Lab 1 - Redwood Data, Stat 215A, Fall 2017

Hongxu Ma

8/31/2017

(PLEASE CHECK THE HTML VERSION FOR ANIMATION AND INTERACTIONS, THANK YOU SO MUCH!)

Introduction

In this report, I analysis the data of temperature, relative humidity, incident photosynthetically active solar radiation(PAR) and reflected PAR from two trees from the redwood forest. The data measured in both spatial and temporal scales with high resolution in both scales. I clean the data according to the physical background and criteria from the paper. I find that the data strongly support the climate system in west coast ecosystem in Northern hemisphere. Such as, the solar radiation is higher in south side than that in north side, and the humidity is higher in west side than that in east side. I also find that temperature is negatively associated with humidity, and most importantly, the ratio of temperature to humidity change along with time in hour scale. This is majorly affected by the air flow from inland to coast during the sunset time period. I find that the PAR data may have some error when calculating the ratio of incident PAR to reflected PAR. The ratios are mostly perfect interger which is very abnormal and unreal in the real world observation.

The Data

Data Collection

The data is collected via the wireless sensor network named macroscope by Tolle et al. They designed a divices collecting data on the environmental synamics around two 70 meters high redwood trees for 44 days. One tree is located on the edge of the redwood forest, while the other tree is located at the interior of the forest.

72 divices are implemented on two trees. The devices measured and recorded the temperature, relative humidity, incident photosynthetically active solar radiation(PAR) and reflected PAR. The data is collected from Apr. 27th to Jun. 10th. The data is collected every 5 minutes. The voltage of the devices of each epoch are also recorded.

Data Cleaning

1.Range cleaning

According to the Table 1 in the paper (Tolle et al, 2005), the rows with temperature less than 6.6 Celsius or greater than 32.6 Celsius are deleted. The rows with relative humidity that less than 16.4% and greater than 100.2% are deleted. The rows with negative value in Incident PAR or Reflected PAR are all discarded, as well as the rows with incident PAR larger than 2154 or reflected PAR larger than 2000. I didn't follow the criteria from the paper strictly.

