VAST challenge 2017 CM2

An exploration of the data set and its properties

Angelo Groot, Hidde Jessen,
Philip Kleinbrahm en Sven Boogmans

14 Juni 2018

Contents

1 Introduction

Near the city of Mistford and the wildlife preserve of Boonsong Lekagul in Thailand, 4 factories are emitting plumes of smoke into the air: Roadrunner Fitness Electronics, Kasios Office Furniture, Radiance ColourTek, and Indigo Sol Boards. Suspicion has it that these factories might be polluting the air in the area, thus 9 sensors have been placed in order to measure 4 types of noxious gasses.

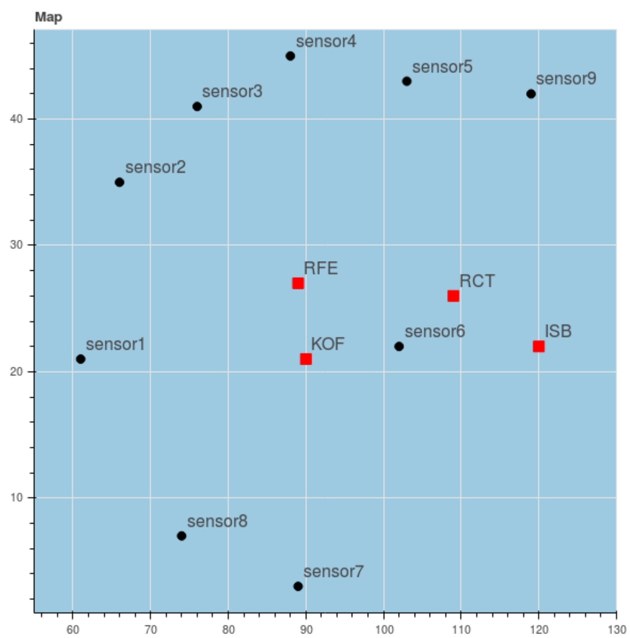

Figure 1: Map of the environment, black dots represent the sensors, red squares are the factories

These gases are: Appluimonia, Chlorodine, Methylosmolene, and AGOC-3A. In order to be able to take productive steps towards a healthier environment, it is of importance that any factory emitting one or more of these gases can be identified. As such, there will be three questions this project will try to answer:

1. Are all the sensors working properly at all times? Are there any unexpected behaviors of the sensors through analyzing the readings they capture?
2. Which chemicals are being detected by the sensor group? What patterns of chemical releases can be seen?
3. Which factories are responsible for which chemical releases?

1.1 The Gases

The effects of the gases are as follows:

Ap: Appluimonia is a gas that produces an unpleasant odor[2]. It is not very likely that Appluimonia causes serious injury, but the precise effects are unknown[2].

Ch: Chlorodine is a corrosive gas which is harmful when inhaled or swallowed due to its highly aggressive character which can destroy body tissue[2]. Additionally it may also damage or destroy metals[2].

Me: Methylosmolene is a name for a family of solvents. Methylosmolene has toxic side effects in vertibrates, and it is illegal to dispose liquid forms of methylosmolene without having chemically neutralized it[1].

AG: AGOC-3A is a solvent produced to have low-VOC. It is less harmful for humans and environmental health than high-VOC solvents[2].

1.2 The Factories

RFE: Roadrunner Fitness Electronics produces sport related consumer electronics, for instance heart rate monitors and GPS-watches[2].

KOF: Kasios Office Furniture produces office furniture made out of composite wood and metal[2].

RCT: Radiance ColourTek produces metallic flake paints. They are proud to have the lowest VOCs in the industry[2]. Or so they claim.

ISB: Indigo Sol produces skateboards and snowboards. It is a small company that has seen modest growth in recent years[2].

1.3 The Chemicals Sensors

The manufacturing companies and the town of Mistford close by, wanted an environmentally sound partnership. For this reason the Mistford Pact was signed, and nine air sampling sensors have been placed to monitor the air quality. The nine air sampling sensors monitor the amount of noxious gas in ppm precisely every full hour. This datacollection was carried out in the months of april, august, and december. It registers the data in the following format:

Chemical	Monitor	Date Time	Reading
Methylosmolene	3	4/1/16 0:00	2.68382
Methylosmolene	7	4/1/16 0:00	2.63064
Chlorodinine	3	4/1/16 0:00	1.25917
Chlorodinine	7	4/1/16 0:00	0.943983
AGOC-3A	3	4/1/16 0:00	0.722303

Table 1: VAST challenge 2017 MC2 Chemical data: first 5 entries

The column 'Chemical' contains which chemical was identified. The column 'Monitor' provides the sensor number of the sensor at which the chemical was measured. The column 'Date Time' contains the date and time when the chemical was found (without a daylight saving time correction), and the 'Reading' column contains the amount of the chemical detected in parts per million.

