GEO1001 Assignment 1

Lars Langhorst

September 20, 2020

1 A1

1.1 Mean, Variance, Standard Deviation

From the given data set (Maiullari & Sanchez, 2020), the mean statistics were computed as can be seen in the tables below, and from these statistics certain observations can be made. Starting by taking a look at the wind direction values, it is clear that the mean values for B and C are from a similar 'lower' direction, whereas E contains a higher value, differing around 40 degrees. The variance for these values also fluctuates, where C and D are significantly lower than the rest. The same applies for the standard deviation.
The Wind speed for sensor E is much lower than the other sensor, where sensor D reads the highest values. These two are approximately 1 m/s apart, which is significant considering the highest mean amongst the sensors is 1.58 m/s. The variance and standard deviation of the wind speed data also result in the lowest value for sensor E, and the highest for D. this could be due to the fact that lower wind speeds also have fluctuations that aren't as high as those on places with higher wind speeds. The same observations can be made looking at the crosswind speed data.
A variable used a lot in this report is the Temperature. The temperature at sensor E is slightly higher than the other values, which are all around the 18 degrees Celsius. The variance of the temperature values of sensor E are also significantly higher than the rest; all around the same value of 16.
Interesting enough, the mean of the globe temperature for sensor E is lower than the rest of the sensors, and highest for sensor B.
The relative Humidity for sensor E is lowest, and sensors A, B, C, and D are all relatively equal around 78, with sensor A reading the highest value. The variance for this variable is highest at sensors B and E, and much lower at A and C, leaving D in the middle.

	mean A	mean B	mean C	mean D	mean E
Direction , True	209.4518998	183.458367	183.5889248	198.3265966	223.9466451
Wind Speed	1.289894907	1.241552142	1.371463217	1.581649151	0.596483428
Crosswind Speed	0.965076799	0.835044462	0.963298302	1.210509297	0.438682296
Headwind Speed	0.164955538	-0.128981407	-0.262894099	-0.300565885	0.195028294
Temperature	17.96932094	18.06544058	17.91313662	17.99636217	18.35379951
Globe Temperature	21.54539208	21.80024252	21.58738884	21.35929669	21.17647534
Wind Chill	17.83835893	17.94587712	17.77299919	17.83536783	18.29389652
Relative Humidity	78.17215036	77.86835085	77.96285368	77.94203719	76.78799515
Heat Stress Index	17.89943411	18.00400162	17.82825384	17.9216249	18.28617623
Dew Point	13.55133387	13.52845594	13.45812449	13.50860954	13.55743735
Psychro Wet Bulb Temperature	15.2692401	15.29409863	15.19664511	15.26018593	15.40582053
Station Pressure	1016.168068	1016.657033	1016.689329	1016.728011	1016.166006
Barometric Pressure	1016.128254	1016.616492	1016.6519	1016.688884	1016.127688
Altitude	-25.98504446	-30.05820534	-30.33872272	-30.65319321	-25.96038804
Density Altitude	137.3148747	135.5699272	129.6228779	132.4110752	150.8306386
NA Wet Bulb Temperature	15.98080032	15.99611964	15.93423605	15.91564268	15.93629749
WBGT	17.2539612	17.32166532	17.22502021	17.17679871	17.18516572
TWL	301.3778901	299.4350849	301.8997575	305.2545675	284.1330639
Direction , Mag	208.9502829	183.2631366	183.0836702	197.8261924	223.8872272

Figure 1: Data means

	variance A	variance B	variance C	variance D	variance E
Direction , True	10114.50092	9982.664865	7703.363096	8133.890057	9311.815187
Wind Speed	1.251947987	1.301903938	1.430920058	1.73981677	0.511289691
Crosswind Speed	0.927312024	0.878718821	1.042574802	1.451502935	0.315991955
Headwind Speed	1.033227515	1.256794191	1.271732179	1.232502712	0.319186756
Temperature	15.87703254	16.64246469	16.10453824	16.10559129	19.05078418
Globe Temperature	68.24565001	66.10191664	67.9413047	61.2022528	63.24082122
Wind Chill	16.27755347	17.04952808	16.54112266	16.55685213	19.14476257
Relative Humidity	376.1128796	408.8273562	374.622643	389.8560405	406.5955574
Heat Stress Index	15.00894428	15.45152866	15.35625356	15.11764378	18.48255852
Dew Point	9.723251354	9.637168096	10.08414949	10.07188298	9.42187956
Psychro Wet Bulb Temperature	6.946926474	6.773244577	7.239313447	7.044402877	6.998502299
Station Pressure	38.50232809	36.87172771	37.69149142	34.98778359	38.95563734
Barometric Pressure	38.49900326	36.85864422	37.67562316	34.95232686	38.95089097
Altitude	2665.78991	2547.766121	2608.534634	2419.723591	2693.440404
Density Altitude	26531.46369	26884.82426	26986.60297	26516.12573	29726.72626
NA Wet Bulb Temperature	10.01948969	9.816580165	10.4802791	9.98743414	9.435131615
WBGT	16.14813837	15.84803831	16.54674535	15.5071849	15.49579684
TWL	815.1452651	790.3664273	766.5335139	616.0098073	1289.654822
Direction , Mag	10111.28058	9980.910513	7704.62017	8135.315513	9271.539643

