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

# Homework 1

**Guilherme Spinoza Andreo (5383994)**

## 1 A1

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

Firstly, due to the high number of variables in our data [1], we cannot examine in depth all of the descriptive statistics. From the results, we can observe that the means of the same variables are quite close yet differ. Big differences in the mean are mostly evident in Wind Speed, Head Wind Speed and Crosswind Speed, especially in sensor E in comparison with all the other sensors.

| Sensor, Variable | Mean | Var | SD |
|---|---|---|---|
| A - True Direction | 209.4063 | 10108.94 | 100.5432 |
| B - True Direction | 183.4124 | 9977.218 | 99.88602 |
| C - True Direction | 183.5889 | 7703.363 | 87.7688 |
| D - True Direction | 198.3266 | 8133.89 | 90.18808 |
| E - True Direction | 223.9564 | 9308.285 | 96.47945 |
| A - Wind Speed | 1.290307 | 1.251154 | 1.11855 |
| B - Wind Speed | 1.242124 | 1.301502 | 1.140834 |
| C - Wind Speed | 1.371463 | 1.43092 | 1.196211 |
| D - Wind Speed | 1.581649 | 1.739817 | 1.319021 |
| E - Wind Speed | 0.596242 | 0.511227 | 0.715001 |
| A - Head Wind Speed | 0.964943 | 0.926593 | 0.962597 |
| B - Head Wind Speed | 0.835622 | 0.878585 | 0.937329 |
| C - Head Wind Speed | 0.963298 | 1.042575 | 1.021066 |
| D - Head Wind Speed | 1.210509 | 1.451503 | 1.204783 |
| E - Head Wind Speed | 0.438505 | 0.315942 | 0.562087 |
| A - Crosswind Speed | 0.16353 | 1.03494 | 1.01732 |
| B - Crosswind Speed | -0.12981 | 1.256719 | 1.121035 |
| C - Crosswind Speed | -0.26289 | 1.271732 | 1.127711 |
| D - Crosswind Speed | -0.30057 | 1.232503 | 1.110181 |
| E - Crosswind Speed | 0.194949 | 0.319073 | 0.564866 |
| A - Temperature | 17.9691 | 15.86427 | 3.982998 |
| B - Temperature | 18.06543 | 16.62907 | 4.077875 |
| C - Temperature | 17.91314 | 16.10454 | 4.013046 |
| D - Temperature | 17.99636 | 16.10559 | 4.013177 |
| E - Temperature | 18.35394 | 19.04313 | 4.363844 |

| | | | |
|---|---|---|---|
| A - Globe Temperature | 21.54459 | 68.19135 | 8.257806 |
| B - Globe Temperature | 21.79943 | 66.04932 | 8.127073 |
| C - Globe Temperature | 21.58739 | 67.9413 | 8.242652 |
| D - Globe Temperature | 21.3593 | 61.20225 | 7.823187 |
| E - Globe Temperature | 21.17616 | 63.2155 | 7.950818 |
| A - Wind Chill | 17.83821 | 16.26445 | 4.03292 |
| B - Wind Chill | 17.94592 | 17.03583 | 4.127448 |
| C - Wind Chill | 17.773 | 16.54112 | 4.067078 |
| D - Wind Chill | 17.83537 | 16.55685 | 4.069011 |
| E - Wind Chill | 18.29402 | 19.13706 | 4.374593 |
| A - Relative Humidity | 78.18477 | 376.0101 | 19.39098 |
| B - Relative Humidity | 77.87831 | 408.623 | 20.21443 |
| C - Relative Humidity | 77.96285 | 374.6226 | 19.35517 |
| D - Relative Humidity | 77.94204 | 389.856 | 19.74477 |
| E - Relative Humidity | 76.79305 | 406.4945 | 20.16171 |
| A - Heat Stress Index | 17.8996 | 14.99685 | 3.872576 |
| B - Heat Stress Index | 18.00428 | 15.43916 | 3.929269 |
| C - Heat Stress Index | 17.82825 | 15.35625 | 3.918706 |
| D - Heat Stress Index | 17.92162 | 15.11764 | 3.888141 |
| E - Heat Stress Index | 18.28642 | 18.47524 | 4.298283 |
| A - Dew Point | 13.55388 | 9.723472 | 3.118248 |
| B - Dew Point | 13.53086 | 9.636518 | 3.104274 |
| C - Dew Point | 13.45812 | 10.08415 | 3.175555 |
| D - Dew Point | 13.50861 | 10.07188 | 3.173623 |
| E - Dew Point | 13.55879 | 9.422585 | 3.069623 |
| A - Psychro Wet Bulb Temperature | 15.27072 | 6.944027 | 2.635152 |
| B - Psychro Wet Bulb Temperature | 15.29552 | 6.770263 | 2.601973 |
| C - Psychro Wet Bulb Temperature | 15.19665 | 7.239313 | 2.690597 |
| D - Psychro Wet Bulb Temperature | 15.26019 | 7.044403 | 2.654129 |
| E - Psychro Wet Bulb Temperature | 15.40667 | 6.997445 | 2.645268 |
| A - Station Pressure | 1016.168 | 38.47127 | 6.202521 |
| B - Station Pressure | 1016.657 | 36.84193 | 6.069756 |
| C - Station Pressure | 1016.689 | 37.69149 | 6.13934 |
| D - Station Pressure | 1016.728 | 34.98778 | 5.915047 |
| E - Station Pressure | 1016.166 | 38.93991 | 6.240185 |
| A - Barometric Pressure | 1016.128 | 38.46795 | 6.202254 |
| B - Barometric Pressure | 1016.616 | 36.82887 | 6.068679 |
| C - Barometric Pressure | 1016.652 | 37.67562 | 6.138047 |
| D - Barometric Pressure | 1016.689 | 34.95233 | 5.912049 |
| E - Barometric Pressure | 1016.128 | 38.93518 | 6.239806 |
| A - Altitude | -25.9871 | 2663.641 | 51.61047 |
| B - Altitude | -30.0582 | 2545.708 | 50.45501 |
| C - Altitude | -30.3387 | 2608.535 | 51.07382 |
| D - Altitude | -30.6532 | 2419.724 | 49.19069 |
| E - Altitude | -25.9612 | 2692.353 | 51.88789 |
| A - Density Altitude | 137.3166 | 26510.04 | 162.8191 |
| B - Density Altitude | 135.5808 | 26863.31 | 163.9003 |

