Sensing Technologies and Mathematics for Geomatics
GEO1001.2020
MSc Geomatics
Delft University of Technology

Homework 1 Georgios Triantafyllou (5381738)

1 A1

1.1 A1.1

From the results i have, there are no many things that could be said. It would be adviseble and needed further analysis in order to reach some conclusion because the amount of data for each sensor are really big.

1.2 A1.2

The number of bins it is really important because if we use too few of them, the histogram does not really portray the data very well (Figure 1). From the other side if we have too many bins, we get a broken comb look, which also does not give a good sense of the distribution (Figure 2). A good example is visible in to the next histograms.

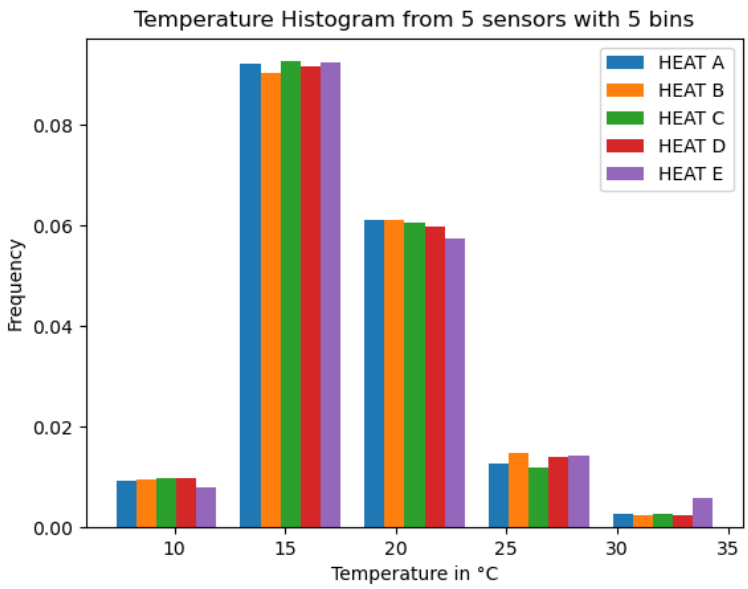

Figure 1: Temperature Histogram from 5 sensors with 5 bins[1]

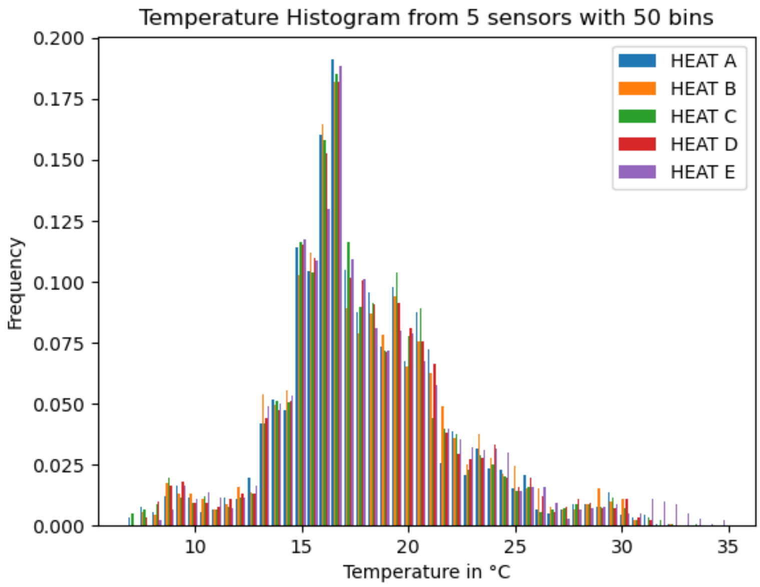

Figure 2: Temperature Histogram from 5 sensors with 50 bins[1]

