the PCs identify atmospheric layers with strong variability, express the extent of this variability, and describe connections between the 10 hPa slabs throughout the atmospheric column.

3.3 PC Retention Analysis

Although an inherently subjective decision, the number of PCs, or dimensions, to retain was based primarily on the percentage of total variance explained by each PC (Table 3), found by dividing the eigenvalues by the total variance. However, three additional factor tests were applied (Figure 4) based on Kaiser's Rule (Kaiser, 1960), Cattell's visual scree plots (Cattell, 1966), and Horn's Parallel Analysis (Horn, 1965). Kaiser's Rule retains any eigenvector with an eigenvalue greater than 1.0, and tends to overestimate the number of significant PCs (Zwick & and Velicer, 1986), making it useful

only as an upper bound. Horn's Monte Carlo simulation process is often considered the most accurate (Zwick ~~&~~ and Velicer, 1986; Thompson ~~&~~ and Daniel, 1996), whereby PCs are significant only if the eigenvalue is larger than the 95th percentile of the distribution of eigenvalues derived from random normal uncorrelated data (Cota et al., 1993; Glorfield, 1995). The parallel analysis test was carried out with a program similar to the one found in Ledesma and Valero-Mora (2007). After this assessment, it was determined that only five PCs needed to be retained to represent a majority of the variance in the tropical MC system.

Despite the reduction in dimensions, the five retained components are capable of reconstructing the tropical atmosphere with an order of magnitude less input data. By applying the vector direct product between the eigenvectors and the PC weights and adding the layer mean, a specific observation can be reconstructed outside of PC-space and in its natural unit of measure. If all PCs are retained, this transformation will flawlessly reproduce the original data. As the number of PCs is reduced, the conversion will become an increasingly coarse approximation of the original observations. Examples of 5-PC reconstructed soundings are included in Figure 5, along with the original radiosonde data for comparison. The fundamental thermodynamic patterns are easily reproduced using just 5-PCs, including soundings that are close to moist adiabatic (Fig. 5a) and relatively broad dry layers (Fig. 5b). In soundings with a more fine-~~s~~ scale structure, the 5-PC approximation cannot reproduce rapid vertical changes in the original data. Strong inversions (Fig. 5c) and fine-~~-~~scale humidity variations (Fig. 5d) may be missed or will be heavily smoothed, and additional PCs would be required to represent these features. Nevertheless, the retained components do capture the bulk characteristics of the soundings, and clearly represent a majority of the variance explained in the tropical thermodynamic system (Table 3). As a result, even in soundings with complex vertical thermodynamic structures the 5-PC approximation is capable of capturing the overall shape of the sounding and reproducing the salient features of the observed atmospheric environment.

3.4 PC Rotation

Because PCs are necessarily orthogonal, each has a strong modal dependency whereby the number of inherent modes corresponds to the PC number (Fig. 6). While the system information is still retained in PC-space, this enforced modality makes assigning physical interpretations difficult, if not impossible, after the first component. To aid understanding, the axes can be rotated to remove the modality signal and make patterns more pronounced. The original PCs contrast the rotated components (RCs) in their depiction of modality. The PCs have varying modes based on the component number, whereas the RCs have five modes each, but result in one accentuated mode with RC weights which are either large or near zero. Figure 7 has been included as an example to highlight the difference in modality in the PCs and RCs. Figure 7 also illustrates the fact that positive and negative eigenvectors are possibilities for both the PCs and RCs, and that the area between the two represents the full spread of variability. It is clear from this plot that the non-rotated PCs are constrained to have increasing numbers of modes of variability with increasing PC number. In contrast, the RCs are allowed multiple modes, with a primary mode highlighted in each component. The specific modes of variability, and their physical interpretation, will be discussed momentarily in Section 4.1.

The total variance explained by the RCs will be identical to that of the original PCs, but the variance explained by each RC is spread amongst the retained components and no longer weights primarily on the first component. Because of this transformation, each RC has virtually equal importance, which is opposite of the variance explained in Table 3. The Varimax rotation method was selected for this application, since it retains orthogonality and the uncorrelated nature of the components. Moving forward, all analysis in this study will focus on the RCs for more interpretable results.

3.5 Interpretability and Limitations

It is important to note that levels at which the structure functions are non-zero indicate layers of variability in the atmosphere. However, due to the reduced dimensionality of PC-space, certain limitations are imposed on the results. For instance, there is no indication of the physical atmospheric phenomenon producing the observable variability, nor can the period of these oscillations be calculated. PCA, like all unsupervised data reduction methods, requires no prior assumptions about the underlying causes of the identified variability, which can be both a strength and a weakness of this approach. To attach meaning to these signals, the structure functions must be compared to existing climatology (Fig. 9), which will be analyzed in Section 4.1. Moreover, PCA does not seek to attribute the cause of the identified variability but only to quantify it. Attribution of specific environmental, dynamical, and microphysical influences that produce variability would need to be studied via factor analysis rather than through PCA.

Furthermore, T, RH, and the U and V wind components are inspected independently in this study. While in reality these variables are coupled, their interconnectedness is outside the scope of this paper. Because of the independent treatment of thermodynamics and wind, the signal magnitude is unique to its respective variable and cannot be compared across variables. This is a consequence of the structure function axes being in a standardized unit. The magnitude on the x-axis reflects the individual variable's descriptive statistics (Fig. 2). As such, while the signal magnitude may be identical in normalized PC-space, the mapping of that signal onto the variable's correct units may vary widely. This is intuitive, because the unit of measure for T, RH, and wind are dissimilar, as are the range of associated values in the tropics.

Nevertheless, looking at results in a lower-dimensional subspace has its advantages. Firstly, the structure functions are orthogonal and independent signals, suggesting unique phenomena are captured by each component. Additionally, within a component the relationships between layers are apparent. This information is invaluable in that it provides a framework for how perturbations at a specific level affect the rest of the atmospheric column, along with the relative sign and magnitude of this effect. Moreover, due to the unit scaling of the eigenvectors, the signal magnitude is comparable across the five RCs for a single variable. While each RC is of equal significance to the total variability of the system, the magnitude is an indication of relative importance.

21

4 Modes of Variability and Analysis

4.1 Case Study: Ranai

Although the MC is comprised of many islands, Ranai's location is unique: the minor island is part of the Indonesian Riau Archipelago in the South China Sea and is distant enough from the major MC islands to be largely free of the influence of
5 land and the diurnal cycle of convection forced by day-time heating of the surface. ~~is distant enough from the major MC islands to be largely free of the influence of land and the diurnal cycle of convection forced by day-time heating of the surface.~~ As an exception, the Ranai site can experience a diurnal convective cycle due to the northerly monsoon winds and the sea / land breeze circulations off of Borneo (Houze et al., 1981). There is some evidence of a diurnal cycle at Ranai in non-monsoon months (Ciesielski and Johnson, 2006), although the signal is not as strong as the surrounding larger islands.
10 Incontestably, no site, maritime, island, or continental, ~~would~~ can be completely free of diurnal effects in the mostly enclosed ~~South China Sea~~SCS region~~.~~ Nonetheless, compared to other land masses in the MC, Ranai ~~experiences relatively fewer~~is more representative of the southern ~~South China Sea~~SCS basin.~~As such, the Ranai site~~ is essentially maritime, and can therefore provide a basic framework for examining tropical over-ocean atmospheric variability. We will use the structure functions obtained from Ranai as a baseline for comparison with Puerto Princesa and Singapore, which are increasingly
15 land-influenced, respectively (Section 4.2, Fig. 10).

Structure functions (RCs) for each sounding variable at Ranai are plotted in Figure 8. The first RC in relative humidity appears to reflect deep convective moistening/drying of the upper troposphere. This strong signal represents the oscillation between a saturated upper atmosphere during convection and a dry region of subsidence preceding and following the wave disturbance. The structure of RC1 also indicates that upper tropospheric humidity covaries with vapor content at lower
20 levels. Increases (decreases) in RH between 300 – 400 hPa are positively correlated with increases (decreases) in RH in a shallow layer between 600 – 850 hPa, and anti-correlated with decreases (increases) in RH around 500 hPa. The two lower tropospheric modes are proportionally smaller than the dominant deep convective signal. RC2 in RH reflects the moistening and drying of the lower free troposphere, RC3 represents deep convective detrainment in the upper troposphere/lower stratosphere, RC4 captures melting level water vapor detrainment, and RC5 reflects variability associated with shallow
25 convection atop the marine boundary layer. It is important to note that more variability may exist within humidity in the upper troposphere and lower stratosphere, but its detection is outside the current limitations of radiosonde sensors (Soden and Lanzante, 1995). RC1 for temperature appears to represent the seasonal shift in the ITCZ, which causes a cyclic change in temperature above 150 hPa. RC2 reflects warming in both the boundary layer and the upper troposphere, perhaps a reflection of the connection between lower tropospheric warming (and moistening) and the incidence of deep convection.
30 RC3 appears to be a deep convective heating / radiative cooling mode, while shallow convective heating is represented in RC4. RC5 appears to reflect the signal of upper-tropospheric heating. Examination of the winds reveals the quasi-biennial oscillation, or cyclic changes in the zonal westerlies aloft (RC5; note the lack of a corresponding structure in the RCs for the meridional wind). Fluctuations possibly associated with deep convective detrainment may be seen in RCs 2 and 4, as well as

lower tropospheric shifts associated with fluctuations in the monsoon trough (RC3). RC1 is intriguing, and suggests covariability between upper tropospheric (~100 hPa) and mid-tropospheric (~400 – 500 hPa) zonal winds. A similar structure may be seen in RC3 for the meridional winds, and it is possible that this reflects simultaneous detrainment from convection in the upper troposphere and around the melting level. The overwhelmingly dominant signal in the V-component of wind is the seasonal monsoon. The MC monsoon is characterized by a complete reversal in the north-south wind component at the surface (RC1, RC2) and a corresponding reversal in the upper troposphere of the opposite sign (RC4, RC5).