| | | | |
|---|---|---|---|
| C - Density Altitude | 129.6229 | 26986.6 | 164.276 |
| D - Density Altitude | 132.4111 | 26516.13 | 162.8377 |
| E - Density Altitude | 150.84 | 29714.93 | 172.3802 |
| A - NA Wet Bulb Temperature | 15.98154 | 10.01211 | 3.164191 |
| B - NA Wet Bulb Temperature | 15.99681 | 9.809254 | 3.131973 |
| C - NA Wet Bulb Temperature | 15.93424 | 10.48028 | 3.237326 |
| D - NA Wet Bulb Temperature | 15.91564 | 9.987434 | 3.16029 |
| E - NA Wet Bulb Temperature | 15.93689 | 9.432184 | 3.071186 |
| A - WBGT | 17.25432 | 16.13526 | 4.016872 |
| B - WBGT | 17.32197 | 15.83536 | 3.979366 |
| C - WBGT | 17.22502 | 16.54675 | 4.067769 |
| D - WBGT | 17.1768 | 15.50718 | 3.937916 |
| E - WBGT | 17.18554 | 15.48987 | 3.935717 |
| A - TWL | 301.3929 | 814.7666 | 28.54412 |
| B - TWL | 299.4517 | 790.0692 | 28.10817 |
| C - TWL | 301.8998 | 766.5335 | 27.68634 |
| D - TWL | 305.2546 | 616.0098 | 24.81954 |
| E - TWL | 284.1153 | 1289.913 | 35.91536 |
| A - Direction, Mag | 208.9051 | 10105.68 | 100.527 |
| B - Direction, Mag | 183.2173 | 9975.447 | 99.87716 |
| C - Direction, Mag | 183.0837 | 7704.62 | 87.77597 |
| D - Direction, Mag | 197.8262 | 8135.316 | 90.19598 |
| E - Direction, Mag | 223.8966 | 9268.008 | 96.27049 |

Table 1: Measures of means, variability and standard deviations.

In Table 1, we can see that the Wind Speed mean for sensor E is clearly lower in comparison with the means of Sensors A, B, C and D. Another clear distinctive mean is the Altitude, which is negative, meaning that the sensors are below sea level. However, for a more profound analysis, we require more representative data.

## 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?*

When the number of bins is changed, the histogram changes as well. With 50 bins, the variability of the temperatures is higher and the distribution is clearer than with 5 bins. Moreover, the higher the number of bins, the clearer it becomes to check the most frequent values, the outliers and distinguish the mean. In general, the more the bins, the clearer the distribution.