1.3 A1.3

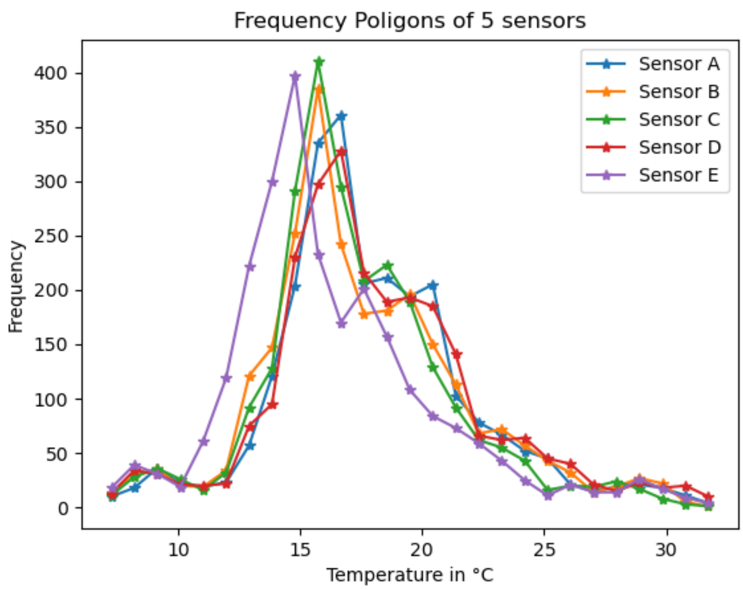

Figure 3: Frequency Poligons of 5 Sensors[1]

1.4 A1.4

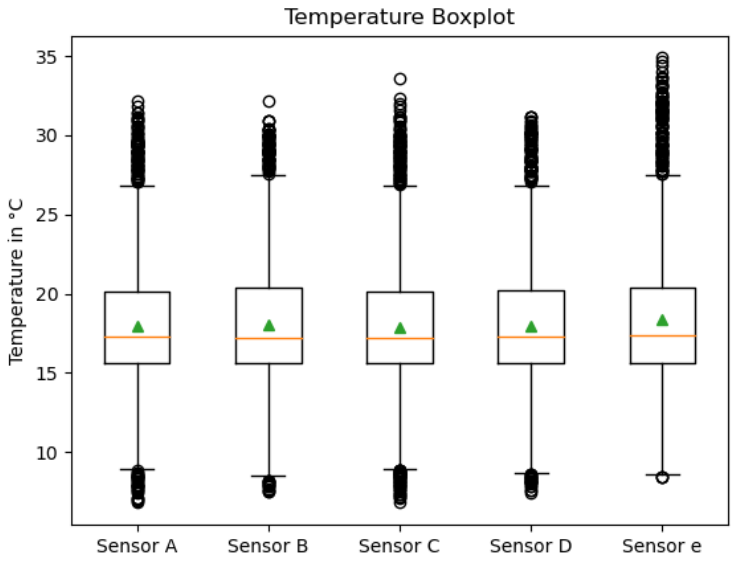

Figure 4: Temperature Boxplot[1]

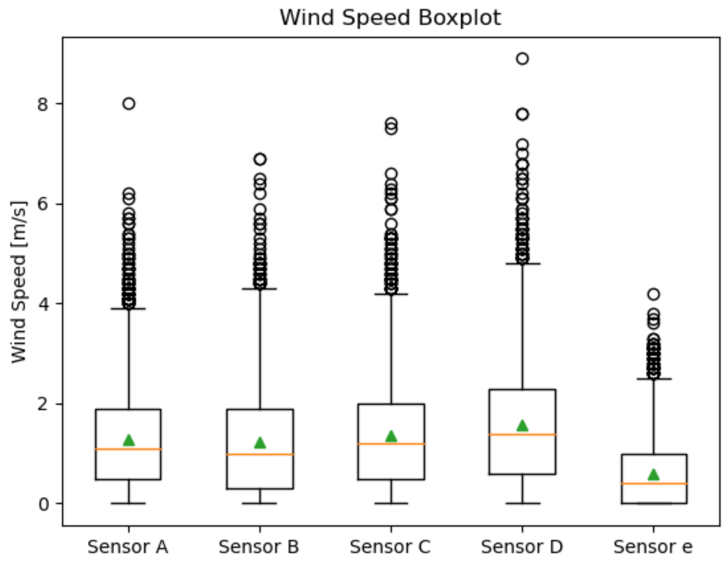

Figure 5: Wind Speed Boxplot[1]

