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

Assignment 01

Simon Pena Pereira (5391210)

Contents

1 After lesson A1:

1.1 Compute mean statistics (mean, variance and standard deviation for each of the sensors variables), what do you observe from the results?

The computed mean statistics of all five sensors show mostly homogeneous results. Nevertheless, by observing the variance and the standard deviation, inferences regarding the respective mean can be drawn. For instance, the temperature's mean of each sensor differs by less than 0.5 °C. Contrary to this, the variance of sensor E shows a spread of 19.04 °C, which is about 2.41 °C higher than the second highest variance of sensor A to D. By looking at the standard deviation, which is almost at the same level as the standard deviation of the other sensors, it can be derived that the mean temperatures are equally representative, despite the higher variance. Potential reasons for the greater spread of temperature values might be outliers. A similar pattern can be observed for the relative humidity, where both the mean and the standard deviation of all sensor are nearly the same. Therefore, it is noticeable that the variances differ from each other, which highlights the different spreads. Nevertheless, the accuracy of the mean is determined by the standard deviation, which is close to zero for the values of Wind Speed, Crosswind Speed and Headwind Speed of sensor E, and therefore particularly accurate.

Sensor 1	**Mean**	**Variance**	**Standard Deviation**
Direction True	209.406300	10108.940308	100.543226
Wind Speed	1.290307	1.251154	1.118550
Crosswind Speed	0.964943	0.926593	0.962597
Headwind Speed	0.163530	1.034940	1.017320
Temperature	17.969103	15.864269	3.982998
Globe Temperature	21.544588	68.191353	8.257806
Wind Chill	17.838207	16.264447	4.032920
Relative Humidity	78.184774	376.010059	19.390979
Heat Stress Index	17.899596	14.996848	3.872576
Dew Point	13.553877	9.723472	3.118248
Psychro Wet Bulb Temperature	15.270719	6.944027	2.635152
Station Pressure	1016.168255	38.471267	6.202521
Barometric Pressure	1016.128433	38.467951	6.202254
Altitude	-25.987076	2663.641045	51.610474
Density Altitude	137.316640	26510.044345	162.819054
NA Wet Bulb Temperature	15.981543	10.012108	3.164191
WBGT	17.254321	16.135258	4.016872
TWL	301.392932	814.766564	28.544116
Direction Mag	208.905089	10105.677049	100.526997

Table 1: Statistics of Sensor 1

Sensor 2	Mean	Variance	Standard Deviation
Direction True	183.412359	9977.217770	99.886024
Wind Speed	1.242124	1.301502	1.140834
Crosswind Speed	0.835622	0.878585	0.937329
Headwind Speed	-0.129806	1.256719	1.121035
Temperature	18.065428	16.629067	4.077875
Globe Temperature	21.799435	66.049317	8.127073
Wind Chill	17.945921	17.035826	4.127448
Relative Humidity	77.878312	408.623008	20.214426
Heat Stress Index	18.004281	15.439157	3.929269
Dew Point	13.530856	9.636518	3.104274
Psychro Wet Bulb Temperature	15.295517	6.770263	2.601973
Station Pressure	1016.657027	36.841934	6.069756
Barometric Pressure	1016.616478	36.828868	6.068679
Altitude	-30.058158	2545.708131	50.455011
Density Altitude	135.580775	26863.310240	163.900306
NA Wet Bulb Temperature	15.996809	9.809254	3.131973
WBGT	17.321971	15.835355	3.979366
TWL	299.451696	790.069221	28.108170
Direction Mag	183.217286	9975.446909	99.877159

Table 2: Statistics of Sensor 2

Sensor 3	Mean	Variance	Standard Deviation
Direction True	183.588925	7703.363096	87.768805
Wind Speed	1.371463	1.430920	1.196211
Crosswind Speed	0.963298	1.042575	1.021066
Headwind Speed	-0.262894	1.271732	1.127711
Temperature	17.913137	16.104538	4.013046
Globe Temperature	21.587389	67.941305	8.242652
Wind Chill	17.772999	16.541123	4.067078
Relative Humidity	77.962854	374.622643	19.355171
Heat Stress Index	17.828254	15.356254	3.918706
Dew Point	13.458124	10.084149	3.175555
Psychro Wet Bulb Temperature	15.196645	7.239313	2.690597
Station Pressure	1016.689329	37.691491	6.139340
Barometric Pressure	1016.651900	37.675623	6.138047
Altitude	-30.338723	2608.534634	51.073816
Density Altitude	129.622878	26986.602970	164.275996
NA Wet Bulb Temperature	15.934236	10.480279	3.237326
WBGT	17.225020	16.546745	4.067769
TWL	301.899757	766.533514	27.686342
Direction Mag	183.083670	7704.620170	87.775966

