Project 1

Sanskriti Purohit & Caleb Woo

2022-10-13

1. Summaries

1.1 Paper

The paper, "A Macroscope in the Redwoods" highlights the potential of the wireless sensor network macroscope in the field of science, particularly its use in observing and monitoring instances that vary over time and space. It also provides a way to analyse the complex data obtained through these sensors and draws a few conclusions about the dynamics of the microclimate around a coastal redwood tree. Therefore, establishing and achieving its motivation of devising a base model for future researches to observe what they weren't able to measure at the time. The redwood tree, proven to have a substantial variation in its microclimate over both space and time, was thus chosen to be the center of this research.

During the time frame of a month, the collected data, measurements of temperature, humidity, incident PAR and reflected PAR, was transferred over a network and also logged through a distributed data logger contained in the suite of sensors placed on the redwood tree, such that each sensor maintained a certain minimum and maximum distance from the ground level, the trunk and other sensors.

By projecting the sensor values onto time and then onto space, the researchers discovered a much higher variation over time than in space which informed them to remove temporal variation and focus on spatial variation.The researchers determined that PAR and temperature correlate with solar movement, although reflected PAR is more noisy because the absolute readings are lower.Humidity, on the other hand, was shown to change considerably over a wide range prior to sunrise and after sunrise spatial variability remains high suggesting that the foliage of the tree has a significant effect on humidity across different heights of the tree.

Thus, multi-dimensional analysis of the data paints a detailed picture of the spatiotemporal variance, previously validated by biologists. Further, it shines light on the shortcomings of the existing wireless sensor networks.

1.2 Data Collection

The data for the study was collected over a duration of approximately 44 days (Tuesday, April 27, 2004 at 5:10pm to Thursday, June 10, 2004 at 2:00pm), where each sensor was sampled every 5 minutes. The sensors were deployed on the west side of the tree, at least 15m above the ground level, reaching a maximum distance of 70m while maintaining a 0.1-1 meter distance from the tree trunk. This was to ensure the capture of the variation caused by the foliage. There was also a 2 meter distance between each node, which proved to have a significant impact on the results obtained.

The package was designed to ensure protection of the components of the node from various natural phenomenon such as rain, dirt, wind e.t.c. while enabling them to have the required sensitivity to measure the variables of interest. The researchers utilized the TinyDB data collection framework to collect the data into table format and represent each sensor as a column and each row as a reading taken at a particular time with additional columns for node ID, sample number, and sample reception time. Further the hardware and the software was calibrated twice to ensure optimum collection of the measurements of temperature,

humidity, incident PAR and reflected PAR. Roof calibration was used on the PAR sensors to ensure that they were producing acceptable readings relative to a high-quality-well-established PAR sensor. Two-point calibration for each humidity and temperature sensor was performed by placing sensors in a controllable weather chamber and cycling through various temperatures and humidity levels. While calibration improved the mean temperature deviation from 0.35 degrees Celsius to 0.18 degrees Celsius and the mean humidity deviation from 1.24% relative humidity to 0.60% relative humidity, the absolute increase in accuracy was small enough to be of minimal benefit.

The data was collected in two ways, through the transfer of the data to the gateway over a network and logging of the data at the site by the data logger installed in each node. **sonoma-data-log.csv** contains the data obtained from the logger while **sonoma-data-net.csv** contains that transferred over the network.

2. Data Cleaning

Voltage is our first inconsistent variable. We can see the histograms of network and log voltage in section A.1. We see that the network data voltage is in ADC units in the left plot while the log data voltage is in volts in the middle plot. Tolle et al. uses volts as the voltage unit, so we converted the network data voltage units from ADC to volts using the MICA2DOT user's manual found at this link: http://www-db.ics.uci.edu/pages/research/quasar/MPR-MIB%20Series%20User%20Manual%207430-0021-06_A.pdf.

In section 6.5 of the user's manual, we see that we can convert ADC to volts by using the following formula:

$$\text{Voltage} = 0.6 * \frac{1024}{ADC}$$

The histogram of the fixed network data voltage in volts can be seen in the Appendix section A.1. Incident PAR is our second inconsistent variable. The histograms of network and log incident PAR can be found in Appendix section A.2. In the top two plots, we see that both network data incident PAR and log data incident PAR are represented as illuminance in lux units rather than as the photosyntehtic photon flux density (PPFD) in micromoles per second per meter squared units used in the paper. So we converted illuminance to PPFD using Apogee Instrument's conversion table found at this link: https://www.apogeeinstruments.com/conversion-ppfd-to-lux/.

In the table on the right hand side, we see that we can convert illuminance in lux to PPFD in micromoles per second per meter squared for sunlight measurements by using the following formula:

$$\text{PPFD} = 0.0185 * \text{illuminance}$$

The histograms of the fixed network and log incident PAR as PPFD in micromoles per second per meter squared are shown in the bottom two plots in Appendix section A.2. Similarly for reflected PAR, both network data reflected PAR and log data reflected PAR are represented as illuminance in lux units rather than as the PPFD in micromoles per second per meter squared units used in the paper. So we converted reflected PAR just like how we converted incident PAR above. The histograms of network and log reflected PAR prior to conversion and after conversion can be found in the Appendix section A.3. The top two plots show the histograms of network and log reflected PAR in lux prior to the conversion. The bottom two plots show the histograms of network and log reflected PAR in micromoles per second per meter squared after the conversion.

The other two relevant variables, humidity and temperature, appear to be consistent with each other and with the units used in the paper. Temperature is in degrees Celsius while humidity is in percent of relative humidity.

After we made all relevant variables in the network and log data consistent with each other and with the units used in the paper, we combined their data and also merged this combined data with the dates information found in "sonoma-dates". There are 8760 observations with missing measurements and these

missing observations occur between April 30 at 8:05 AM and May 25 at 9:15 PM inclusive. This information can be referenced in Appendix section A.4. We removed these 8760 missing observations.

When we initially combined the network data and log data, we noticed that they essentially have identical columns except for voltage. This must be because network voltage data had to be converted from ADC to volts so the conversion may not be exact. Since network data is missing for some rows, log data is not missing for any rows, and log data has essentially identical values as network data (other than voltage but log data voltage is more accurate anyways), we removed network data columns from the combined data frame. In addition, the result_time column from the combined data and the epochDates column merged from the dates information found in "sonoma-dates" represent the same information. However, epochDates from "sonoma-dates" is more accurate so we also removed the result_time column from the combined and merged data. Finally, I merged the locations data found in "mote-location-data.txt" with the rest of the combined and merged data described above to get a total of 16 variables that includes network and log data, dates data, and locations data. After merging the locations data, we found that there are 6091 observations with missing values in the locations data so we removed these observations as well.

Since Tolle et al. found a correlation between battery failure and outliers, we first utilized their automatic outlier rejection strategy to get rid of most outliers before investigating for further outliers. We removed all data with voltages greater than 3 volts and less than 2.4 volts as described by Tolle et al. This removed 26,011 observations from our data.

We can find the quantiles of humidity, adjusted humidity, temperature, incident PAR, and reflected PAR after removing outlier voltages as well as their corresponding histograms in Appendix section A.5. While the quantiles of humidity and adjusted humidity are very similar, the maximum value of adjusted humidity at 100.223 is closer to the maximum humidity value of 100.2 found in the paper so we will use adjusted humidity as our humidity measurement going forward. Comparing the quantiles of incident PAR with the quantiles of the paper and visually inspecting the incident PAR histogram revealed that there are some abnormally high values of incident PAR. Looking at the observations with an incident PAR higher than the maximum incident PAR of 2154 given in the paper, we found that node 40 is solely responsible for all of those readings. So we removed all observations recorded by node 40 to remove incident PAR outliers. Comparing the other quantiles with the quantiles in the paper and visually inspecting the other histograms do not reveal any further noticeable outliers in adjusted humidity, temperature, and reflected PAR.

We can find the quantiles and histograms of the relevant variables after removing outlier voltages and removing node 40 in Appendix section A.6. The quantiles match up well with the quantiles found in the paper and the histograms do not reveal any further noticeable outliers. Removing node 40 resulted in removing 98 observations recorded by node 40. Therefore, the final number of observations after all data cleaning is 271,776. Considering that log data contained 301,056 observations and network data contained 114,980 observations, our final cleaned data retains most of the information found in the log and network data.

