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

Homework 1

Konstantinos Pantelios, (5374367)

The data that were used in the following report were extracted from TU Dleft data portal. [1]

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?*

For 1.1, the code computes the mean, the variance and the standard deviation of all the different variables (19) from sensors A, B, C, D and E. The results (Figure 1) are displayed separately for each of the sensors and each variable takes occupies different rows.

```
 === VALUES FOR SENSOR: A ===========================================================================================
True Direction->          mean: 209.40630048465266        variance: 10104.857537040565    standard deviation: 100.52292045618535
Wind Speed->              mean: 1.290306946688207         variance: 1.2506491788047323    standard deviation: 1.1183242726529423
Crosswind Speed->         mean: 0.9649434571890144        variance: 0.9262185347673694    standard deviation: 0.9624024806531669
Headwind Speed->          mean: 0.16352988691437803       variance: 1.034522111788517     standard deviation: 1.0171146011087036
Temperature->             mean: 17.96910339256866         variance: 15.857862039390751    standard deviation: 3.9821931192988056
Globe Temperature->       mean: 21.544588045234246        variance: 68.1638115831204      standard deviation: 8.256137812750003
Wind chill->              mean: 17.838206785137317        variance: 16.257877882926497    standard deviation: 4.032105886869354
Relative humidity->       mean: 78.18477382875606         variance: 375.8581970813183     standard deviation: 19.387062621277064
Heat stress index->       mean: 17.899596122778675        variance: 14.990791436236988    standard deviation: 3.8717943432260173
Dew point->               mean: 13.553877221324719        variance: 9.719544740983556     standard deviation: 3.1176184405702307
Psychro Wet Bulb Temp->   mean: 15.270718901453955        variance: 6.9412225849577585    standard deviation: 2.6346200076970794
Station pressure->        mean: 1016.1682552504037        variance: 38.45572894292478     standard deviation: 6.201268333407673
Barometric pressure->     mean: 1016.1284329563813        variance: 38.4524145072175      standard deviation: 6.201001089115974
Altitude->                mean: -25.98707592891761        variance: 2662.5652610782413    standard deviation: 51.60005097941514
Density Altitude->        mean: 137.31663974151857        variance: 26499.337542182006    standard deviation: 162.7861712252672
NA Wet Bulb Temperature->mean: 15.981542810985461         variance: 10.008064017149449    standard deviation: 3.1635524362889025
WBGT->                    mean: 17.25432148626817         variance: 16.128741421686968    standard deviation: 4.016060435512266
TWL->                     mean: 301.39293214862676        variance: 814.4374985107435     standard deviation: 28.538351362872092
Direction , Mag->         mean: 208.90508885298868        variance: 10101.595596074496    standard deviation: 100.50669428488082

 === VALUES FOR SENSOR: B ===========================================================================================
True Direction->          mean: 183.41235864297255        variance: 9973.18819944488      standard deviation: 99.86585101747684
Wind Speed->              mean: 1.242124394184168         variance: 1.300975939291838     standard deviation: 1.1406033224972818
Crosswind Speed->         mean: 0.8356219709208401        variance: 0.8782302674332721    standard deviation: 0.9371394066163647
Headwind Speed->          mean: -0.12980613893376414      variance: 1.25621175563275      standard deviation: 1.1208085276409838
Temperature->             mean: 18.065428109854604        variance: 16.622350826415005    standard deviation: 4.077051732123962
Globe Temperature->       mean: 21.799434571890146        variance: 66.02264103731852     standard deviation: 8.125431744671696
Wind chill->              mean: 17.945920840064623        variance: 17.028945395995418    standard deviation: 4.126614277588277
Relative humidity->       mean: 77.87831179321486         variance: 408.4579746943844     standard deviation: 20.21034326018201
Heat stress index->       mean: 18.00428109854604         variance: 15.432921898366484    standard deviation: 3.928475772913266
Dew point->               mean: 13.530856219709205        variance: 9.632626245886195     standard deviation: 3.103647248945375
Psychro Wet Bulb Temp->   mean: 15.295516962843294        variance: 6.767528367644411     standard deviation: 2.6014473601525
Station pressure->        mean: 1016.6570274636512        variance: 36.82705481507772     standard deviation: 6.068529872636183
Barometric pressure->     mean: 1016.6164781906298        variance: 36.81399341269066     standard deviation: 6.067453618503454
Altitude->                mean: -30.05815831987076        variance: 2544.679977868311     standard deviation: 50.444821120391644
Density Altitude->        mean: 135.58077544426493        variance: 26852.460761272676    standard deviation: 163.86720465447831
NA Wet Bulb Temperature->mean: 15.996809369951535         variance: 9.805292727795887     standard deviation: 3.131340404331009
WBGT->                    mean: 17.321970920840062        variance: 15.82895992807201     standard deviation: 3.978562545451813
TWL->                     mean: 299.45169628432956        variance: 789.7501304021025     standard deviation: 28.102493312909132
Direction , Mag->         mean: 183.2172859450727         variance: 9971.418053377041     standard deviation: 99.85698800473125
```

```
 === VALUES FOR SENSOR: C ==========================================================================================
True Direction->          mean: 183.58892481810832        variance: 7700.24936804366      standard deviation: 87.75106476871754
Wind Speed->              mean: 1.3714632174616006        variance: 1.4303416746777642    standard deviation: 1.1959689271372247
Crosswind Speed->         mean: 0.9632983023443816        variance: 1.0421533895929145    standard deviation: 1.0208591428757028
Headwind Speed->          mean: -0.2628940986257074       variance: 1.2712181399570899    standard deviation: 1.1274830996325798
Temperature->             mean: 17.91313662085691         variance: 16.09802872264436     standard deviation: 4.012234878798145
Globe Temperature->       mean: 21.587388843977365        variance: 67.91384257555865     standard deviation: 8.240985534240346
Wind chill->              mean: 17.77299919159256         variance: 16.53443667987        standard deviation: 4.0662558551903745