Table 3: Statistics of Sensor 3

Sensor 4	Mean	Variance	Standard Deviation
Direction True	198.326597	8133.890057	90.188082
Wind Speed	1.581649	1.739817	1.319021
Crosswind Speed	1.210509	1.451503	1.204783
Headwind Speed	-0.300566	1.232503	1.110181
Temperature	17.996362	16.105591	4.013177
Globe Temperature	21.359297	61.202253	7.823187
Wind Chill	17.835368	16.556852	4.069011
Relative Humidity	77.942037	389.856040	19.744772
Heat Stress Index	17.921625	15.117644	3.888141
Dew Point	13.508610	10.071883	3.173623
Psychro Wet Bulb Temperature	15.260186	7.044403	2.654129
Station Pressure	1016.728011	34.987784	5.915047
Barometric Pressure	1016.688884	34.952327	5.912049
Altitude	-30.653193	2419.723591	49.190686
Density Altitude	132.411075	26516.125733	162.837728
NA Wet Bulb Temperature	15.915643	9.987434	3.160290
WBGT	17.176799	15.507185	3.937916
TWL	305.254568	616.009807	24.819545
Direction Mag	197.826192	8135.315513	90.195984

Table 4: Statistics of Sensor 4

Sensor 5	Mean	Variance	Standard Deviation
Direction True	223.956364	9308.285080	96.479454
Wind Speed	0.596242	0.511227	0.715001
Crosswind Speed	0.438505	0.315942	0.562087
Headwind Speed	0.194949	0.319073	0.564866
Temperature	18.353939	19.043132	4.363844
Globe Temperature	21.176162	63.215503	7.950818
Wind Chill	18.294020	19.137062	4.374593
Relative Humidity	76.793051	406.494463	20.161708
Heat Stress Index	18.286424	18.475240	4.298283
Dew Point	13.558788	9.422585	3.069623
Psychro Wet Bulb Temperature	15.406667	6.997445	2.645268
Station Pressure	1016.166101	38.939913	6.240185
Barometric Pressure	1016.127798	38.935177	6.239806
Altitude	-25.961212	2692.353386	51.887892
Density Altitude	150.840000	29714.927502	172.380183
NA Wet Bulb Temperature	15.936889	9.432184	3.071186
WBGT	17.185535	15.489872	3.935717
TWL	284.115313	1289.913383	35.915364
Direction Mag	223.896566	9268.007890	96.270493

Table 5: Statistics of Sensor 5

1.2 Create 1 plot that contains histograms for the 5 sensors Temperature values. Compare histograms with 5 and 50 bins, why is the number of bins important?

A histogram with 50 bins represents the data in greater detail than a histogram with 5 bins. By grouping less values per bin, as it is done with 50 bins, it is easier to determine outliers in the data set and to assign certain frequencies to a smaller range of temperature values. However, the number of bins to compare depends also on the number of sample sets. In this case, each bin shows five sample sets, which are in total almost 250 columns for 50 bins and 25 columns for 5 bins. Consequently, a high number of bins tends to be more unclear than a small number of bins. Thus, five bins in a histogram are easier to read but also less detailed due to a larger amount of values within a bin, which makes the comparison within a bin more even. In summary, it can be stated that the number of bins is important to represent a data set in the most efficient and clear way.

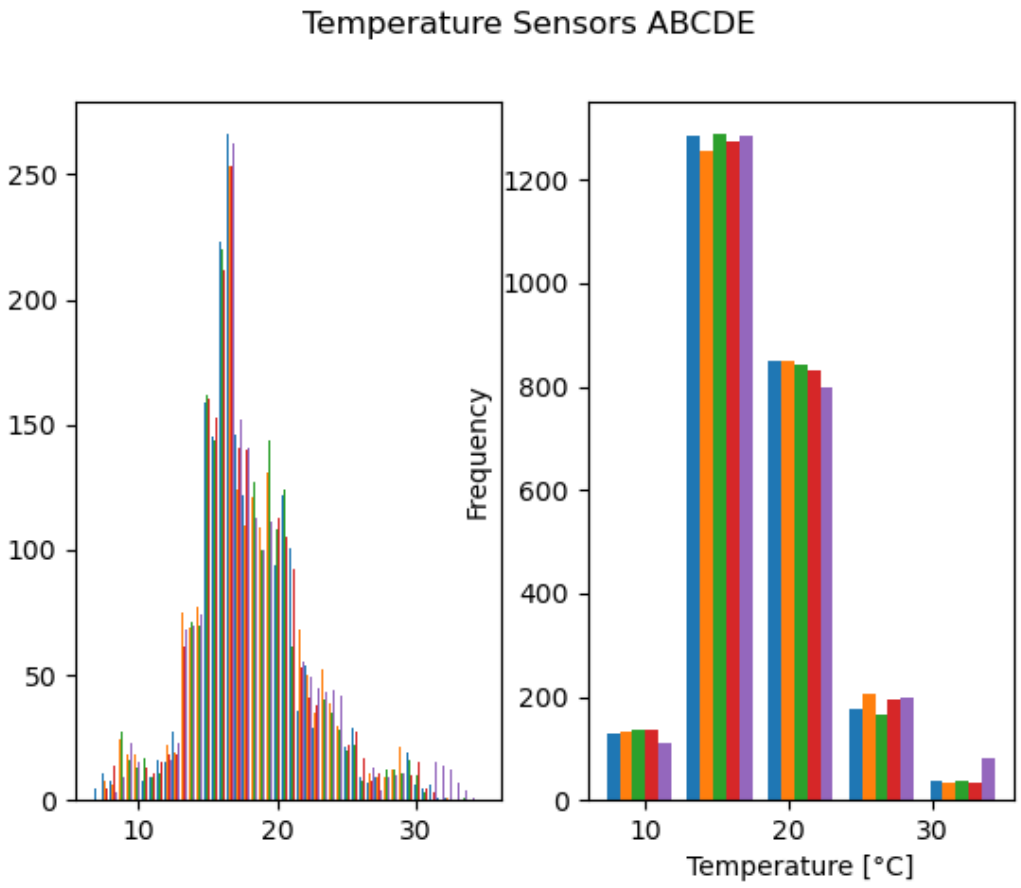

Figure 1: Comparing histograms with 50 and 5 bins

1.3 Create 1 plot where frequency poligons for the 5 sensors Temperature values overlap in different colors with a legend.