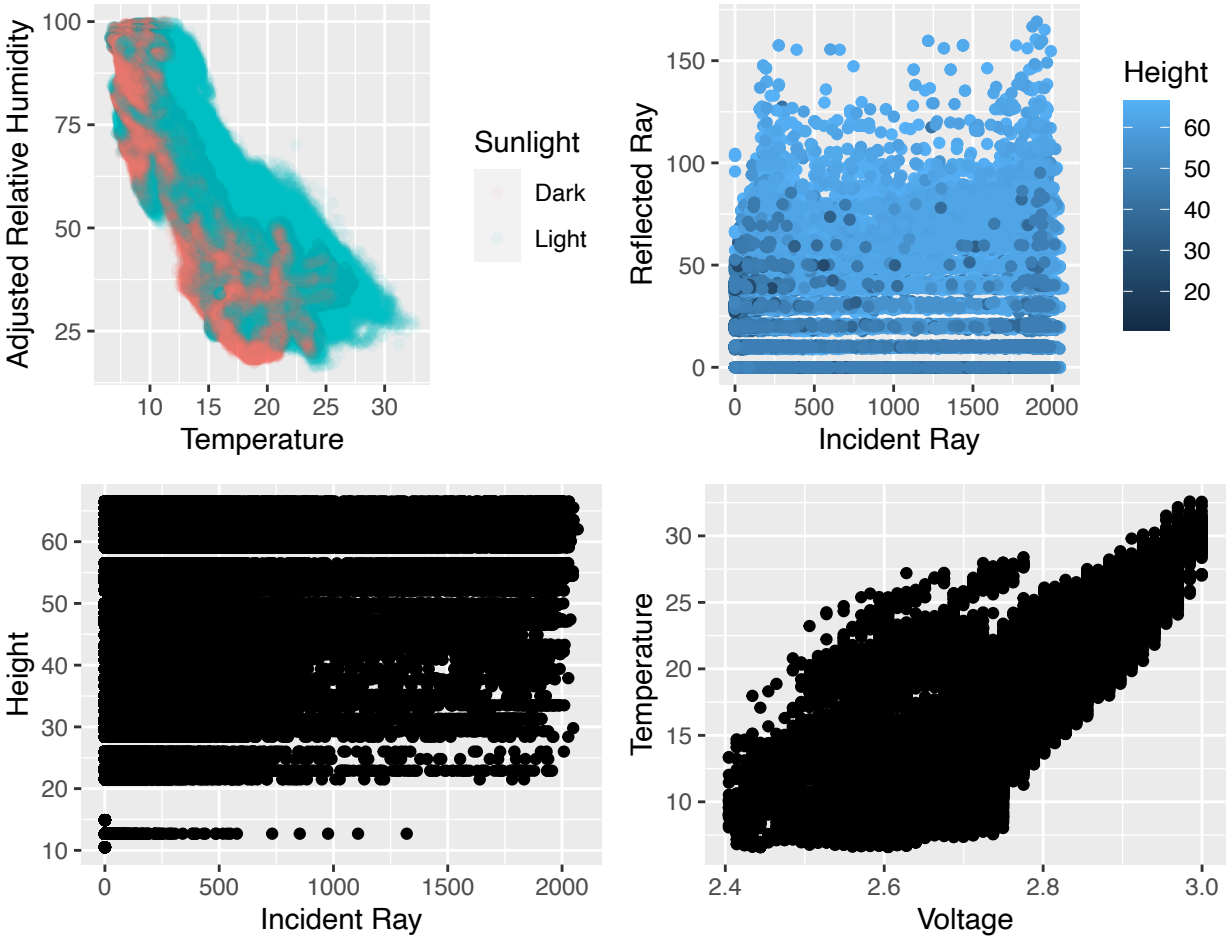
3. Data Exploration

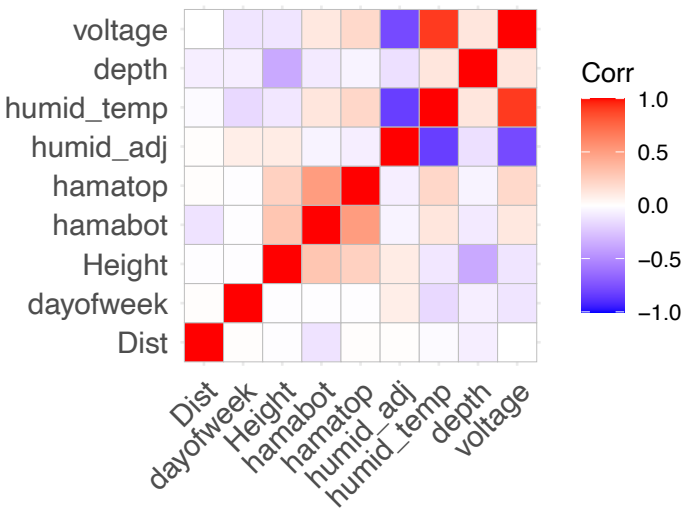
In order to gain insights into the data, data points between 27th April and 27th May were chosen as they constitute more than 90% of our information. Through the scatterplots of the corresponding data, we were able to gain some knowledge regarding the correlation between the variables and their behavior in the presence/absence of sunlight. A strong negative correlation between humidity and temperature can be observed through the scatterplot, which is also verified by obtaining a significant correlation coefficient of -0.827. Coloring the graph with the presence/absence of sunlight also confirms our belief of the existence of higher relative humidity and lower temperatures in the dark and an opposite behavior in the presence of sunlight.

By plotting the incident and reflected ray against each other, and coloring by height, we can observe the presence of higher readings in both at higher heights, suggesting that the level of absorption of sunlight is more or less uniform throughout the height of the tree. This can also be confirmed from the relatively strong positive correlation, 0.51, obtained from the test.

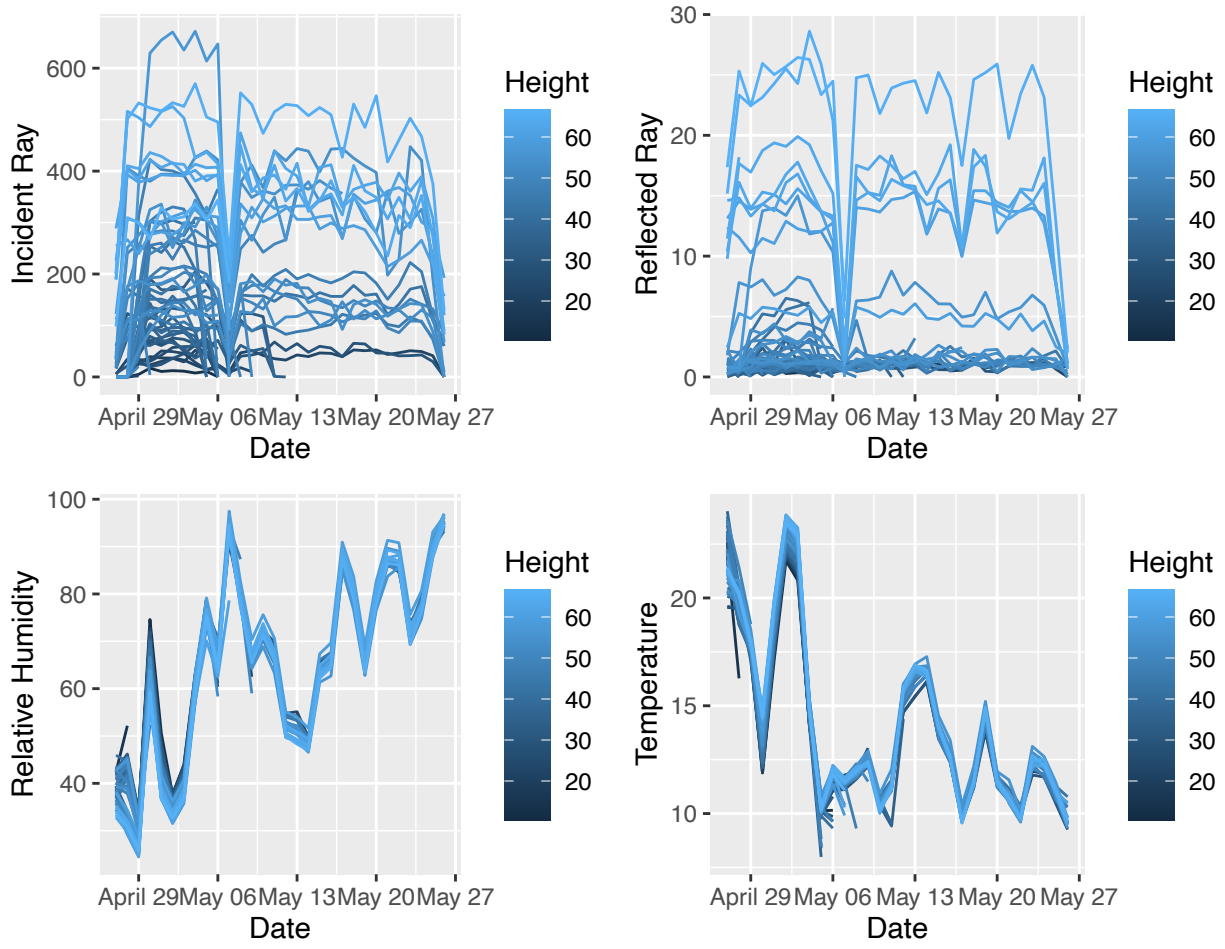
A positive correlation between height and incident ray, though not much visible through the figure, can also be concluded through the coefficient test. We can also observe a resultant positive correlation between height and temperature, perhaps due to presence of more sunlight in higher heights.

A strong positive correlation of 0.91 between voltage and temperature suggests a drop in voltage with the drop in temperature, which could further suggest that some readings below 2.4V and above 3V could be due to actual drop or rise in temperature and not due to node failure.