Relative humidity->       mean: 77.96285367825384         variance: 374.47121918069183    standard deviation: 19.35125885260935
Heat stress index->       mean: 17.82825383993533         variance: 15.350046506300938    standard deviation: 3.917913539921592
Dew point->               mean: 13.458124494745352        variance: 10.080073442868077    standard deviation: 3.174913139420995
Psychro Wet Bulb Temp->   mean: 15.196645109135003        variance: 7.236387289573896     standard deviation: 2.690053399019041
Station pressure->        mean: 1016.689329021827         variance: 37.67625638083111     standard deviation: 6.13809875945566
Barometric pressure->     mean: 1016.6518997574777        variance: 37.660394531584316    standard deviation: 6.136806541808558
Altitude->                mean: -30.338722716248988       variance: 2607.4802547953855    standard deviation: 51.063492387373834
Density Altitude->        mean: 129.62287793047696        variance: 26975.694884846056    standard deviation: 164.2427924897956
NA Wet Bulb Temperature->mean: 15.934236054971707         variance: 10.476042928918309    standard deviation: 3.2366715818751692
WBGT->                    mean: 17.22502021018593         variance: 16.540057093366812    standard deviation: 4.066946900731163
TWL->                     mean: 301.8997574777688         variance: 766.2236781950229     standard deviation: 27.68074562209304
Direction , Mag->         mean: 183.08367016976555        variance: 7701.505933821689     standard deviation: 87.75822430873183

 === VALUES FOR SENSOR: D ==========================================================================================
True Direction->          mean: 198.32659660468877        variance: 8130.602307980361     standard deviation: 90.16985254496295
Wind Speed->              mean: 1.5816491511721908        variance: 1.739113529289902     standard deviation: 1.3187545371637217
Crosswind Speed->         mean: 1.2105092966855295        variance: 1.450916232128608     standard deviation: 1.2045398424828495
Headwind Speed->          mean: -0.3005658852061439       variance: 1.2320045302185576    standard deviation: 1.109956994760859
Temperature->             mean: 17.99636216653193         variance: 16.099081349837828    standard deviation: 4.012366053818847
Globe Temperature->       mean: 21.359296685529507        variance: 61.177514622256783    standard deviation: 7.821605629445135
Wind chill->              mean: 17.835367825383994        variance: 16.55015978790578     standard deviation: 4.068188760112513
Relative humidity->       mean: 77.94203718674213         variance: 389.6984592290132     standard deviation: 19.740781626597595
Heat stress index->       mean: 17.92162489894907         variance: 15.11153317215288     standard deviation: 3.887355550004769
Dew point->               mean: 13.50860953920776         variance: 10.067811890385965    standard deviation: 3.172981545862813
Psychro Wet Bulb Temp->   mean: 15.26018593371059         variance: 7.041555503019602     standard deviation: 2.6535929422237317
Station pressure->        mean: 1016.7280113177042        variance: 34.973641396799955    standard deviation: 5.913851654953813
Barometric pressure->     mean: 1016.6888843977364        variance: 34.938198997953734    standard deviation: 5.910854337399437
Altitude->                mean: -30.653193209377527       variance: 2418.745529415378     standard deviation: 49.18074348172644
Density Altitude->        mean: 132.41107518189168        variance: 26505.40781590138     standard deviation: 162.80481508819506
NA Wet Bulb Temperature->mean: 15.915642683912694         variance: 9.983397181945264     standard deviation: 3.1596514336149903
WBGT->                    mean: 17.1767987065481          variance: 15.500916833369384    standard deviation: 3.9371203732384643
TWL->                     mean: 305.254567502021          variance: 615.7608138186043     standard deviation: 24.814528281202612
Direction , Mag->         mean: 197.8261924009701         variance: 8132.027187846571     standard deviation: 90.17775328675344

 === VALUES FOR SENSOR: E ==========================================================================================
True Direction->          mean: 223.95636363636365        variance: 9304.524156473828     standard deviation: 96.459961416506
Wind Speed->              mean: 0.5962424242424242        variance: 0.5110202240587696    standard deviation: 0.7148567856982051
Crosswind Speed->         mean: 0.4385050505050505        variance: 0.3158143307825732    standard deviation: 0.5619736032791693
Headwind Speed->          mean: 0.19494949494949496       variance: 0.3189441893684318    standard deviation: 0.5647514403420604
Temperature->             mean: 18.353939393939395        variance: 19.035438016528925    standard deviation: 4.362962069114161
Globe Temperature->       mean: 21.176161616161615        variance: 63.18996102438527     standard deviation: 7.949211346063537
Wind chill->              mean: 18.2940202020202          variance: 19.129329898581776    standard deviation: 4.373708940771182
Relative humidity->       mean: 76.7930505050505          variance: 406.3302224115906     standard deviation: 20.157634345616813
Heat stress index->       mean: 18.286424242424246        variance: 18.46777529476584     standard deviation: 4.297414954919509
Dew point->               mean: 13.558787878787879        variance: 9.418778328741965     standard deviation: 3.0690028231889857
Psychro Wet Bulb Temp->   mean: 15.406666666666666        variance: 6.994618181818182     standard deviation: 2.6447340474645427
Station pressure->        mean: 1016.1661010101009        variance: 38.92418015141312     standard deviation: 6.23892459895238
Barometric pressure->     mean: 1016.127797979798         variance: 38.91944545413724     standard deviation: 6.238545139224148
Altitude->                mean: -25.96121212121212        variance: 2691.26556620753      standard deviation: 51.877409015943826
Density Altitude->        mean: 150.84    variance: 29702.92147070707     standard deviation: 172.3453552339229
NA Wet Bulb Temperature->mean: 15.93688888888889          variance: 9.428372543209877     standard deviation: 3.0705655086986625
WBGT->                    mean: 17.185535353535354        variance: 15.483612996224876    standard deviation: 3.934922235092439
TWL->                     mean: 284.11531313131314        variance: 1289.3922059120498    standard deviation: 35.908107801888555
Direction , Mag->         mean: 223.89656565656566        variance: 9264.263240730537     standard deviation: 96.25104280334077
```