To aid in attributing meaning to the RCs, pressure versus time contour plots of the radiosonde observations are included (Fig. 9) for the entire time period (2008-2016) and for a subset of that range (2013-2015). To smooth the time series, if radiosonde data was missing for less than 5 consecutive days it was interpolated for use only in Figure 9. Along with the time series, MERRA2 outgoing longwave radiation (OLR) and daily precipitation are included to illustrate the general convective environment. Lastly, the active (Phases 3-5) and suppressed (Phases 6-8) MJO amplitudes for Ranai are plotted. The distinction between active and suppressed phases was determined based on recent studies by Peatman et al. (2015) and Birch et al. (2016). Because the MJO propagates eastward, its effects have a spatial and temporal nature. Depending on location, the convectively enhanced phases of the MJO, or active phases, and the convectively suppressed phases will be different. MJO Phases 1-2 are transition phases for Ranai, do not significantly impact convection, and were therefore not incorporated. Note that the active and suppressed phases are unique to the Ranai release upper-air sounding site and are not representative of the entire MC.

Examination of contour time-height plots of RH, T, U, and V help to illustrate the variability seen in the vertical structure functions. Periodic oscillations in RH in the boundary layer, lower free troposphere, mid-troposphere, and upper troposphere are clearly visible in the top row in Figure 9. Close inspection of the right column of Fig. 9 reveals significant moistening around January 2014, with mid-tropospheric warming prior and cooling following. While this moistening is most likely dominated by the monsoon cycle, it is also enhanced by an active MJO phase. Cooling in the upper troposphere (anvil-level) with warming below is evident in late fall 2013 as well. The monsoon reversal in the upper-tropospheric meridional wind can clearly be seen in the 2013 – 2015 v-direction winds. Examination of the precipitation, OLR, and MJO index plots for 2013-2015 indicates low precipitation and high OLR associated with periods during which the MJO was in a suppressed phase and vice versa for the active phases. Finally, an examination of the u-component winds from 2008 – 2016 reflects activity associated with the QBO; downward propagating easterly wind anomalies can be seen starting in mid-2010, early 2013, and early 2015.

4.2 Release Upper-Air Sounding Site Comparison

As mentioned above, the sounding release upper-air sounding site at Ranai is located on a small island generally representative of over-ocean conditions and broadly , free of irregular mountainous and jagged coastal terrain. Overall, Ranai island is flat and the upper-air sounding site sits at 2 meters above mean sea level (AMSL). Conversely, the sites at

23

Singapore and Puerto Princesa are surrounded by, or located in close proximity to, larger areas of land (Fig. 1). The Singapore city-state resides on the Malay Peninsula and is located in the island lowlands with the release site at 16 meters AMSL. Nevertheless, the topography north of Singapore on the Malay Peninsula encompasses the Titiwangsa Mountains, which peak at over 2,000 meters AMSL at Mount Korbu. Singapore is also adjacent to the island of Sumatra, whose mountainous topography and biomass burning events exert an influence on the regional meteorology. Like Singapore, the release site at Puerto Princesa is also at 16 meters AMSL and is surrounded by mountainous terrain on the island of Palawan, which also surpasses 2,000 meters AMSL at Mount Mantalingajan. While Palawan island is long and narrow, it has enough topography that it may reasonably be expected to exhibit orographic precipitation enhancement. In addition, Palawan's location in the northeastern portion of the MC makes it subject to periodic tropical cyclones as well as a portion of the southwest monsoon.

While both Singapore and Puerto Princesa reside in the island lowlands, the sites are still influenced by the surrounding terrain. Ranai will still be influenced by the larger islands in the area, but is further removed from high topography. while the sites at Singapore and Puerto Princesa are surrounded by, or located in close proximity to, larger areas of land (Fig. 1). Singapore, at the Malay Penninsula is adjacent to the island of Sumatra, whose topography and biomass burning events exert an influence on the regional meteorology. Puerto Princesa is located on Palawan Island, which is bordered on the west by the South China Sea and on the east by the Sulu Sea. While Palawan island is long and narrow, it has enough topography that it may reasonably be expected to exhibit orographic precipitation enhancement. In addition, its location in the northeastern portion of the MC makes it subject to periodic tropical cyclones as well as a portion region of the southwest monsoon. Comparison among the thermodynamic and wind structures obtained from the three different sounding sites allows an assessment of the possible influence of land as well as TCs and different portions of the monsoon trough. A cursory examination of the RCs for each site (Fig. 10) indicates similar patterns of structural variability across the MC, a reflection of the similarity in the meteorological drivers in the region (primarily convection, convectively coupled waves, and the seasonal monsoon).

Closer inspection reveals important differences among the three sites. First, note that the ordering of RCs is rarely the same among the soundings. Recall that the PCs are arranged in decreasing order according to the fraction of variance explained, while the RCs have weighting that is approximately similar to one another. Even so, the order of the RCs does reflect decreasing variability, and as such it is useful to compare RC ordering among the sounding sites. To facilitate the comparison, we have chosen to plot similar structures together, regardless of the order of the RC (with Ranai the default). The dominant variability in RH is contained in the mid-tropospheric moist mode, indicative of detrainment from congestus and deep convection, while the main mode at Ranai is the deep convective mode coupled with shallow moistening from trade cu. Ranai also exhibits more variability in the deep upper tropospheric moistening/drying (RC3), while the shallow moistening is more prominent at Puerto Princesa and Singapore. The rank ordering of RCs of temperature and RH is identical between Singapore and Puerto Princesa with different ordering at Ranai. All three sites exhibit maximum variability in temperature in the upper troposphere, likely reflecting the seasonal oscillation in the height of the tropopause

with the meridional shift in the ITCZ. The secondary mode of variability is also similar among the different sites, with lower tropospheric cooling/warming that positively co-varies with smaller magnitude cooling/warming in the upper troposphere. Overall, Ranai and Puerto Princesa appear to be more similar in structure to each other than to Singapore. The largest differences appear in RCs 3-5; Singapore has larger amplitude variability in the upper troposphere relative to Ranai and Puerto Princesa, and the upper-tropospheric mode (far right panel, Fig. 10) is located higher than at the other two sites. Interestingly, the u-direction wind RCs (Fig. 10; second row from bottom), while exhibiting similar structures among all three sites, have different rank order among the sites. Finally, the ordering is identical between Ranai and Singapore for the meridional winds (Fig. 10, bottom row), while RCs 1 and 5 are reversed for Puerto Princesa. Recall that Ranai RC1 reflects the low-level oscillation in the meridional wind, while RC5 reflects the oscillation in meridional winds in the upper troposphere. It appears that, at Puerto Princesa, there is more variability in the meridional upper tropospheric flow than there is at low levels.

While the rank ordering of the RCs varies among the three sites, we wish to point out that the form of the vertical structures obtained from the rotated PCA at three rather widely dispersed sounding sites are very similar. The implication is that (1) similar patterns of variability in both thermodynamic and dynamic vertical structure occur across the MC, and (2) the drivers of this variability (e.g., the spectrum of convective waves and seasonal, intraseasonal, annual, and inter-annual modes) are expressed similarly across the MC.

5 Conclusions and Applications

This study applied Principal Component Analysis to 2008-2016 radiosonde time series data to transform vertical patterns in temperature, relative humidity, and winds into structure function profiles for the Maritime Continent. The main benefits and conclusions from this decomposition of observations into a lower dimensional subspace can be summarized as follows:

1. PCA reduces system noise and condenses a large spectrum of meteorological fluctuations to a few coherent signals. The resulting principal components and rotated components are orthogonal, and represent independent signals of variability in the thermodynamic variables and winds.

2. Structure functions obtained from PCA approximate a majority of tropical thermodynamic variability, and the vertical location of this variability, in a much lower-dimensional sub-space. Specifically, layers with large temporal variability are highlighted, while redundant correlations between levels are ignored, leaving only the fundamental relationships between layers.

3. In most cases, the rotated structure function signals can be attributed to physical atmospheric oscillations and phenomena. Specific signals that are of particular interest include:

   - Rotated components in relative humidity show coherent structures indicative of: tropopause level detrainment from deep convection, moistening/drying of the upper free troposphere, melting-level detrainment, and shallow convective moistening/drying.

- Upper tropospheric humidity co-varies with vapor content at lower levels: increases (decreases) in RH between 300 – 400 hPa are positively correlated with increases (decreases) in RH between 600 – 850 hPa, and anti-correlated with decreases (increases) in RH around 500 hPa.

- Temperature structure functions reflect the influence of deep convection and radiative cooling in the upper troposphere, as well as convective heating in the lower free troposphere.

- Vertical structure functions in the u and v components of the wind reflect the seasonal variability in the monsoon, as well as more local effects of convective detrainment at various levels.