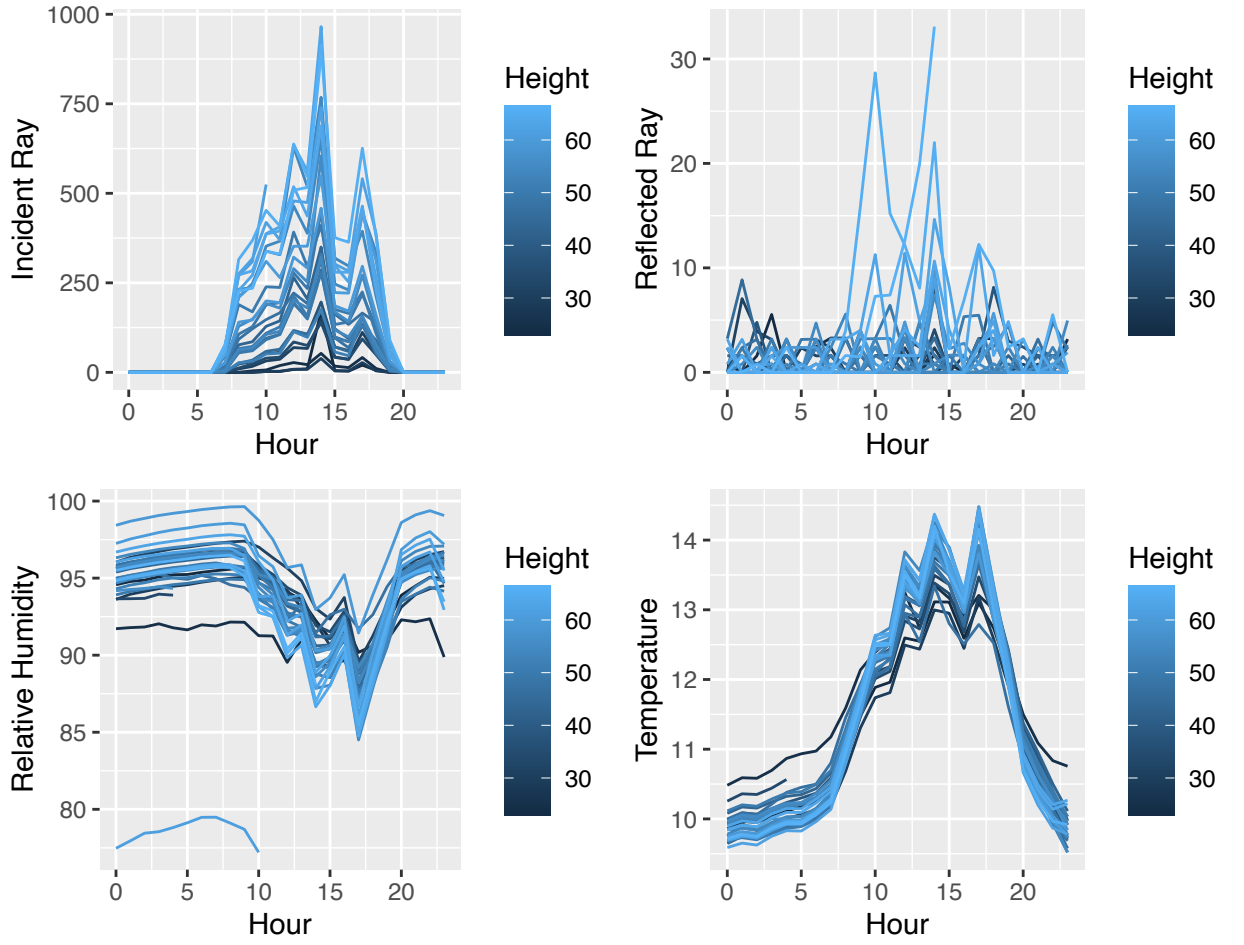




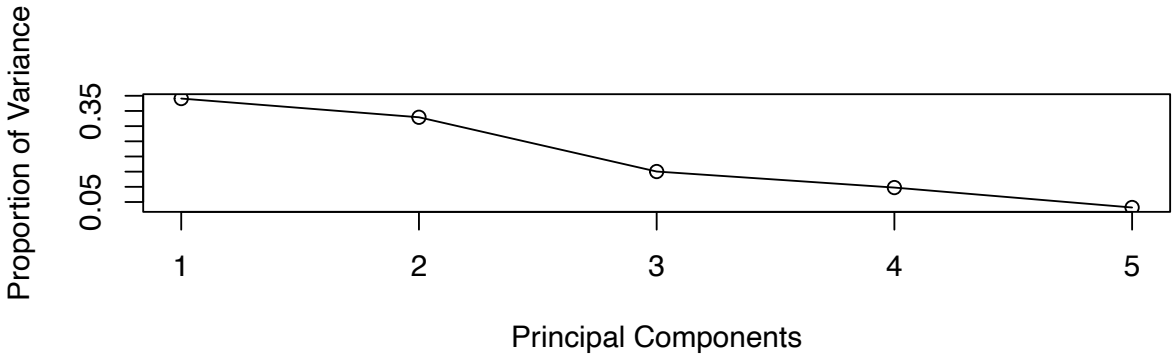
A temporal view, with coloring by space, also assisted in deriving further insights and confirming our previous suggestions. At first, we took a broader look at the relationship between the values, time and height by taking the average of each value for each day for each height. Through these temporal graphs, we can strengthen our belief of the previously established correlation between the values and height. We also observed that the days that saw a drop in incident ray also saw a drop in temperature and a rise in humidity, suggesting a cloudy/rainy day.



To delve deeper into the data, we took a look at one of the dates, May 7th, corresponding to a drop in the average incident ray. We took the average of the values by hour and colored them by height. It can be observed that the entirety of the tree had high values of RH, resulting in lower temperatures, especially in the early morning and late hours of night. The average amount of incident ray and reflected ray are also much lower than those observed on other days. Therefore confirming that May 7th was a day with cloudy day with spells of rainfall.



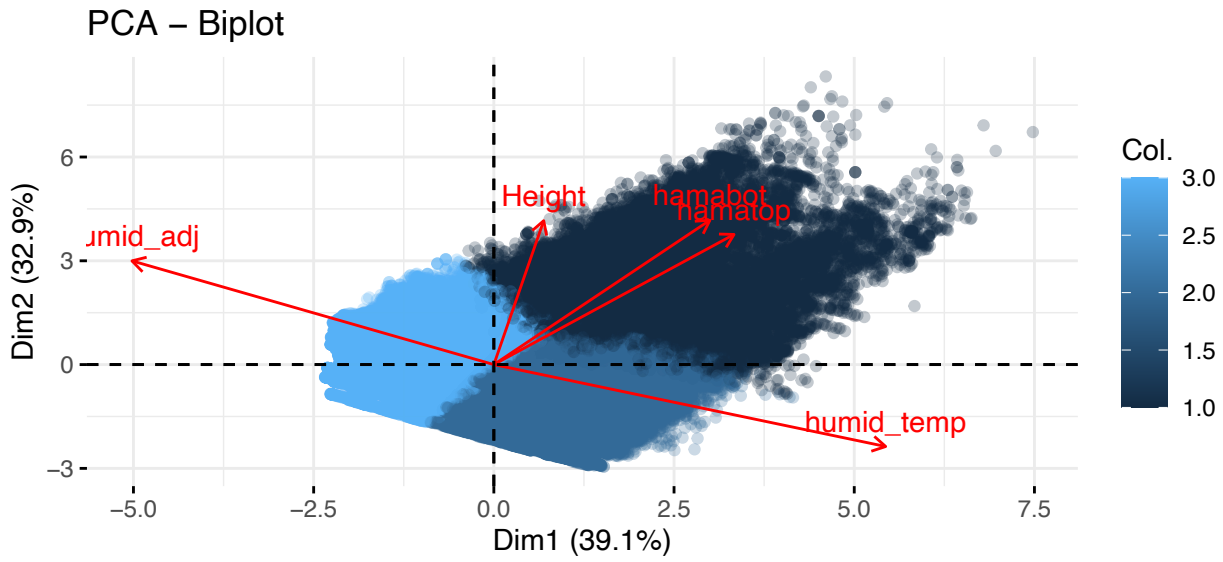
PCA was also performed on the data in order to check if it was possible to obtain a low dimensional projection. The values of interest: adjusted relative humidity, tempature, incident ray, reflected ray along with height were selected as the variables for the PCA. Through the scree-plot and the summary of the PCA, we can observe that approximately 72% of the variance of the data can be explained by the first two Principal Components.



4. Interesting Findings

Through the k-means clustering of the Principal Components, it can be observed that while plotting the components that explain a high variability of the data, humidity and temperature appear in opposite directions, hence confirming their negative correlation with each other.