Figure 2: Data variances

	std A	std B	std C	std D	std E
Direction , True	100.5708751	99.91328673	87.7688048	90.18808157	96.49774706
Wind Speed	1.118904816	1.141010052	1.196210708	1.319021141	0.715045237
Crosswind Speed	0.962970417	0.937400033	1.021065523	1.204783356	0.562131617
Headwind Speed	1.016477995	1.121068326	1.127711035	1.110181387	0.564966155
Temperature	3.98459942	4.079517703	4.013046005	4.013177206	4.364720401
Globe Temperature	8.261092543	8.130308521	8.242651558	7.82318687	7.952409774
Wind Chill	4.034545014	4.12910742	4.067077902	4.069011198	4.37547284
Relative Humidity	19.39362987	20.21947962	19.35517096	19.74477249	20.16421477
Heat Stress Index	3.874137876	3.930843251	3.918705598	3.88814143	4.299134625
Dew Point	3.118212846	3.104378858	3.17555499	3.173623006	3.069508032
Psychro Wet Bulb Temperature	2.635702273	2.602545788	2.690597229	2.654129401	2.645468257
Station Pressure	6.205024423	6.072209459	6.139339657	5.915047218	6.241445132
Barometric Pressure	6.204756503	6.071132037	6.138047178	5.912049294	6.241064891
Altitude	51.63128809	50.47540115	51.07381554	49.19068602	51.8983661
Density Altitude	162.8848173	163.9659241	164.2759963	162.8377282	172.4144027
NA Wet Bulb Temperature	3.165357751	3.133142219	3.237325918	3.160290199	3.071665935
WBGT	4.018474632	3.980959471	4.067769088	3.937916314	3.936470099
TWL	28.55074894	28.11345634	27.68634165	24.81954486	35.9117644
Direction , Mag	100.5548636	99.90450697	87.77596579	90.19598391	96.28883447

Figure 3: Data standard deviations

1.2 Histogram

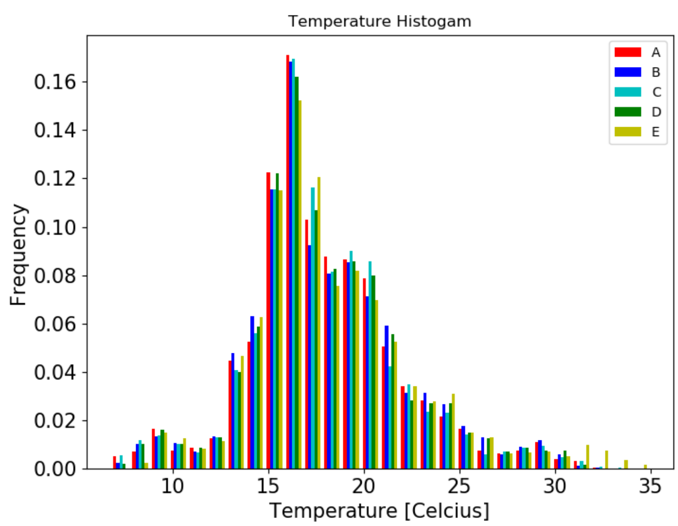
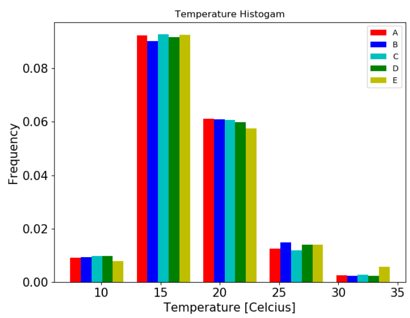
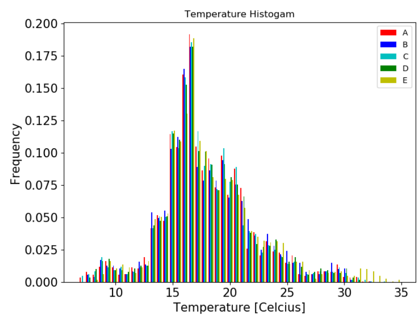


Figure 4: Histogram with the Temperature readings for all sensors

When plotting the histogram for a data set, the number of bins used to show the data is crucial to get a complete overview of the data. If the number of bins is too small, it can hide important information, small peaks in the data for example. On the other hand, if the bins are too small, it will result in a lot of noise in the plot, and loose a good overview of the data. Without this clear overview, important information can also be missed. To choose the number of bins, the Rice's rule of thumb can be used. For 2746 data points, this results in a bin number of 28, which can be seen in figure 4. To compare this clear result to the same data set with too many and too few bins, in the figures below, the data is shown with 5 and 50 bins.