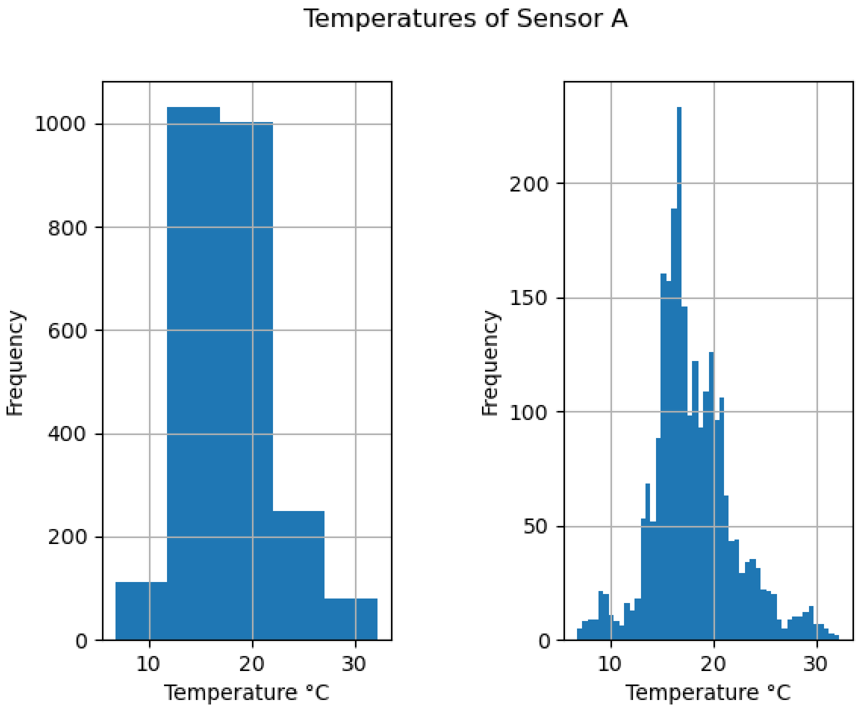

Figure 1: Temperature Histograms Sensor A: bins = 5, bins = 50

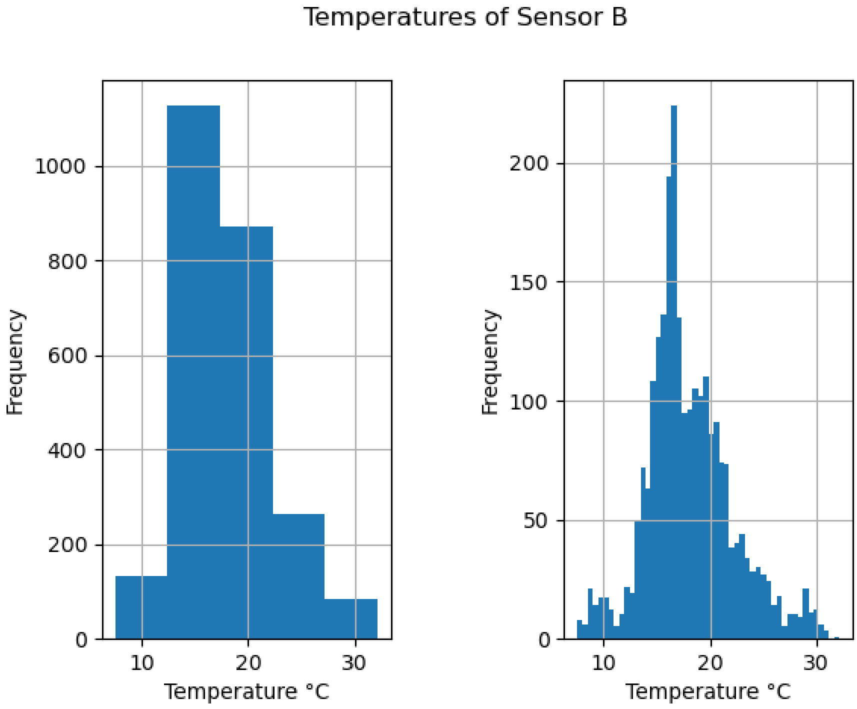

Figure 2: Temperature Histograms Sensor B: bins = 5, bins = 50

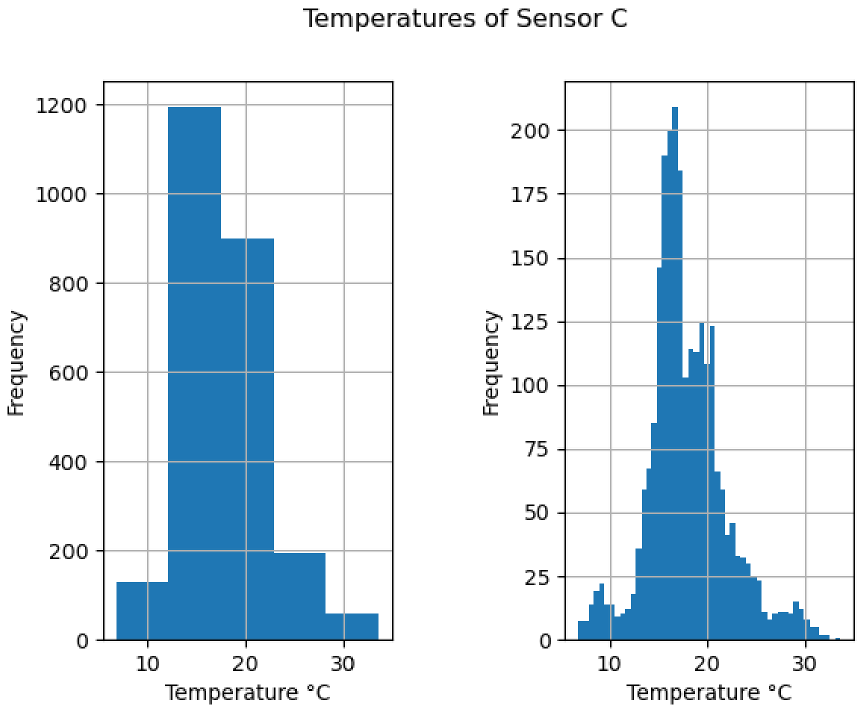

Figure 3: Temperature Histograms Sensor C: bins = 5, bins = 50

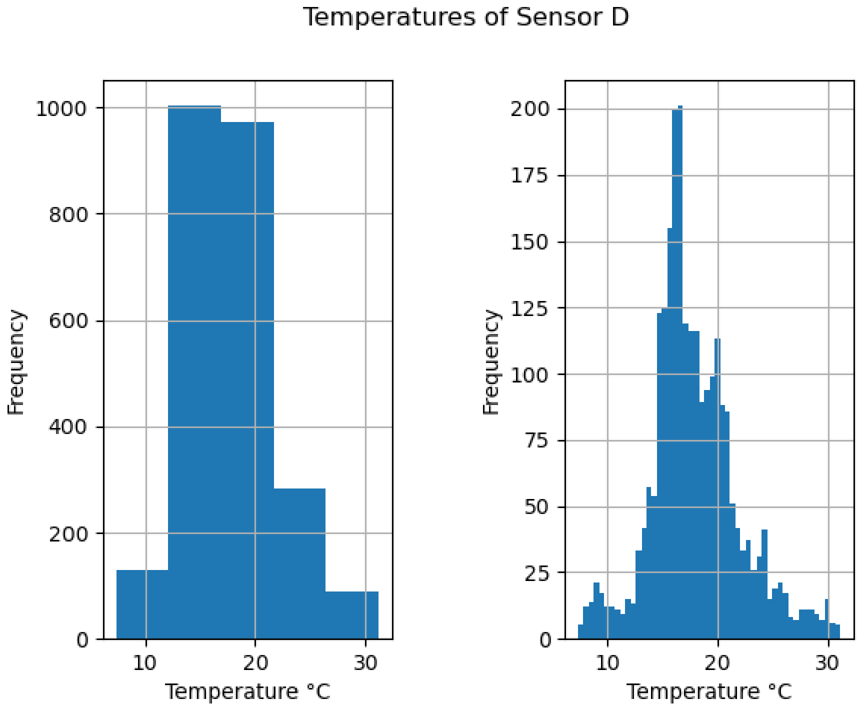

Figure 4: Temperature Histograms Sensor D: bins = 5, bins = 50

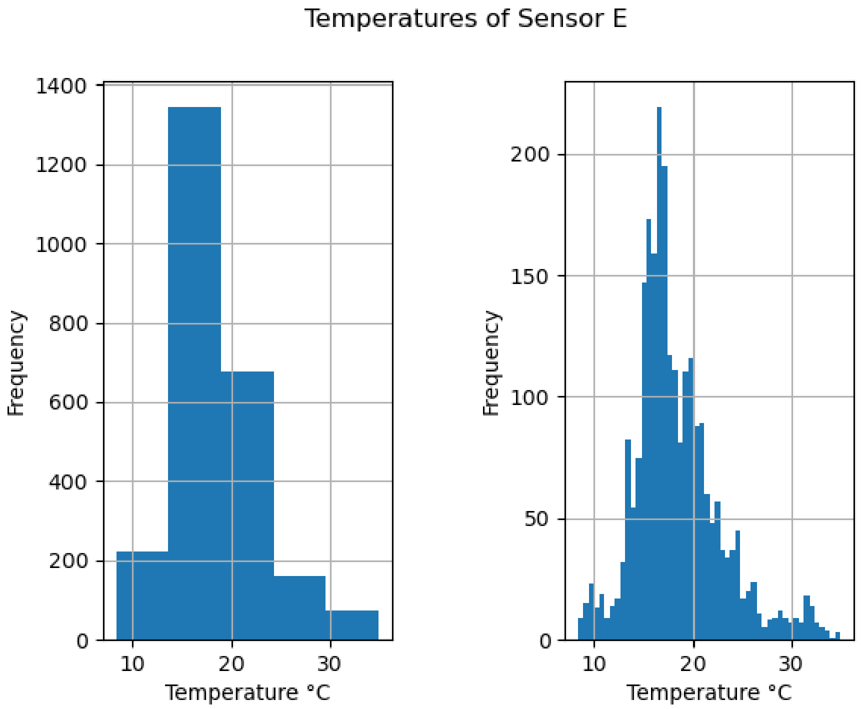

Figure 5: Temperature Histograms Sensor E: bins = 5, bins = 50

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