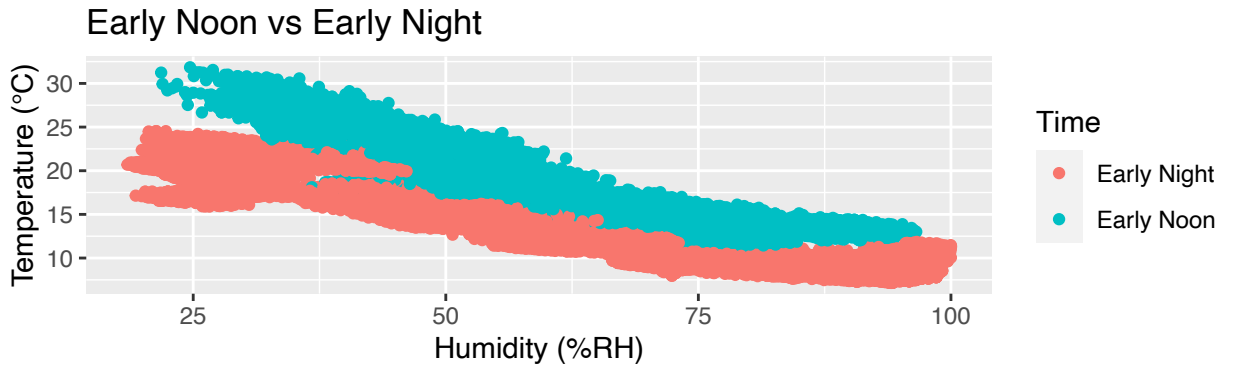
Also, when we consider the 3 clusters in the plot of PC1 against PC2, which cumulatively explain approx. 72% of our data, we can observe that the darkest cluster (represented by 1.0) represents values that contain high values of height, incident PAR and reflected PAR and relatively high values of temperature, leading to lower values of relative humidity. The second cluster (represented by 2.0) -> Lower values of height with higher temperature, higher incident PAR, higher reflect PAR and lower humidity. Third cluster (represented by 3.0) -> lower values of height, higher humidity, lower temperature, lower incident and reflected PAR.



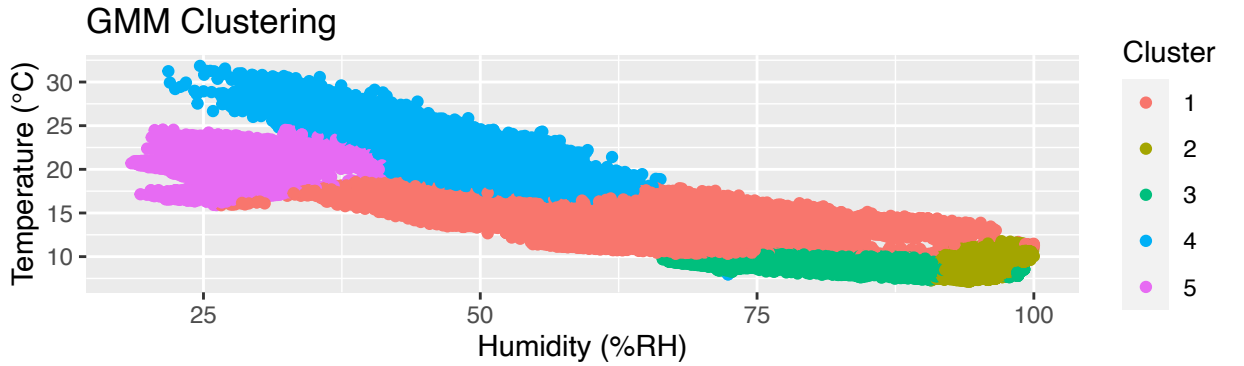
GMM Clustering

Please reference Appendix section A.7 to get a full description of our time period variable creation process. Below are the observations of the early noon and early night time periods projected onto humidity and temperature. Since humidity generally moves inversely with temperature, we thought that two very different time periods such as early noon and early night would show up as well separated groups when projected onto

humidity and temperature. Early noon and early night appear to be well separated with early noon having higher temperature values on average when compared to early night at corresponding humidity values. This separation checks out with our box plots of temperature over time shown above where early noon quantiles are noticeably greater than early night quantiles.



Using Gaussian mixture model clustering, we grouped these early noon and early night observations into 5 different clusters. The clustering results are shown below. The Gaussian mixture model clustering algorithm appears to separate the observations into 5 distinct groups with each cluster representing a particular section of either the early noon or early night observations. Cluster 1 is somewhat ambiguous since its observations with lower humidity correspond to early night while its observations with higher humidity correspond with early noon. However, the other 4 clusters appear to clearly represent either early noon or early night.



Below is a table of accuracy percentages and means of the humidity and temperature values for each cluster. We computed the accuracy percentage for each cluster as the percentage of the majority of points in that cluster. So cluster 1 has about 60% of observations being early noon, cluster 4 has about 89% of observations being early noon, and cluster 5 has about 99% of observations being early night. Clusters 2 and 3 both have 100% of observations being early night. Besides cluster 1, all other clusters appear to group sections of early noon or early night observations quite accurately. The mean humidity and mean temperature are also quite distinct between clusters.

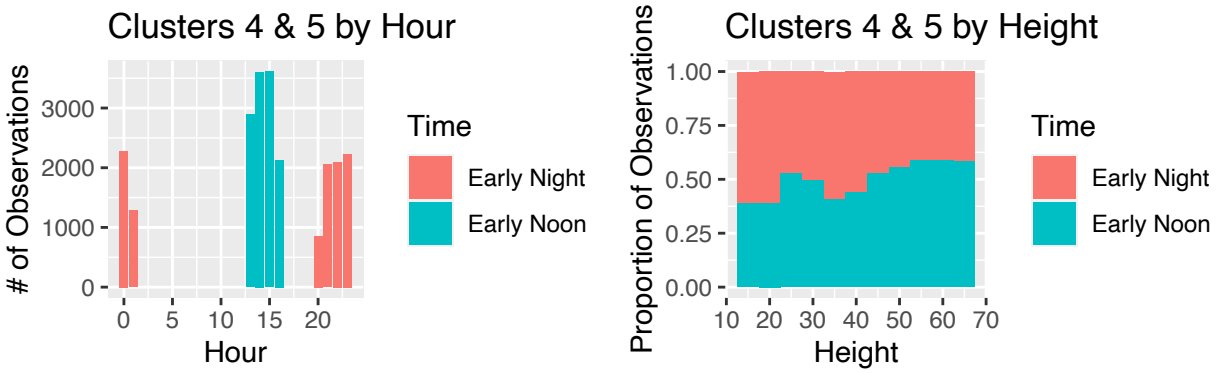
Table 1: GMM Accuracy and Cluster Means

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Accuracy %	39.937	100.000	100.000	11.113	1.202
Humidity Mean	65.560	94.944	83.222	51.169	30.107
Temperature Mean	13.887	9.028	8.946	20.031	20.942

Let us first compare clusters 4 and 5. Although clusters 4 and 5 both have similar values of low humidity and high temperature, about 89% of observations in cluster 4 are early noon while about 99% of observation in

cluster 5 are early night. The histogram below on the left suggests a clear temporal trend where early noon observations with low humidity and high temperature range from about 1:00 PM to about 4:00 PM while early night observations with low humidity and high temperature range from about 8:00 PM to about 1:00 AM. This temporal trend is expected since these ranges are practically the ranges of each time period when we created the time period variable. The histogram on the left allows us to conclude that during May, 2004 in Sonoma, California, early noon and early night time periods shared very similar values of low humidity and high temperature.

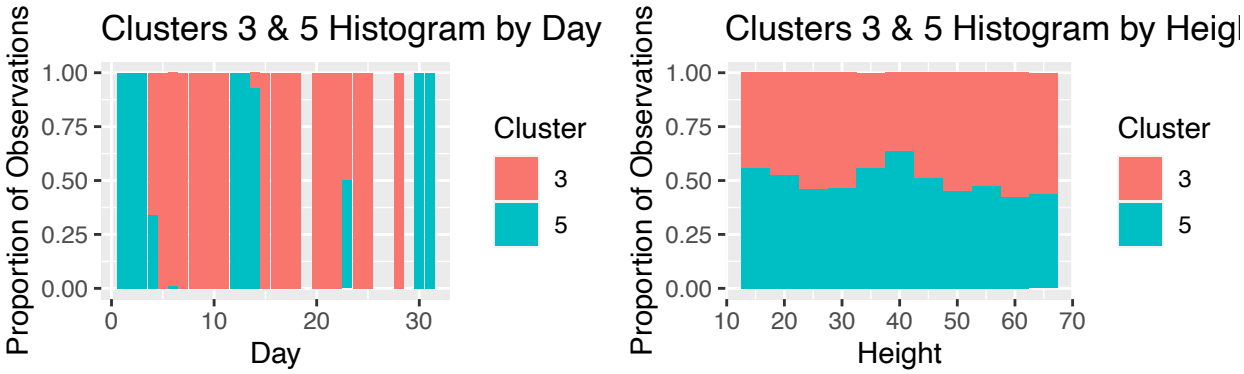
The histogram below on the right suggests a spatial trend where the early night time period makes up the majority of observations at low to middle heights while the early noon time period makes up the majority of observations at high heights. This spatial trend is more interesting than the expected temporal trend discussed above and perhaps it reveals a new insight. We believe that since sensors at high heights are more directly exposed to the sun, low humidities and especially high temperatures are captured more often during the early noon hours when the sun has a more direct effect on the sensor readings compared to the early night hours when the sun no longer has any direct effect on the sensor readings. We also believe that since sensors at low heights are more shaded from the sun, low humidities and high temperatures are captured more often during the early night hours compared to the early noon hours due to the delay in the effect of the sun on the sensor readings. Perhaps it takes a while for the sun to have any direct effect at low heights because of the heavy shading. So it is not until later in the day and into the early night when sensors at low heights are able to record the low humidities and high temperatures which were recorded during the early noon hours by sensors at higher heights.