4. Structure functions generated from MC sites in very different contexts (remote island, larger mountainous island, and proximal to a larger land mass) are remarkably similar. This indicates that the dominant modes of variability are the result of the propagation of tropical waves and the monsoon, rather than topographic or latitudinal influences, which may exist in higher order components. The most ~~discernable~~discernible difference among ~~release~~ upper-air sounding sites is the position of the tropopause, and the fact that different structures explain a different fraction of the overall variance among the three different sites.

Overall, the application of PCA to the vertical atmospheric column and subsequent transformation of vertical thermodynamic and wind information into principal components expresses the important variability and relationships among layers. This results in a robust statistical, and therefore mathematical identification and quantification of the vertical height and magnitude of variability. While this study focused on the Maritime Continent, the same methodology can be applied to different variables, time durations, and ~~radiosonde release~~ upper-air sounding sites. In addition, these structures can be used to examine co-variability between thermodynamic and dynamic structures and measures of convective activity (e.g., precipitation rate and OLR). They can also be used to control for variations in meteorology when examining temporal and spatial variability in aerosol content and composition.

Undoubtedly, the future applications of this method are far reaching. With short-term tropical meteorological variability constrained, it is possible to identify relatively elusive system signals in this region. Furthermore, because of their unique properties, the structure functions can be used as boundary conditions in numerical models. Especially important for modeling is the representation of the interconnectedness between levels and the scaling of perturbations throughout the atmospheric column. In all, decomposing vertical patterns into modes of variability will allow for more realistic future model simulations and analysis of atmospheric processes over the MC.

Acknowledgements

References

5   Atwood, S. A., J. S. Reid, S. M. Kreidenweis, L. E. Yu, S. V. Salinas, B. N. Chew, and R. Balasubramanian, 2013: Analysis of source regions for smoke events in Singapore for the 2009 El Nino burning season, Atmos. Environ.,78, 219-230, doi:10.1016/j.atmosenv.2013.04.047.

Atwood, S. A., et. al. 2016: Size resolved aerosol and cloud condensation nuclei (CCN) properties in the remote South China Sea: Measurement and sources. Atmos. Chem. Phys., Submitted.

10   Birch, C. E., S. Webster, S. C. Peatman, D. J. Parker, A. J. Matthews, Y. Li and M. E. E. Hassim, 2016: Scale Interactions between the MJO and the Western Maritime Continent. Journal of Climate, 29, 2471-2492.

Brown, R. G., and C. Zhang, 1997: Variability of Midtropospheric Moisture and Its Effect on Cloud-Top Height Distribution during TOGA COARE. J. Atmos. Sci., 54, 2760–2774.

Cattell, R. B., 1966: The Scree test for the number of factors. Multivariate Behavioral Research, 1, 245-276.

15   Chen, S. S. and R. A. Houze, 1997: Diurnal variation and life-cycle of deep convective systems over the tropical pacific warm pool. Q. J. Roy. Meteor. Soc., 123, 357–388. doi: 10.1002/qj.49712353806

Chew, B. N., J. R. Campbell, S. V. Salinas, C. W. Chang, J. S. Reid, E. J. Welton, B. N. Holben, S. C. Liew (2013), Aerosol particle vertical distributions and optical properties over Singapore, Atmos. Environ., 79, 599-613.

Ciesielski, P. E., and R. H. Johnson, 2006: Contrasting Characteristics of Convection over the Northern and Southern South China Sea during SCSMEX. Mon. Wea. Rev., 134, 1041-1062.

20   Ciesielski, Paul E., H. Yu, R. H. Johnson, K. Yoneyama, M. katsumata, C. N. Long, J. Wang, S, M. Loehrer, K. Young, S. F. Williams, W. Brown, J. Braun, and T. van Hove, 2014: Quality-Controlled Upper-Air Sounding Dataset for DYNAMO/CINDY/AMIE: Development and Corrections. Journal of Atmospheric and Oceanic Technology, 31, 741-764.

25   Cota, A. A., R. S. Longman, R. R. Holden, G. C. Fekken, and S. Xinaris, 1993: Interpolating 95th percentile eigenvalues from random data: An empirical example. Educational and Psychological Measurement, 53, 585-596.

Dai, A., 2001: Global Precipitation and Thunderstorm Frequencies. Part II: Diurnal Variations. J. Climate, 14, 1112–1128.

Folkins, I., S. Fueglistaler, G. Lesins, and T. Mitovski, 2008: A Low-Level Circulation in the Tropics. J. Atmos. Sci., 65, 1019–1034.

30   Glorfeld, L. W., 1995: An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. Educational and Psychological Measurement, 55, 377-393.

Horn, J. L., 1965: A rationale and test for the number of factors in factor analysis, Psychometrika, 30, 179–185.

Houze, R. A., S. G. Geotis, F. D. Marks, and A. K. West, 1981: Winter Monsoon Convection in the Vicinity of North Borneo. Part I: Structure and Time Variation of the Clouds and Precipitation. Mon. Wea. Rev., 109, 1595-1614.

Huang, H. L. and P. Antonelli, 2000: Application of Principal Component Analysis to High-Resolution Infrared Measurement Compression and Retrieval. Journal of Applied Meteorology, 40, 365-388.

Jourdain, N. C., A. S. Gupta, A. S. Taschetto, C. C. Ummenhofer, A. F. Moise and K. Ashok, 2013: The Indo-Australian monsoon and its relationship to ENSO and IOD in reanalysis data and the CMIP3/CMIP5 simulations. Climate Dynamics, 41(11), 3073-3102.

Kaiser, H. F., 1960: The application of electronic computers to factor analysis. Educational and Psychological Measurement, 20, 141-151.

Ledesma, R. D. and P. Valero-Mora, 2007: Determining the Number of Factors to Retain in EFA: An easy-to-use computer program for carrying out Parallel Analysis. Practical Assessment Research and Evaluation, 12, 1–11.

Misra, V. and H. Li, 2014: The seasonal predictability of the Asian summer monsoon in a two tiered forecast system, Cli. Dyn., 42, 2491-2507, doi:10.1007/s00382-013-1838-1.

Moron, V., A. W. Robertson, J-H. Qian and M. Ghil, 2015: Weather types across the Maritime Continent: from the diurnal cycle to interannual variations. Front. Environ. Sci., 2, 1-19.

Neale, R., and J. Slingo, 2003: The Maritime Continent and Its Role in the Global Climate: A GCM Study. Journal of Climate, 16, 834-848.

Nuret, M., J.-P. Lafore, F. Guichard, J.-L. Redelsperger, O. Bock, A. Agusti-Panareda and J.-B. N'Gamini, 2008: Correction of humidity bias for Vaisala RS80-A sondes during the AMMA 2006 observing period. J. Atmos. Oceanic Technol., 25, 2152–2158, doi:10.1175/2008JTECHA1103.1.

Peatman, S. C., A. J. Mathews, and D. P. Stevens, 2014: Propagation of the Madden-Julian Oscillation through the Maritime Continent and scale interaction with the diurnal cycle of precipitation, Q. J. Roy. Meteor. Soc., 140, 814-825, doi:10.1002/qj.2161.

Peatman, S. C., A. J. Matthews and D. P. Stevens, 2015: Propagation of the Madden–Julian Oscillation and scale interaction with the diurnal cycle in a high-resolution GCM. Climate Dynamics, 45, 2901-2918.

Reid., J. S., P. Xian, E. J. Hyer, M. K. Flatau, E. M. Ramirez, F. J. Turk, C. R. Sampson, C. Zhang, E. M. Fukada, and E. D. Maloney, 2012: Multi-scale meteorological conceptual analysis of observed active fire hotspot activity and smoke optical depth in the Maritime Continent, Atmos. Chem. Phys., 12, 1–31, doi:10.5194/acp-12-1-2012.

Reid, J. S., N. D. Lagrosas, H. H. Jonsson, E. A. Reid, W. R. Sessions, J. B. Simpas, S. N. Uy, T. J. Boyd, S. A. Atwood, D. R. Blake, J. R. Campbell, S. S. Cliff, B. N. Holben, R. E. Holz, E. J. Hyer, P. Lynch, S. Meinardi, D. J. Posselt, K. A. Richardson, S. V. Salinas, A. Smirnov, Q. Wang, L. Yu, and J. Zhang, 2015: Observations of the temporal variability in aerosol properties and their relationships to meteorology in the summer monsoonal South China Sea/East Sea: the scale-dependent role of monsoonal flows, the Madden–Julian Oscillation, tropical cyclones, squall lines and cold pools, Atmos. Chem. Phys., 15, 1745-1768, doi:10.5194/acp-15-1745-2015.

28

Reid, J. S., P. Lynch, B. N. Holben, E. J. Hyer, E. A. Reid, S. V. Salinas, J. Zhang, J. R. Campbell, B. N. Chew, R. E. Holz, A. P. Kuciauskas, N. Lagrosas, D. J. Posselt, A. L. Walker, and E. J. Welton, 2016a: Aerosol meteorology of the Maritime Continent for the 2012 7SEAS southwest monsoon intensive study: Part I regional scale phenomena. Atmos. Chem. Phys., Submitted.