Figure 1: *'Mean, variance and standard deviation for each variable of the 5 sensors'*

From a first glance we can see that in most variables the statistical indicators hold similar values between all the sensors.
There is a notable observation that can be derived from the variables Altitude and Density Altitude. The difference between mean and the standard deviations is such that could be only possible if the sensors are NOT fixated on the ground. So, we can hypothesize that all the sensors are strapped on balloons or drones and the measurement are taken from different heights Looking at the variable Wind Direction, True, it is apparent that the wind was rarely blowing from the North (in comparison to the other directions), which is information that can be used to calculate various correlations.
In general, there many different but related variables and the statistical indicators can reveal, on a surface level, possible correlations and interesting underlying facts or questions.

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?*

For 1.2, the codes computes and plots the Temperature histograms of 5 and 50 bins for each of the 5 different sensors. The following figure (Figure 2) illustrates the above comment in 10 different plots where each row of plots corresponds to each sensor and each column correspond to the histogram with the respective number of bins.

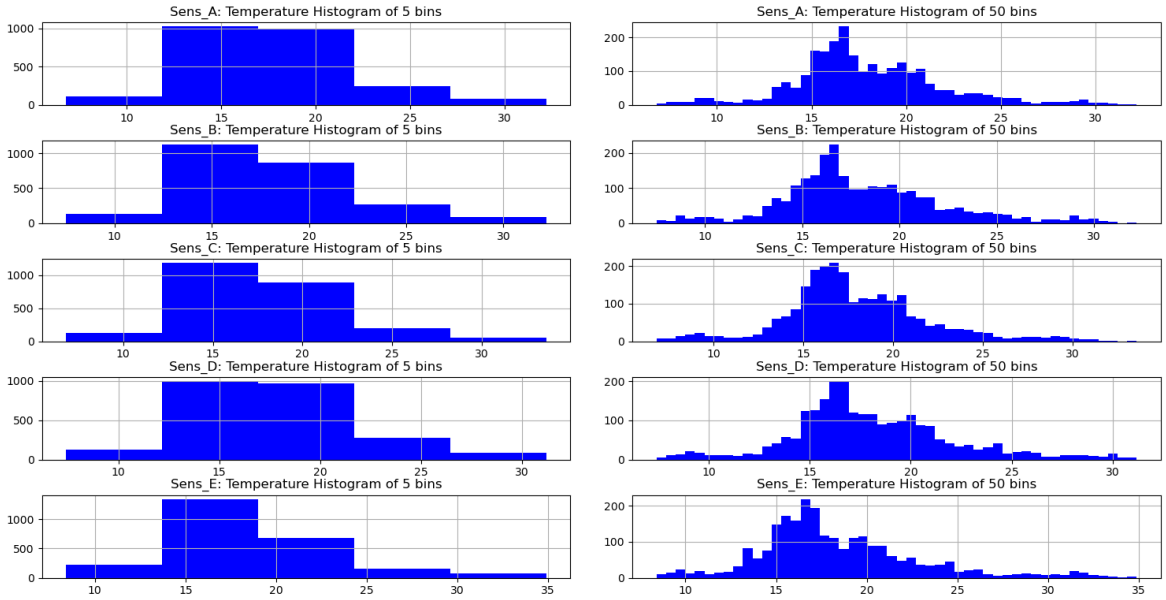

Figure 2: *'Bin comparison of 5 sensor Temperature Histogram'*

It is immediately apparent that the number of bins in a histogram plays a major role in the transmission of the message that plot tries to convey. Basically, the number of bins depend on the scale of data (values) that need to be illustrated. As such, five (5) bins are not enough to be able to get substantial information from almost 2.500 different values and we lose important levels of detail. On the other hand, using 50 bins allows the graph to show much more detailed results. However, too many bars can also hinder the overall illustration of the figure.
Having that in mind I continue to code and plot the rest of the graphs using 30 bins when needed.