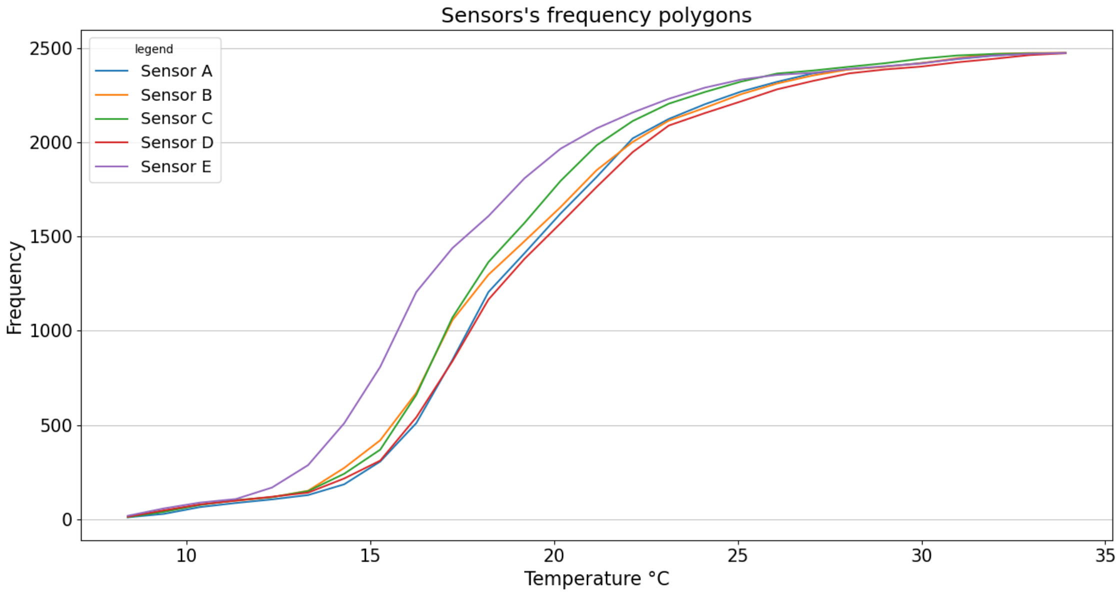

Figure 6: Frequency Polygons for Temperature Values

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

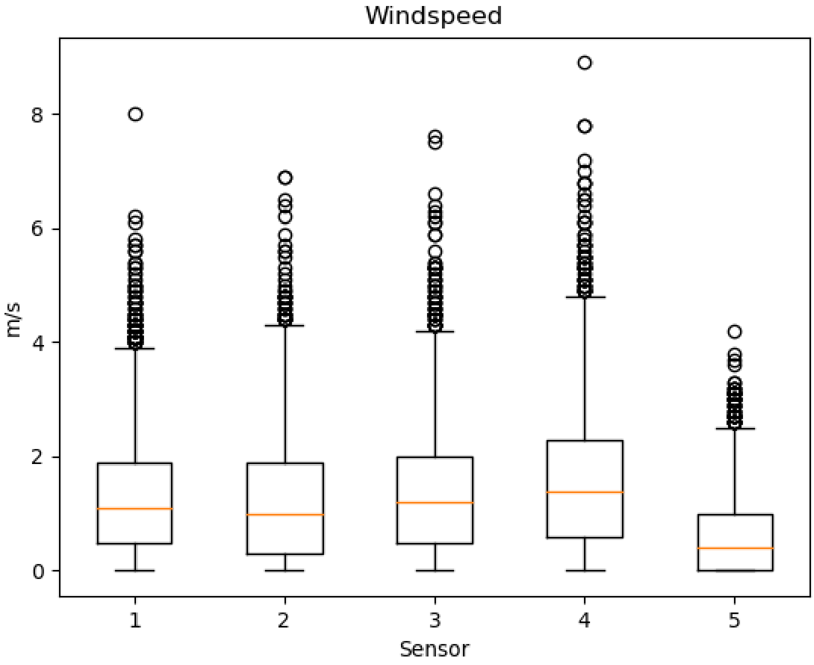

Figure 7: Boxplot for Wind Speed

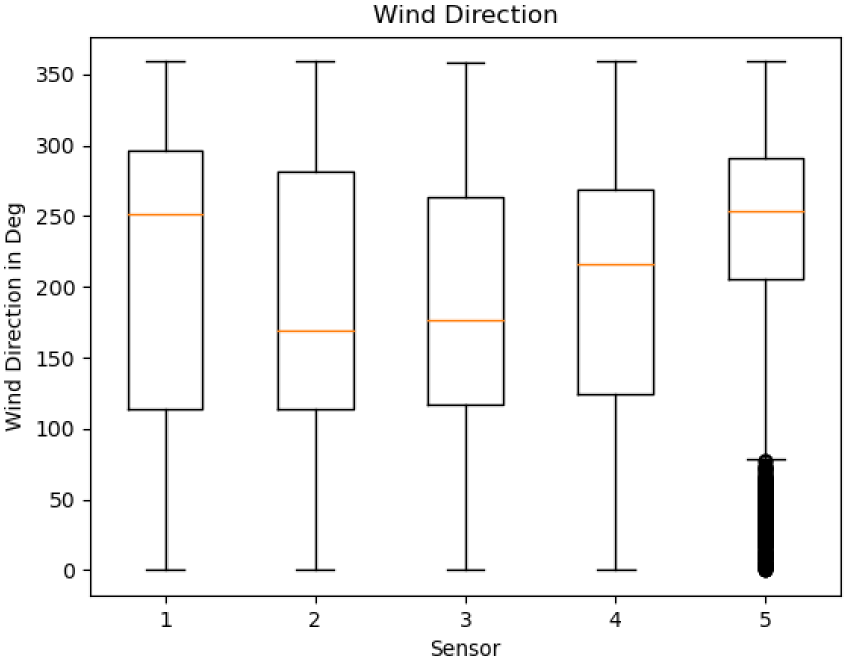

Figure 8: Boxplot for Wind Direction

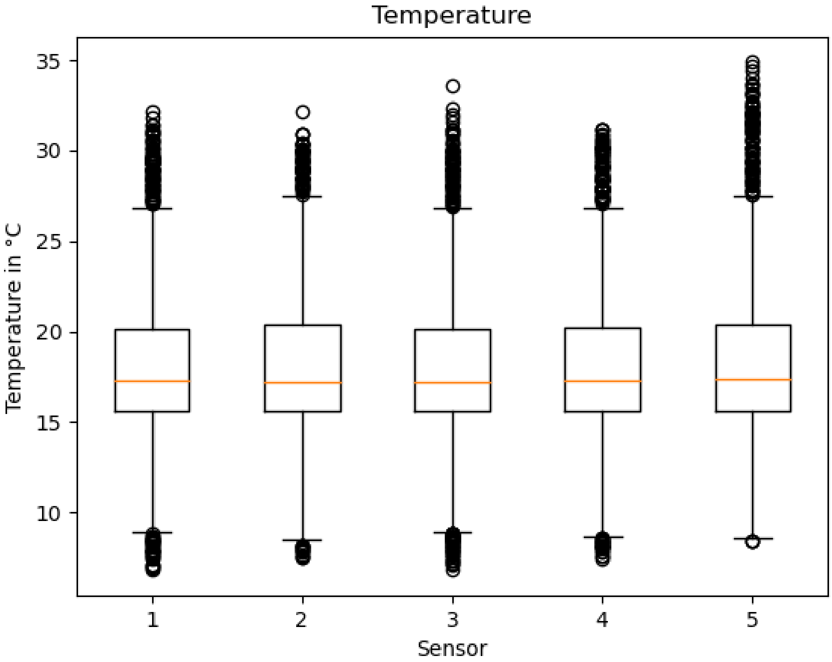

Figure 9: Boxplot for Temperature

# 2 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?*

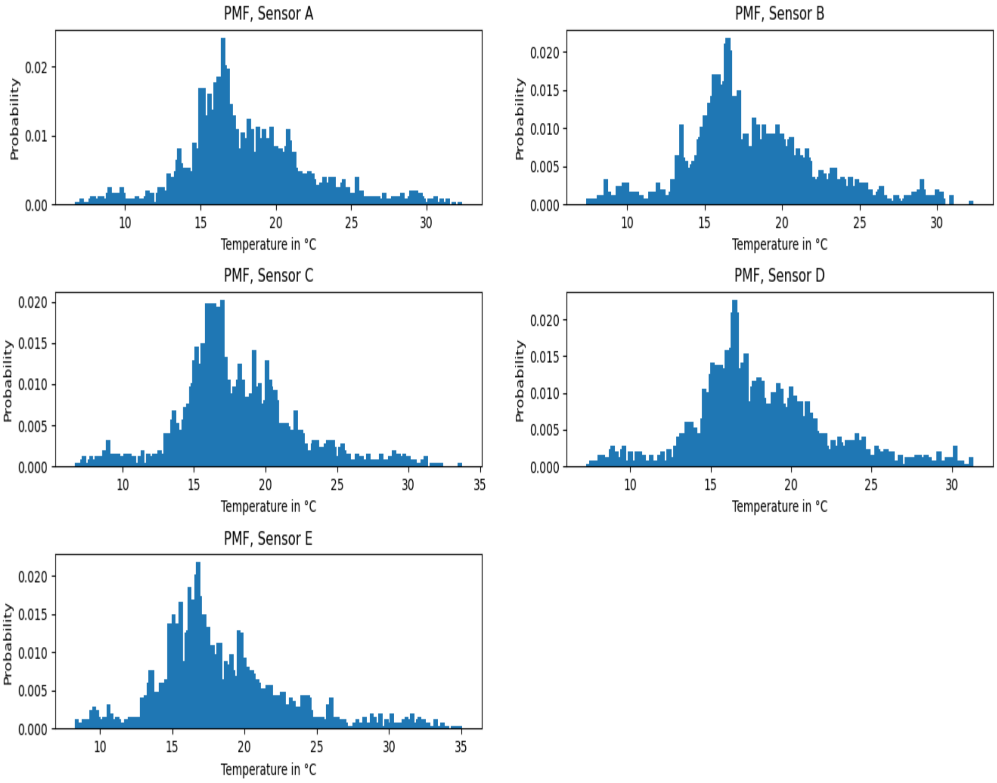

Figure 10: Probability Density Function for Temperature

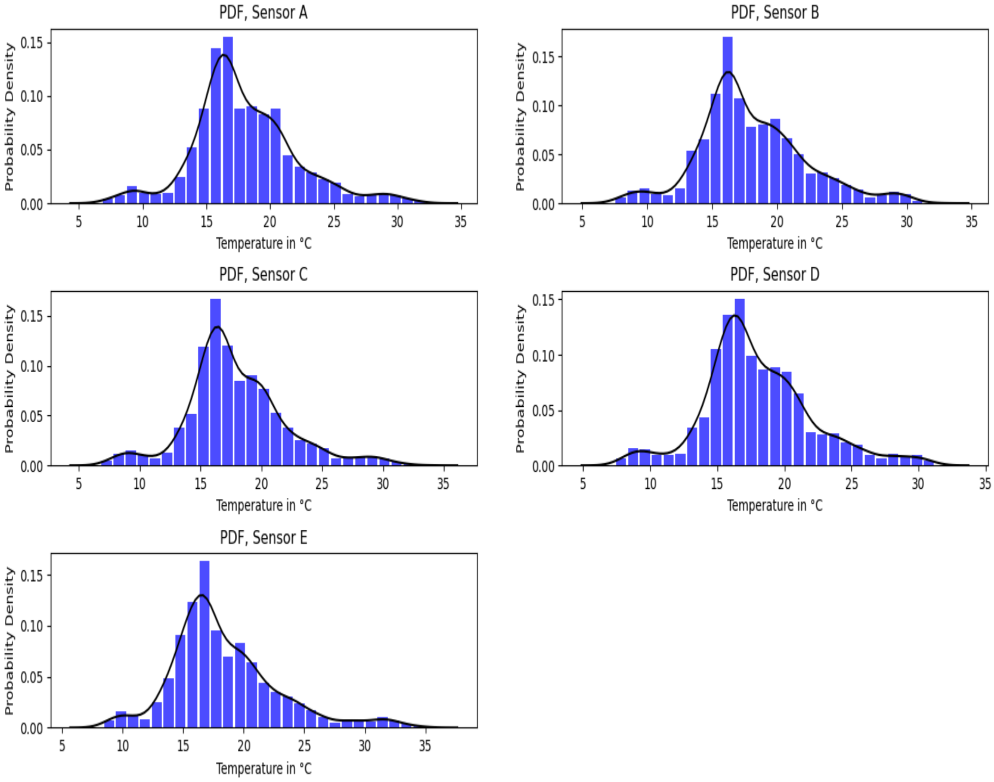