Reid, J. S., N. Lagrosas, H. H. Jonsson, E. A. Reid, S. A. Atwood, T. J. Boyd, V. P. Ghate, P. Lynch, D. J. Posselt, J. B. Simpas, S. N. Uy, K. Zaiger, D. R. Blake, A. Bucholz, J. R. Campbell, B. N. Chew, S. S. Cliff, B. N. Holben, R. E. Holz, E. J. Hyer, S. M. Kreidenweis, A. P. Kuciauskas, S. Lolli, M. Oo, K. D. Perry, S. V. Salinas, W. R. Sessions, A. Smirnov, A. L. Walker, Q. Wang, L. Yu, J. Zhang, and Y. Zhao, 2016b: Aerosol meteorology of Maritime Continent for the 2012 7SEAS southwest monsoon intensive study: Part II Philippine receptor observations of fine scale aerosol behavior. Atmos. Chem. Phys., Submitted.

Reid, J. S., R. E. Kuehen, R. E. Holz, E. W. Eloranta, K. C. Kaku, S. Kuang, M. J. Newchurch, A. M. Thompson, C. R. Trepte, J. Zhang, S. A. Atwood, J. L. Hand, B. N. Holben, P. Minnis, and D. J. Posselt, 2016c: Ground based high spectral resolution lidar observation of aerosol vertical distribution in the summertime Southeast United States. Atmos. Chem. Phys., Submitted.

Serra, Y. L., X. Jiang, B. Tian, J. Amador-Astua, E. D. Maloney and G. N. Kiladis, 2014: Tropical Intraseasonal Modes of the Atmosphere. Annual Review of Environment and Resources, 39, 189-215.

Soden, B. J. and J. R. Lanzante, 1996: An Assessment of Satellite and Radiosonde Climatologies of Upper-Tropospheric Water Vapor. Journal of Climate, 9, 1235-1250.

Smith, J. A. and J. P. Taylor, 2003: Initial Cloud Detection Using the EOF Components of High-Spectral-Resolution Infrared Sounder Data. Journal of Applied Meteorology, 43, 196-210.

Sui, C.-H., K-M. Lau, Y. N. Takayabu, and D. A. Short, 1997: Diurnal Variations in Tropical Oceanic Cumulus Convection during TOGA COARE. J. Atmos. Sci., 54, 639–655

Teo, C-K., T-Y. Koh, J. C-F. Lo and B. C. Bhatt, 2011: Principal Component Analysis of Observed and Modeled Diurnal Rainfall in the Maritime Continent. Journal of Climate, 24, 4662-4675.

Tobin, D. C. and coauthors, 2007: Hyperspectral data noise characterization using principle component analysis: application to the atmospheric infrared sounder, Journal of Applied Remote Sensing, 1, 1-11.

Thompson, B. and L. G. Daniel, 1996: Factor Analytic Evidence for the Construct Validity of Scores: A Historical Overview and Some Guidelines. Educational and Psychological Measurement, 56, 197 – 208.

Wang, B., J-Y. Lee and B. Xiang, 2014: Asian summer monsoon rainfall predictability: a predictable mode analysis. Climate Dynamics, 44(1), 61-74.

Wang, J., and L. Zhang, 2008: Systematic errors in global radiosonde precipitable water data from comparisons with ground-based GPS measurements. J. Climate, 21, 2218–2238, doi:10.1175/2007JCLI1944.1.

Wang, J., L. Zhang, A. Dai, F. Immler, M. Sommer & and H. Vömel, 2013: Radiation dry bias correction of Vaisala RS92 humidity data and its impact on historical radiosonde data. J. Atmos. Oceanic Technol., 30, 197–214, doi:10.1175/JTECH-D-12-00113.1.

Wilks, Daniel S., 2011: Statistical Methods in the Atmospheric Sciences. Academic Press, 704 pp.

Yang, G.-Y., and J. Slingo, 2001: The Diurnal Cycle in the Tropics. Mon. Wea. Rev., 129, 784–801.

Yanto, B. Rajagopalan & and E. Zagona, 2016: Space–time variability of Indonesian rainfall at inter‑annual and multi‑decadal time scales. Climate Dynamics, 1-15.

Zhang, C. 2005: Madden-Julian Oscillation, Rev. Geophys., 43, RG2003, doi: 0.1029/2004RG000158.

Zhang, C., and S. M. Hagos, 2009: Bi-modal Structure and Variability of Large-Scale Diabatic Heating in the Tropics. J. Atmos. Sci., 66, 3621–3640.

Zhang, C., 2014: Madden-Julian Oscillation: Bridging Weather and Climate, Bull. Amer. Meteor. Soc., 94, 1849-1870, doi: 10.1175/BAMS-D-12-00026.1.

Zwick, W. R. & and W. F. Velicer, 1986: Comparison of five rules for determining the number of components to retain. Psychological Bulletin, 99(3), 432-442.

Sounding Quality ~~Control Flags~~Checks		
Variables	**Level**	**Threshold**
Temperature	1000-hPa	$15° C < T < 40° C$
Relative Humidity	1000-hPa	$30 \% < RH < 100 \%$
Temperature	480-hPa	$T < 0° C$
Temperature	300-hPa	$T < -20° C$
Temperature	200-hPa	$T < -40° C$
Temperature ~~&~~ and Dew Point	All	$T > Td$

Table 1: Quality control flags enacted during the sounding dataset creation as a first ~~checkpoint in~~pass at removing erroneous soundings or raw data decoding. This is not a comprehensive list, as additional subjective quality control was applied.~~This list is not comprehensive, as a hand analysis was performed after these flags were applied.~~

Number of Soundings (2008-2016)			
Variable	**Ranai**	**Puerto Princesa**	**Singapore**
Temperature	3429	2326	4143
Relative Humidity	1401	2235	2878
U and V Wind	3380	2257	4096

Table 2 Total number of quality controlled radiosonde observations for each of the ~~release~~ upper air sounding sites. To increase the number of data points, each variable was given independent consideration during the quality control process.

Percent Variance Explained by Principal Components											
	Ranai				Singapore				Puerto Princesa			
PC #	T	RH	U	V	T	RH	U	V	T	RH	U	V
1	35.44	53.86	42.58	43.80	36.02	38.14	36.53	47.81	33.60	64.38	51.17	45.15
2	21.16	13.29	19.74	14.52	22.23	17.35	19.59	12.03	20.50	10.27	25.68	21.43
3	10.00	10.34	9.58	10.29	10.49	10.16	10.38	10.32	8.51	5.82	5.41	8.50
4	6.73	4.83	7.72	8.61	6.20	6.67	9.28	7.62	8.16	3.68	4.98	6.79
5	5.52	3.96	6.41	5.43	5.19	5.40	7.77	5.61	6.29	3.06	3.49	4.38
Sum	78.85	86.28	86.02	82.65	80.13	77.72	83.55	83.39	77.06	87.20	90.73	86.25

Table 3 Percent of total variance explained for each variable's PCs from the three ~~release~~ upper-air sounding sites. The first PC represents the highest variance and each subsequent component explains a decreasing amount of the residual variance.