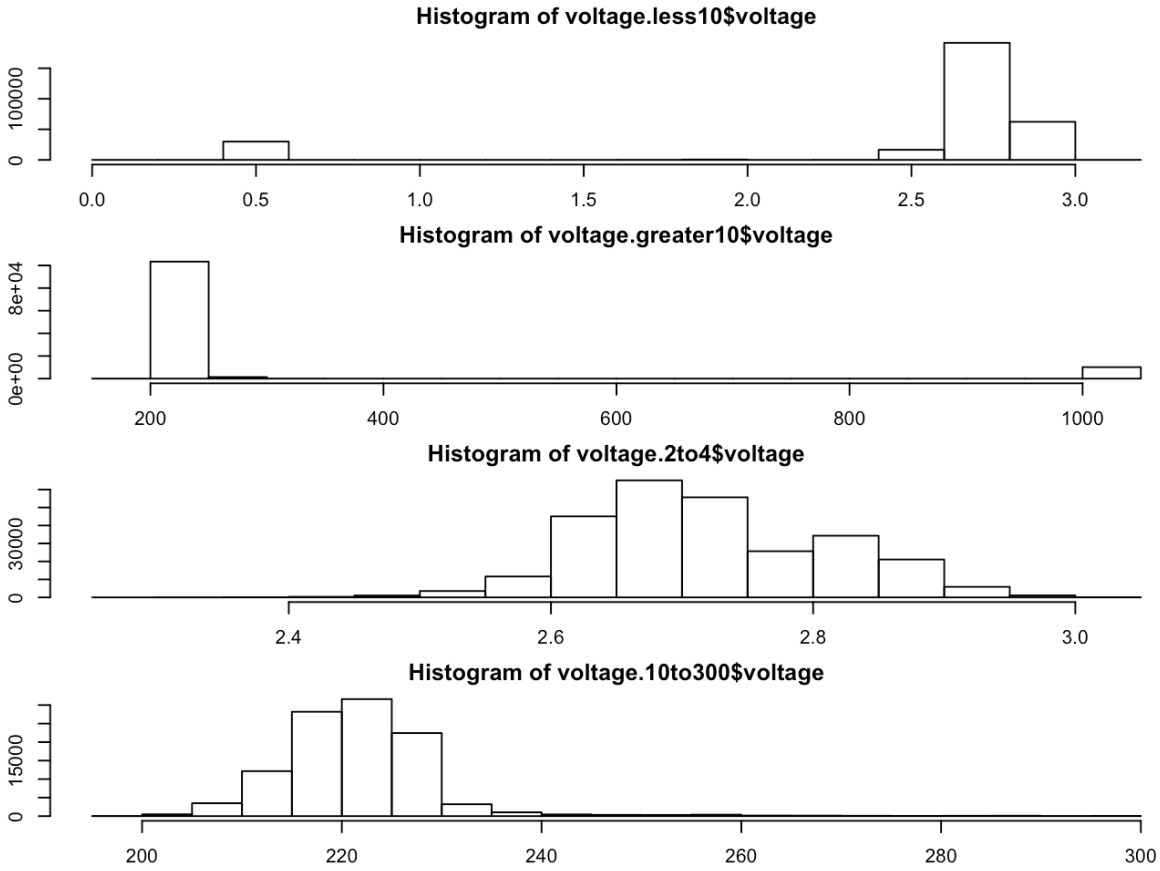
2.Voltage cleaning

According to the paper, the abnormal voltage cause the data measuring accuracy. They recommend to delete the rows with voltage less than 2.4 or greater than 3.0. Howerver, for most data collected for the network (denoted as redwood_net_orig) have voltage range around 220V. The criteria is not applicable to those data, which is about 25% of the whole data. So I divide the data according to the voltage range into two subset. One subset with voltage less than 10 will be trim by the criteria from the paper. The other subset with voltage greater than 10. From the figure below, we could easily eyeball there are

two outlier clusters located around 0.5 and 1000. After deleted the outliers, we compare the distribution of two subset by ploting the histograms. The historgrams shows that the criteria range from 2.6 to 3.0 actually coverred 97% data with voltage ranging from 2.4 to 3.0. To keep the cleaning criteria consistent between two subsets, we pick up 97% data from the data with voltage ranging from 200 to 300, which is roughly range from 203 to 237. So, we will delete the data with voltage sit outside the range (2.6, 3.0) and range(203, 237).

3.Time Stamps

The resulting time in the raw data doesn't recorded the real time when the variables are measured. We need to find the real time according to epoch that linking to number in sonoma-dates files.