Now we will compared clusters 3 and 5. Although clusters 3 and 5 both represent early night with very high accuracies, cluster 3 observations have high humidity and low temperature while cluster 5 observations have low humidity and high temperature. The histogram below on the left suggests a temporal trend where early night observations transition from low humidity and high temperature values to high humidity and low temperature values and vice versa throughout the month. For example, within the first 10 days of the month, early night observations clearly transition from cluster 5 to cluster 3 which suggests that the weather transitioned from hot, dry days to cold, wet days. This temporal trend agrees with the temporal trend found by Tolle et al. where week one includes warm, dry days and cold, wet days but the three following weeks contain predominantly cold, wet days.

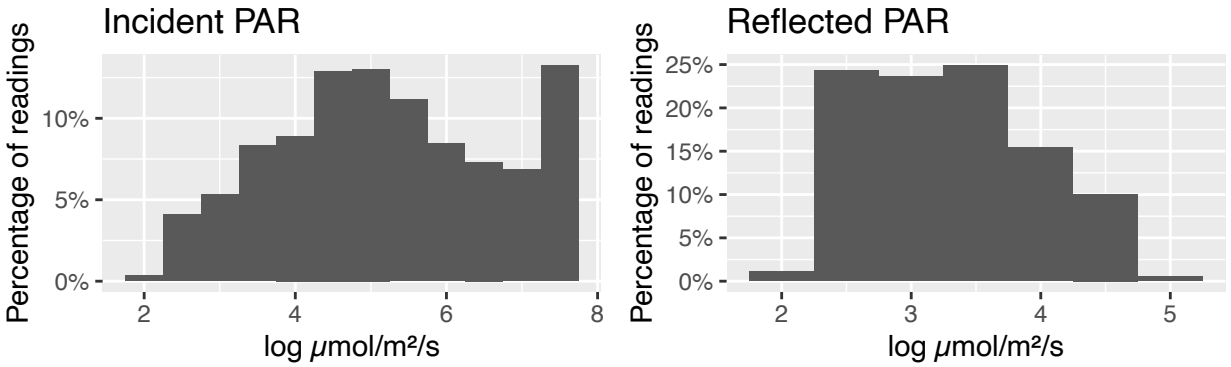
The histogram below on the right suggests a subtle spatial trend where low to middle heights tend to have low humidity and high temperature during the early night time period while high heights tend to have high humidity and low temperature during the early night time period. We believe that this spatial trend may have a similar mechanism to the spatial trend above. If the sun has a delayed effect on temperatures at low heights, then it will not be until later in the day and into the early night that sensors at low heights are able to record the high temperatures that sensors at higher heights were able to record earlier in the day. A potential explanation for the high humidity at high heights may be due to the transpiration process of such a massive redwood tree. Since a gigantic redwood tree would lose a ton of water vapor via transpiration, the resulting humidity recordings would be much higher at high heights compared to low heights since the water vapor is rises. However, most plants are believed to primarily transpire during the daytime, so further research into the transpiration rates of redwood trees would be necessary before drawing such conclusions.

This spatial trend is also more subtle than the spatial trend observed when comparing clusters 4 and 5 above so perhaps this trend is more of a coincidence than a finding.

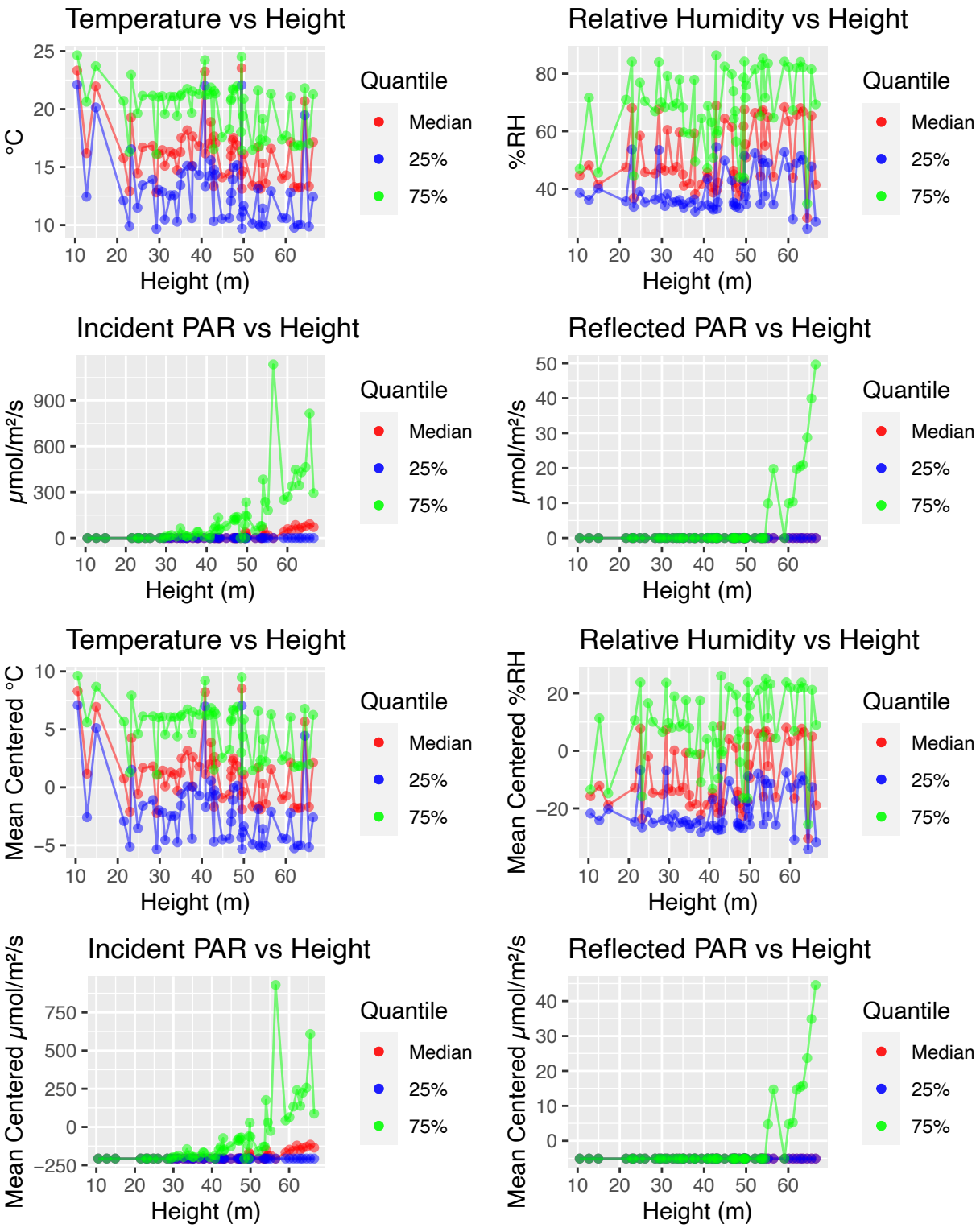


5. Graph Critique in the paper

The main difficulty in reading the full information from the long tails of the incident and reflected PAR histograms in figure 3[a] is that there are so many 0 values compared to the other values of incident and reflected PAR. This makes the first bin in each histogram extremely tall while all remaining bins at greater PAR values are extremely short in comparison. The long tails combined with the short heights of the incident and reflected PAR histograms makes reading the full information from the histograms extremely difficult. We believe that it would be best to plot all values other than the 0 values on the log scale to be able to more clearly differentiate the different values of incident and reflected PAR. About 55.8% of incident PAR values are 0 and about 84.9% of reflected PAR values are 0. The percentages on the y axes below represent the relative percentages of all non-zero PAR values plotted. So the incident PAR histogram represents percentages of the remaining 44.2% of values that are non-zero. The reflected PAR histogram represents percentages of the remaining 15.1% of values that are non-zero.