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

For 1.3, the code generates a plot of the frequency polygons for the variable Temperature, for each of the 5 sensors. The frequency polygons graph (Figure 3) derives from a Temperature CDF stepped histogram of 50 bins and the 5 sensors are overlapping each other with different colours.

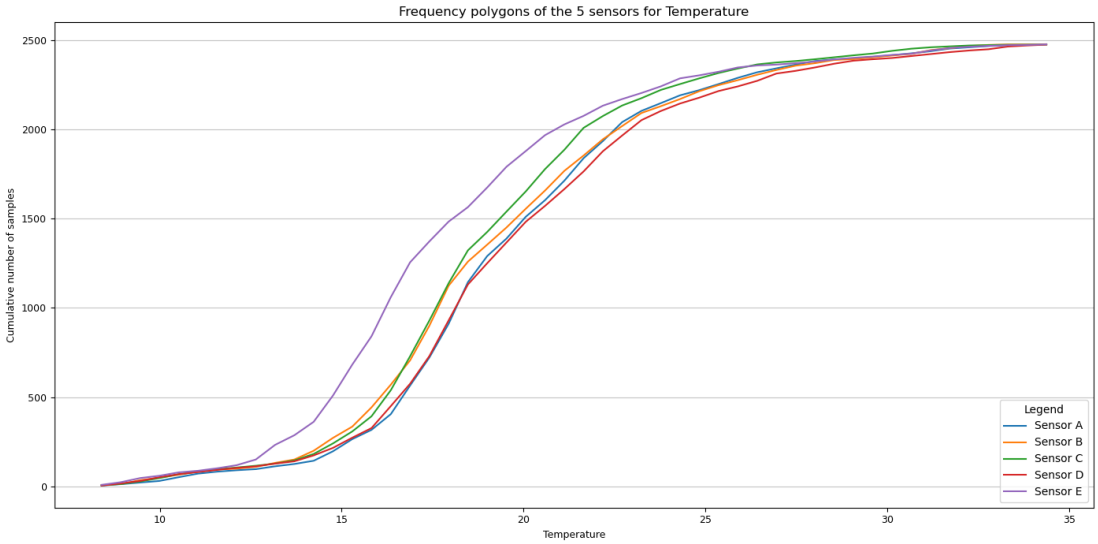

Figure 3: *'Frequency polygons of the 5 sensors for Temperature'*

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

For 1.4, the code creates box-plots for the variables Wind Speed, Wind Direction and Temperature. The figure (Figure 4) shows the 3 different plots side by side with their respective variables' values for the x axis and the 5 different sensors as the boxes. The plot provides visualization of min, max, 25th, 50th,75th percentiles, mean and outliers.

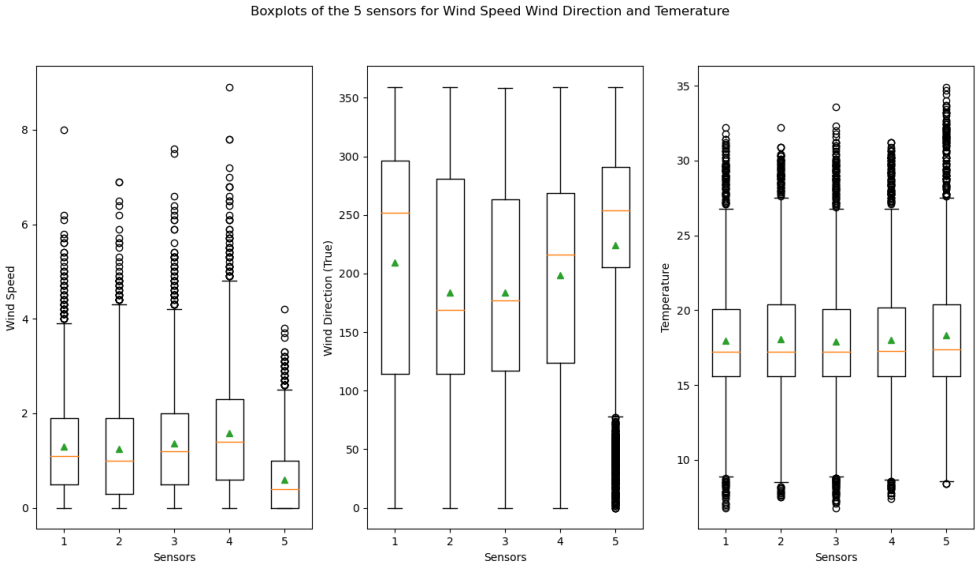

Figure 4: *'Boxplots of the 5 sensors for Wind Speed Wind Direction and Temerature'*

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?*

For 2.1, the code computes and plots the PMF, PDF and CDF for the variable Temperature and for each of the 5 sensors. The figures (Figure 5, Figure 6, Figure 7) illustrate the above by combining the 5 sensors in one figure every time in order to display the 3 distributions separately.