Figure 11: Probability Density Function for Temperature

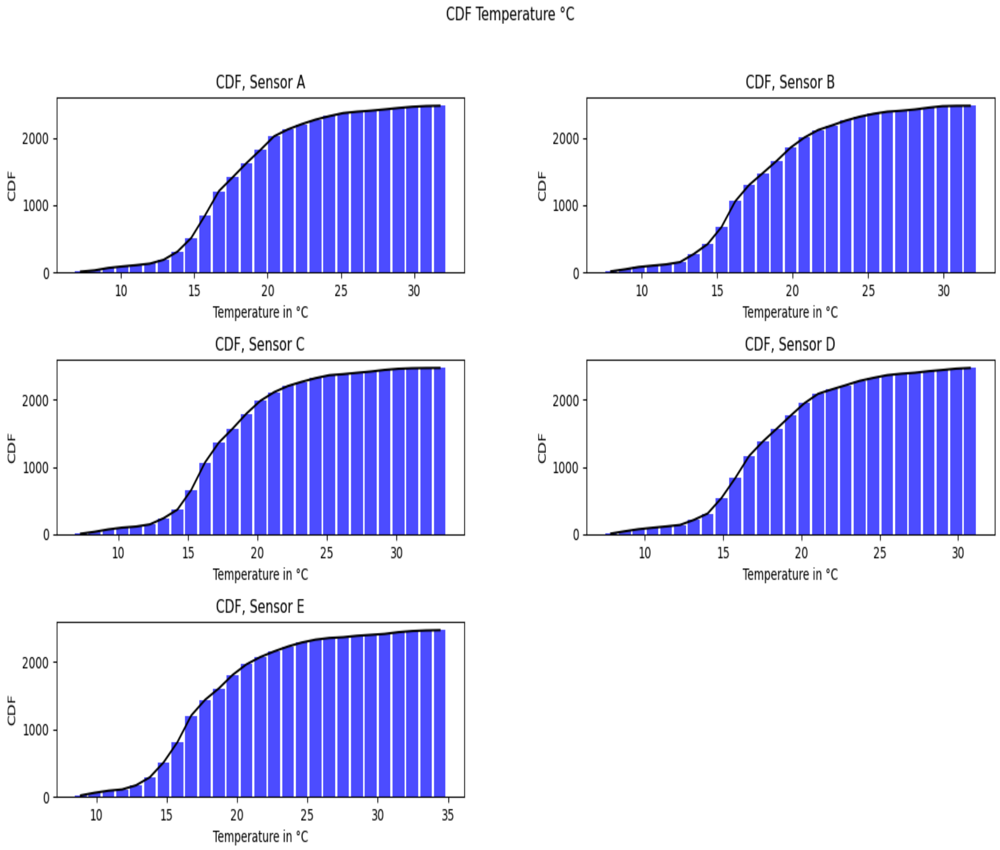

Figure 12: Cumulative Distribution Function for Temperature

First of all, we can observe that the PMF and PDF results are very similar because of the high amount of data to our samples [1], furthermore, most of the data is contained in a range of 0 to 35 °C. The PMF and PDF are both asymmetrical distributions with long right tails. Nevertheless, PDF and PMF of sensor E seems to have the longest right tail, which could indicate that there more outliers in the higher temperature values.

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

Kernel and PDF graphs of the same sensors are almost identical. The Kernel graphs offer a smoother picture of the data visualization. We can observe that A and B are very similar, however, E has definitely lower wind speed values than all of the other sensors. Most of the values in sensors A, B , C and D are contained in in the range of 0 to 4, however sensor E has very few values above 3. This could mean that sensor is at a location where it receives less wind comparatively to the other sensors.