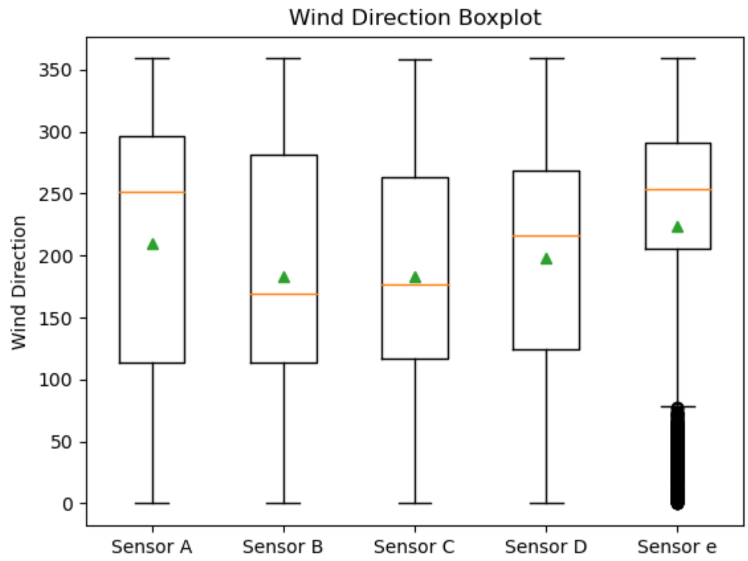

Figure 6: Wind Direction Boxplot[1]

2 A2

2.1 A2.1

From the figures below we can realize that the behavior of the distributions for each Sensor's Temperature look pretty similar when we look the PDF and PMF (Figures 7,8) while the CDF is totally diferent because it is comulative distribution (Figure 9). As about their tails i could talk only for PMF and PDF where it could be said that in PDF the tail is longer somehow and thats because its more smooth than in PMF. More can be seen in the Figures below:

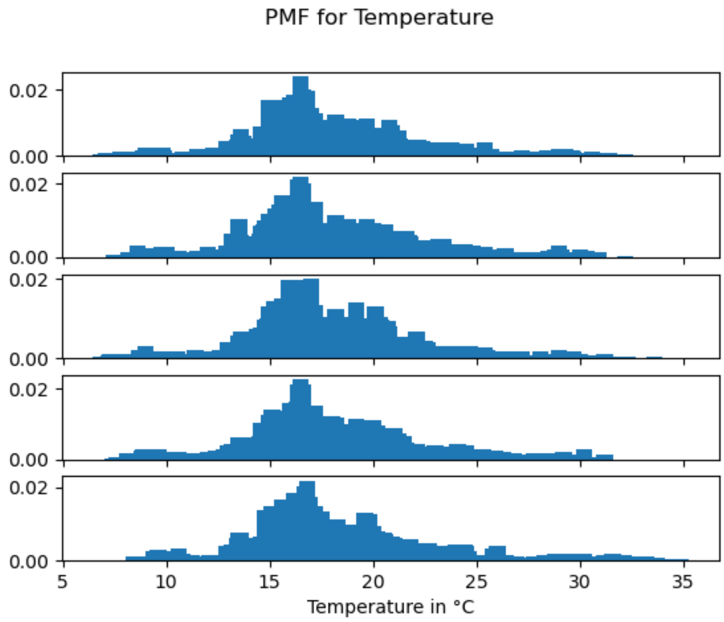

Figure 7: PMF for Temperature[1]

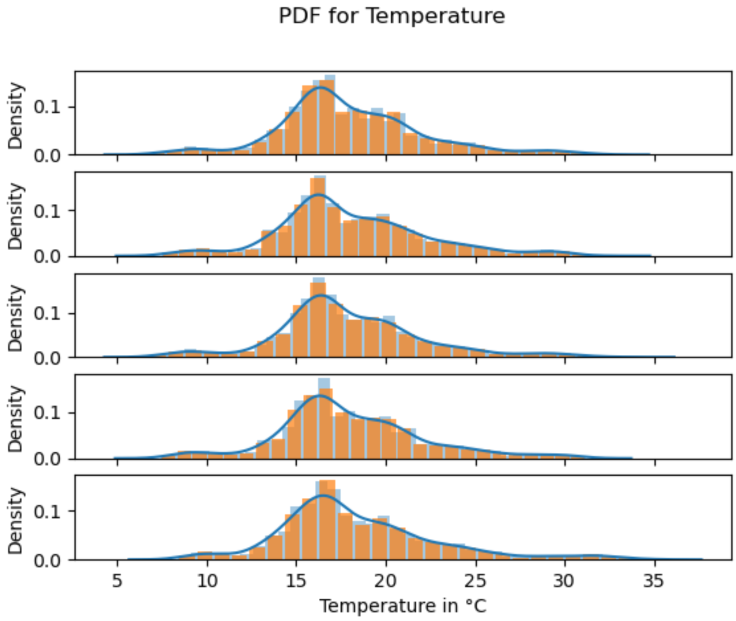

Figure 8: PDF for Temperature[1]