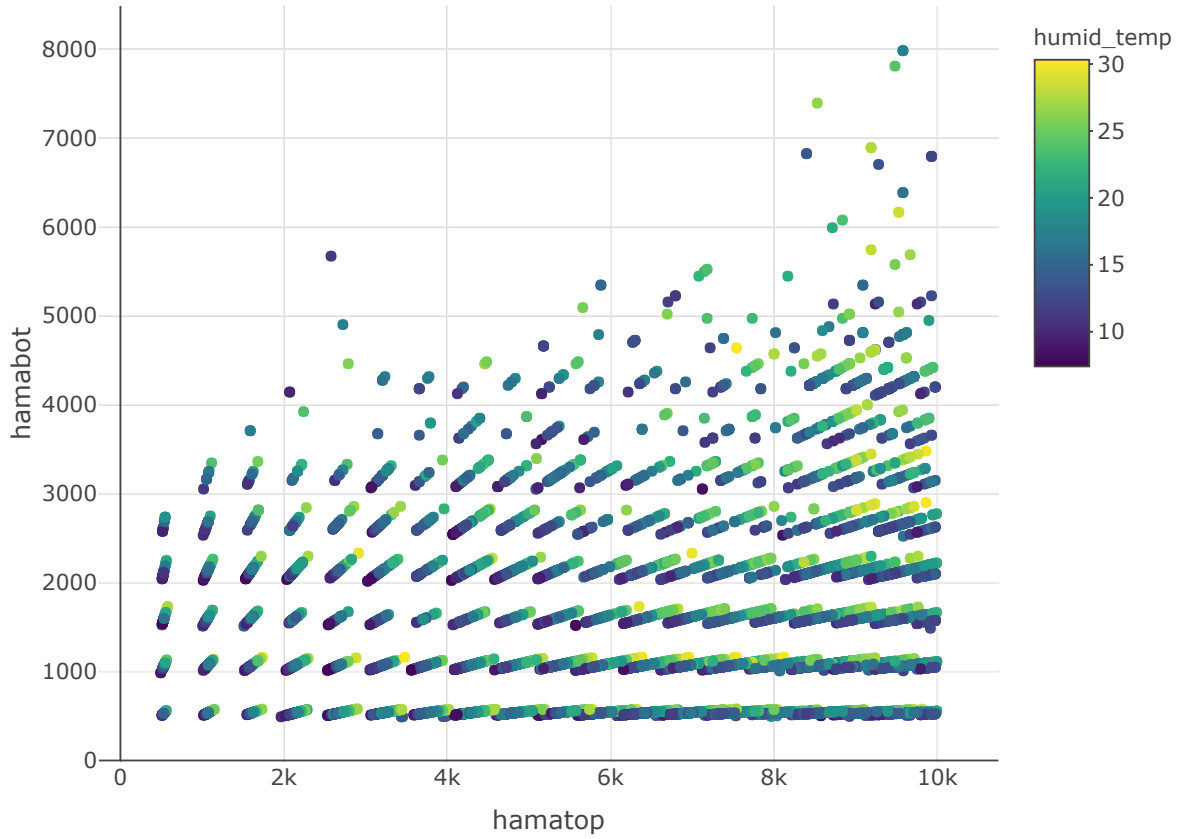


Data Exploration

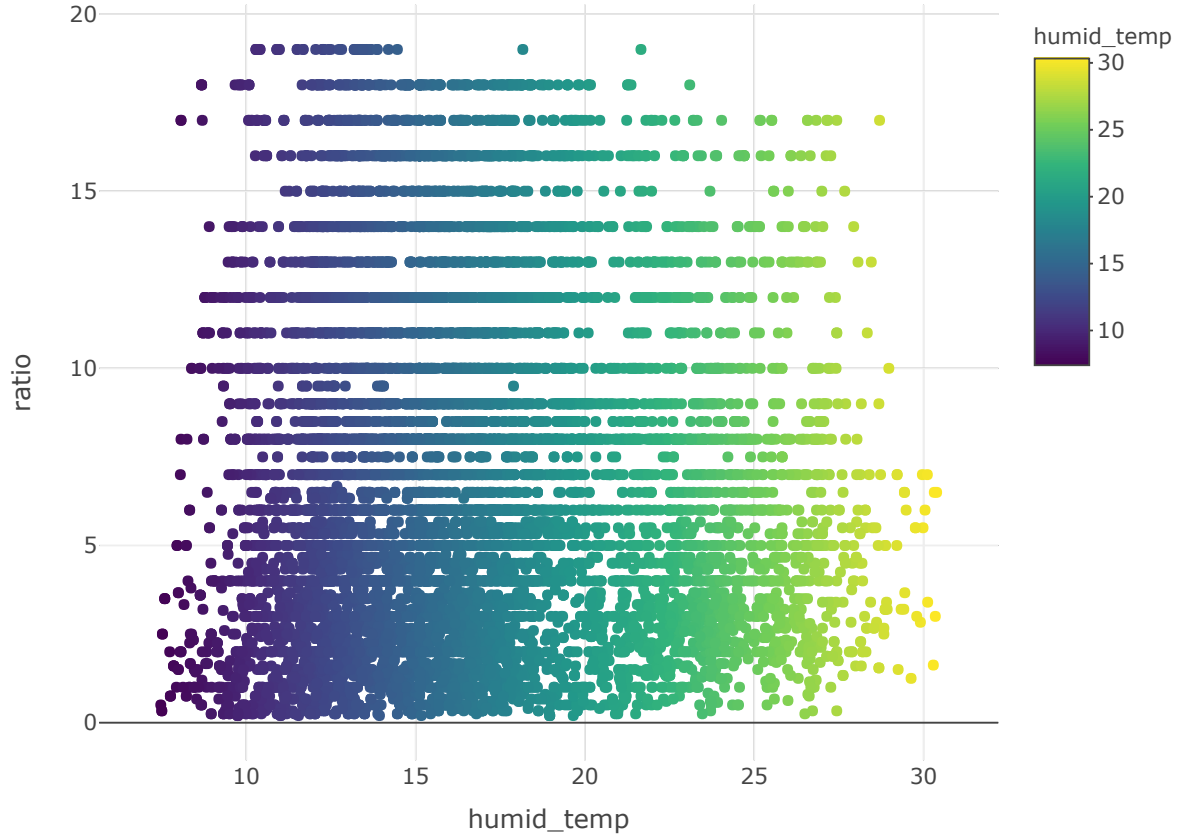
1. Is there any relationship between incident PAR and reflected PAR?