(a) histogram with 5 bins



(b) histogram with 50 bins

1.3 Frequency Polygons

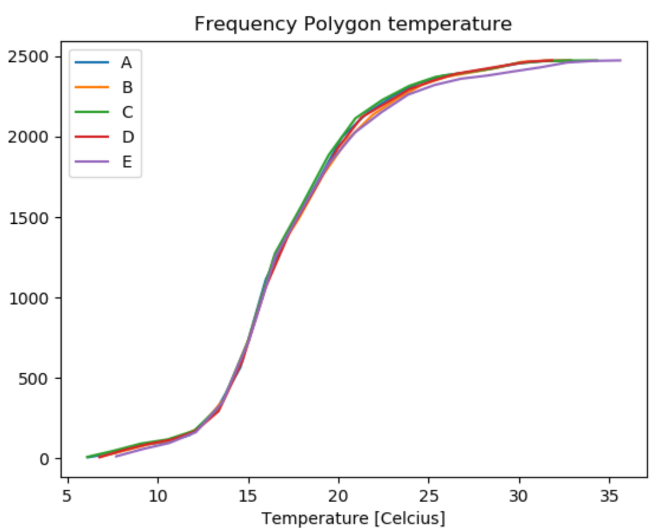


Figure 6: Frequency Polygons for Temperature values of all sensors

1.4 Boxplots

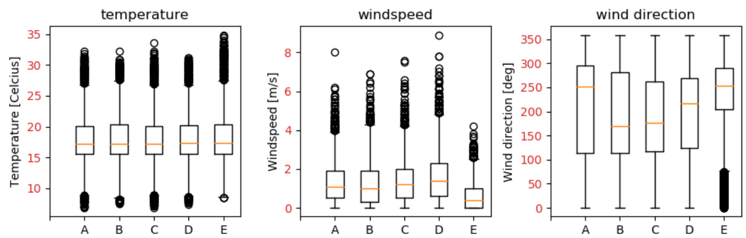


Figure 7: Boxplots for all sensors, for the Temperature, windspeed, and wind directions data

2 A2

2.1 Temperature PMF, PDF, CDF

NOTE: The PMF plot layout is improper, this is an error in the python script which could not be found, however the data representation itself is correct.

In the graphs below, the PDF, PMF, and CDF for the temperature values of all sensors are shown. The distributions of the data points for all sensors are relatively similar, however, some differences can be seen. Sensor B has much steeper peaks at both tails of the graph. Sensors C and D contain peaks at the tails of the graph at the low end of the x axis, and are both quite flat on the higher-end tail. All graphs have 2 main peaks, one around 16 and one around 20 degrees Celsius. What must also be noticed, is that for sensor E, the volume of the data seems to be leaner towards the lower side of the temperature axis, in general, the graphs mainly contain a positive skew starting around 16 degrees Celsius.
From the CDF plots, its visible that sensors B and C have less steep inclines, meaning the data is more spread out amongst the temperature around the temperature peaks. For all sensors, it is clear that there's a very steep peak around the 15 degrees Celsius. This is most profound for sensor E.