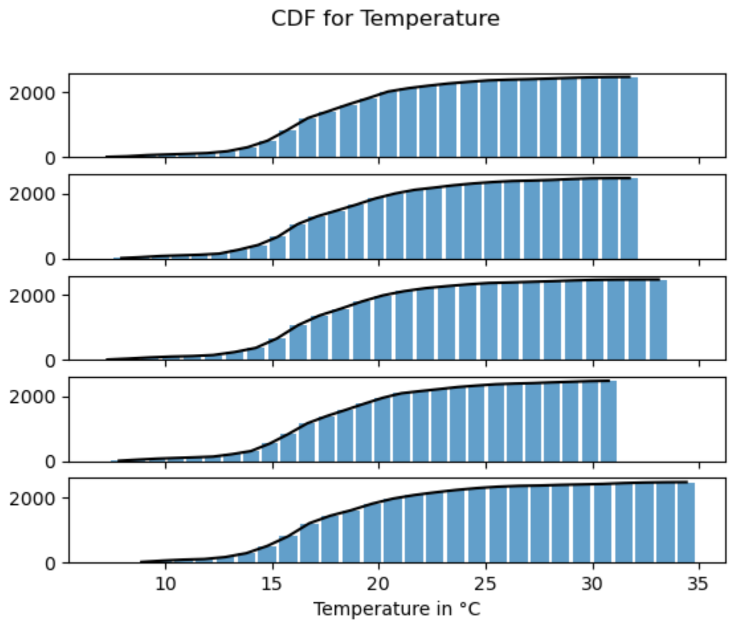

Figure 9: CDF for Temperature[1]

2.2 A2.2

From the figures below we can see that there is no actual diference between the PDF and the Kernel Density Estimation (KDE) for the Wind Speed and that happens because the KDE is actually an algorithm that takes a sample and finds an appropriately smooth PDF that fits the data. So the only diference is that the KDE shows less information in the graph and makes it easier for the audience to understand it.

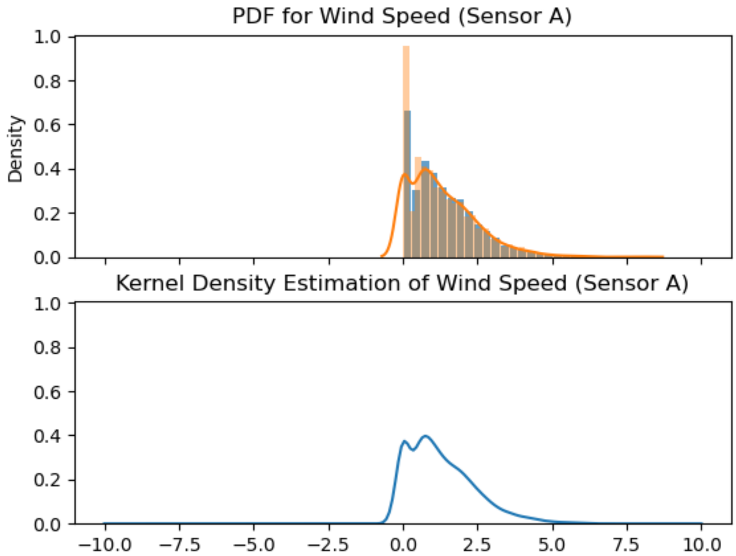

Figure 10: PDF for Wind Speed (Sensor A)[1]

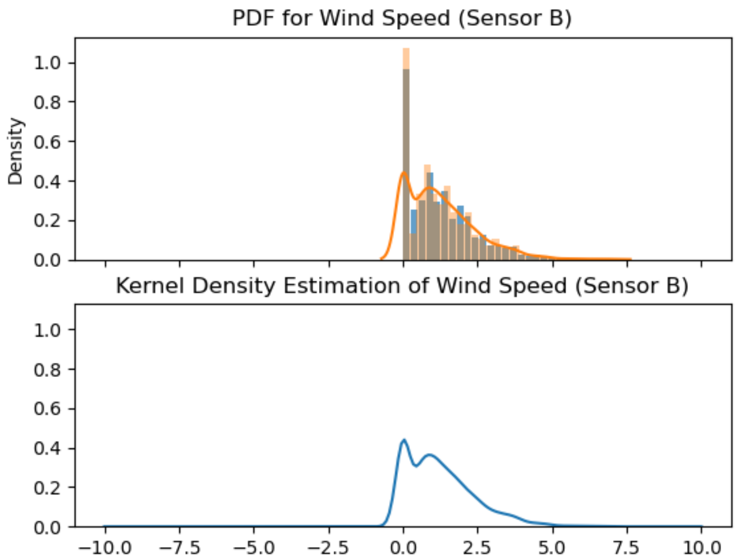

Figure 11: PDF for Wind Speed (Sensor B)[1]