TRMM: 1998-2016 Seasonal Average Precipitation Rate
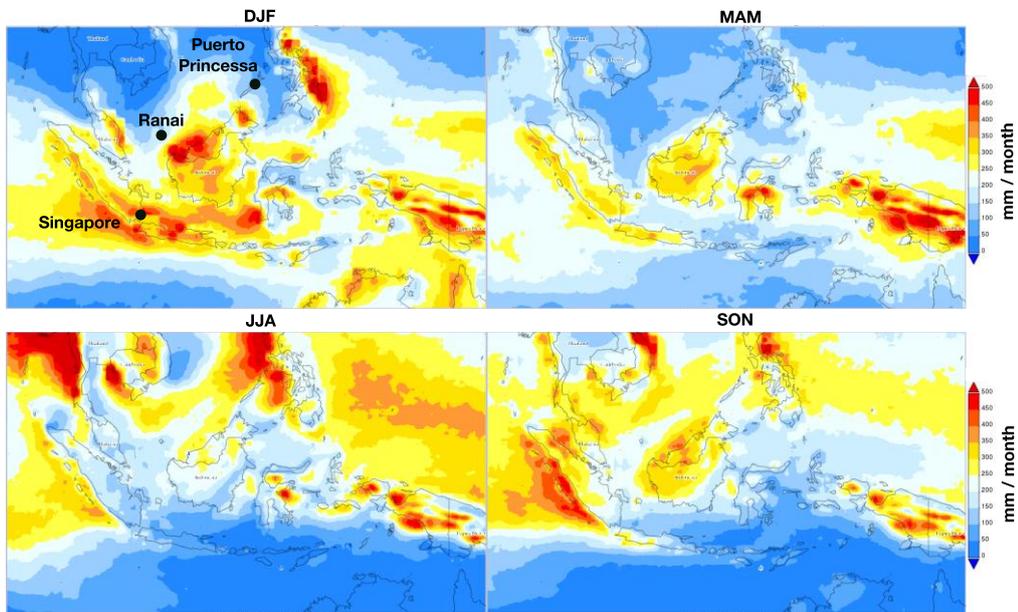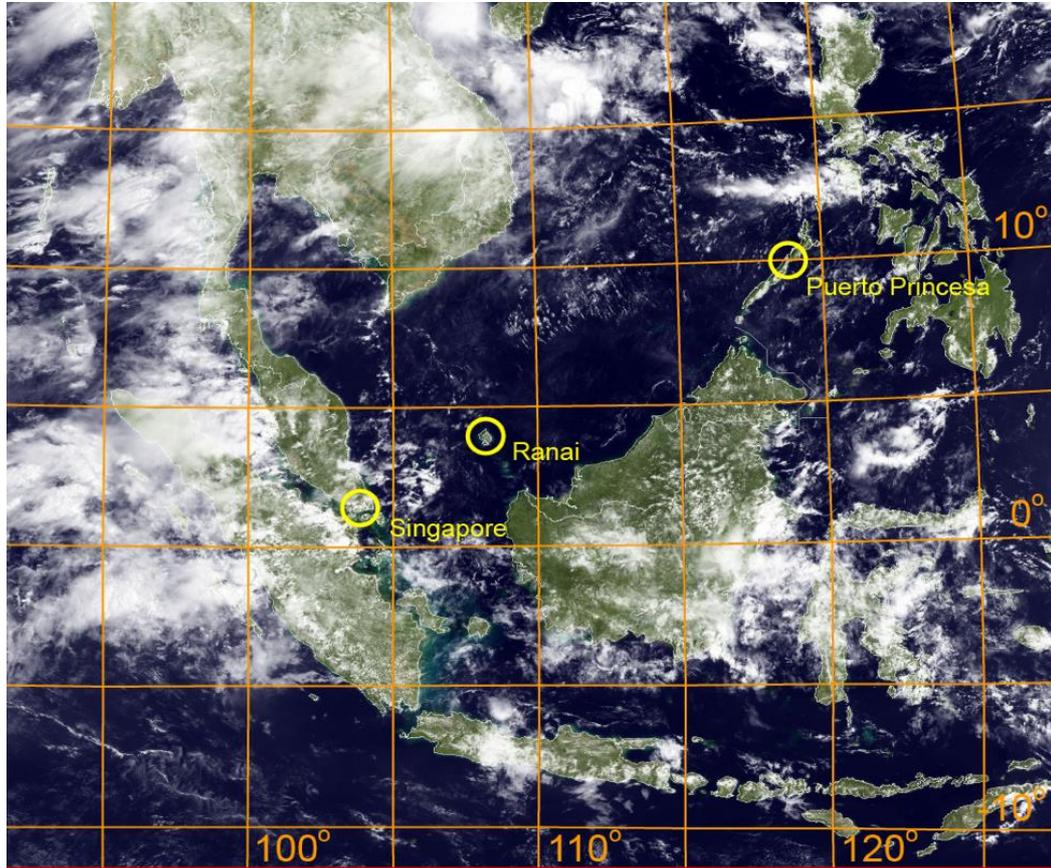
Figure 1: ~~MTSAT visible image over the~~TRMM 1998-2016 seasonally averaged precipitation for the~~–~~ Maritime Continent. Seasonal variability is apparent and highly localized. Upper-air ~~Maritime Continent obtained at 0332 UTC 14 September 2011.~~ ~~s~~Sounding sites used in the analysis are labeled and circled in black in the first panel, and are most representative of the South China Sea. ~~.~~
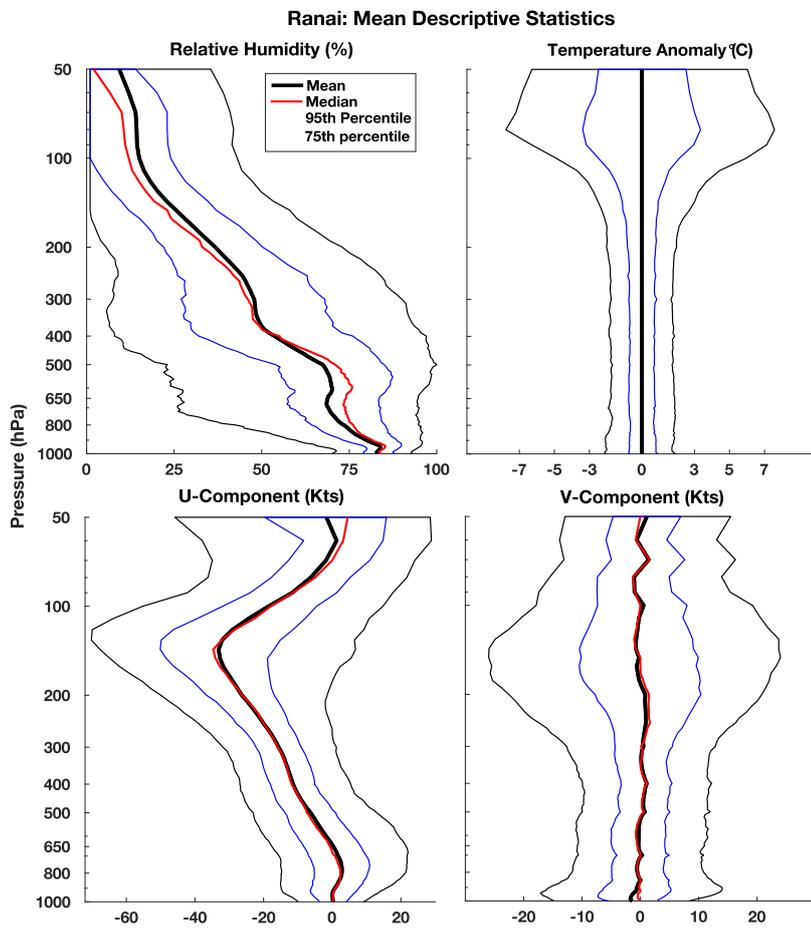
Figure 2: Mean, median, interquartile range, and 95th percentile range of the Ranai sounding dataset. Relative humidity and the U-component show signs of skewness. The range of values for each variable's unit of measure at each level is apparent.
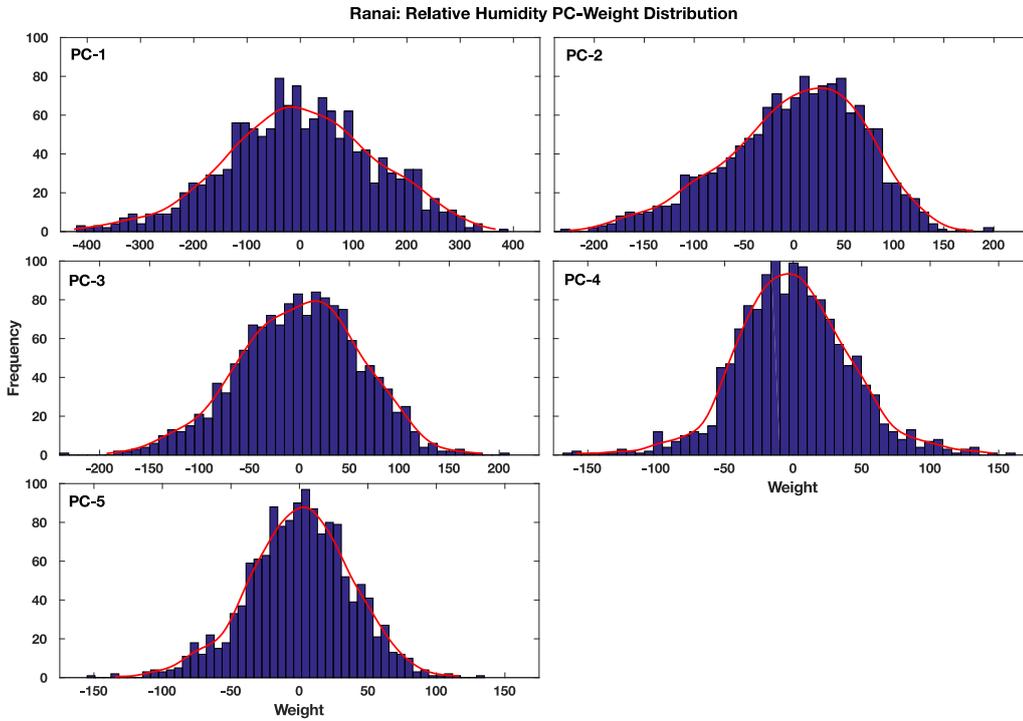
Figure 3: Histograms of the relative humidity PC-weights for Ranai in blue, with distribution fits in red. The original descriptive statistics (Fig. 2) showed signs of skewness for this variable. Because the original input data was not standardized, the PCA algorithm retains the inherent skewness, which is observable in PC-2, PC-3, and PC-4.