(see below)

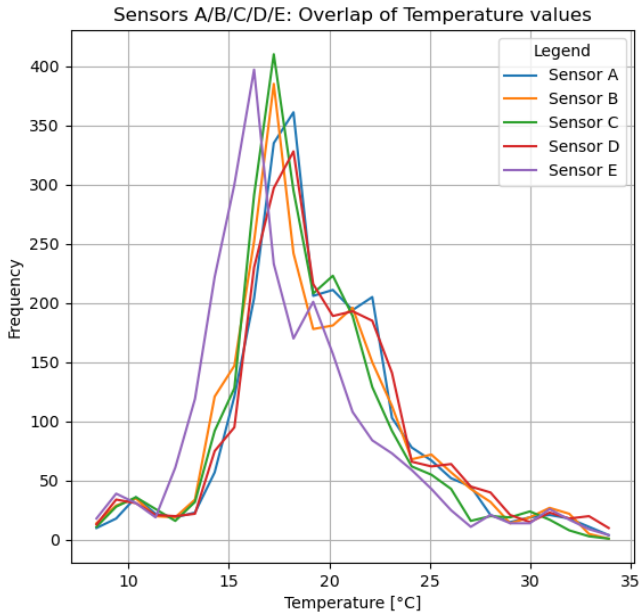

Figure 2: 5 sensors Temperature values overlap

1.4 Generate 3 plots that include the 5 sensors boxplot for: Wind Speed, Wind Direction and Temperature.

(boxplots are attached to the end)

2 2. After lesson A2:

2.1 Plot PMF, PDF and CDF for the 5 sensors Temperature values in independent plots (or subplots). Describe the behaviour of the distributions, are they all similar? what about their tails?

(plots are attached to the end)
The PMFs of temperature values are very similar regarding the distribution of values on the x-axis. All plots show a unimodal distribution and a similar basic scope of the probabilities of temperatures. Apart from that, most significant variances and some outliers are mainly located in the PMFs' centres. By comparing the PMFs' tails, the long tail of sensor 5 is the most noticeable one, because it is the only tail going till 35 °C. Therefore, the centre of the dataset moves more to the left on the x-axis. The comparison between the PMFs and the PDFs shows that the PDF's distribution is similar to the shape of the PMF's distribution. By comparing both in greater detail, the probability of each value seems to be higher in the PMFs than in the PDFs. This is due to fact that the PMFs use discrete values, in this case temperature values which are rounded to one decimal place and assigned to a probability in a range of [0, 1]. For the PDFs, continuous values are assigned to a probability density. In order to evaluate the probability density, it is helpful to analyse the PDF's antiderivative, which is the CDF. The behaviour of the CDF's

distribution determines the distribution of the PDFs, which explains why all plots of the CDF look very similar as well. They begin with a flat rise due to the PDF's tails. From about 12 °C the CDF starts to rise strongly, because from there it accumulates most of its distribution, which is represented by the PDF's centre. Therefore, the highest gradient is found at about 16 °C where the peaks of all the PDF are located. Accordingly, after the strong rise there is also a flat and long accumulation in sensor 5 until it reaches a cumulative probability of 1 (the plots show non-normalized CDFs, which do not sum up to 1).

2.2 For the Wind Speed values, plot the pdf and the kernel density estimation. Comment the differences.

(plots are attached to the end)
The PDFs as well as the KDEs have a bimodal distribution and are left-skewed. Usually KDEs are useful to smooth the distribution of a small sample sizes, which is not necessary for the wind speed sample, because its size is great enough. Therefore, both distributions look the same.

3 After lesson A3:

3.1 Compute the correlations between all the sensors for the variables: Temperature, Wet Bulb Globe Temperature (WBGT), Crosswind Speed. Perform correlation between sensors with the same variable, not between two different variables; for example, correlate Temperature time series between sensor A and B. Use Pearson's and Spearmann's rank coefficients. Make a scatter plot with both coefficients with the 3 variables.

(plots are attached to the end)

3.2 What can you say about the sensors' correlations?

The comparison of the sensors' correlations shows that both the Pearson's and the Spearman's correlation display very similar results for all sensors except for the sensors monitoring crosswind speed. Both plots show their correlations on different scales. Furthermore, the relations between the points of the sensors A-D and B-D differ. That is why I do not rely on their validity for the following subtask. In general, it is noticeable that all scatter plots show a less significant correlation between sensor E and all the other sensors than between all the other sensors among themselves. Between these, strong correlations exist especially for the respective temperature and WBGT values.

3.3 If we told you that that the sensors are located as follows, hypothesize which location would you assign to each sensor and reason your hypothesis using the correlations.