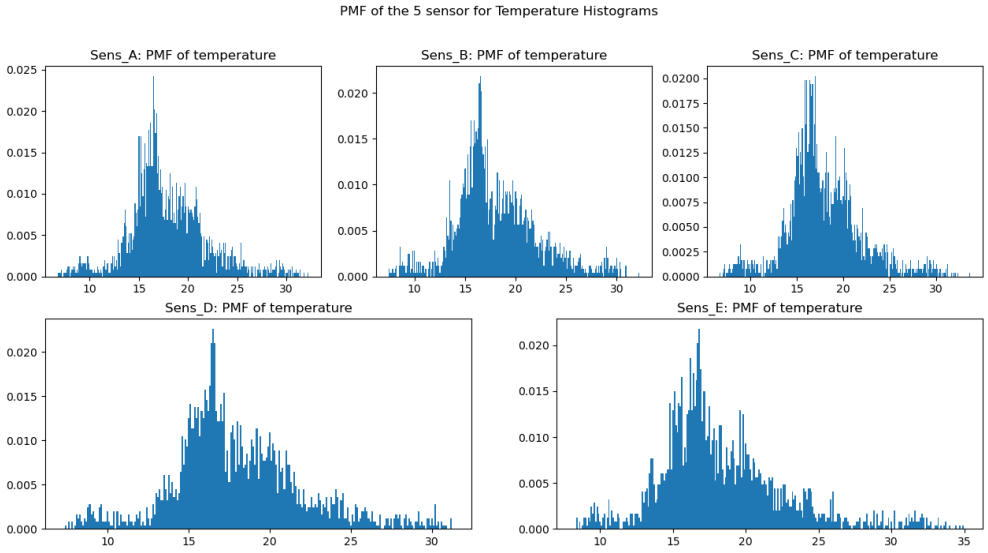

Figure 5: *'PMF of the 5 sensor for Temperature Histograms'*

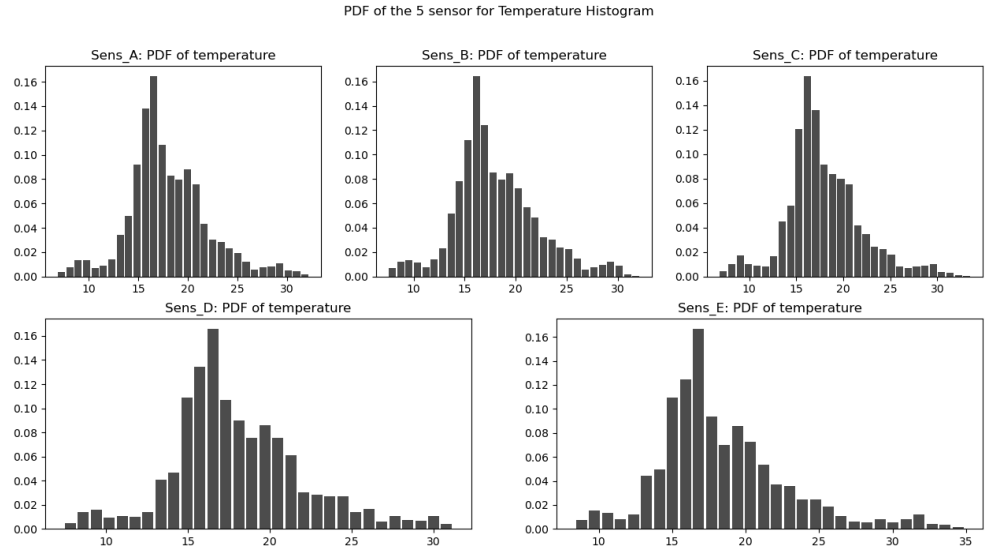

Figure 6: *'PDF of the 5 sensor for Temperature Histograms'*

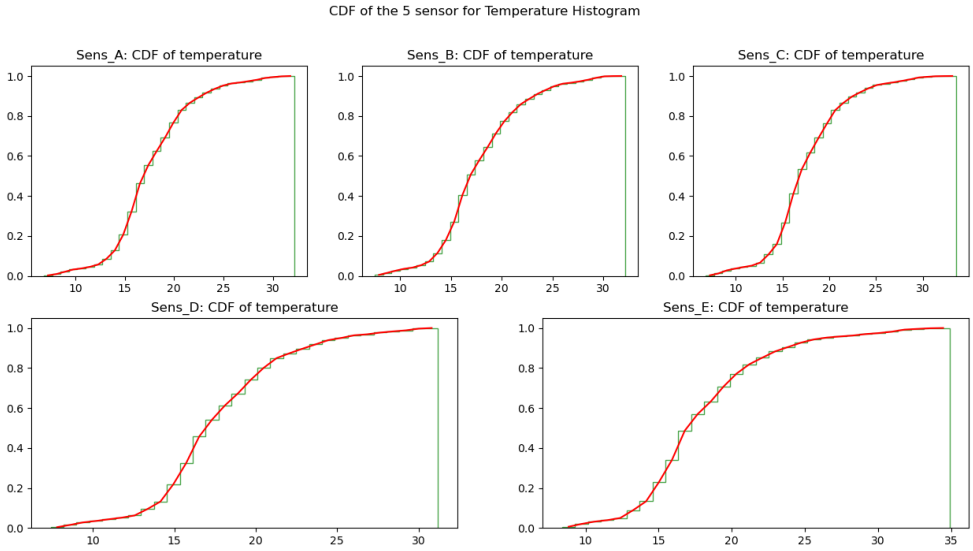

Figure 7: *'CDF of the 5 sensor for Temperature Histograms'*

Comparing the 3 distribution we can immediately figure out that there are major differences between each other as they use different methods. However, when comparing between sensors for each distribution the pattern remains basically the same. As for their tails, they seem to be right skewed.