1.4 The Metereological Sensor

The meteorological sensor measures the wind speed in meters per second, and wind direction every three hours, during the months of April, August, and December. It registers the data in the following format:

Date	Wind Direction	Wind Speed (m/s)
4/1/16 0:00	190.5	4
4/1/16 3:00	203.3	5
4/1/16 6:00	201.1	5.2
4/1/16 9:00	204.9	4.1
4/1/16 12:00	207	3.6

Table 2: VAST challenge 2017 MC2 Meteorological data: first 5 entries

The Date column contains the date and time of the measurement (without a daylight saving time correction). The column Wind Direction contains the direction where the wind is originating from, in degrees from the absolute north. 000.0 degrees and 360.00 degrees being north, 90 degrees east, 180 degrees south, etc. The column Wind Speed (m/s) contains the wind speed in meters per second.

2 Methods and Results

2.1 Initial analysis: peculiar properties and readings

In order to take a first look at the consistency of patterns of chemical release provided by the sensors, a scatterplot was crafted for each sensor-chemical pair.

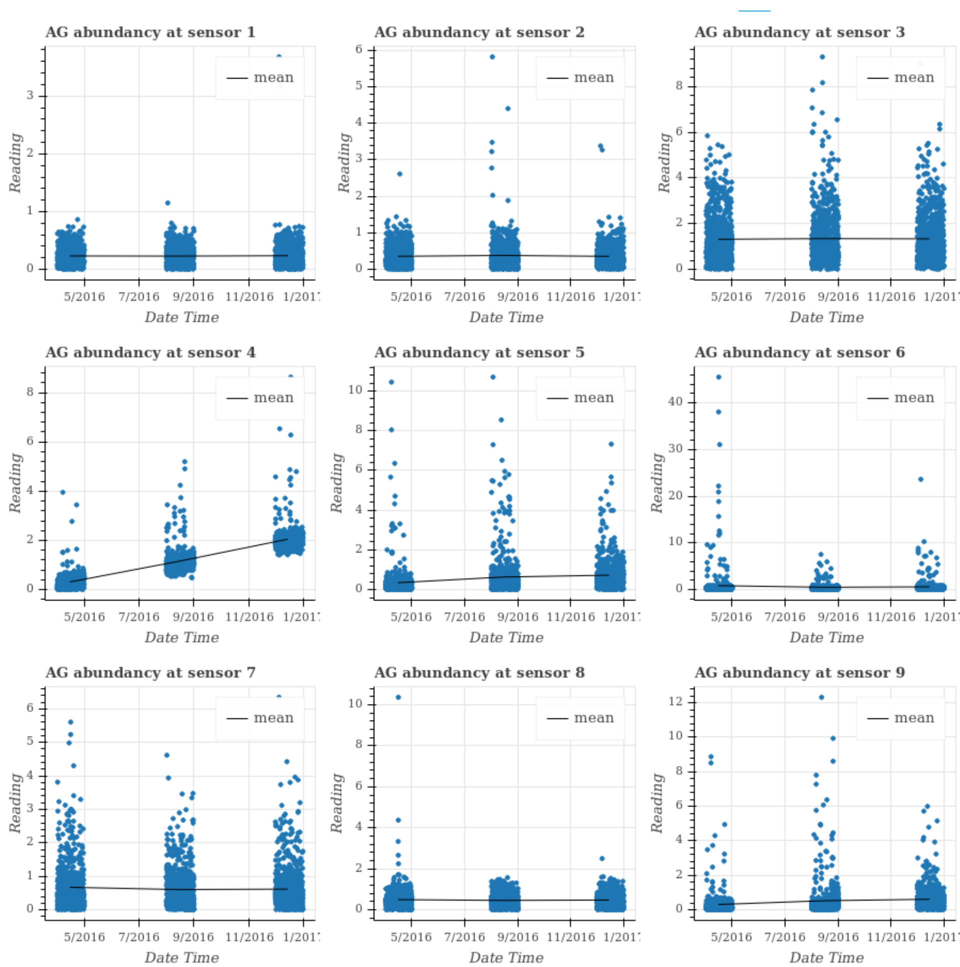

Figure 2: AG (AGOC-3A) abundance (blue) at the sensors and their average measurement per month (black)

2.1.1 Sensor 3 behaviour

It looks like sensor 3 has its readings more spread out than the other sensors (Figure 2). The other chemicals display similar results for sensor 3. The underlying error might be caused by a defect in the sensor, causing the sensor to make less precise measurements. But this is only speculation. If we look at the standard deviations of the sensor readings for each test period (April, August, December), we can see that these standard deviations are much higher than deviations measured by other sensors. The standard deviations of sensor 3 (table 3):

Month	Standard deviation sensor 3	Average deviation all sensors
april	0.90	0.38
august	0.86	0.41
december	0.90	0.42

Table 3: Standard deviation of sensor 3 for each month and the mean standard deviation of all sensors for each month

2.1.2 Sensor 4 behaviour

The values for all chemicals measured by sensor 4 seem to drift upward (Figure 2). The suspected reason for this drift is an error, because no other sensor seems to display this pattern. This potential error might be caused by a broken air filter, that wears down over time, and lets more of the chemicals through over time, causing higher readings over time, but this is only speculation. Figure 3 contains the bottom part of the complete scatter plot of sensor 4 and 1.