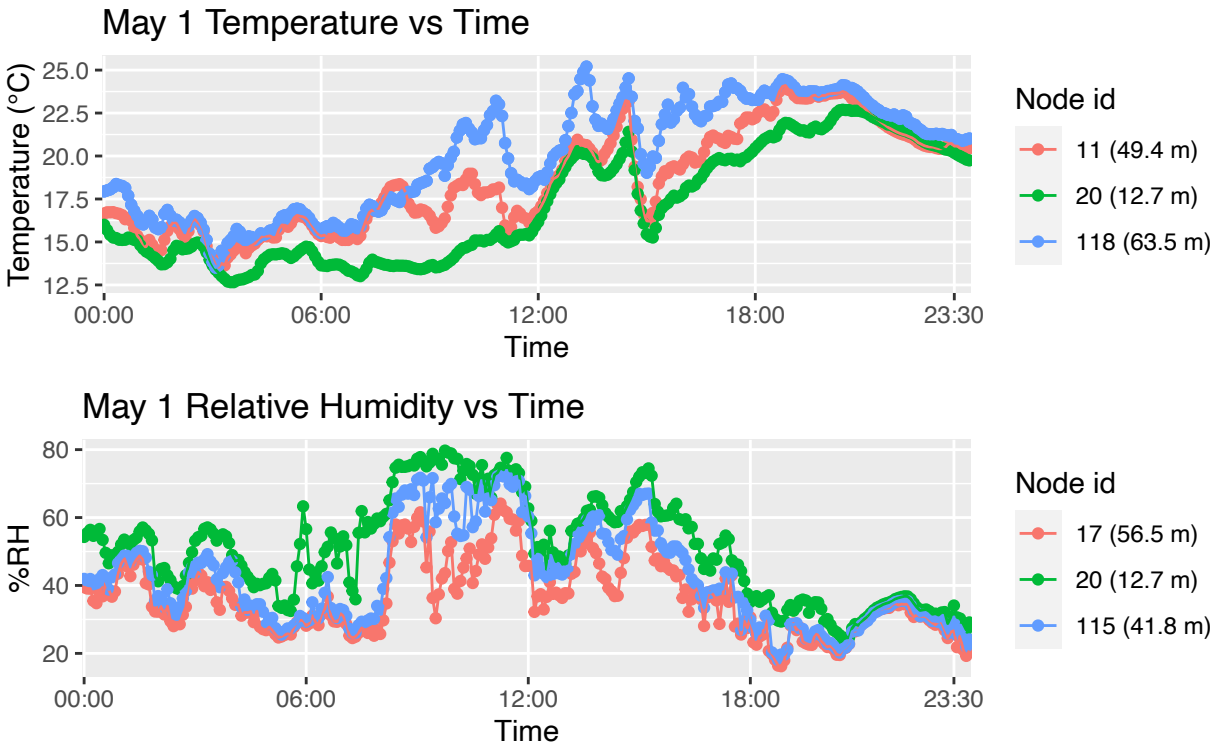


The box plots in figure 3[c] and 3[d] represent distributions and mean centered distributions respectively of temperature, humidity, incident PAR, and reflected PAR at varying levels of height. The box plots try to convey the differences across height of the four variables. However, we think that comparing the median and quartiles vertically and the visualization of the whiskers make it difficult to truly compare the differences of the distributions and mean centered distributions across height of the four variables. Using figure 4 as inspiration, we decided to plot height horizontally on the x axis, the four variables vertically on the y axis, and only the median, upper quartile, and lower quartile of the box plots to more clearly visualize the differences across height.



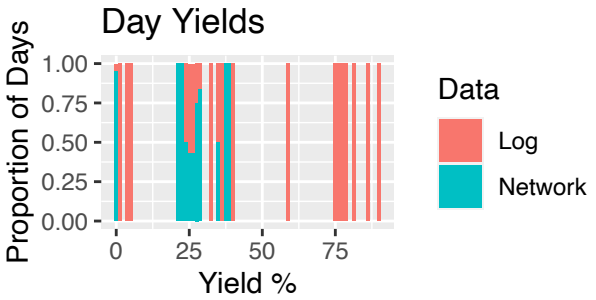
The first two plots in figure 4 are not very effective because it is so difficult to distinguish any of the individual sensor readings across time since all sensor readings are overlayed on top of each other. It is very difficult to distinguish any one color so it is very difficult to distinguish any particular sensor in these two plots. We thought it would be more effective to plot just a few individual sensors to more clearly distinguish between sensors across time on May 1, 2004. In order to do so, for each plot we found and plotted the node with the

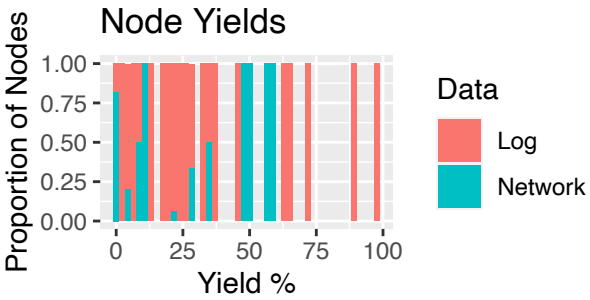
highest average value, the node with the lowest average value, and the node with approximately the middle most average value. By doing this, the three nodes across time are fairly well separated so it is much easier to distinguish between the different line colors and thus distinguish between the different nodes.

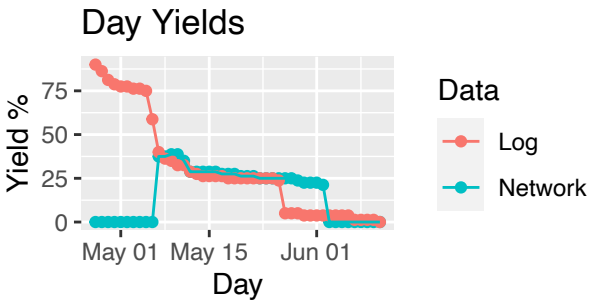


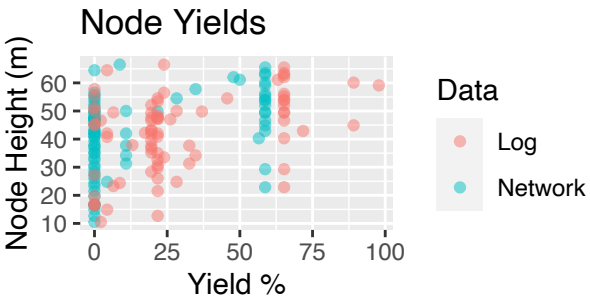
Although figure 7 clearly displays the different yields between the network and log data, it is difficult to directly compare them because the network and log data are plotted separately. We thought it would be easier to highlight the difference between network and log data by combining these plots and differentiating the network and log data by color. We found this to be most helpful for the second plot which is the plot of yield percentage over each day connected by a line. The two different colors representing network and log data in this combined second plot clearly show the difference in the change in yield percentage over the days between network and log data. The combined third plot, which is the plot of yield percentage over height, is more effective than when plotted separately because it now shows a clear difference in the yield percentage at each height between the network and log data. The combined fourth plot, which is the plot of the presence of any yield at each combination of day and height, is more effective than when plotted separately because the transparency of the points and the resulting shading of non-overlapping or overlapping points allow us to see where yield is present for network data, log data, or both at each combination of day and height.

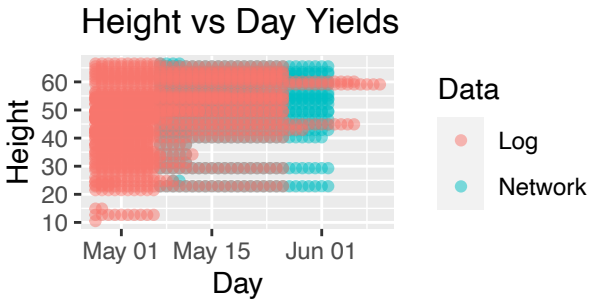
One additional concern is that it is unclear whether the original first plot in figure 7 represents the day yield or the node yield. Yield can be calculated for each day by summing and dividing by the total number of nodes. Yield can be calculated for each node by summing and dividing by the total number of days. The original first plot does not make this clear so we also thought it would be more effective to plot both day yield and node yield to highlight the difference in the two yields as well. Similar to the other three plots, we plot the day yield and the node yield with the combined network and log data and make each bar a proportion. This allows us to more directly compare the difference in yield percentages between network and log data for both day yield and node yield.


Day Yields
Proportion of Days
1.00
0.75
0.50
0.25
0.00
0
25
50
75
Yield %
Data
Log
Network


Node Yields
Proportion of Nodes
1.00
0.75
0.50
0.25
0.00
0
25
50
75
100
Yield %
Data
Log
Network


Day Yields
Yield %
75
50
25
0
May 01
May 15
Jun 01
Day
Data
Log
Network


Node Yields
Node Height (m)
60
50
40
30
20
10
0
25
50
75
100
Yield %
Data
Log
Network

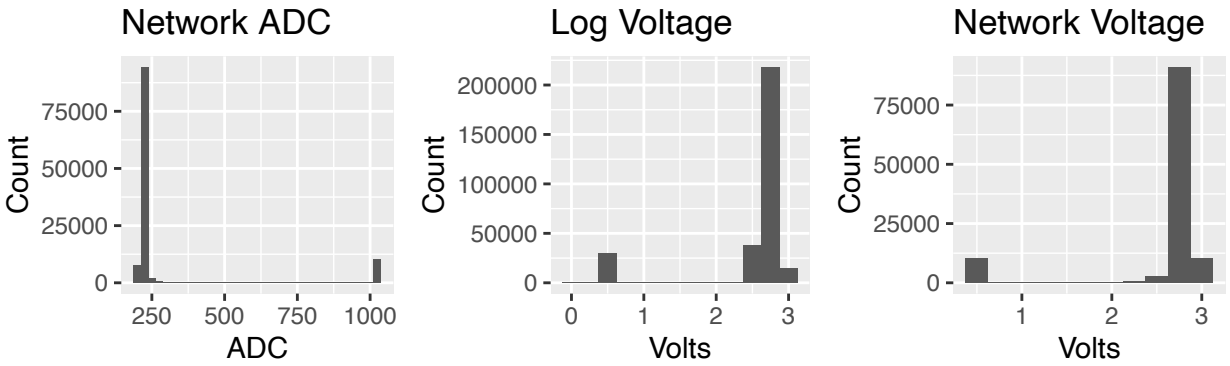

Height vs Day Yields
Height
60
50
40
30
20
10
May 01
May 15
Jun 01
Day
Data
Log
Network