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

For 2.2, the code outputs 5 plots of the PDF and KDE of the variable Wind Speed for each of the sensors. The figure(Figure 8) is a comparison between the PDF and the KDE of the above variable.

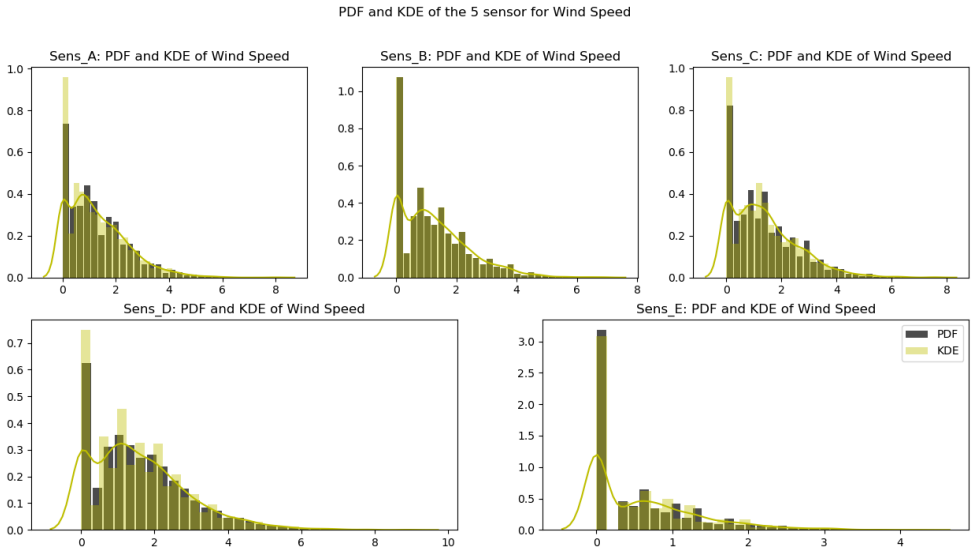

Figure 8: *'PDF and KDE of the 5 sensor for Wind Speed'*

As it is expected the KDE resembles the respective PDF as it smooths it out

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 Spearman's rank coefficients. Make a scatter plot with both coefficients with the 3 variables.*

For 3.1, the codes computes Pearson's and Spearman's coefficients between all the of the 5 sensors for the variables Temperature, Wet Bulb Globe Temperature (WBGT), Crosswind Speed. The figure(Figure 9) illustrates the above mentions with 6 scatter plots (Pearson and Spearman for each variable). The sensor pairs are 10 in total (disregarding the symmetry).

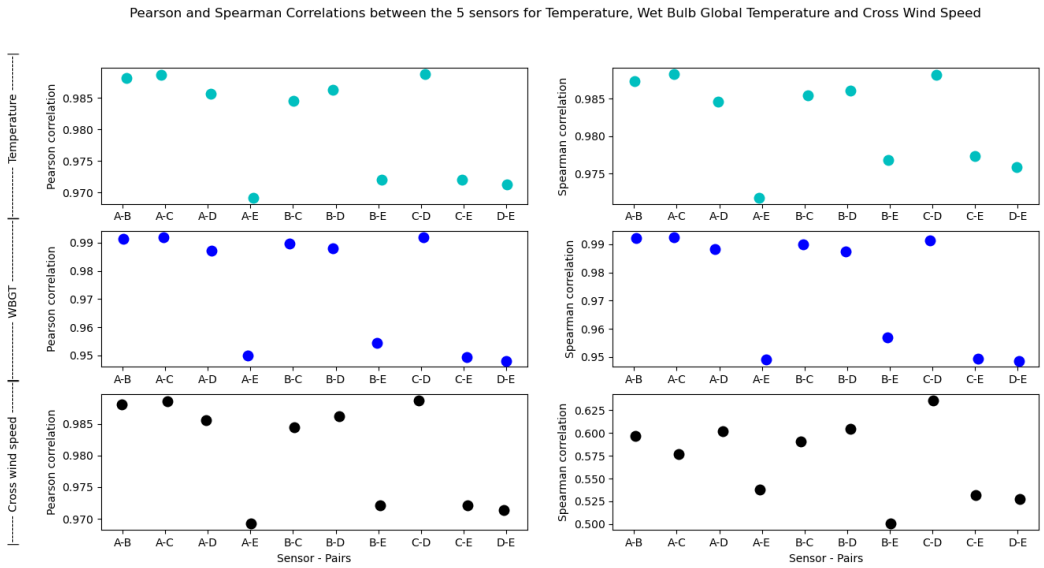

Figure 9: *'Pearson and Spearman Correlations between the 5 sensors for Temperature, Wet Bulb Global Temperature and Cross Wind Speed'*

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

For 3.2, almost all of the correlations of the sensors tend to follow the same patterns. Overall, is seems that sensor E is considerably distant (in term of correlation) from the other sensors. In detail, sensor pairs A-B, A-C, A-D, B-C, B-D and C-D show that relate the highest, ranging from 98.4 to almost 100 percent. While sensor pairs A-E, B-E, C-E and D-E seem to fall behind by 1-3%. The above correspond to all of the 3 variables for both correlation methods. However, for the variable Cross Wind Speed, Spearman's method reveals an outlier in the overall correlation of the sensors for this variable. The

pattern stays somewhat the same as the others, but with much lower overall correlation (from just 50 to 65%).