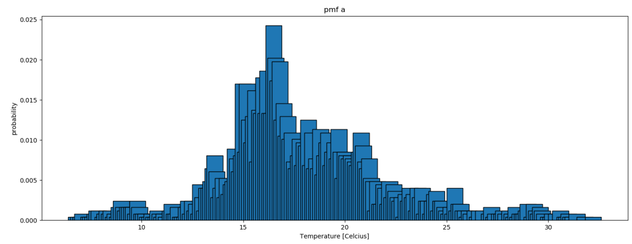


Figure 8: PMF A

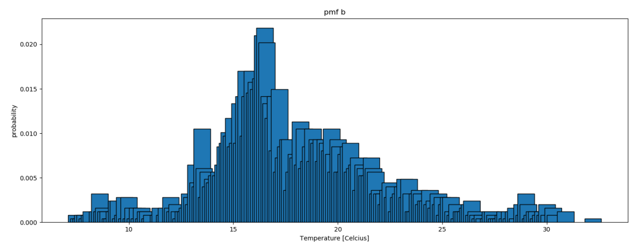


Figure 9: PMF B

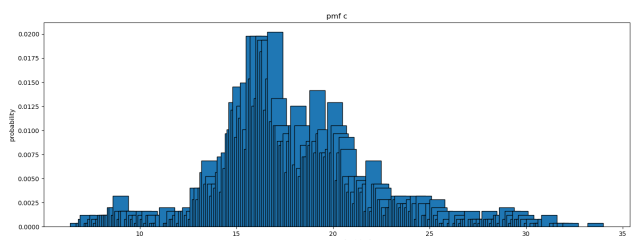


Figure 10: PMF C

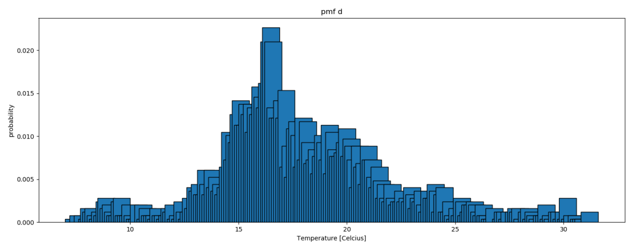


Figure 11: PMF D

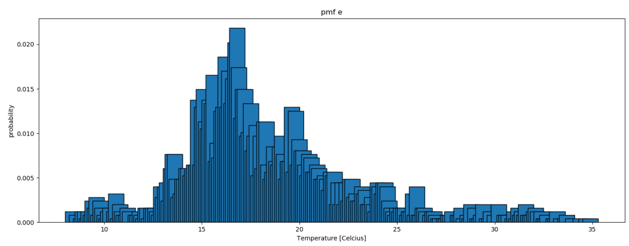


Figure 12: PMF E

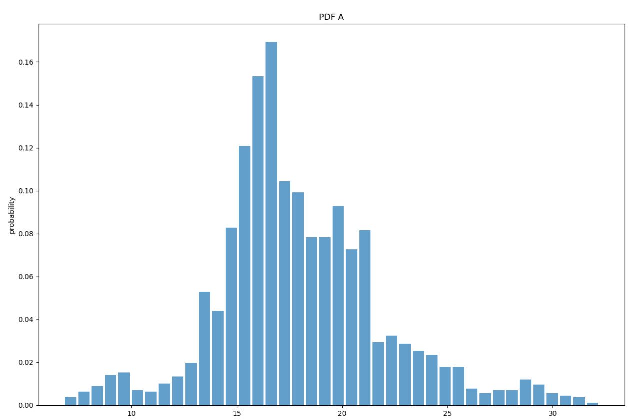


Figure 13: PDF A

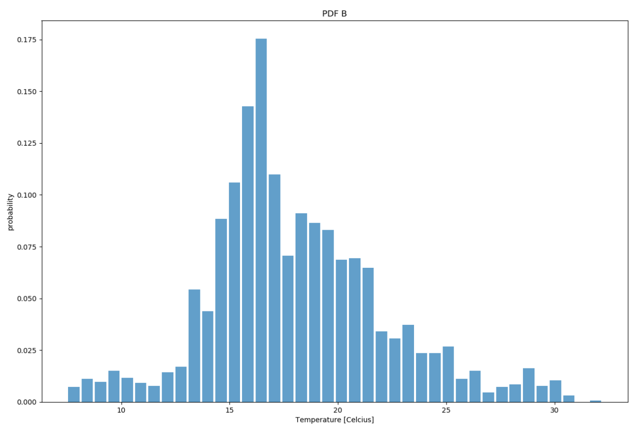


Figure 14: PDF B

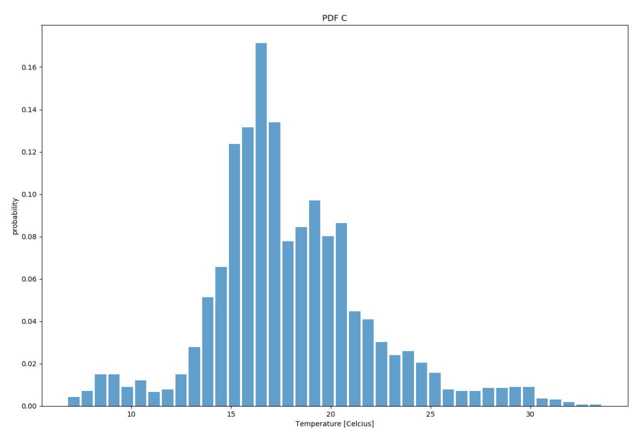


Figure 15: PDF C

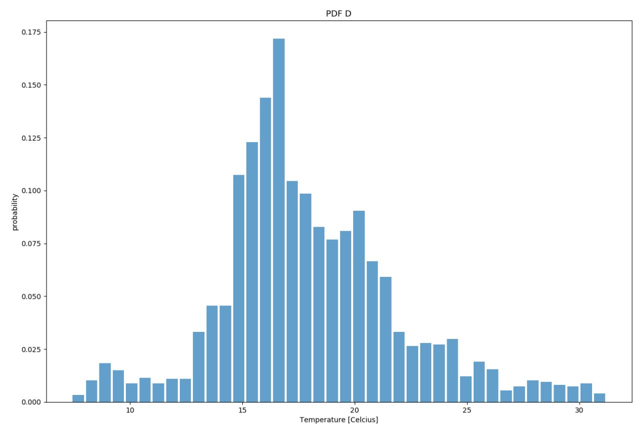


Figure 16: PDF D

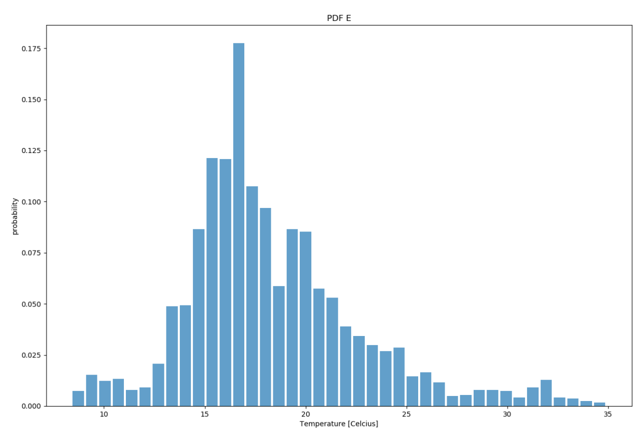


Figure 17: PDF E

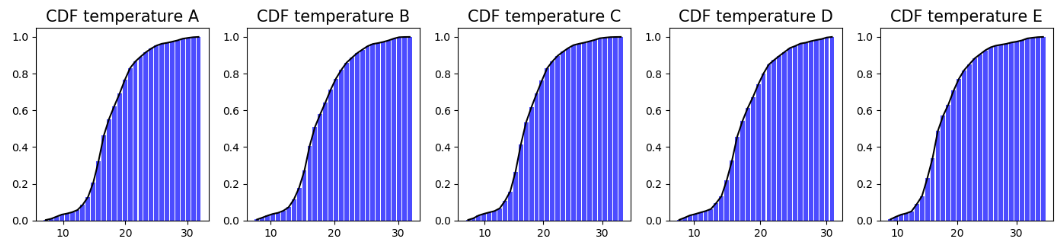


Figure 18: CDF graphs for A, B, C, D, and E

2.2 Windspeed PDF and kernel density