According to the crosswind speed correlations, it stands to reason that sensors C and D are in immediate proximity. Thus, I would assign the sensors C and D to the points at the bottom on the left of the image. Since all correlations between E and all the other sensors are displayed as the lowest ones, I would assign the sensor E to the point at the top on the right of the image. This is reasoned by the fact that sensor E seems to be isolated from all other sensors as well as the mentioned point, which is isolated in a backyard with great distance to the other points. Regarding the WBGT correlation, the sensors A and B as well as the sensors C and D are strongly linked together. Therefore, I would assign the sensors A and B to the points in the centre of the image. Due to the strong correlation between sensors A and C regarding the temperatures and WGBT values, I would assume that both sensors are in a similar environment. That is why I would assign the sensors A and C respectively to the lower point of the previous defined pairs, because they are both located on dried-out grass, compared to the upper football pitch and the paving.

4 After lesson A4:

4.1 Plot the CDF for all the sensors and for variables Temperature and Wind Speed, then compute the 95% confidence intervals for variables Temperature and Wind Speed for all the sensors and save them in a table (txt or csv form).

(plots are attached to the end)

	Start	End
Temperature	1.781.214.113.267.340	18.126065652463858
Temperature	1.790.472.689.963.890	18.226129320070267
Temperature	17.754.926.235.060.200	18.071347006653575
Temperature	1.783.814.660.824.380	18.15457772482005
Temperature	18.181.933.946.027.700	18.525944841851015
Wind Speed	1.246.227.038.990.970	1.3343868543854427
Wind Speed	11.971.663.346.979.200	1.287082453670411
Wind Speed	13.243.037.885.948.900	1.418622646328308
Wind Speed	15.296.480.419.653.700	1.633650260379006
Wind Speed	0.5680599051948441	0.6244249432900044

4.2 Test the hypothesis: the time series for Temperature and Wind Speed are the same for sensors:E,D; D,C; C,B; B,A What could you conclude from the p-values?

In order to conclude a statement from the p-values, the hypothesis needs to be considered as a null hypothesis. The smallest level of significance to reject the null hypothesis is $p < 0.05$. Apart from the sensor pair E and D, the p-values of the time series for temperature

are all above the level significance. Therefore, the null hypothesis can be accepted for the sensor pairs D and C, C and B as well as for B and A. For the sensor pair E and D, the null hypothesis needs to be rejected. With regards to the time series for wind speed, the sensor pairs E and D, D and C as well as C and B obtain a p-value below the level of significance. Therefore, the null hypothesis needs to be rejected for them. The sensors B and A obtain a p-value above 0.05, which is why the null hypothesis can be accepted for them. Furthermore, it is noticeable that the p-value for the sensor pairs D and C (circa 0.46) as well as A and B (circa 0,40) are the highest ones in comparison the other p-values. Therefore, they both show the strongest evidence for accepting the null hypothesis. This observation also corresponds to the results of the previous done calculations for sensors' correlations.

	Pairs	**t-value**	**p-value**
Temp.:	**E,D**	3.0002339815514034	0.002711172129731209
Temp.:	**D,C**	0.7293907701134738,	0.4657972008220813
Temp.:	**C,B**	-1.3242344224224623,	0.18548636717619374
Temp.:	**B,A**	0.8408449326559486,	0.4004754260262924
WS:	**E,D**	-32.67316852220387,	3.3729639501474365e-212
WS:	**D,C**	5.871152992711887,	4.610149126224334e-09
WS:	**C,B**	3.8926626715412143,	0.00010045473692816457
WS:	**B,A**	-1.500613919591207,	0.13351922750703515