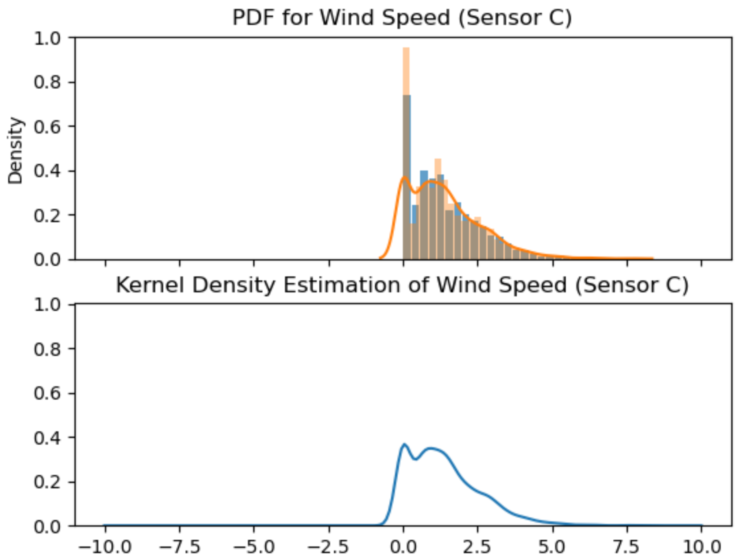

Figure 12: PDF for Wind Speed (Sensor C)[1]

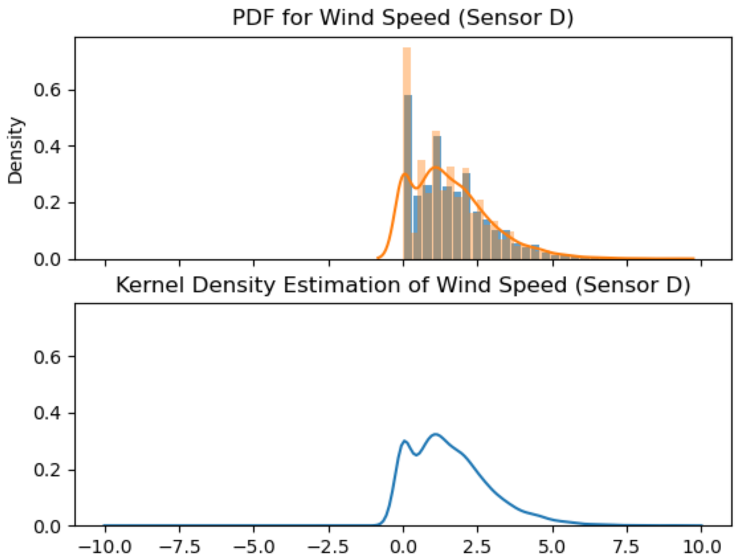

Figure 13: PDF for Wind Speed (Sensor D)[1]

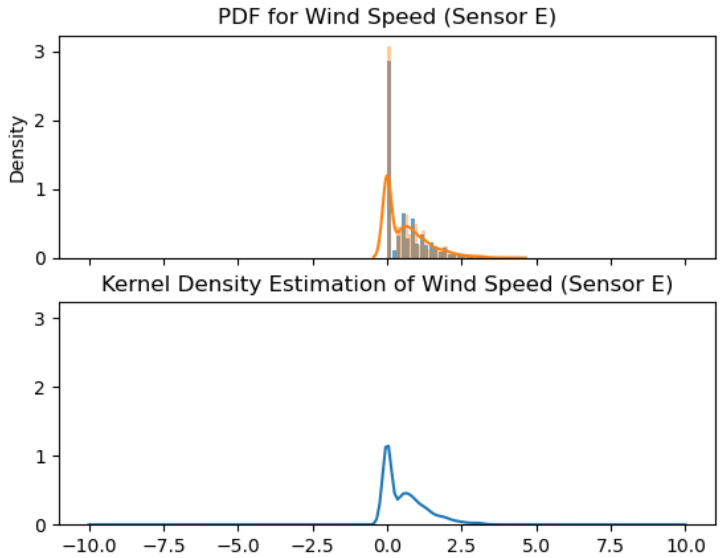

Figure 14: PDF for Wind Speed (Sensor E)[1]