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.*

For 3.3, using the correlation of the sensors in combination with the proximity of the different sensors on the picture, we can guess their exact positions as below(Figure 10). The discrepancies in coefficients were such that the relative positions cannot be estimated with high certainty. However, sensor E seem to have the lowest values with every other sensor and as such we can assume that is going to be located the furthest away from the rest.

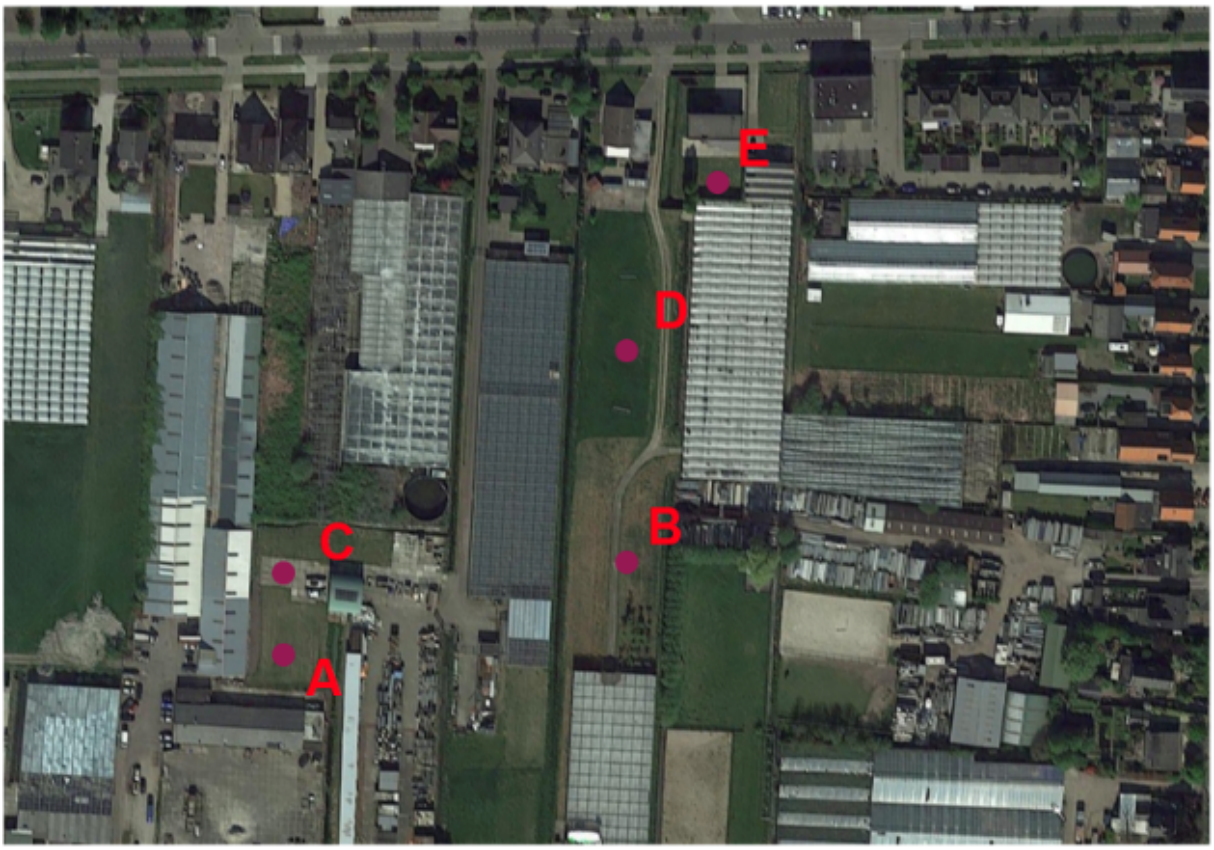

Figure 10: *'Possible Sensor Positions'*

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).*

For 4.1, the code creates the file confidence.txt where the 95% confidence intervals for variables Temperature and Wind Speed for all the sensors are stored. Additionally, it shows a figure(Figure 11) of 10 plots for the CDFs (stepped histogram) of Temperature and Wind Speed for all the sensors.