References

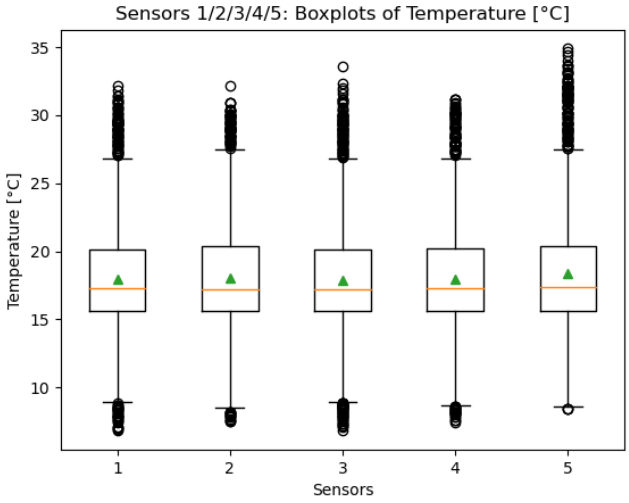

Figure 3: Boxplots of sensors 1-5: Temperatures °C

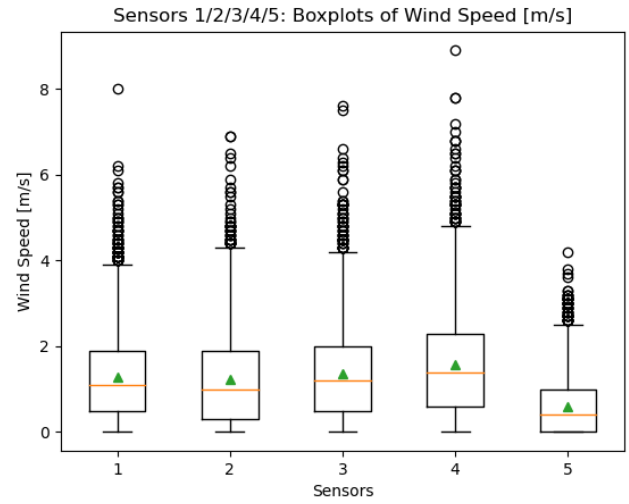

Figure 4: Boxplots of sensors 1-5: Wind Speed m/s

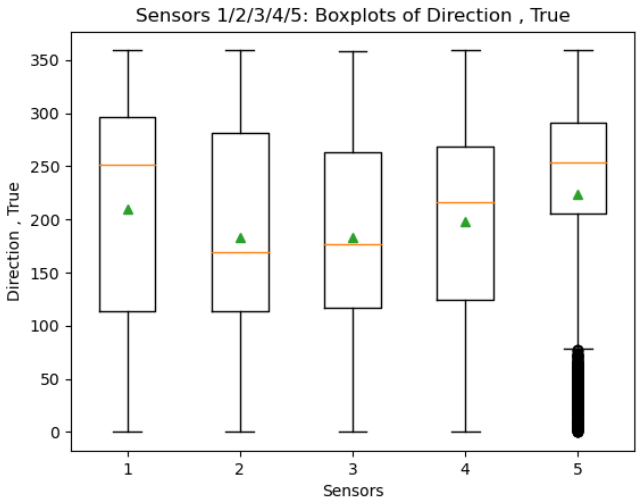

Figure 5: Boxplots of sensors 1-5: Direction , True

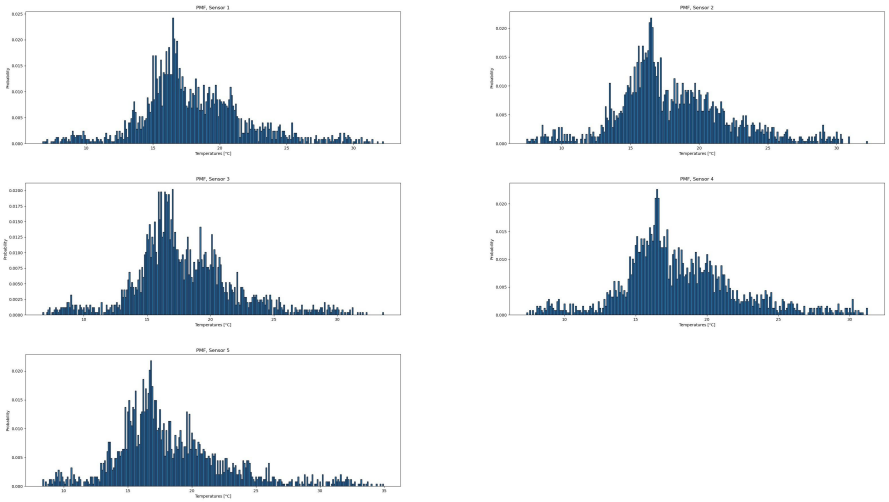