Project 1 Appendix

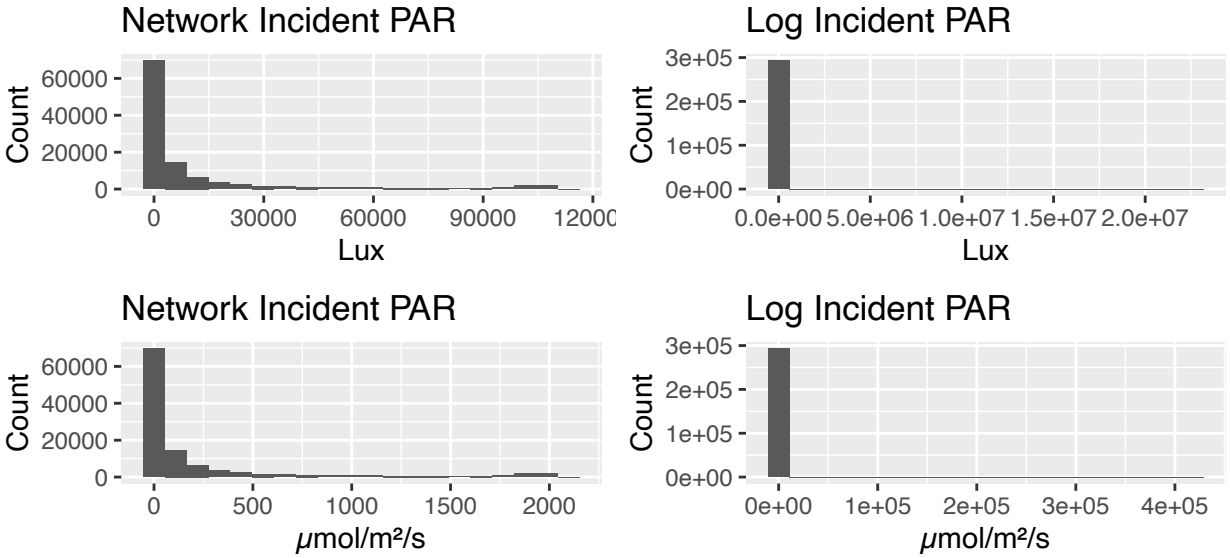
Sanskriti Purohit & Caleb Woo

2022-10-13

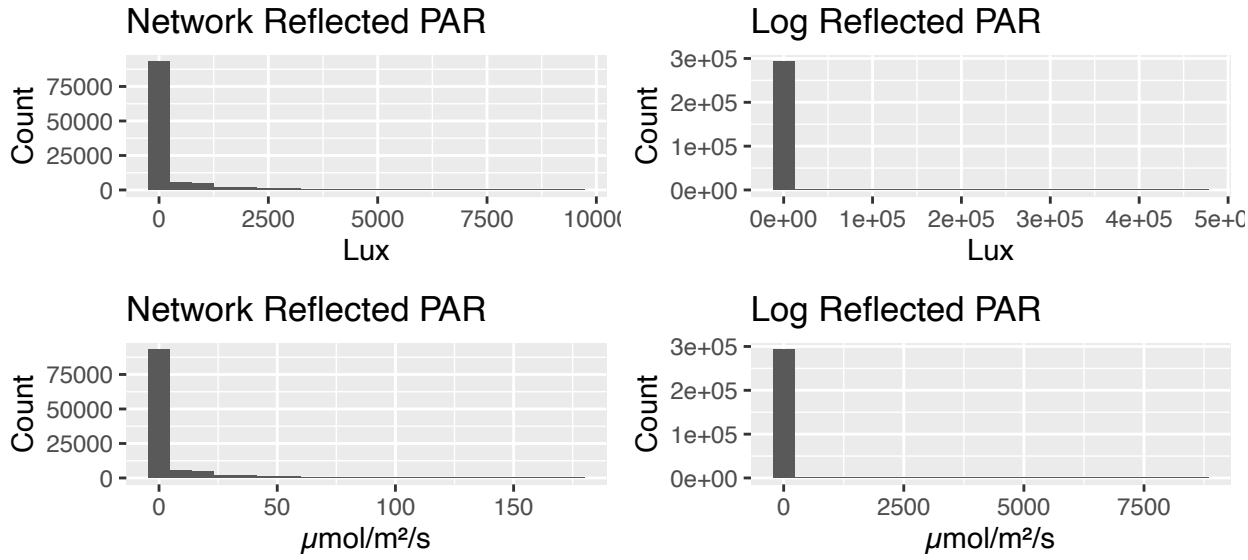
A.1: Voltage Before and After Conversion



A.2: Network and Log Incident PAR Unit Conversion



A.3: Network and Log Reflected PAR Unit Conversion



A.4: Missing Observations

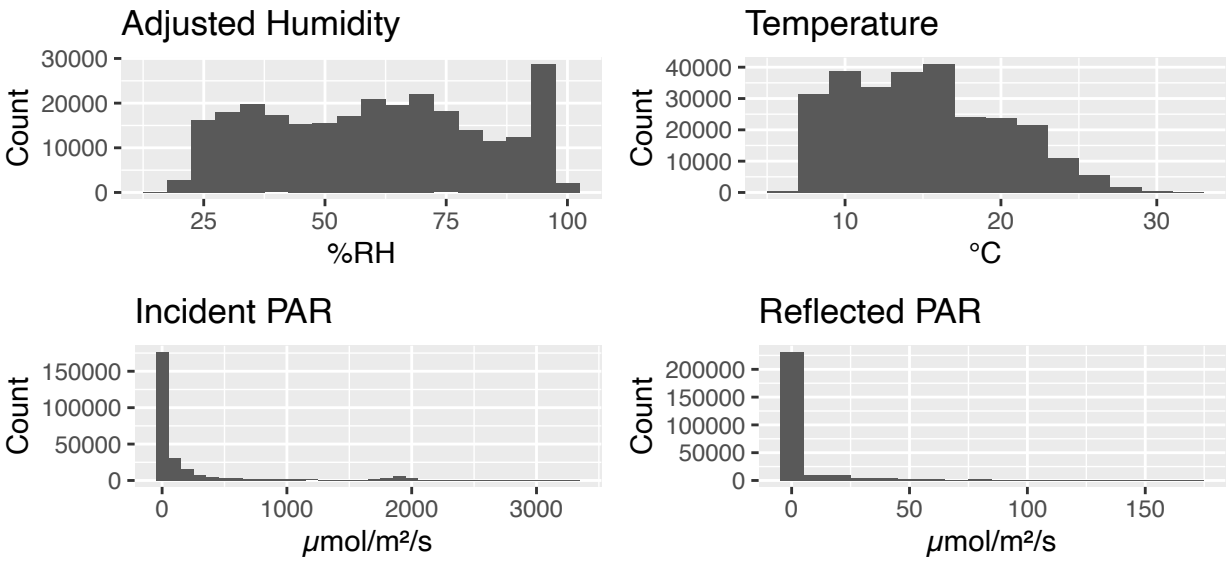
Table 1: 8760 Missing Observations Time Period