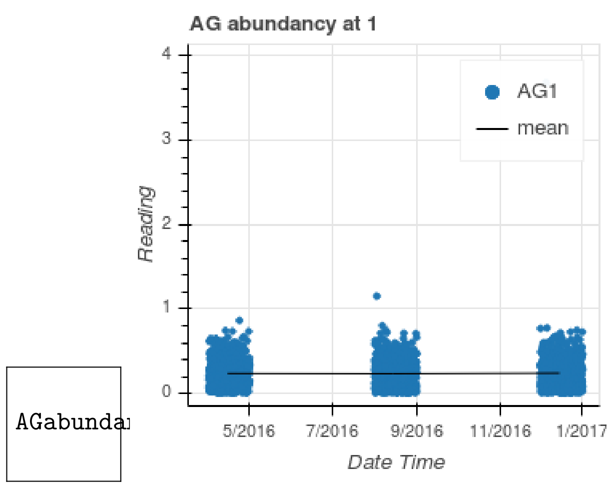

Figure 3: abundance of AGOC-3A at sensor 4, and sensor 1 (for comparison)

2.1.3 Sensor 5 behaviour

The readings for all chemicals measured by sensor 5 seem to slowly spread upward. (Figure 2). The suspected reason for this upward spread is an error, because no other sensor seems to display this pattern. The error might be caused by a slowly changing environment (slowly felling a patch of woodland) making it more and more likely for chemicals to reach the sensor, while less and less trees are in the way, causing a bigger likelihood of higher readings. This is only speculation. If we look at the standard deviation of the sensor, we can see that the standard deviation gets a lot bigger over time, while the mean standard deviation of all the sensors more or less stays the same. The standard deviations of sensor 5 and the average standard deviation of all the sensors:

Month	Standard deviation sensor 5	Average deviation all sensors
april	0.27	38
august	0.46	41
december	0.52	42

Table 4: Standard deviation of sensor 5 for each month and the mean standard deviation of all sensors for each month

2.1.4 Sensor 9 behaviour

The readings for all chemicals measured by sensor 9 seem to suddenly spread upward midway through the second test period (Figure 4). The sudden spread takes place between the 22nd and 25th of august. The suspected reason for this upward spread is an error, because no other sensor seems to display this pattern. The error might be caused by a changing environment (the removal of a building for instance) making it more likely for chemicals to reach the sensor, causing a bigger likelihood of higher readings, but this is again only speculation. If we look at the standard deviation of the sensor, we can see that the deviation before the 22nd of august is 0.27, and the standard deviation after the 25th of august is 0.44.

2.2 Cleaning the data

2.2.1 Handling missing entries

The source data delivered for this study contains suspicious missing chemical data:

1. All chemicals are missing their readings at 5 seperate hours.

2. Methylosmolene has the most missing data points. Whereas the amount of missing data points for chemicals other than Methylosmolene is either 5 or 6, Methylosmolene misses a total of 246 data points at all sensors.

Remarkable is that most of these data points are missing at sensor 3, 4, 5, 6 and 9: respectively 33, 44, 53, 49 and 44 readings.

The chemical sensors save their readings within an hourly interval while the meteorological data is measured within 3-hour intervals. In order to be able to still use all data points, the wind direction and wind speed have been interpolated over every hour. There are also 6 instances in which not a single sensor did a reading at that given hour and the 1st to the 5th of august are missing from the meteorological data in their entirety, these days will be ignored for the utter lack of any certainty that the interpolated data will bear any semblance to reality. (based on the hourly fluctiation in these readings which can already vary more than 10% it should be relatively safe to say that we can not say anything about the wind direction in 5 hours time)

The interpolation of the wind direction has been done by first calculating the sine and cosine of the wind direction, then seperately interpolating them and then deriving the degrees from these 2 newly generated data sets. This was done using two techniques: Linear interpolation and Cubic rom-spline interpolation.