5
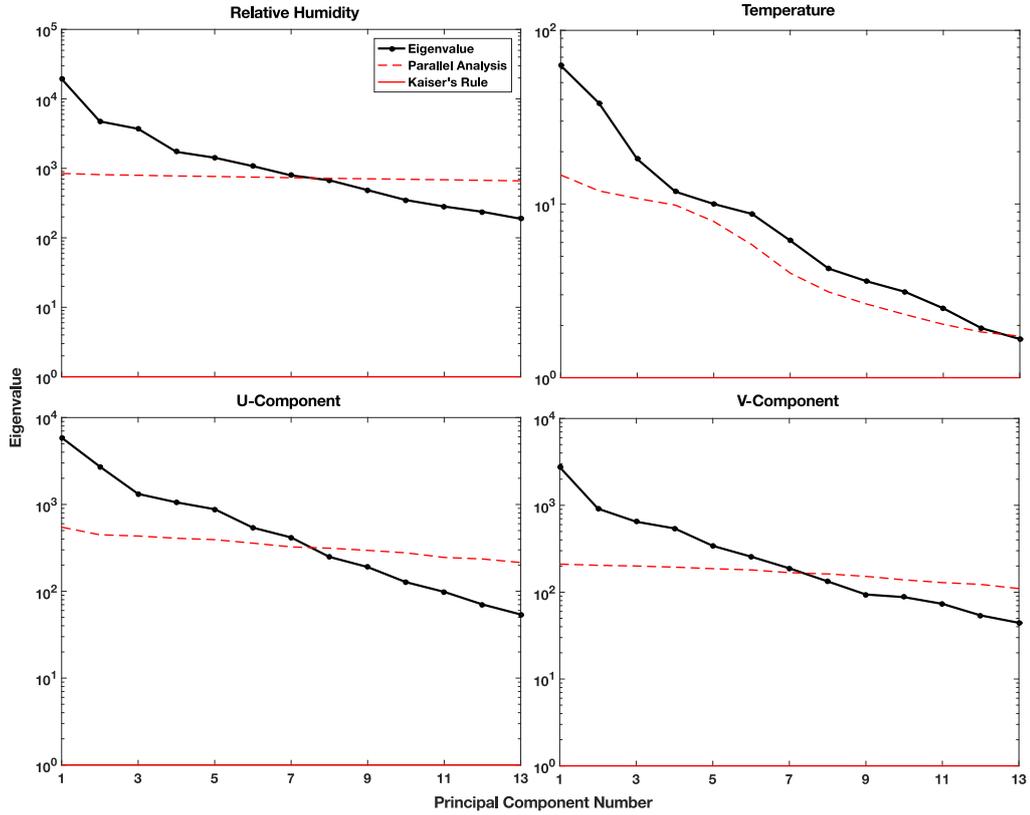
Figure 4: Scree plots for the Ranai ~~release~~ upper-air sounding site for the four variables of interest. The calculated eigenvalues are in black and the PC retention criteria are in red. Generally, Kaiser's Rule (solid red) overestimates the number of PCs to retain, whereas Horn's Parallel Analysis (dashed red) is more reliable.

5
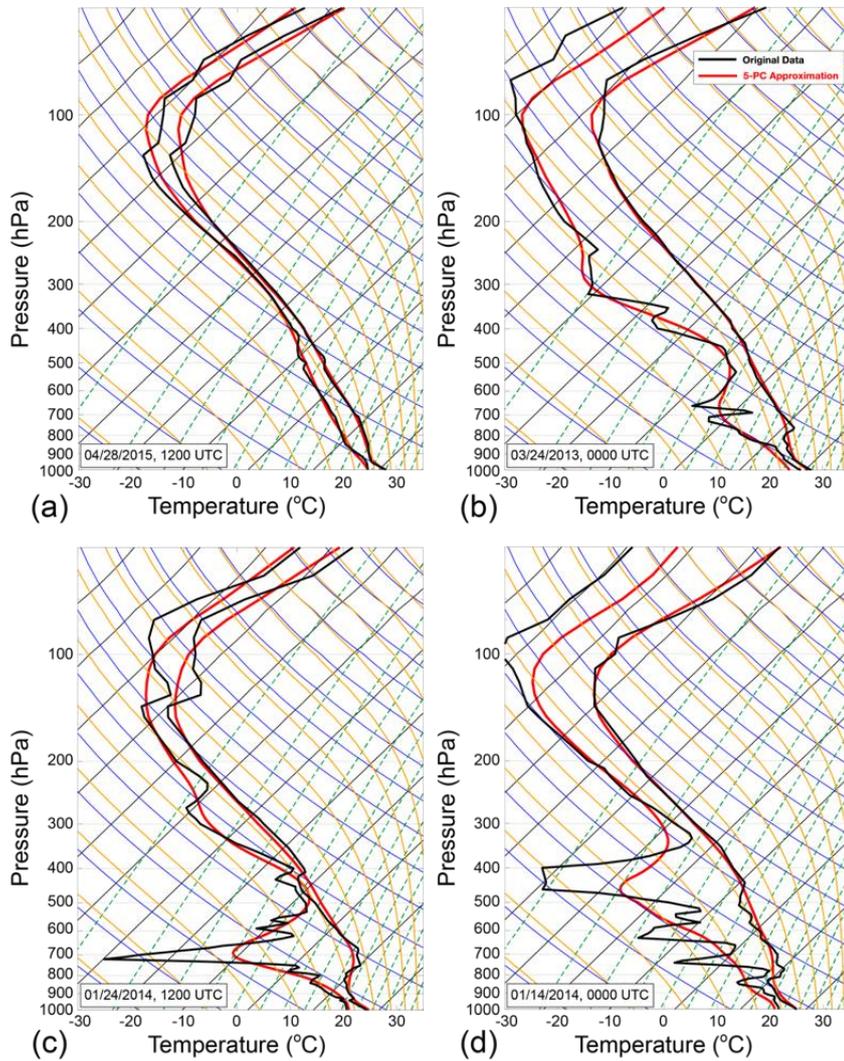
Figure 5: Original radiosonde observations (black) from the Ranai ~~release~~ **upper-air sounding site and reconstructions of the same radiosonde data using only 5 principal components (red). Top rows (a and b) represent thermodynamic structures for which the 5-PC approximation performs well. The bottom row (c and d) depicts soundings with more complex vertical structure that are more difficult to reproduce with a small number of retained components.**
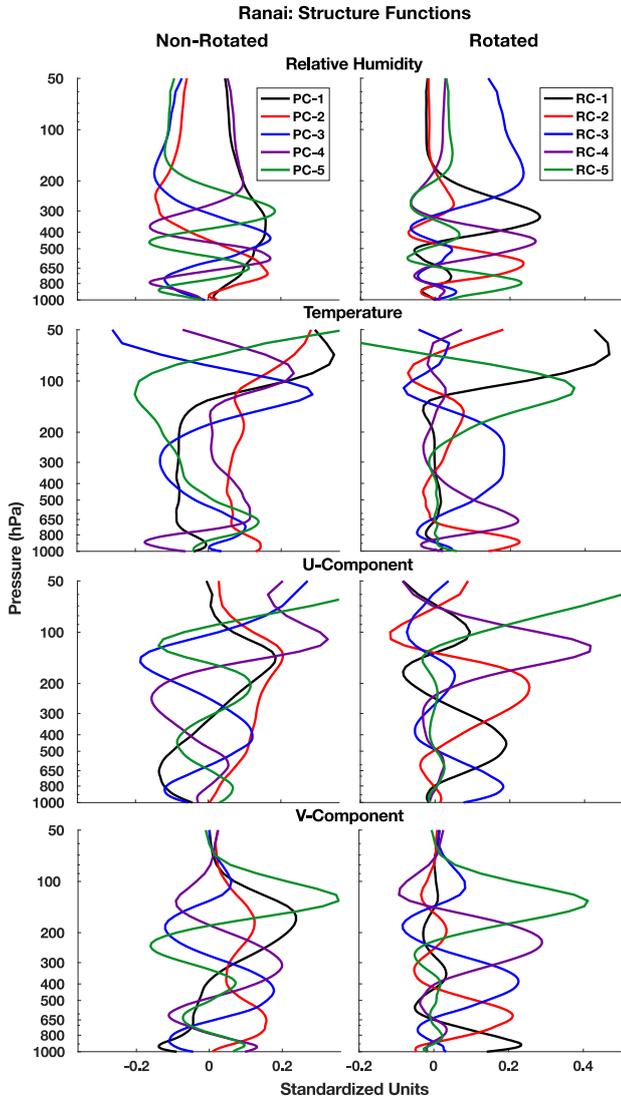
Figure 6: Non-rotated PCs (left) and rotated components (right) for all variables at the Ranai ~~release~~ upper-air sounding site. Only one sign of the eigenvector is plotted. The strong modality (sign change) is apparent in the PCs, whereas the RCs have five modes total due to the Varimax rotation, but only one mode contains a strong signal.