Based on the physical understanding, an increase in incident PAR will cause an increase in relfected PAR. Meanwhile, the ratio of the increasement of incident PAR to relfected PAR could tell us more about the local environment. The high ratio represents the high radiation scattering ability, which is an idication of the ability and effectiveness of photosynthesis (Sivak et al. 1985; Rao et al. 1986). The ratio also reflect the dust concentration in the air, which indicate one of the major fertilization way in coast ecosystem (Meskhidze et al.2005).

To better describe the correlation between incident PAR and reflected PAR, we only plot the data with non-zero ones.



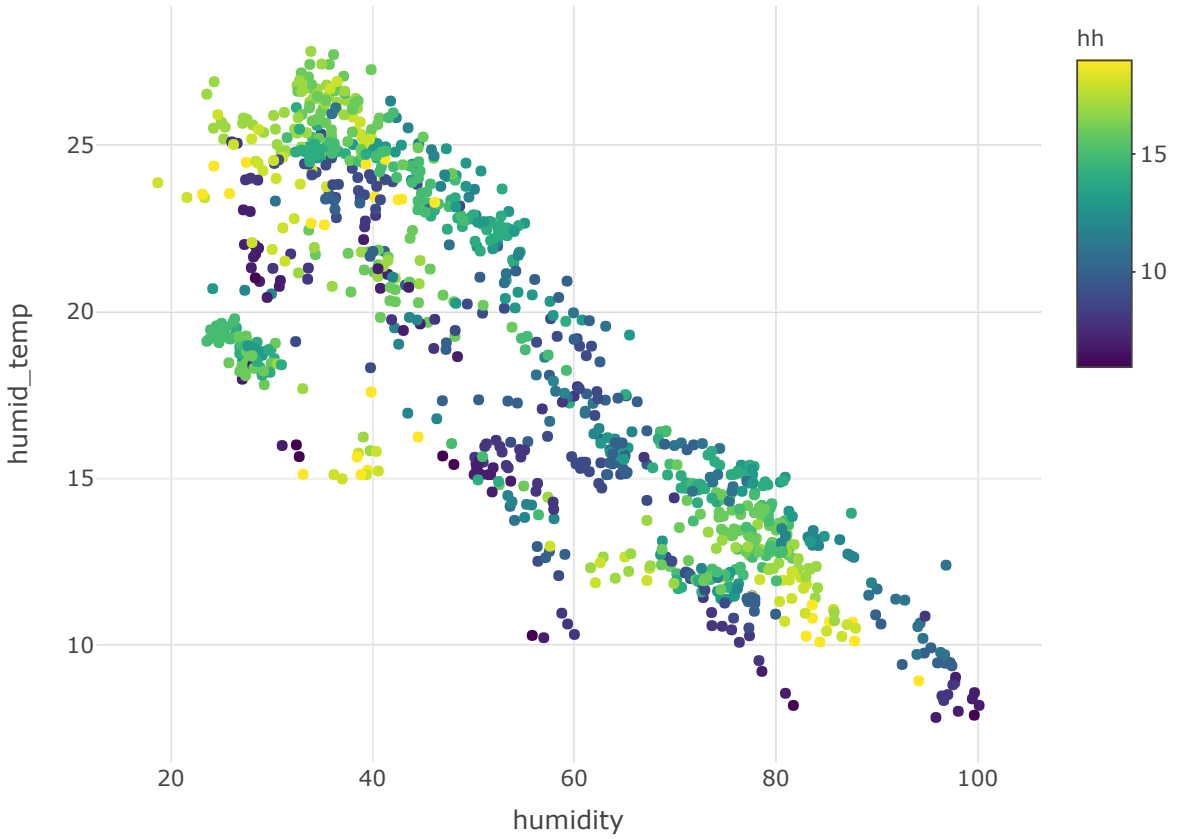
We could find a really interesting pattern between incident PAR and relfected PAR. The ratio of incident PAR to relfected PAR seem have some constant values. To exam if the constant relation is whether real or caused by errors, I plot all the data with PAR larger than 0, including the outliars we deleted at begining. To further check the ratio of incident PAR to relfected PAR.