Figure 6: PMFs for the 5 sensors Temperature values

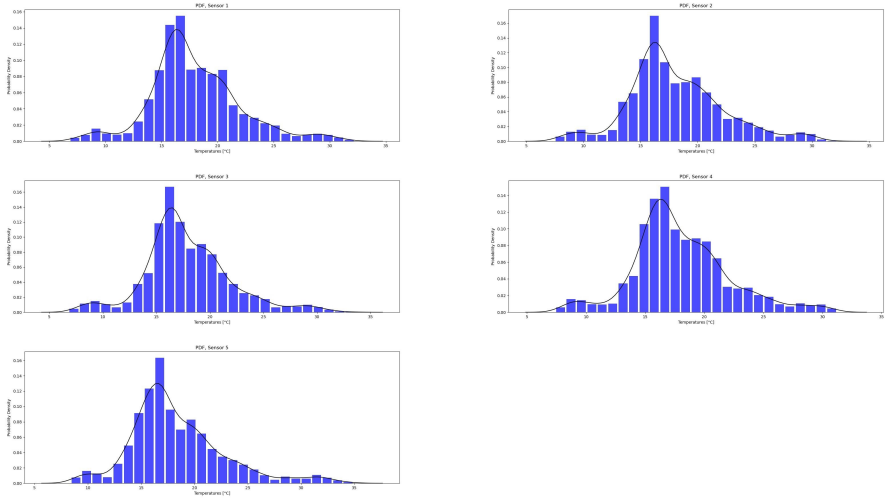

Figure 7: PDFs for the 5 sensors Temperature values

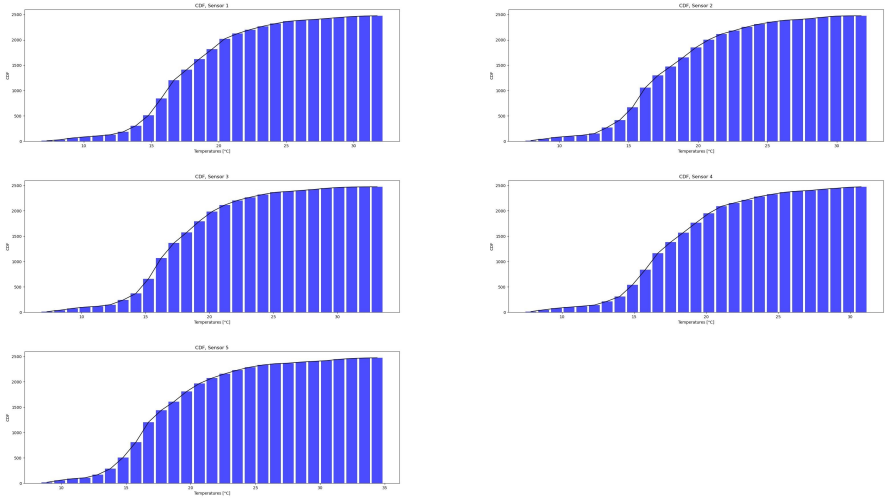

Figure 8: CDFs for the 5 sensors Temperature values

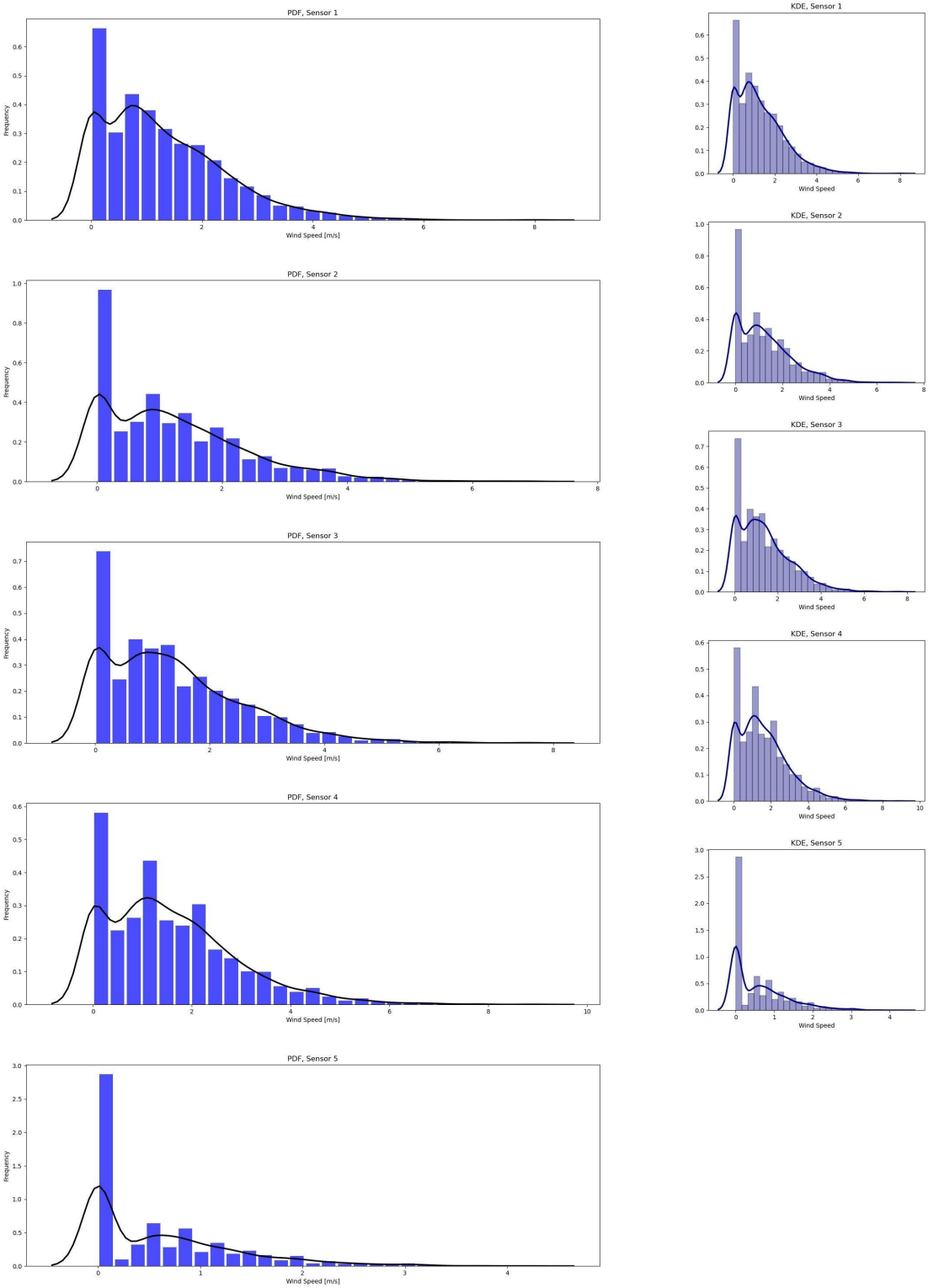