5

Ranai: Relative Humidity Modes
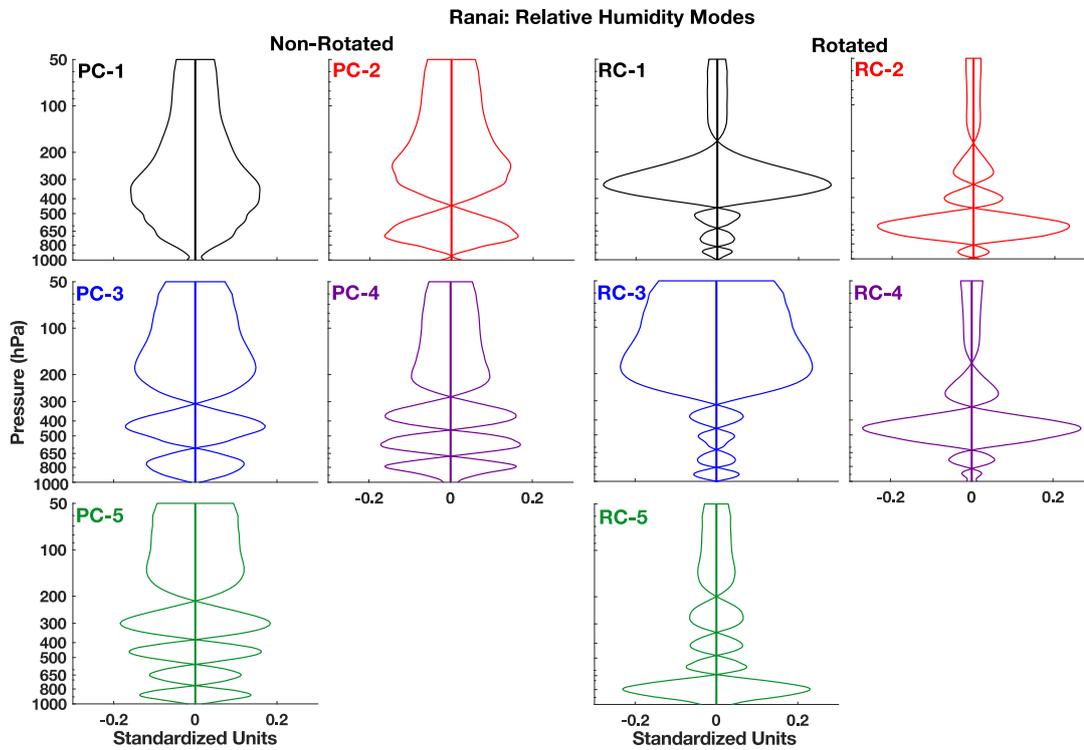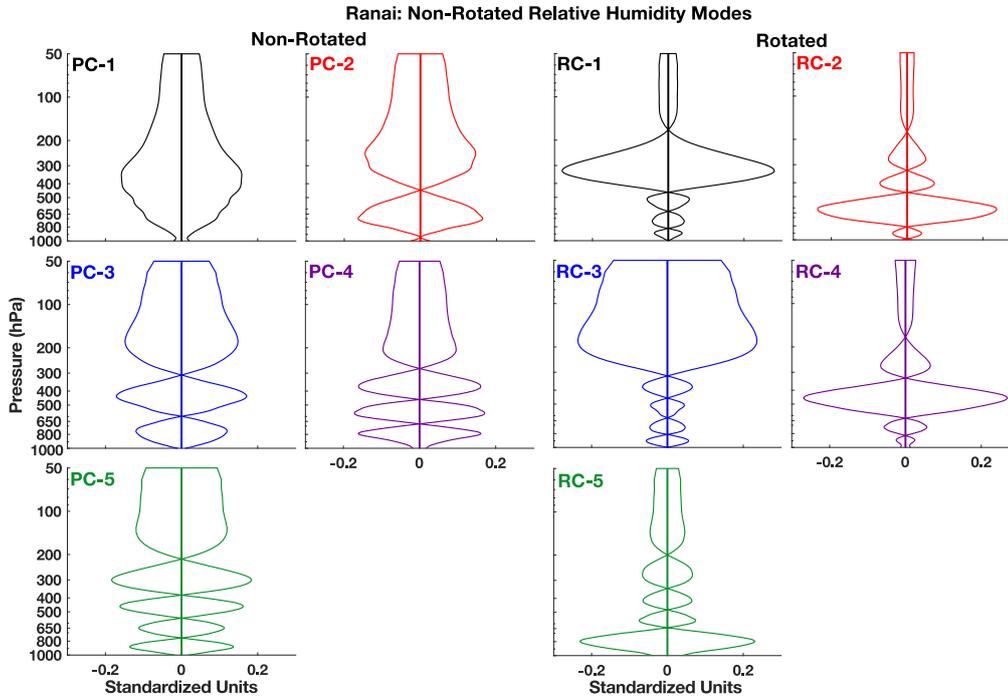
Figure 7: Non-rotated PCs (left) and rotated relative humidity RCs (right) at the Ranai ~~release~~ upper-air sounding site. Bold lines indicate the zero line and the positive and negative values of the eigenvectors. The shaded area represents the total spread of variability.

5
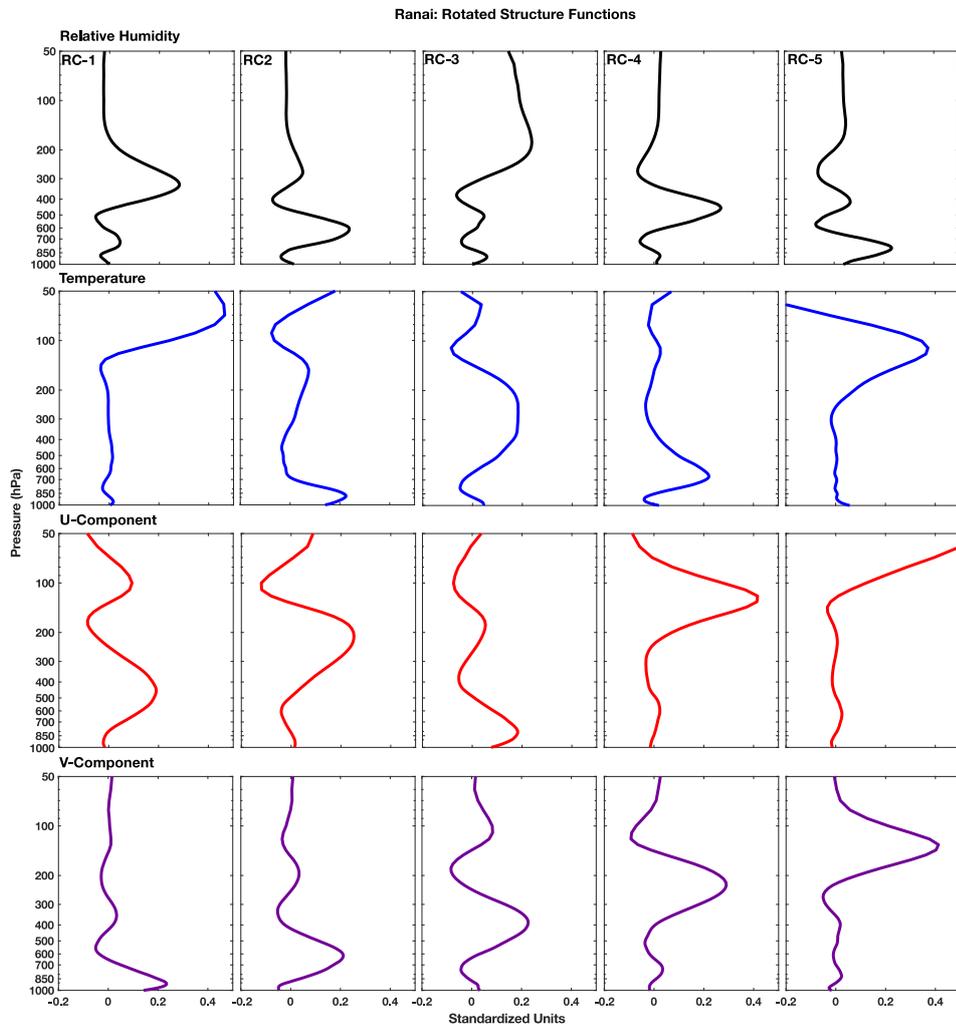
Figure 8: The first five RCs at the Ranai ~~release~~ <u>upper-air sounding</u> site for relative humidity (black), temperature (blue), U-component (red), and V-component (purple). The y-axis is in log-pressure (hPa) and the x-axis is in standardized units. For simplification, only one sign of the eigenvectors is included in this figure.