By ploting the ratio againest temperature, we could confidently say there are constant ratio of incident PAR to reflected PAR. However, limited by variables collected, we are not able to answer what dominant the ratio. However, for the higher ratio that

above 6, the ratio constant numbers are all intergers. The perfect interger is very impossible exist in the real world, especially not only one by chance. This may refelect the devises errors.

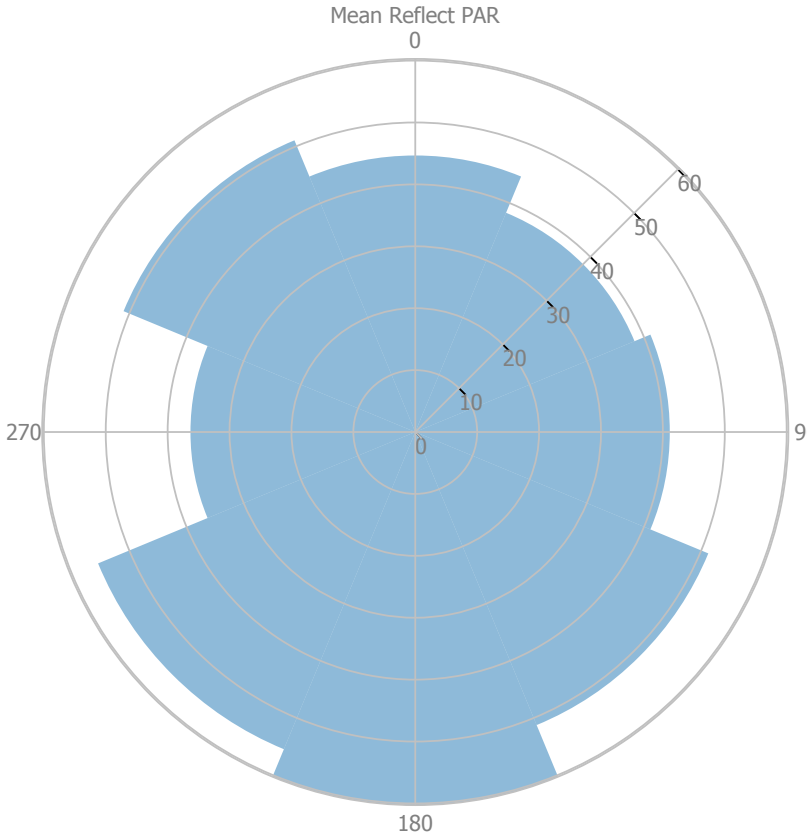
2.The relationship between temperature and humidity



The plots shows that humidity has a significant negative correlationship with temperature. There are two major clusters in this figure, one located at (temperature = 25, humidity = 40%), the other located at (temperature = 13, humidity = 80%). Interestingly, these two clusters are all formed by the data measured around 5 - 7 p.m., which is the sunset period. More details and discussions see Finding #1 section and Discussion #1 section.

3. How direction of the motes affect the data

As the redwood forest located near the ocean, the west side of the tree usually affect by the ocean fog more than the east side. Meanwhile, the instant radiation usually higher in south side than that in north side in Northern hemisphere.



The figure also suggest that south side received much more radiation than the north side. The difference between west side and east side is not significant.

Graphical Critique

For figure 3(a), they try to describe what the distribution and the range of each variables. For incident PAR and Reflected PAR, I think it could be better to narrow the bins to support their arugment that PARs are bimodal distributed. For figure 3(b), they are trying to describe how the data change along with time. With a line delineate the trends will helps a lot. Meanwhile, the figures only describe how the data change along with time in day scale. It is also important to know how the data change in hour scale. Check finding 2 section for further details and discussion. For figure 3(c), they are trying to describe how data change along with height. To answer the question for PAR, it is also know how the data change for non-zero data. Because the zero in PAR mostly measured during the night, which doesn't help with understanding the photosysthesis process. People will be curious how the data change in day time. For firgure 3(d), they are trying to describe how the distribution of measurements away from the mean change along with height. But people will also curious that distribution of PAR looks like in day time with non-zero data. For firgure 4. The trends of temperature and humidity is intuitive and full of information. But the trends of PAR doesn't tell much. It would be great to describe the trends of PAR by using data from a short time range instead of just a single time point.