The Kernel density lines for all sensors are shown in the graphs below. The kernel density for the wind speed data has certain similar characteristics for all sensors; all graphs are positively skewed, and contain 2 main peaks. One peak is on or close to zero. The other peak is closer to 1 m/s for most sensors. What must also be pointed out is that the skew for A, B, and C is much steeper than D and E. This is because the wind speed at sensor D has a peak around a density of 0.3, and skews out until 5 to 6 m/s. Sensor E has much lower wind speeds in general, and thus a less steep skew.

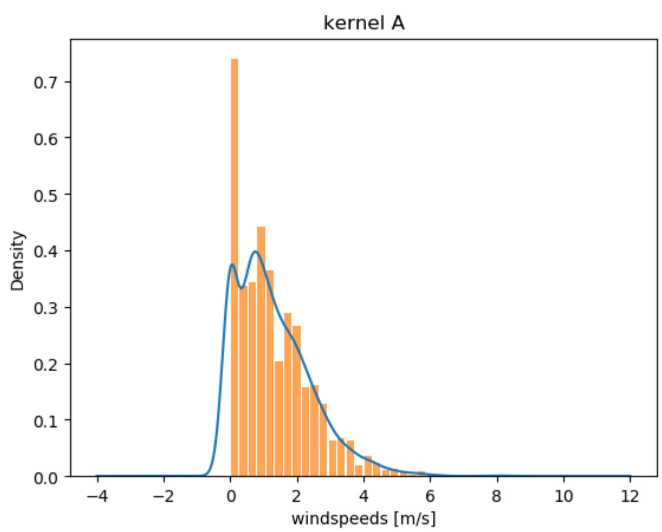


Figure 19: Kernel density plot with PDF for sensor A

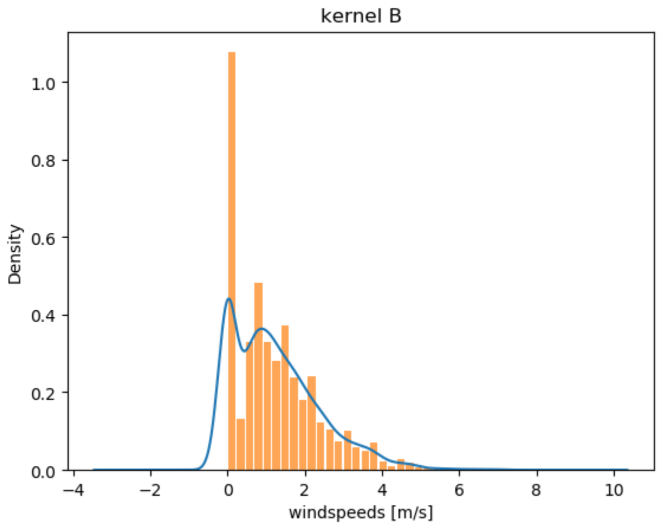
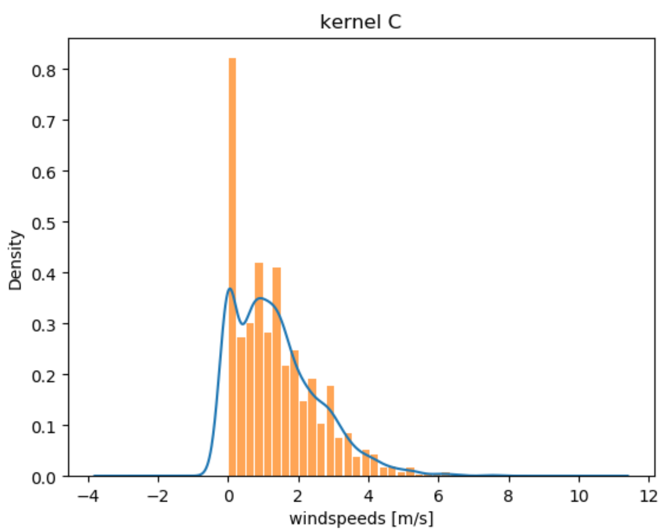