5

Ranai: Radiosonde Observations and Convection Time Series
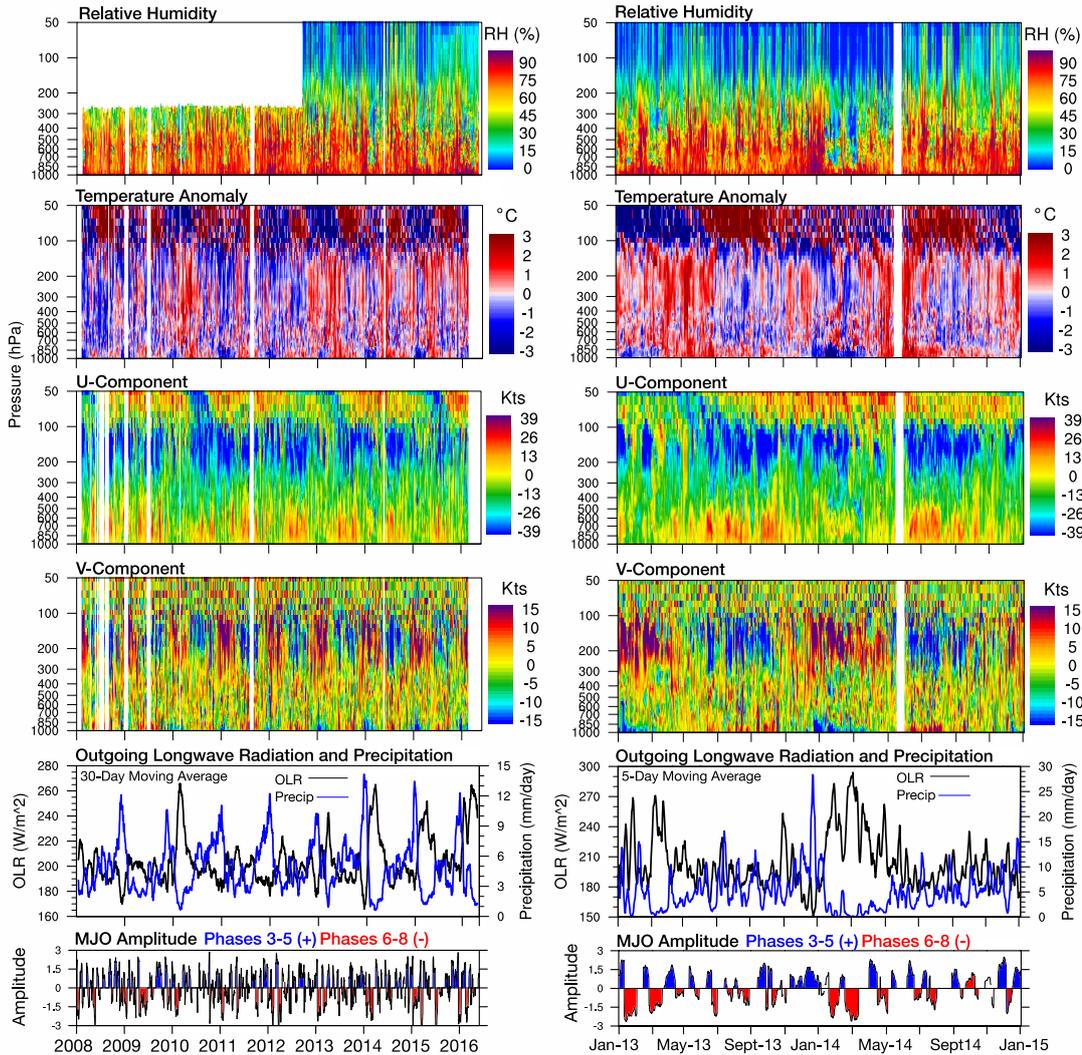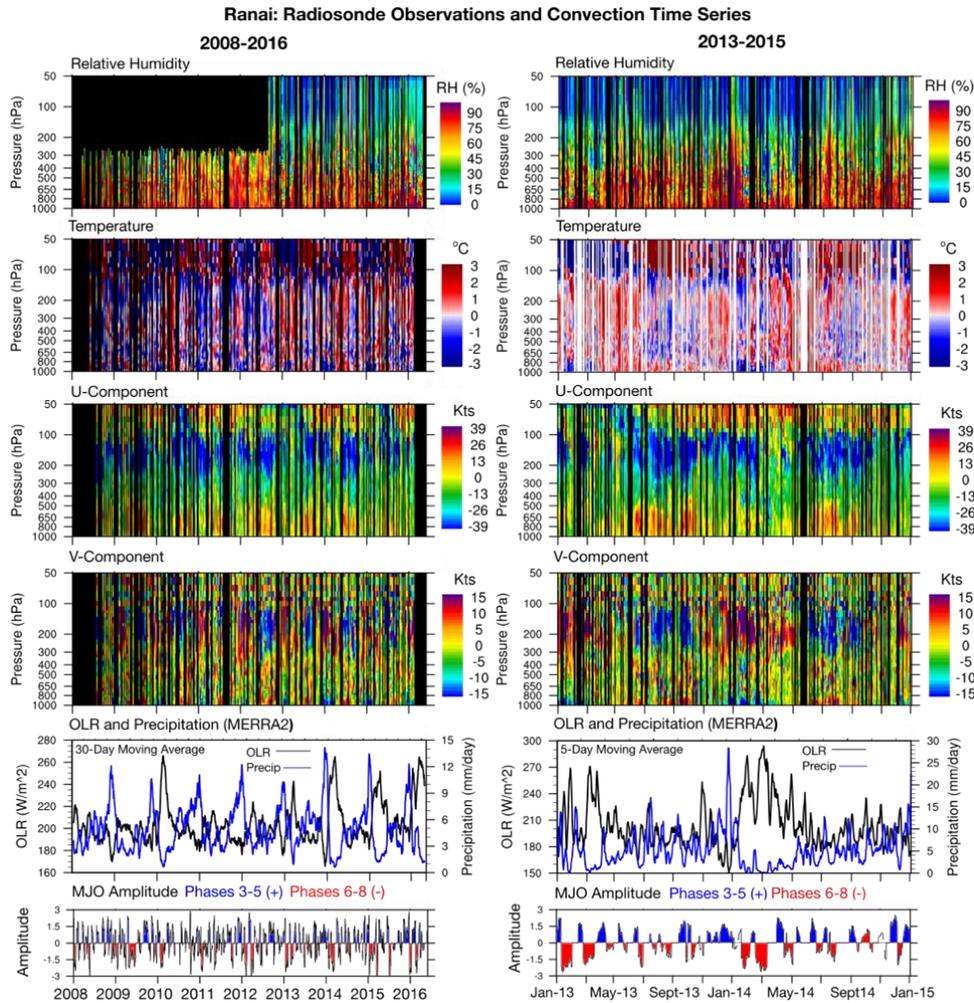
Figure 9. Time series of temperature, relative humidity, U and V winds, outgoing longwave radiation, precipitation, and the MJO active and suppressed phases at Ranai. The left column is the entire study period 2008-2016, and the right column is a subset of that time period from 2013-2015. T, RH, and wind data in the first four rows is from the radiosonde observation dataset in Section 2 and was interpolated if data was missing for no more than 5 consecutive days. The Ranai radiosonde equipment reported RH only below the 300 hPa level until 2012 and is therefore missing. Areas of the time series shaded in ~~black~~ white in the first four rows represent extended periods of missing observations or dates that did not pass quality control. OLR (black) and daily precipitation (blue) in the 5th and 6th rows come from the MERRA2 dataset. To reduce noise, OLR and precipitation have moving average filters applied. The MJO amplitude was split into active phases (~~red~~blue), which enhance precipitation in Ranai (Phases 3-5), and suppressed phases (~~blue~~red), which suppress precipitation in Ranai (Phases 6-8).

45
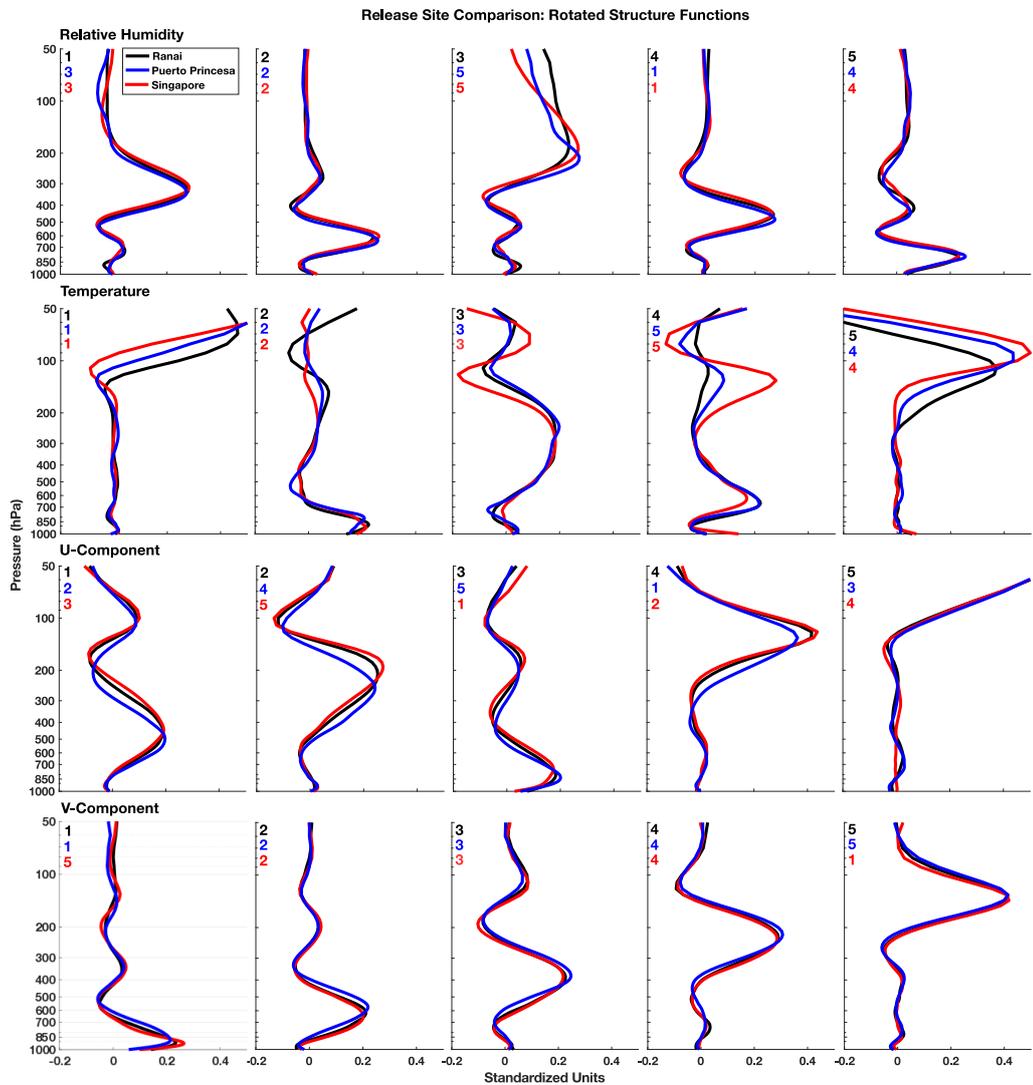
Figure 10. RCs for the three radiosonde ~~release~~ sites in Ranai (black), Puerto Princessa (blue), and Singapore (red). The panel numbers represent the RC number for that specific signal and are color coded to match their respective ~~release~~ upper-air sounding site.

5