Figure 9: Comparison of PDFs and KDEs for 5 sensors Temperature values

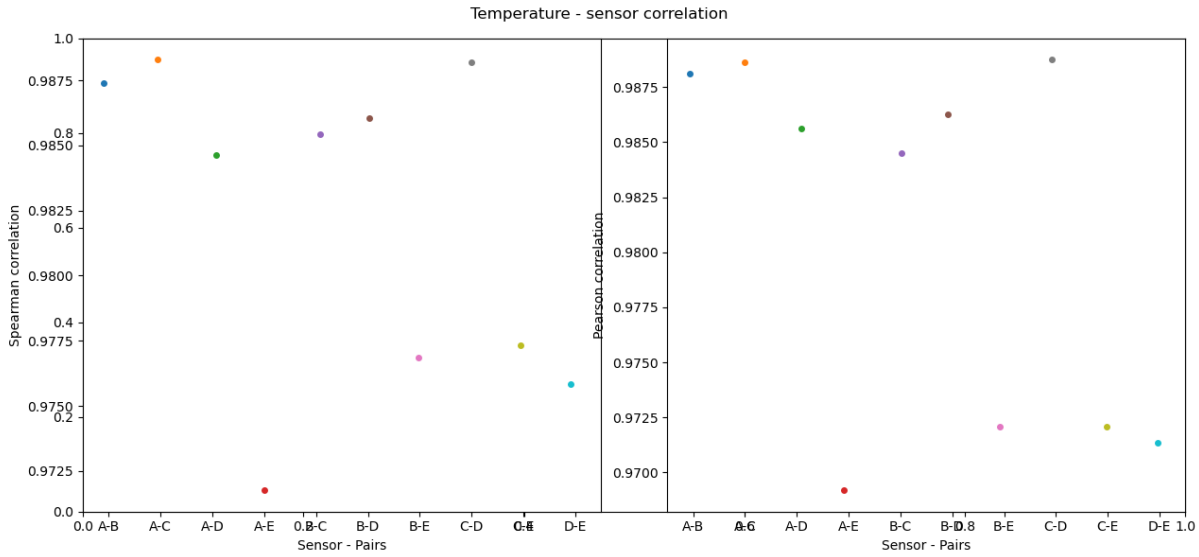

Figure 10: Scatter plot with Spearmann's and Pearson's rank coefficients for Temperature

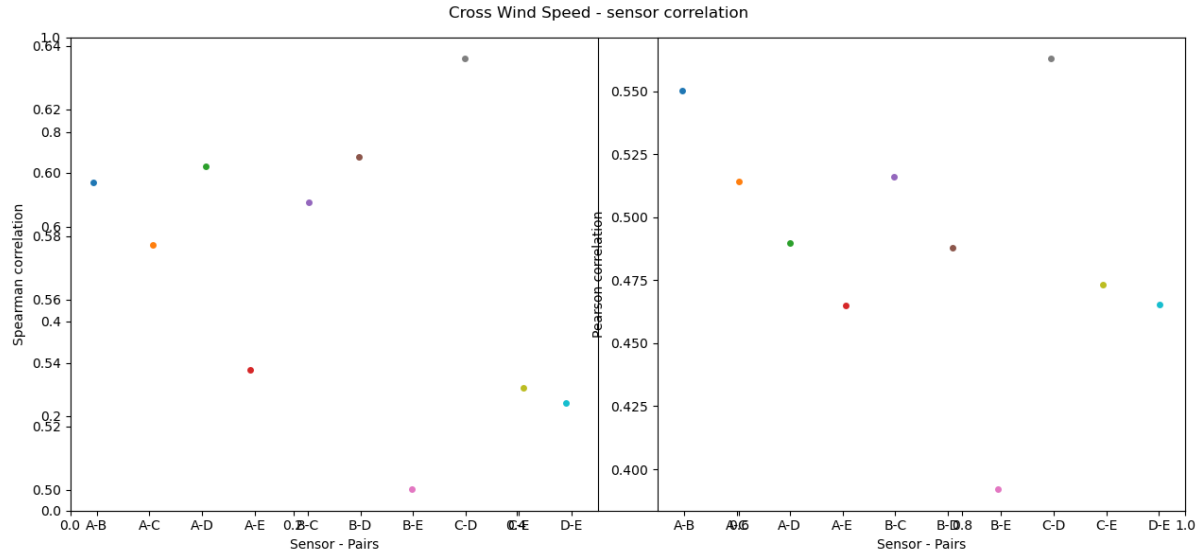

Figure 11: Scatter plot with Spearmann's and Pearson's rank coefficients for Crosswind Speed