Figure 20: Kernel density plot with PDF for sensor B



n

Figure 21: Kernel density plot with PDF for sensor C

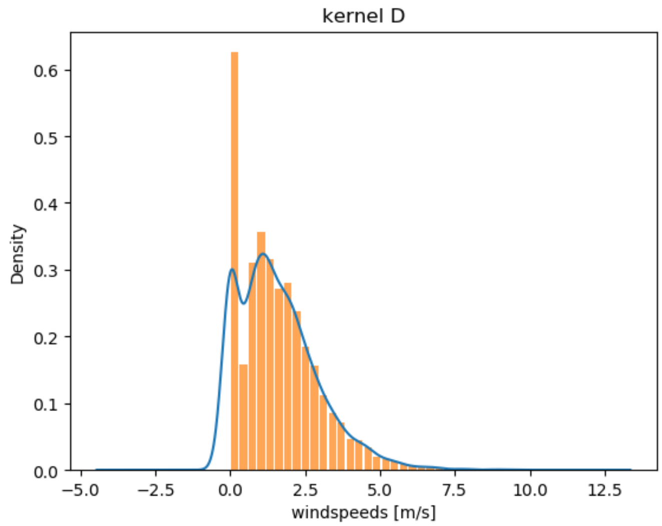


Figure 22: Kernel density plot with PDF for sensor D

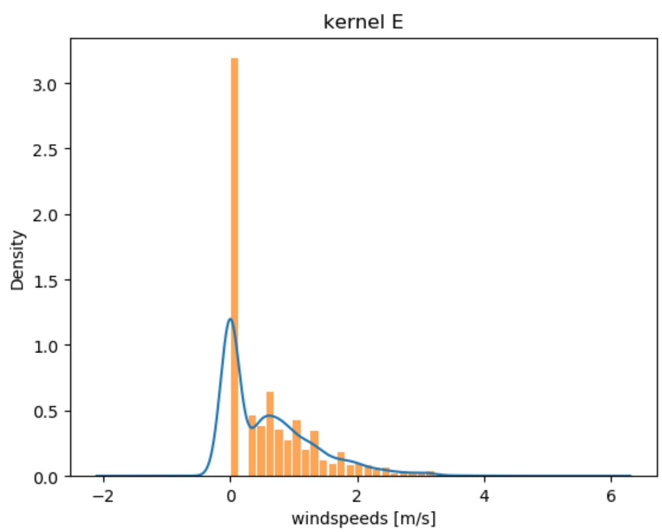
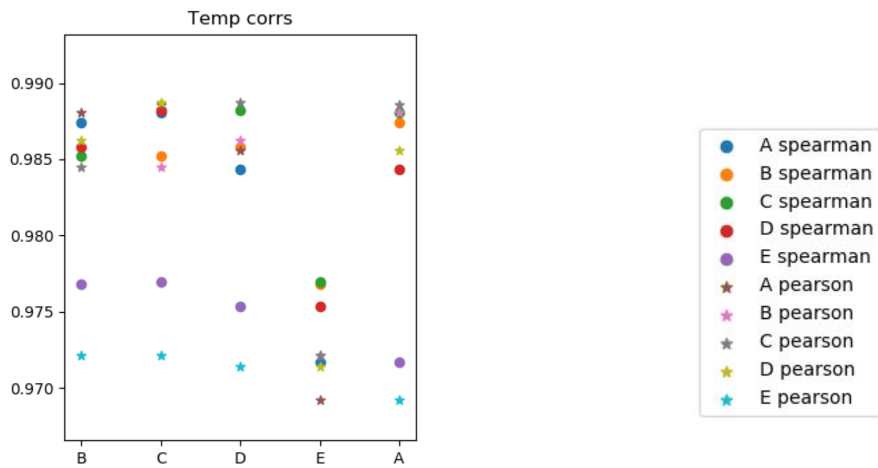


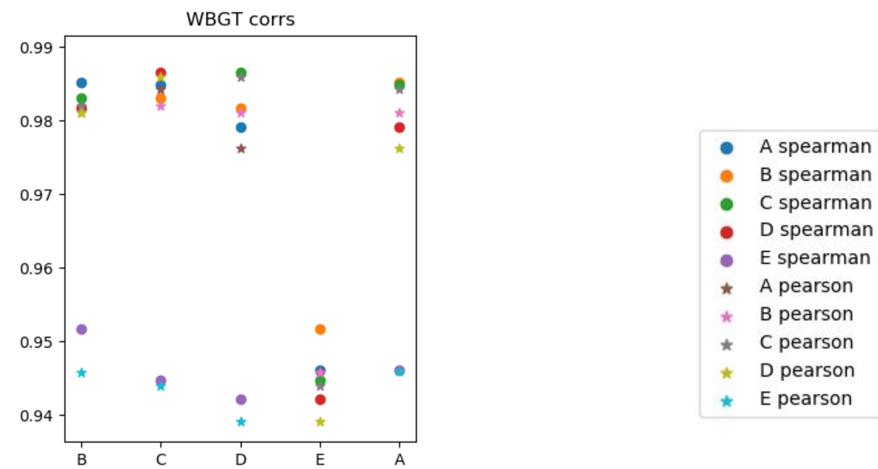
Figure 23: Kernel density plot with PDF for sensor E