# 3 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.*

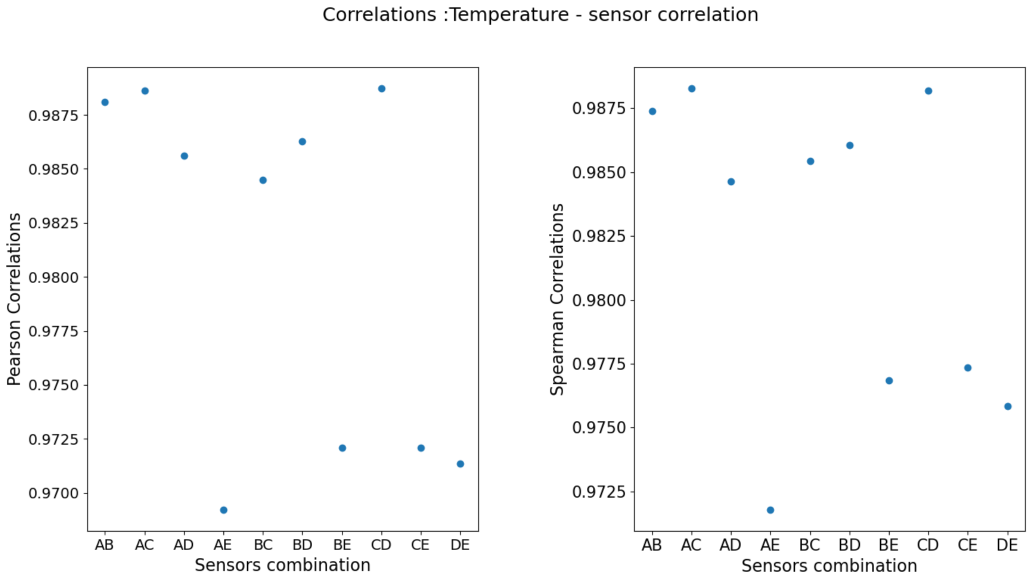

Figure 13: Scatterplots for Temperature

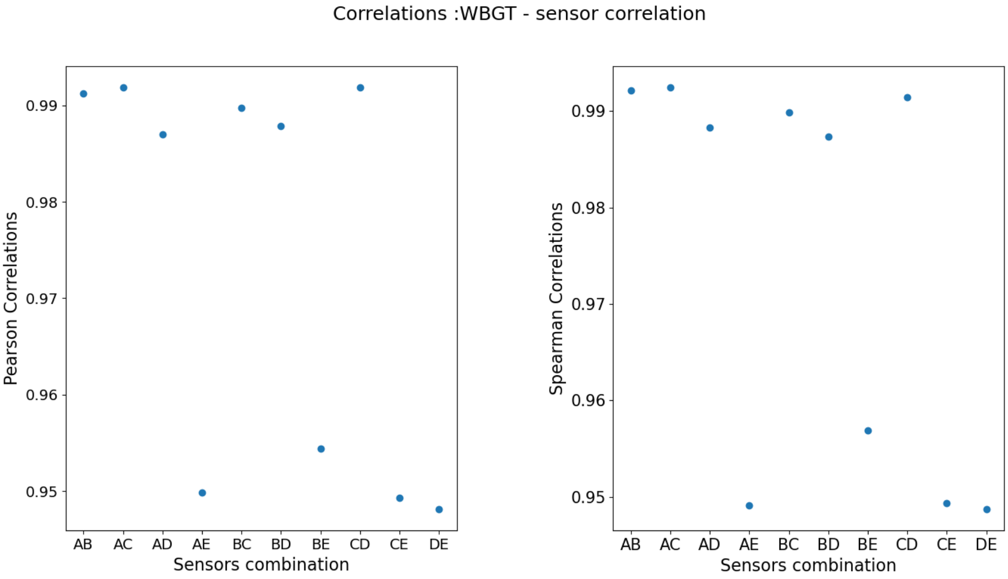

Figure 14: Scatterplots for Wet Bulb Globe Temperature

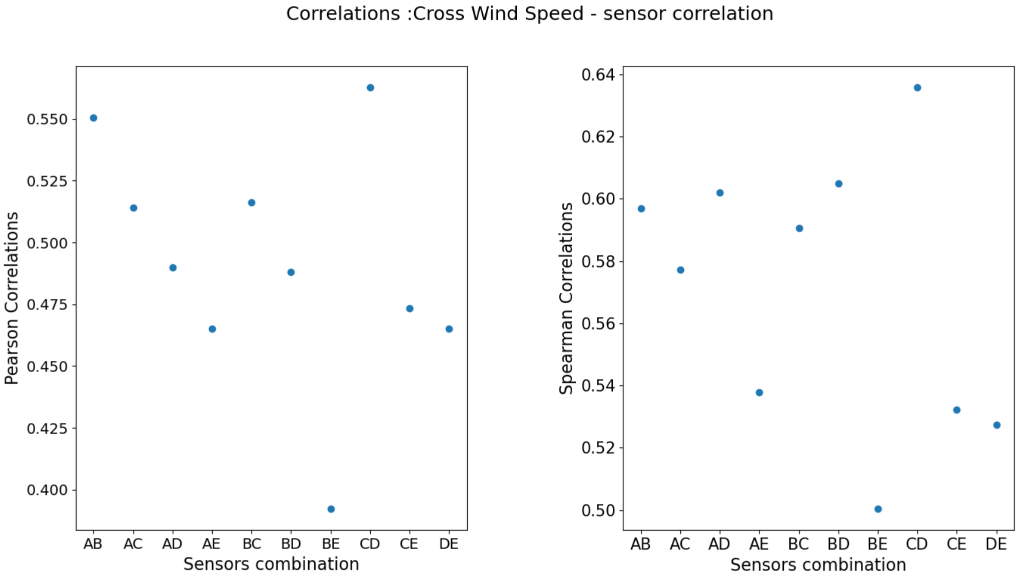

Figure 15: Scatterplots for Crosswinds

### 3.2 *What can you say about the sensors' correlations?*

With Temperature and WBGT, we can see that all of the sensors have a very high positive correlation in both the Pearson and Spearman correlation graphs, which is greater than or equal to 95% with AB, AC and CD possessing the highest correlations. Sensor E has the lowest correlation with the other sensors, yet still very high. the correlation between AE appears to have the lowest in temperature, along with DE in WBGT.

Cross wind speed on the other hand, has much more modest positive correlations in both correlation graphs, ranging from 40% to 64%. Clearly, the most correlated crosswind speed is between sensors C and D with 64% in the Spearman correlation, while the least correlated crosswind speeds are between sensors B and E which is lower than 40% in the Pearson correlation. Similarly with the previous variables, correlations of sensor E with the rest of the sensors are lower than the rest.

All in all, Sensor E seems to be less correlated with all the rest four sensors in all three different variables under examination. Next to that, BE has the biggest correlation among E correlations with other sensors in WBGT and in Temperature, along with CE, but the least correlated in cross wind speed.

### 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.*

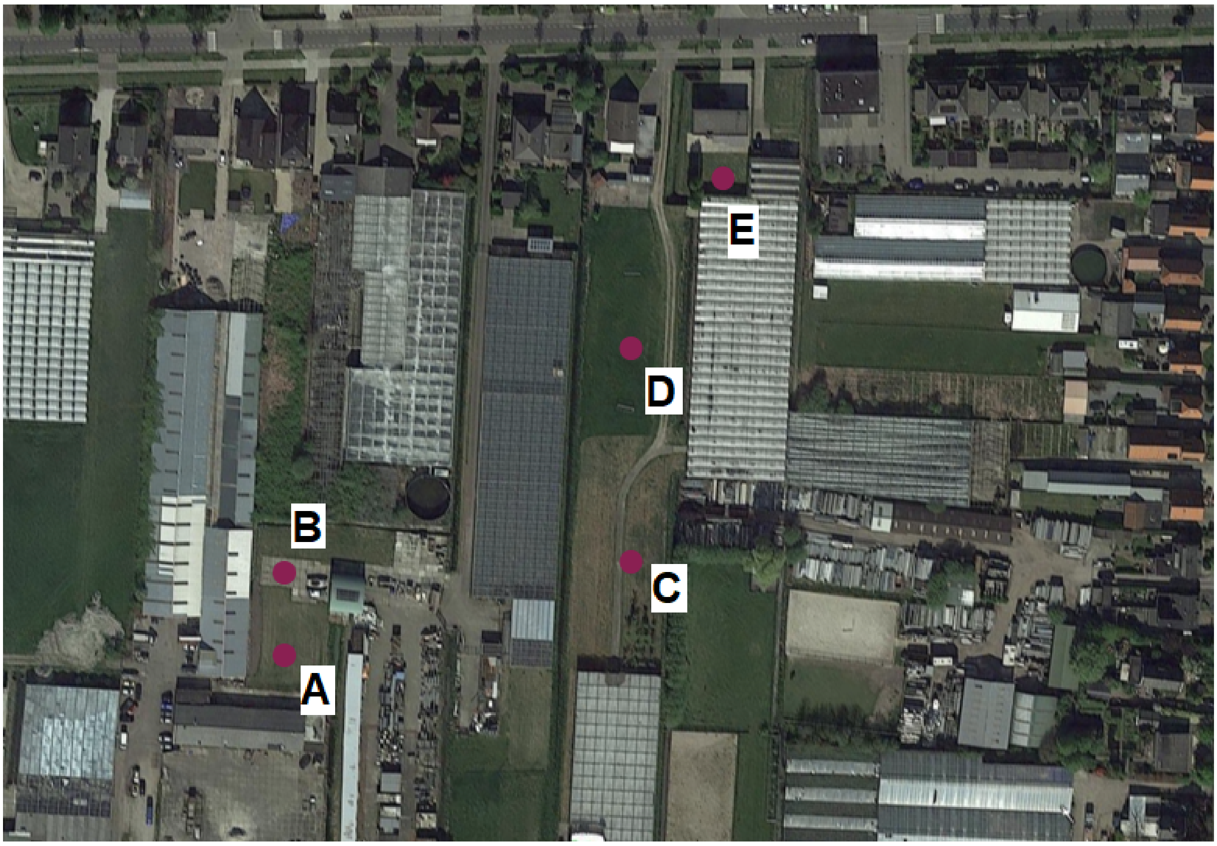

Figure 16: Sensors possible location

According to the to the Spearman and Pearson correlations, It is very certain that E would be the sensor on the upper right. Because sensor E's results are the most discrepant in comparison with the others sensors, therefore, it is reasonable to assume that it is isolated. Furthermore, the location on the upper right is also covered with walls, which would effectively explain the weak Wind Speed results of Sensor E with our other representative data.

I would also hypothesize that Sensors A and B are together, while B and C are together, due to the high value correlations presented in the previous section. Moreover, we can observe that D has the highest correlation with C and A has a extremely high correlation with C in all of the graphs. However, to precisely determine where the sensors are located, we would need to produce more relevant representative data to make a firm conclusion.