3 A3

3.1 A3.1

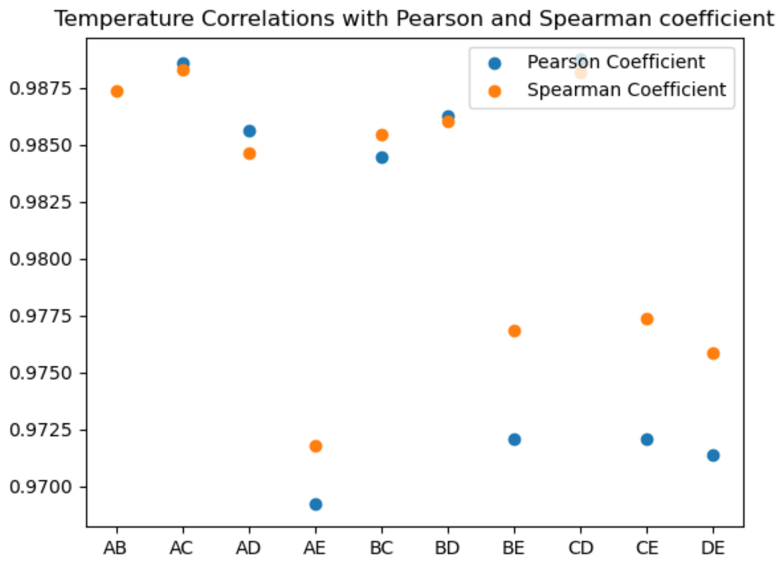

Figure 15: Temperature Correlations with P and Sp coeff[1]

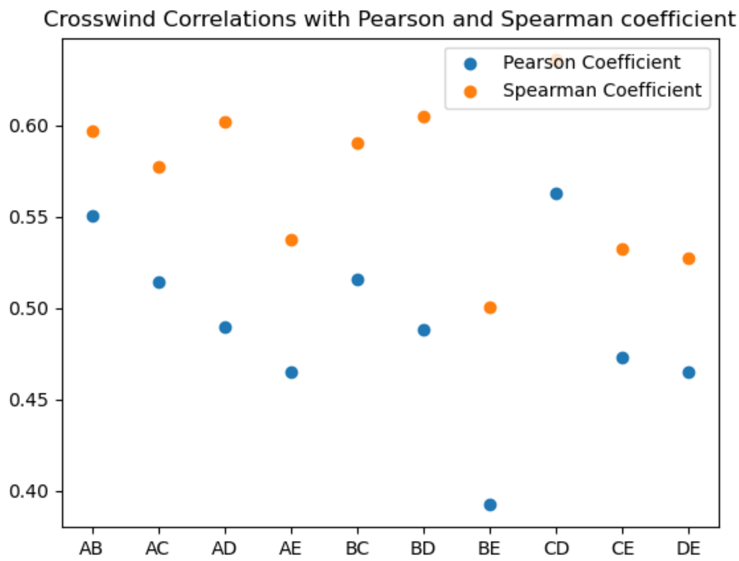

Figure 16: Crosswind Correlations with P and Sp coeff[1]

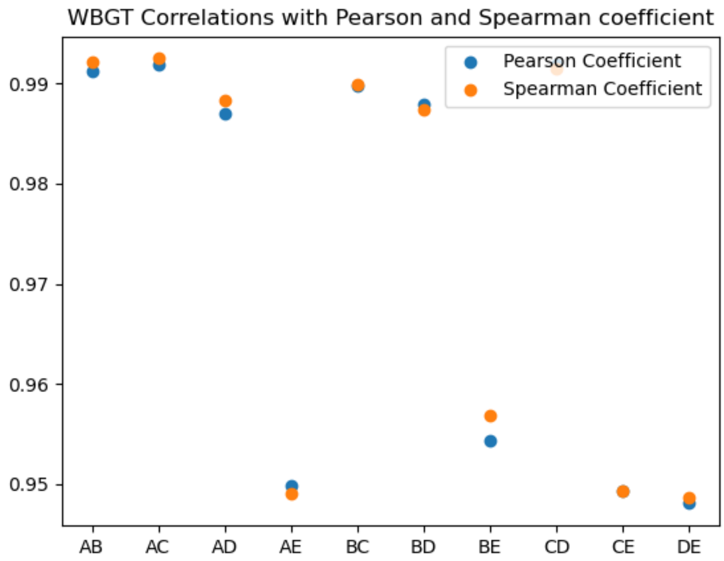

Figure 17: WBGT Correlations with P and Sp coeff[1]