3 A3

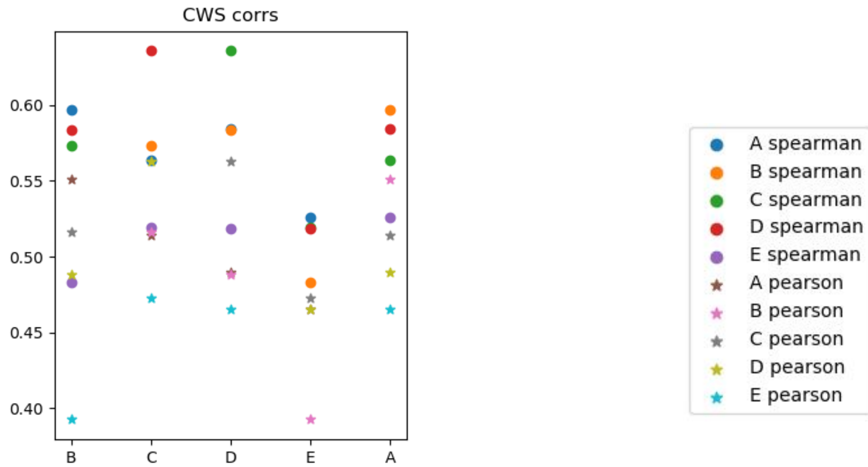
3.1 Scatterplot



(a) Scatterplot showing correlations between sensors for Temperature



(a) Scatterplot showing correlations between sensors for Wet Bulb Temperature



(a) Scatterplot showing correlations between sensors for Cross Wind Speed

3.2 correlation

When observing the correlations between the sensors for the temperature data, it can very clearly be seen that sensor E is correlated much worse with all other sensors than the rest. The largest difference between sensors is between A and E. The wet bulb globe temperature also clearly indicates that Sensor E has less correlation with the rest of the sensors, however for this variable the largest differences are between E and D. Sensors C and D have the closest correlations with each other. The cross wind speed correlation data shows that B and E have the lowest correlation to each other. C and D have the highest correlation value, and the rest of the values are between 0.46 and 0.60.

3.3 Hypothesis

In order to make any hypothesis about the locations of the sensors on the map shown in the figure below, certain characteristics about the landscape must first be pointed out. The locations of the buildings around the sensors are crucial, since these will both influence the wind direction and wind speed. Note here that the wind direction differs between the sensors, ranging from South-West to South. In addition, the type of ground the sensor is places above will also influence the temperature, as well as the humidity at that location. On the map, it is clearly visible that this ranges from grown, wet cropped soil, to dry un-cropped soil, and pavement between the greenhouse and a civilian house. Also an important characteristic is that sensor 4 and 5 are in line from each other, and downwind, as well as 2 and 3. // Based on the correlations, given the far more unique position of the sensor 1 on the map, it can be concluded that this is sensor E, since this sensor has low correlations for the variables with all sensors. Sensors 4 and 5 are in line with each other, behind a building, and therefore are expected to correlate fairly well. Same can be expected for 2 and 3. Sensor 5 is directly downwind of a building, and must therefore be expected to have lower wind speed values. Given their open position, it can be hypothesised that sensors 2 and 3 are equal to sensors B and A, given that sensor B correlates very poorly in wind speed values with sensor E, and these have very different positions. Sensors 4 and 5 therefore must be equal to C and D. The common surroundings of sensors 1, 3 and 5, and the correlation of sensors E and A indicate that position 3 is Sensor A, and position 5 is sensor C. Sensors C and D correspond almost exactly the same with the rest however, thus it is difficult to conclude. Therefore the hypothesis is that positions 1, 2, 3, 4, and 5 correlate with E, B, A, D, and C, respectively.