Findings

First finding

The ratio of temperature to humidity around the sunset time period is highly donimated by inland air supply.

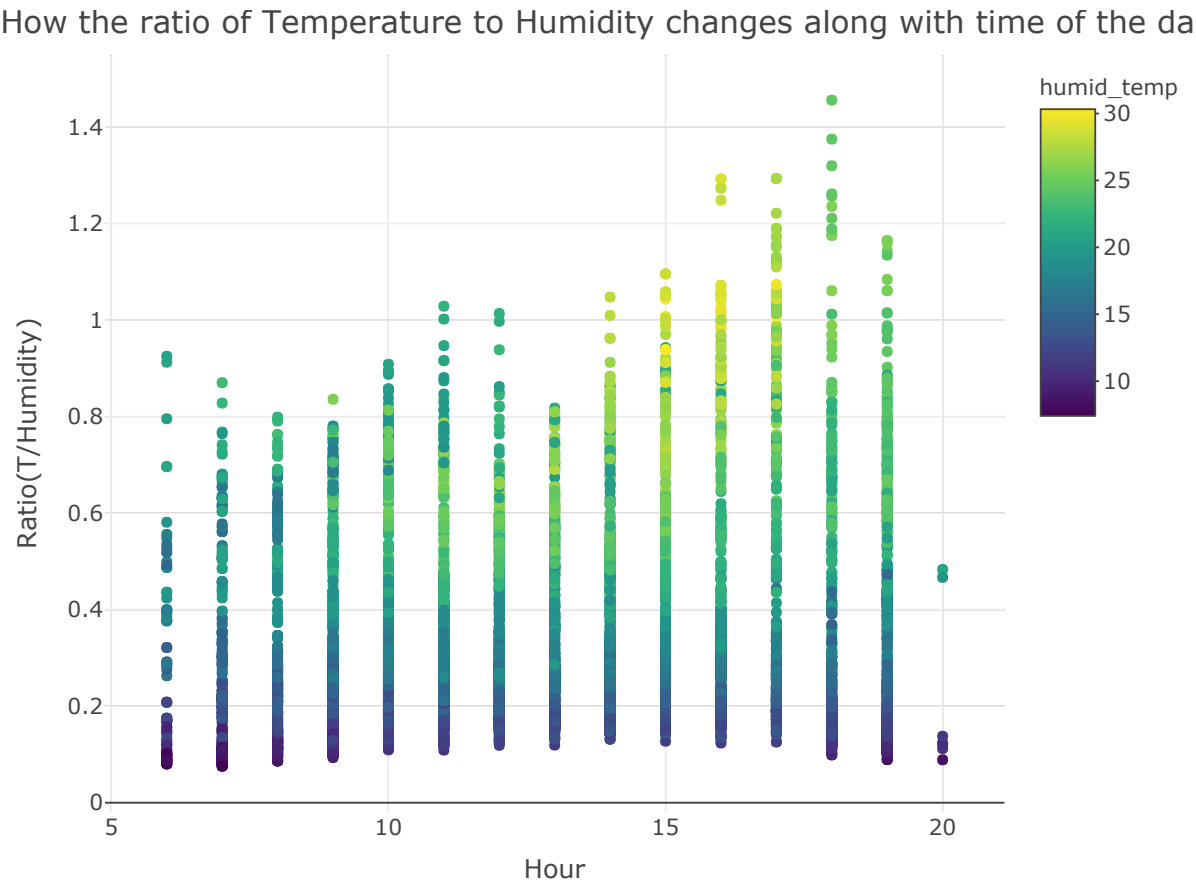
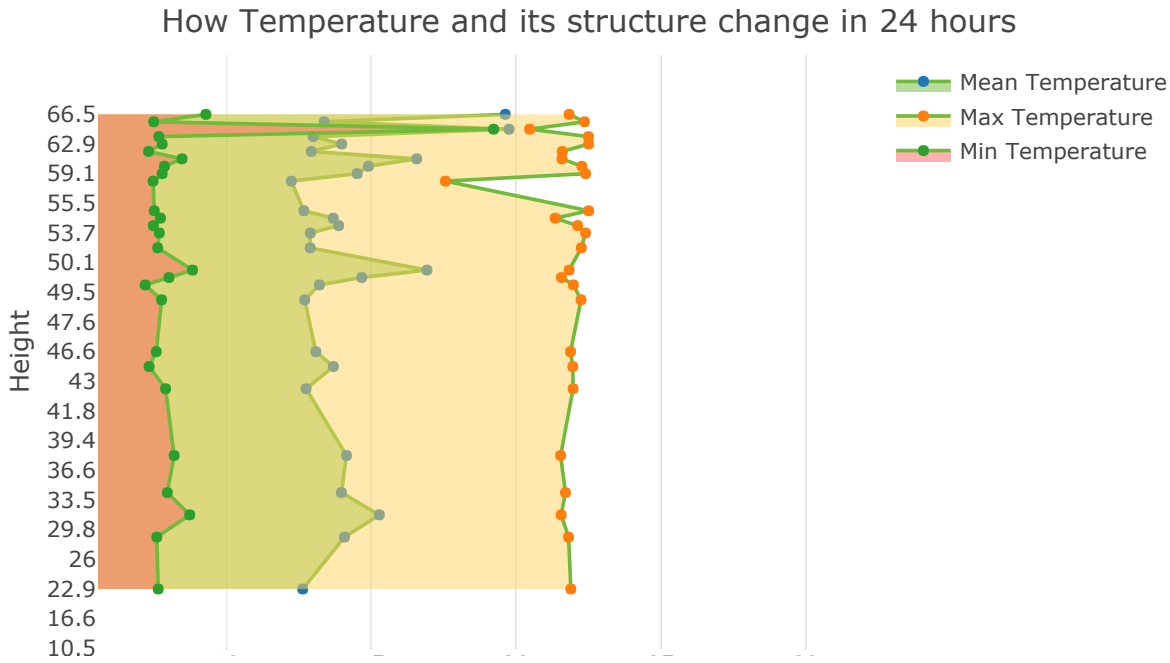