Sensors Relationships	Variables	Pearson Coefficient	Spearman Coefficient
	Temperature	0.98810313	0.987378955
AB	Crosswind	0.550352585	0.596982562
	WBGT	0.991259553	0.992132436
	Temperature	0.988608719	0.988292007
AC	Crosswind	0.51405088	0.577228891
	WBGT	0.99189585	0.992472018
	Temperature	0.985613462	0.984627239
AD	Crosswind	0.489895013	0.601889059
	WBGT	0.987013949	0.988291923
	Temperature	0.969204792	0.9717698
AD	Crosswind	0.465124685	0.537844665
	WBGT	0.949828692	0.949127535
	Temperature	0.98448517	0.985440109
BC	Crosswind	0.516102417	0.590683619
	WBGT	0.989729694	0.989863576
	Temperature	0.986265403	0.986048723
BD	Crosswind	0.488029338	0.604818597
	WBGT	0.987864209	0.987374811
	Temperature	0.972089738	0.976859613
BE	Crosswind	0.39214871	0.500281016
	WBGT	0.95440893	0.956900474
	Temperature	0.988742872	0.988185589
CD	Crosswind	0.562888199	0.635906168
	WBGT	0.991820559	0.991421934
	Temperature	0.972097215	0.977342412
CE	Crosswind	0.473233228	0.532232093
	WBGT	0.949269532	0.949345587
	Temperature	0.971365706	0.975848255
DE	Crosswind	0.465192078	0.527325327
	WBGT	0.948090212	0.94870202

Table 1: Correlations between all the sensors for the variables: Temperature, Wet Bulb Globe Temperature (WBGT), Crosswind Speed[1]

3.2 A3.2

What we could say about the sensors correlations is that mostly they have high correlation since most of them are really near to 1 as about the Temperature and WBGT which means the temperature values they have measure are pretty similar to each other and that could happen because the sensors are in the same area and not really far from each other. About Crosswind since the correlation is around 0.5 means obviously that its not so high like the temperature values and that could mean either they are not so near or the winds in that area are a lot and gives us these results. The results could be seen the Table 1 and to Figures 15,16,17 above.

3.3 A3.3

With a good look in the correlations of the sensors we could say that from the temperatures the A-B and A-C sensors appear to have the highest correlation but at the same time they have low correlation as about the Crosswind. That is something which happens with all the sensors pairs actually. In that case it would unsafe to hypothesize which is the correct place for the sensors because all of them are quite near and so that means the temperature should be similar and therefore the correlation high. As about the crosswind correlation it could easily be low because the sensors are in an semi urban area and the crosswind could affect the correlation. In that case no specific hypothesize could be done.

Figure 18: Sensors Location

4 A4

4.1 A4.1,2

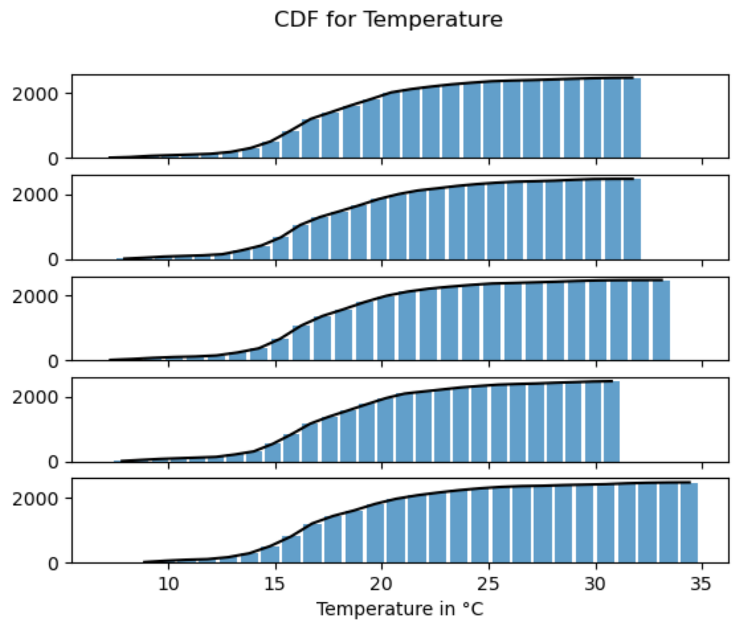

Figure 19: CDF for Temperature[1]