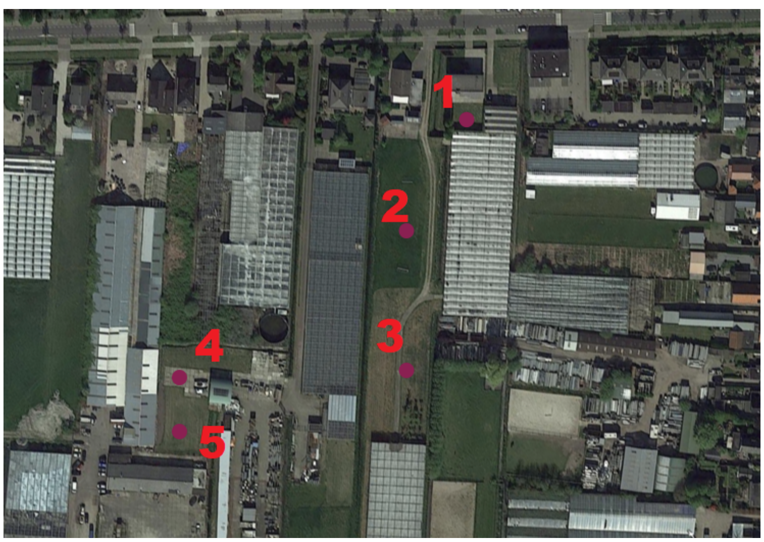


Figure 27: Map with sensor positions

4 A4

4.1 Confidence interval

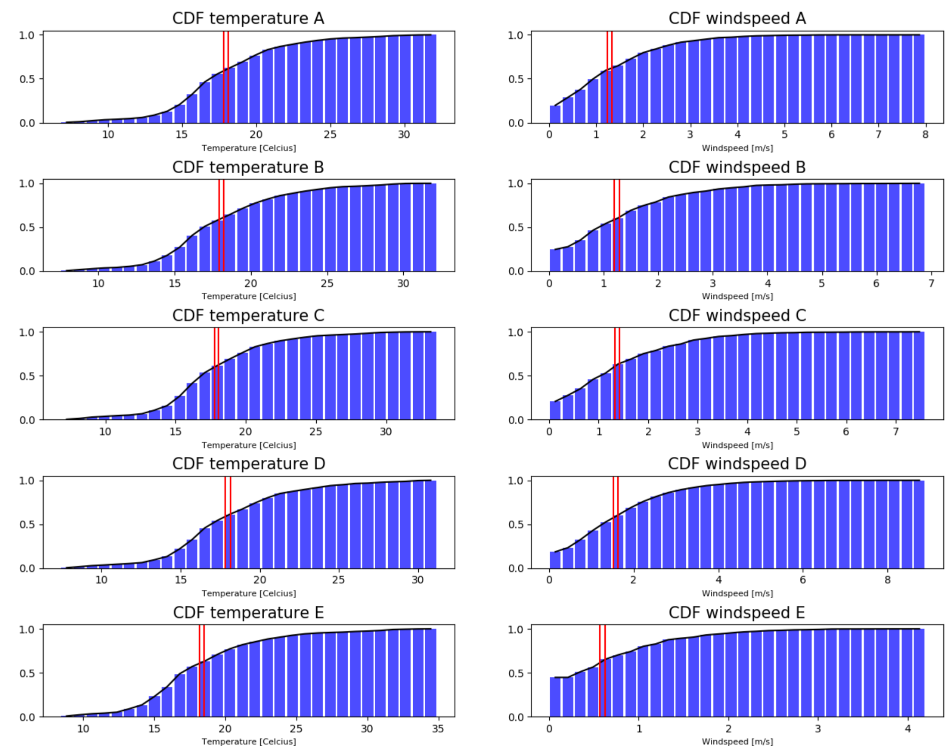


Figure 28: CDF for temperature and windspeed values of all sensors, with the 95 percent confidence intervals given in red

interval	lower	upper
Temperature A	17.81	18.13
Temperature B	17.90	18.23
Temperature C	17.75	18.07
Temperature D	17.84	18.15
Temperature E	18.18	18.53
Windspeed A	1.25	1.33
Windspeed B	1.20	1.29
Windspeed C	1.32	1.42
Windspeed D	1.53	1.63
Windspeed E	0.57	0.62

4.2 test hypothesis

To test the hypothesis, 3 things must first be specified. To start, the significance level α , which in this case is 0.05. Secondly, note that this will be a two tailed test. Lastly, the null hypothesis must be specified. in this case the null hypothesis states that the time series are the same.

p tests	
Temperature E-D	0.002727
Temperature C-D	0.465797
Temperature C-B	0.185627
Temperature A-B	0.401858
Windspeed E-D	4.899592e-212
Windspeed C-D	4.610149e-9
Windspeed B-C	9.400752e-5
Windspeed A-B	0.132479

4.3 conclusion based on p-values

For a 95 percent confidence interval, with an α of 0.05, it can be seen that the null hypothesis for the temperature data can only be rejected between sensors E and D. For the wind speed data, the null hypothesis can be rejected for the sensors E-D, C-D, and B-C. From these results, it can be concluded that for the temperature values, the data is not statistically significant the way the sensors are currently placed. The wind speed p-tests have given more statistically significant results, where only the p test for A-B returns statistically insignificant. This suggest that the data from sensor A is statistically insignificant for the wind speed.

5 Bonus Question

To identify the hottest days, a program is used in python to find the data points for all sensors where the temperature is above a certain value. for instance, if we take 30 degrees Celsius, the data points that were equal to or higher than 31 degrees for sensor A are: [1056, 1060, 1063, 1203, 1204, 1205, 1206, 1207]. this is done using the following line of code:

```
Ahot = (i for i, n in enumerate(Temp[0]) if n >= 31)
print('sensor A',  list(Ahot))
```

Notice that these points are clumped together, indicating a clear two days; once around data point 1060, and once around 1205. To find the exact moment on which it was hotter than 31 degrees, simply find the index of the 1060th or 1205th elements in the temperature data column, being the 24th and the 26th of June 2020, respectively.

References

Maiullari, D., & Sanchez, C. G. (2020, 8). Measured Climate Data in Rijsenhout. Retrieved from https://data.4tu.nl/articles/dataset/Measured$_{C}limate_{D}ata_{i}n_{R}ijsenhout/$12833918 doi: 10.4121/12833918.v1