# 4 A4

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

| **Variable, Sensor** | **Lower Interval Limit** | **Higher Interval Limit** |
|---|---|---|
| Temperature, A | 17.81214113267346 | 18.126065652463858 |
| Temperature, B | 17.90472689963894 | 18.226129320070267 |
| Temperature, C | 17.754926235060246 | 18.071347006653575 |
| Temperature, D | 17.83814660824381 | 18.15457772482005 |
| Temperature, E | 18.181933946027776 | 18.525944841851015 |
| Wind Speed, A | 1.246227038990971 | 1.3343868543854427 |
| Wind Speed, B | 1.1971663346979249 | 1.287082453670411 |
| Wind Speed, C | 1.3243037885948932 | 1.418622646328308 |
| Wind Speed, D | 1.5296480419653757 | 1.633650260379006 |
| Wind Speed, E | 0.5680599051948441 | 0.6244249432900044 |

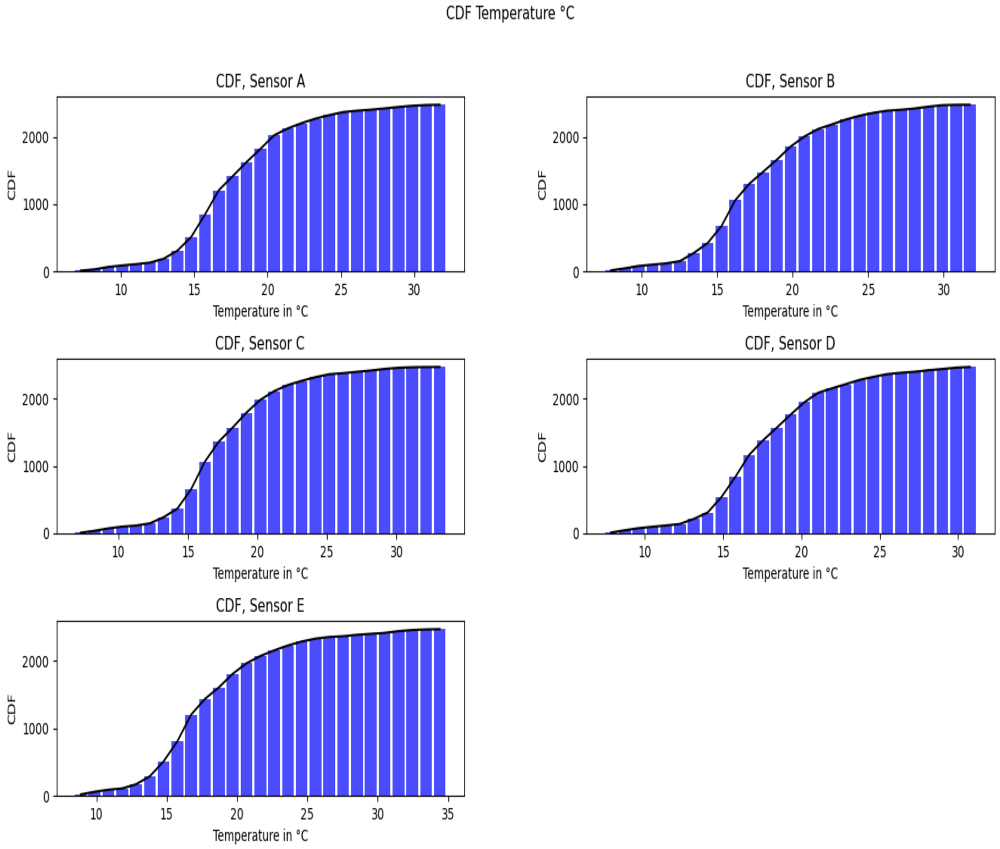

Figure 17: Cumulative Distribution Function for Temperature

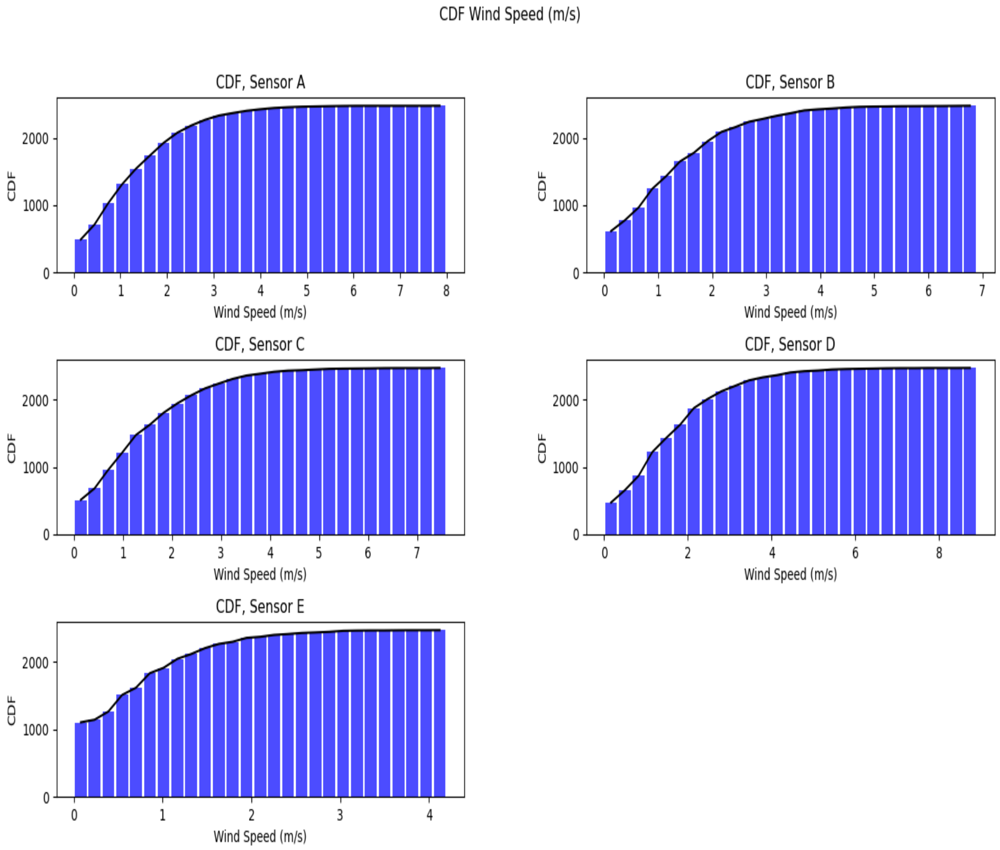

Figure 18: Cumulative Distribution Function for Temperature

## 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.*

## 4.3 *What could you conclude from the p-values?*

| **Sensors and Variable** | **T-Value** | **P-Value** |
|---|---|---|
| E-D, Temperature | 3.000 | .0027 |
| D-C, Temperature | 0.729 | .4657 |
| C-B, Temperature | -1.324 | .1854 |
| B-A, Temperature | 0.840 | .4004 |
| E-D, Wind Speed | -32.673 | .3729 $e^{-212}$ |
| D-C, Wind Speed | 5.871 | .6101 $e^{-09}$ |
| C-B, Wind Speed | 3.892 | .0001 |
| B-A, Wind Speed | 1.500 | .1335 |

We can observe that the p-values for the sensor correlation ED for temperature and ED for wind speed are below 0.05, in this case, we can reject the null hypothesis that there's no difference between the time series for the sensor pairings and we can provide support that there is a significant statistical difference. It's also logical to make the same conclusion for the correlations for DC and CB for Wind Speed.
For other time series correlations that have p-values above 0.05, in this case, DC, CB, BA

in relation to temperature and BA in relation to wind speed are not statistically significant because the p-values are greater than 0.05, therefore, we cannot say that the null hypothesis is rejected, only that the values are not statistically significant and indicates strong evidence for the null hypothesis.

# References

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