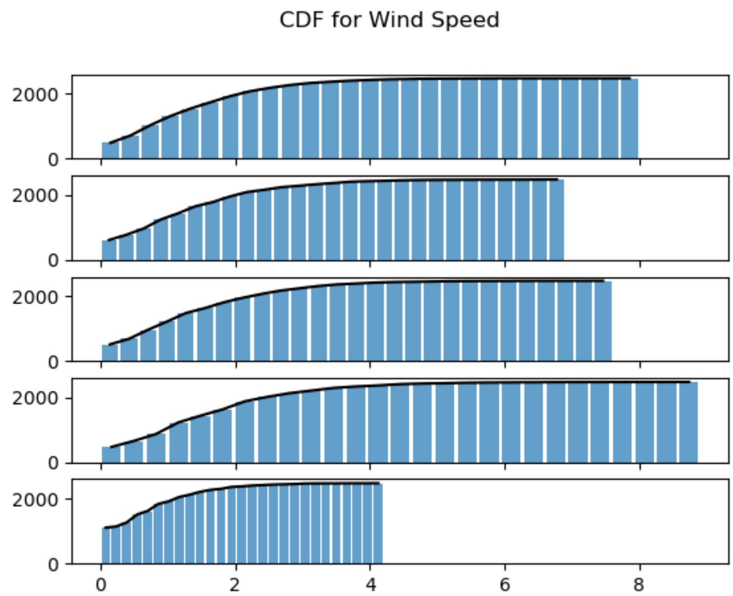

Figure 20: CDF for Wind Speed[1]

Variables	Confidence Intervals			Sensors
	m-h	m	m+h	
Temperature	17.8121	17.9691	18.1261	A
	17.9047	18.0654	18.2261	B
	17.7549	17.9131	18.0713	C
	17.8381	17.9964	18.1546	D
	18.1819	18.3539	18.5259	E
Wind Speed	1.2462	1.2903	1.3344	A
	1.1972	1.2421	1.2871	B
	1.3243	1.3715	1.4186	C
	1.5297	1.5817	1.6337	D
	0.5681	0.5962	0.6244	E

Table 2: 95/100 confidence intervals for variables Temperature and Wind Speed for all the sensors [1]

Sensors	Student Test	p value	Variables
ED	3.00023	0.00271	Temperature
DC	0.72939	0.46580	
CB	-1.32423	0.18549	
BA	0.84084	0.40048	
ED	-32.67317	0.00000	Wind Speed
DC	5.87115	0.00000	
CB	3.89266	0.00010	
BA	-1.50061	0.13352	

Table 3: Student Test and p-values

4.2 A4.3

So in the Table 3 below is visible the p-values from the requested sensors. We can conclude that most of them as about the temperature values (3 out of 4) are way above the 0.05 so that can strengthen the Hypothesis. And totally the opposite is happening with the Wind speed where we could say for the ED, DC, CB that we totally reject the hypothesis and that means the time series are different. Concluding what could be said is that obviously the p-value is not enough to reach a clear conclusion but probably none of the time series are completely the same so the Hypothesis could be rejected.

Sensors	Hot/Cold	Temperature (Celsius)	Date
A	Hottest	25.1833	6/26/2020
	Coolest	14.1556	6/10/2020
B	Hottest	24.9292	6/26/2020
	Coolest	14.3278	6/10/2020
C	Hottest	24.8722	6/26/2020
	Coolest	14.2667	6/10/2020
D	Hottest	24.8750	6/26/2020
	Coolest	14.3708	6/10/2020
E	Hottest	25.9111	6/25/2020
	Coolest	14.4903	7/8/2020

Table 4: Highest and Lowest Temperature for every Sensor

5 Bonus Question

In this question it is asked to identify the hottest and coolest day of the measurement time series provided. In order to do that a python function created named (averagetemperature) which takes the data and calculates the maximum and minimum average temperatures in order to find which day is the hottest and coolest during the days that the data was acquired. In order to do that the function creates a table and sort it from the biggest to smaller value which in our case mean highest and coolest temperature. In that way we are able to find out the hottest and coolest days. From what the Table 4 above show us, the hottest day for the 4 out of 5 sensors was the 26th of June 2020 and the coolest the 10th of the same month. Finally only the E sensor had different days as the results show us.

References

[1] Daniela Maiullari and Clara Garcia Sanchez. Measured Climate Data in Rijsenhout. 8 2020.