Figure Finding #1

This figure describe how the ratio of temperature to humidity changes along with time of the day. The x-axis is the hour of the day ranging from 0 to 24. The y axis is the ratio. The color represent the temperature. Further details see discussion section.

Second finding

How temperatue change along with height and time in hour scale



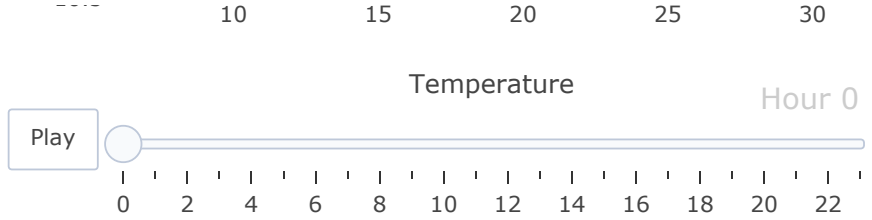


Figure Finding #2 This figure describe how the temperature and its spatial structure change along with time in hour scale. The x-axis is temperature and the y-axis is height of the motes. Please check the html file for interactings. You can click “Play” button to see the animation. You may also slide the bar on the bottom to check the hour you are interested in. Further details see discussion #2

Third finding

How humidity change along with direction and location of the trees.

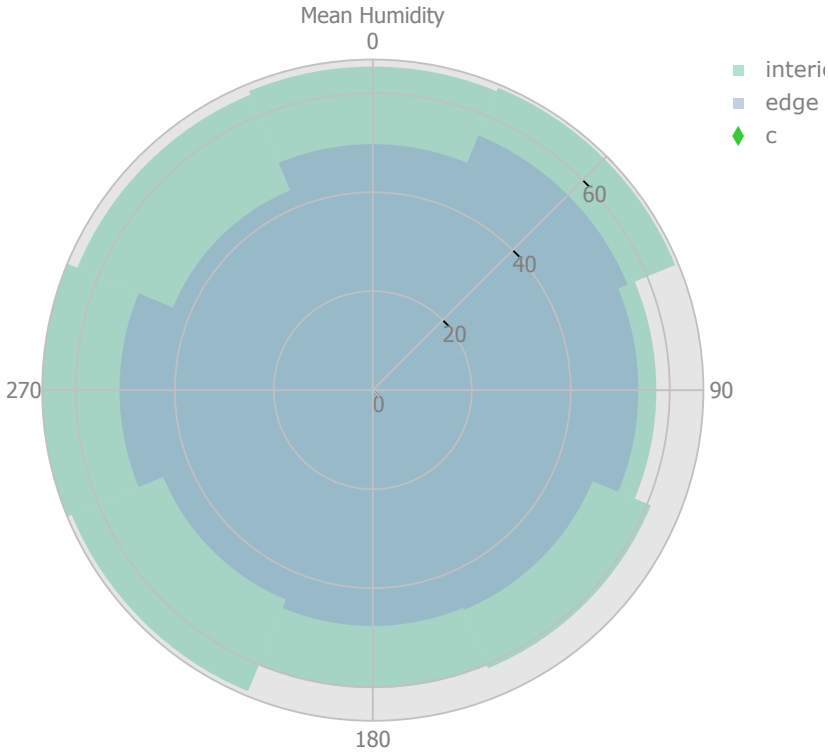


Figure Finding #3 This figure shows how humidity change along with direction. The number on the edge of the circle represents the degree different to North in clockwise direction. So the 0 represents North and 90 represents East. The light green region represents the interior tree and the light blue region is the data from tree located at edge of the forest.

Discussion

Discussion of Finding 1

By comparing the ratio of temperature to humidity against to the hour when the data is measured. We find that the ratio has more variation during the sunset period. This is because when the radiation from the sun decrease, the ocean has higher specific heat capacity than the land. So the land will loss heat much more quickly than the ocean. The air pressure on the ocean part will be lower than the land part because of convection process. The air pressure difference will attract the air from the inner land flow to the forest region. With more uncertainty influence introduced from the outer system, the more uncertainty of correlation between humidity and temperature, which will increase the variation of the ratio.

Discussion of Finding 2

By looking into how the temperature and temperature spatial structure change along with time in hour scale, we find that during the night time, there is not temperature difference between the top and the bottom. But during the day time, the temperature is relatively higher than that on the bottom.

Discussion of Finding 3

By ploting the humidity in direction, we find that the west side of the both trees has higher mean humidity than that in east side. Which is majorly caused by the ocean fog comes from the west side. Meanwhile, the process discussed in #1 section also partially explain this difference, as most of the air bring from the inland are relatively dry. The interior tree has significantly higher humidity than the tree located in the edge. Even the paper didn't tell us the exact location of the edge tree, whether it is on the west edge or east edge. We could make a confident guess that the edge tree is located at the east side of the forest by analysing this figure.

Conclusion

The data measure the temperatue, relative humidity, incident PAR and reflected PAR every 5 minutes for 44 days with 72 motes located at different height in two trees. The data have relatively high spatial and temporal resolution. It offered a great oppotunities for the ecologist and data scientist to analysis the spatial and temporal structure in redwood forest. We find that 1) the ratio of temperature to humidity increase during the sunset time period. It is majorly caused by the air pressure difference between ocean and land during the sunset period. 2)The ratio of incident PAR to reflect PAR seems unreasonable and unreal. It may exists some errors in the devices. 3)The temperature spatial structure change along with time in hour scale. It is relatively flat during the night time. 4)The PAR is relatively higher in south side than that in north side, which is consistant to the physical background. 5)The humidity on the west side is relatively higher than that in east side, which is majorly caused by ocean fog effect. The pressure difference will attract air flow from ocean to land during the daytime which bring air with high humidity, and the pressure difference during the sunset period will attract air flow from inland to ocean, which bring air with low humidity.

Bibliography

Sivak M N, Heber U, Walker D A. Chlorophyll a fluorescence and light-scattering kinetics displayed by leaves during induction of photosynthesis[J]. Planta, 1985, 163(3): 419-423. Rao I M, Abadia J, Terry N. Leaf phosphate status and photosynthesis in vivo: changes in light scattering and chlorophyll fluorescence during photosynthetic induction in sugar beet leaves[J]. Plant science, 1986, 44(2): 133-137. Meskhidze N, Chameides W L, Nenes A. Dust and pollution: a recipe for enhanced ocean fertilization?[J]. Journal of Geophysical Research: Atmospheres, 2005, 110(D3).