Start	End
2004-04-30 08:05:00	2004-05-25 21:15:00

A.5: Quantiles and Histograms After Removing Voltages

Table 2: Quantiles After Removing Outlier Voltages

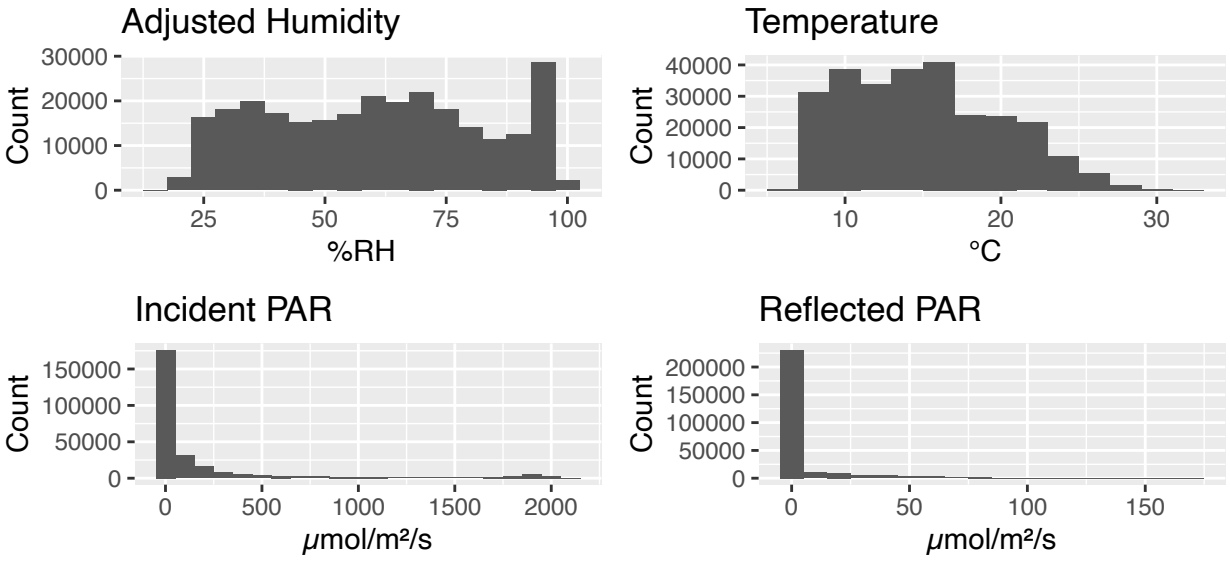
	Humidity	Adjusted Humidity	Temperature	Incident PAR	Reflected PAR
0%	16.265	16.228	6.582	0.000	0.000
25%	41.074	40.581	10.796	0.000	0.000
50%	62.453	60.781	14.579	0.000	0.000
75%	80.805	77.786	18.616	132.848	0.000
100%	104.405	100.223	32.581	3334.717	169.143



A.6: Quantiles and Histograms After Removing Node 40

Table 3: Quantiles After Removing Node 40

	Adjusted Humidity	Temperature	Incident PAR	Reflected PAR
0%	16.228	6.582	0.000	0.000
25%	40.598	10.796	0.000	0.000
50%	60.794	14.579	0.000	0.000
75%	77.794	18.607	132.218	0.000
100%	100.223	32.581	2068.670	169.143



A.7: Creating the Time Period Variable

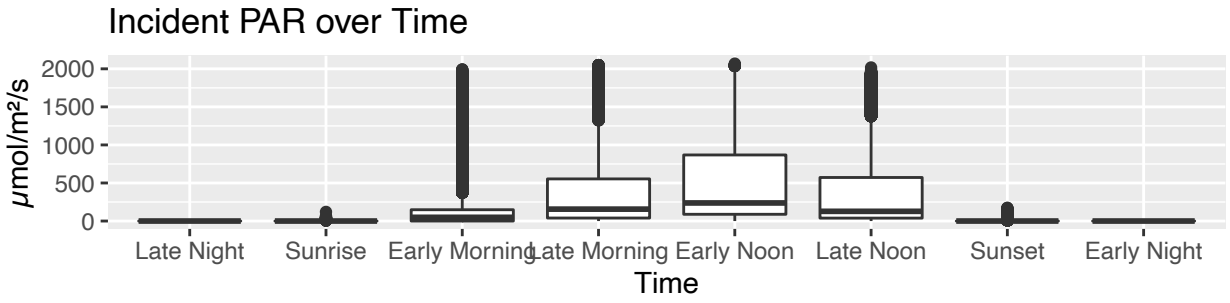
Since the data contains so many different time stamps per day, we wanted to look at bigger periods of time to see if there are any noticeable differences in humidity, temperature, incident PAR, and reflected PAR between the time periods. In order to do so, we wanted to split each day into different time periods based on the sunrise and sunset times. We web scraped the sunrise, sunset, civil twilight, and solar noon times of all 31 days from May, 2004 in Sonoma, California. Observations from May 1, 2004 to May 31, 2004 makes about 78.2% of our total data so working with only May data still accounts for most of our total cleaned data. After web scraping and merging the scraped sun data with our May data, we assigned each observation a time period out of 8 categories: sunrise, sunset, early morning, late morning, early noon, late noon, early night, and late night.

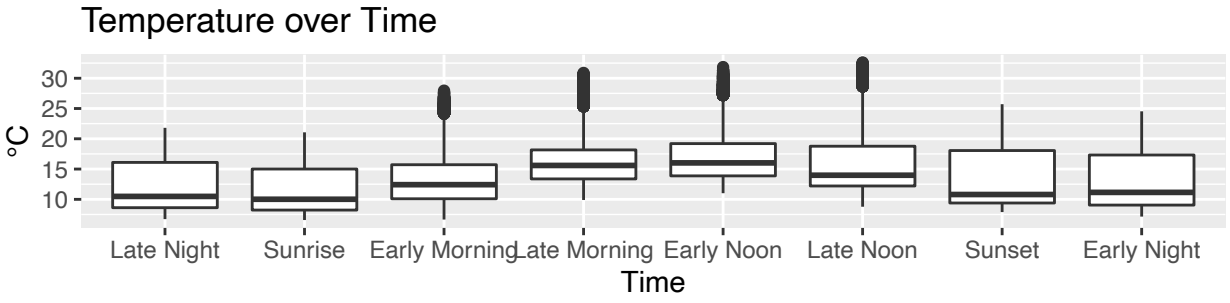
We assigned sunrise as the time period from the start of civil twilight to the number of minutes after sunrise that it took between the start of civil twilight to sunrise. For example, for May 1, civil twilight starts at 5:44 AM and sunrise occurs at 6:12 AM. Since it took 28 minutes between the start of civil twilight to sunrise, we made the end of sunrise 28 minutes after sunrise. So our sunrise time period for May 1 was 5:44 AM to 6:40 AM. Our sunrise time period varies by the start of civil twilight time and sunrise time for that day.

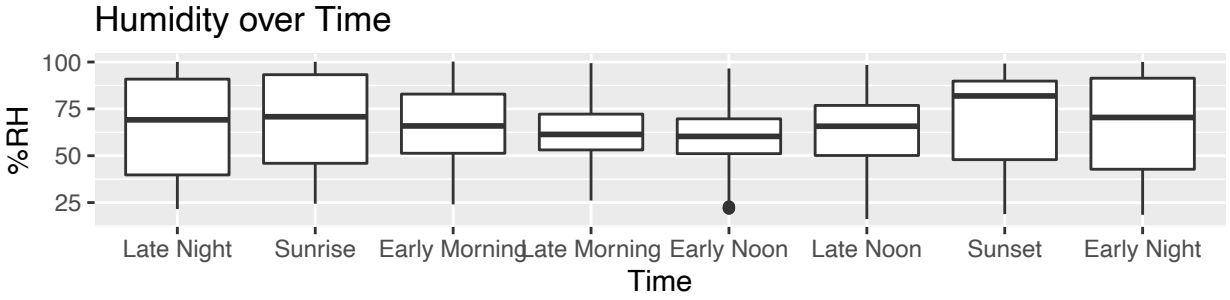
We assigned sunset as the time period from the number of minutes before sunset it took between sunset to the end of civil twilight to the end of civil twilight. For example, for May 1, civil twilight ends at 8:30 PM and sunset occurs at 8:01 PM. Since it took 29 minutes between sunset to the end of civil twilight, we made the start of sunset 29 minutes before sunset. So our sunset time period for May 1 was 7:32 PM to 8:30 PM. Our sunset time period varies by the end of civil twilight time and sunset time for that day.

The remaining time periods fill in quite naturally. Early morning is the time period from the end of sunrise as defined above to 10:00 AM. Late morning is the time period from 10:00 AM to the solar noon time of that day. Early noon is the time period from the solar noon time of that day to 4:30 PM. Late noon is the time period from 4:30 PM to the beginning of sunset time as defined above. Early night is the time period from the end of sunset time as defined above (end of civil twilight time) to 1:30 AM of the next day. Late night is the time period from 1:30 AM to the beginning of sunrise time as defined above (start of civil twilight time).

Looking at the box plots of incident PAR, temperature, and humidity over these time periods, shown below, revealed some general trends that are confirmed by the paper. Both incident PAR and temperature start low and gradually increase before peaking at noon and gradually decreasing which suggests that these two measurements follow the normal movement of the sun as confirmed by Tolle et al. Humidity on the other hand starts high and gradually decreases before hitting its lowest point at noon and gradually increasing which suggests that humidity moves inversely with temperature which Tolle et al. alludes to. We also see that variability is reduced once the sun has risen and humidity has decreased which is confirmed by the paper's finding that temporal variability is reduced once the sun has risen. Since our box plots over our newly defined time periods reveal similar trends to that of the paper, we are fairly confident in how we split up our observations over time periods. Next, we used these newly defined time periods with GMM clustering to reveal further trends.


Incident PAR over Time
µmol/m²/s
2000
1500
1000
500
0
Late Night
Sunrise
Early Morning
Late Morning
Early Noon
Late Noon
Sunset
Early Night
Time


Temperature over Time
°C
30
25
20
15
10
Late Night
Sunrise
Early Morning
Late Morning
Early Noon
Late Noon
Sunset
Early Night
Time


Humidity over Time
%RH
100
75
50
25
Late Night
Sunrise
Early Morning
Late Morning
Early Noon
Late Noon
Sunset
Early Night
Time