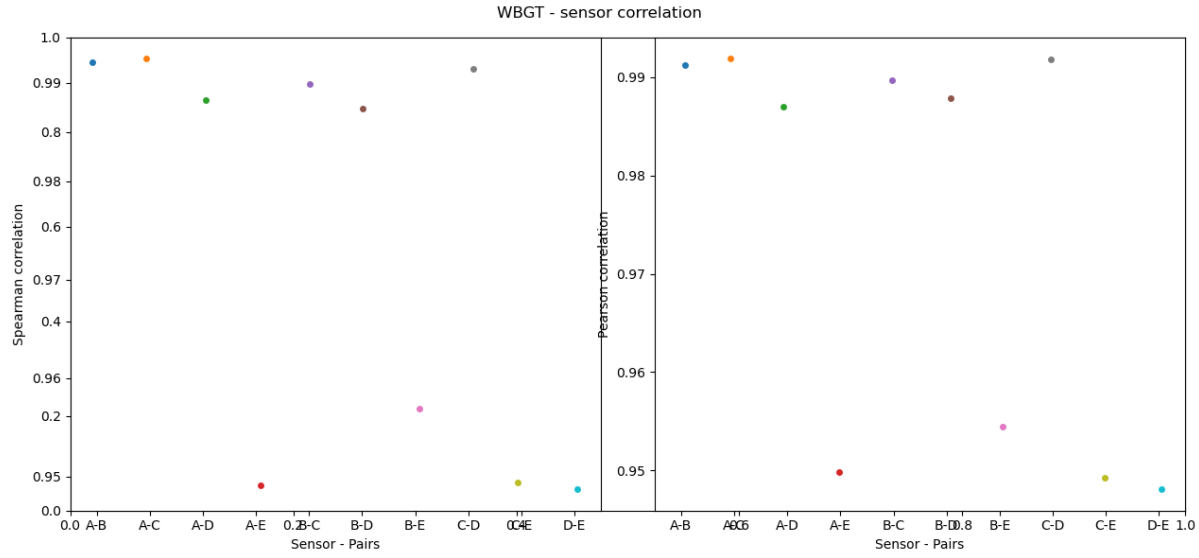

Figure 12: Scatter plot with Spearmann's and Pearson's rank coefficients for WBGT

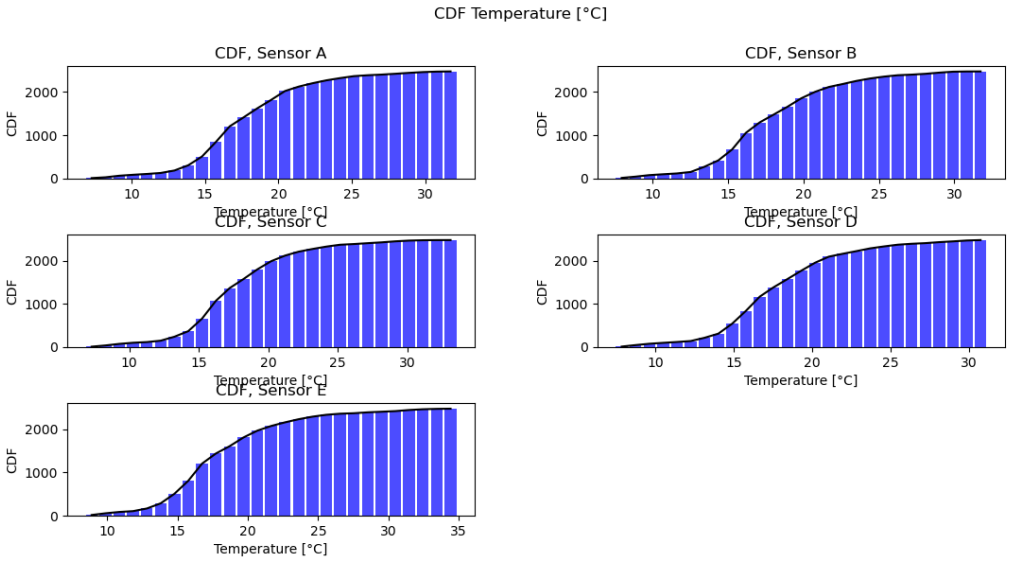

Figure 13: CDF for all the sensors and for variables Temperature

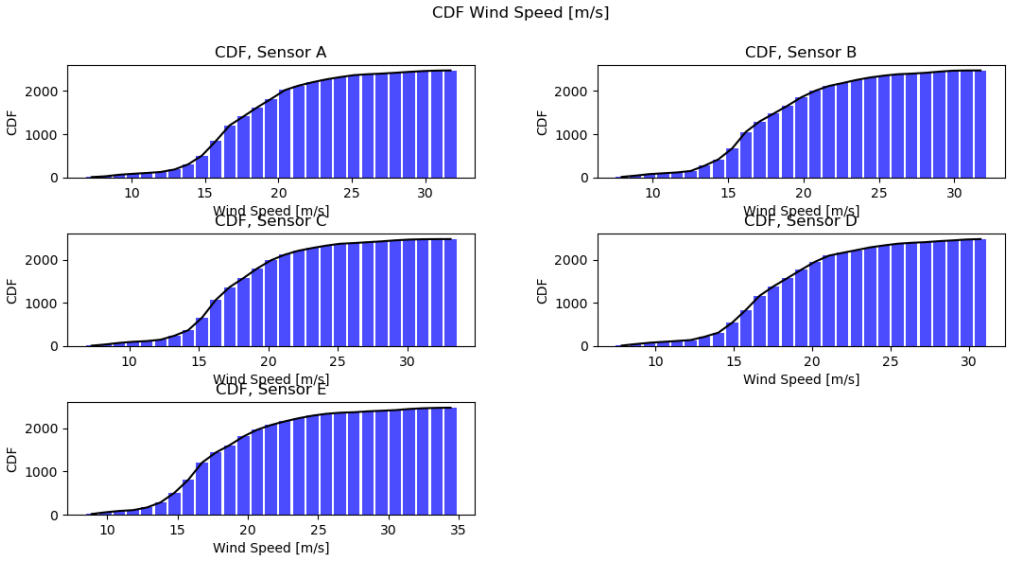

Figure 14: CDF for all the sensors and for variables Wind Speed