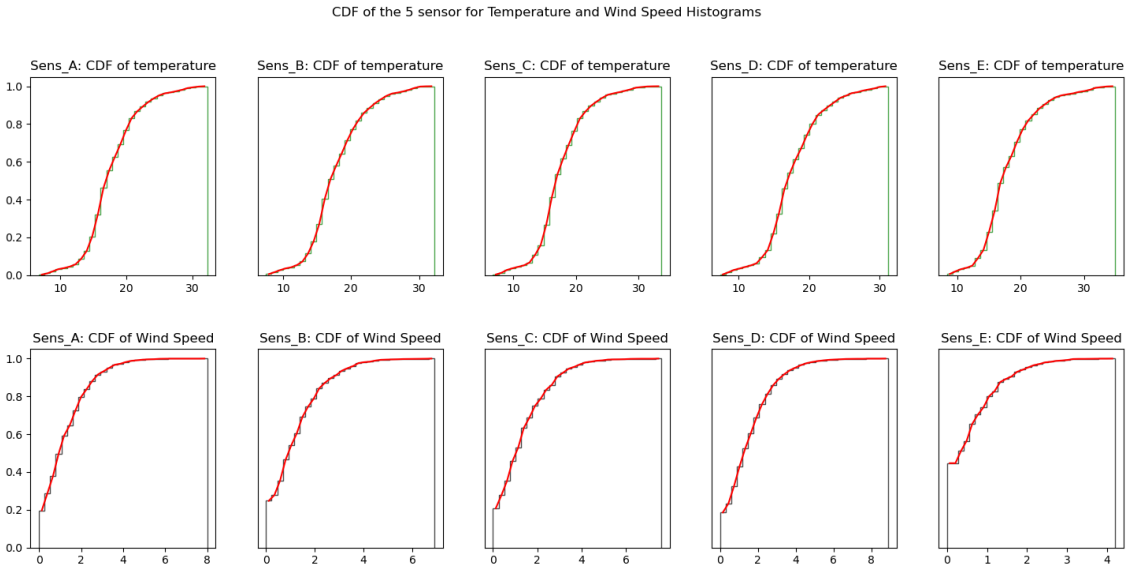

Figure 11: *'CDF of the 5 sensor for Temperature and Wind Speed Histograms'*

confidence.txt

```
1.246227038990971, 1.3343868543854427
1.1971663346979249, 1.287082453670411
1.3243037885948932, 1.418622646328308
1.5296480419653757, 1.633650260379006
0.5680599051948441, 0.6244249432900044
17.81214113267346, 18.126065652463858
17.90472689963894, 18.226129320070267
17.754926235060246, 18.071347006653575
17.83814660824381, 18.15457772482005
18.181933946027776, 18.525944841851015
```

Figure 12: *'95% confidence intervals for Wind Speed(lines 1-5) and Temperature(lines 6-10) of all the sensors'*

4.2 *Test the hypothesis: the time series for Temperature and Wind Speed are the same for sensors*

- 1) E, D;
- 2) D, C;
- 3) C, B;
- 4) B, A;

For 4.2, to test the hypothesis, the code computes with "ttest" the p-values of the sensor pairs E-D, D-C, C-B, B-A (Figure 13).

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

hypothesis_p_only.txt

```
1   0.002711172129731209
2   0.4657972008220813
3   0.18548636717619374
4   0.4004754260262924
5   3.3729639501474365e-212
6   4.610149126224334e-09
7   0.00010045473692816457
8   0.13351922750703515
9
```

Figure 13: *'p-values from 't-test' for Temperature(lines 1-4) and Wind Speed(lines 5-8)'*

Judging from the p-values, the conclusion for the temperature is:

- E-D is statistically significant, null hypothesis rejected
- D-C is statistically insignificant, null hypothesis accepted
- C-B is statistically insignificant, null hypothesis accepted
- B-A is statistically insignificant, null hypothesis accepted

The conclusion for Wind Speed is:

- E-D is statistically significant, null hypothesis rejected
- D-C is statistically significant, null hypothesis rejected
- C-B is statistically significant, null hypothesis rejected
- B-A is statistically insignificant, null hypothesis accepted

The condition that was used to test the hypothesis was:
p-value<0.05 -> Reject
p-value>0.05 -> Accept

5 Bonus Question:

Your "employer" wants to estimate the day of maximum and minimum potential energy consumption due to air conditioning usage. To hypothesize regarding those days, you are asked to identify the hottest and coolest day of the measurement time series provided. How would you do that? Reason and program the python routine that would allow you to identify those days.

The measurements are taken every 20 minutes for every day of 24 hours (72 measurements pre day). So, the first step is to group the measurements for each day. To do this, the code selects all the values of temperature that correspond to each different day, computes the mean of these values and matches them inside a data-frame. After obtaining the mean temperatures for each day, the code finds the days where the maximum and minimum temperatures occurred (Table 1).

Sensors	Min	Max
Sensor A	10-06-2020	26-06-2020
Sensor B	10-06-2020	26-06-2020
Sensor C	10-06-2020	26-06-2020
Sensor D	10-06-2020	26-06-2020
Sensor E	08-07-2020	25-06-2020

Table 1: 'Days that maximum and minimum temperature measurements occure for each of the 5 sensors'

From the table we can observe a major disparity for sensor E, getting the minimum temperature in a completely different day than the others. It gets the maximum temperature a day before than the other sensors. It is an observation that was already identified from the graphs of this report. That said, the rest of the sensors agree on their values and as such we could hypothesize that the day with the least energy consumption for air-conditioning is the 10th of June 2020 and the day with the highest consumption, the 26th of June 2020.

The calculations were made with the assumption that the term "Day" relates to a 24 hour day.

References

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