2.2.2 Linear and Cubic Rom-Spline interpolation

For the interpolation of the missing meteorological data, two techniques have been applied: Linear interpolation and Cubic Spline interpolation. When using Linear interpolation, the missing data values are filled using a formula that linearly bridges the gaps between data points. When using Cubic Spline interpolation, the data points are connected by using the coefficient of the incoming angle when comparing to previous data points by using a formula for a cubic polynomial that fits these readings. The seperate methods yield the folowing patterns:

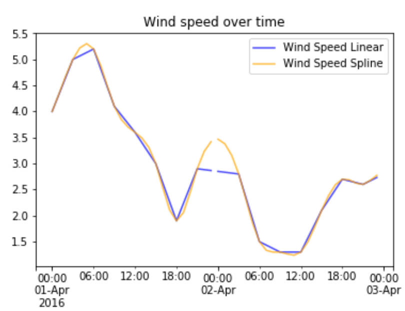

Figure 4: Linear interpolation (blue) vs Cubic spline interpolation (orange) of the wind speed

Since the wind speed and direction are both a result of the changes in pressure and movement of particles, the Cubic Spline interpolation method will be used in further analysis because it more closely portrays dynamic changes in natural systems.

2.3 Analysis of the data

2.3.1 Cumulative plots of chemical readings

A cumulative plots was made in order to discover relations between the readings of different chemicals. In the following part, three informative plots have been included in order to gain insight into the internal relations between chemicals.

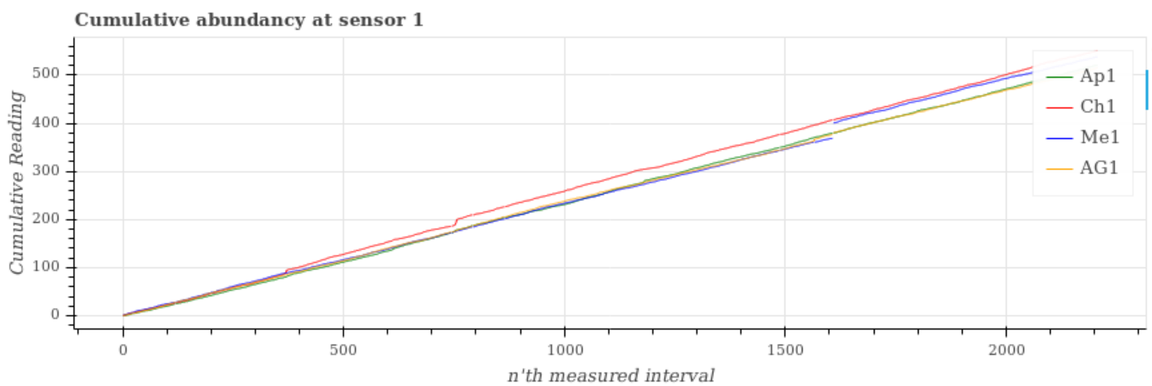

Figure 5: Cumulative plot of chemical abundance at sensor 1

The first plot is of sensor 1, we can see a small, but sudden rise of the chemical Chlorodine around the 750th measured interval.

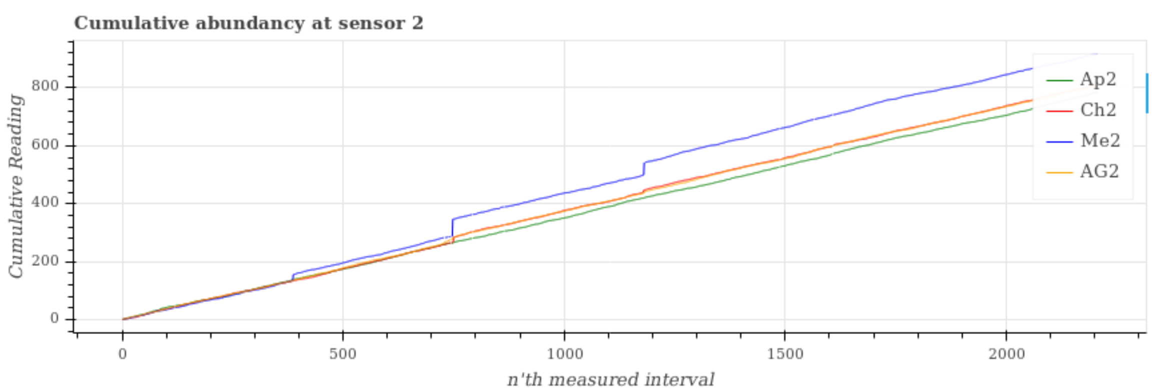

Figure 6: Cumulative plot of chemical abundance at sensor 2

If we look at the readings of sensor 2, we can spot the same sudden rise for the chemical Chlorodine around the 750th interval. This time the rise seems to be pared with a rise in the readings of the chemical Methylosmolene. This seems to indicate a connection between the two chemicals. They could be emitted by the same factory, that increased its production.

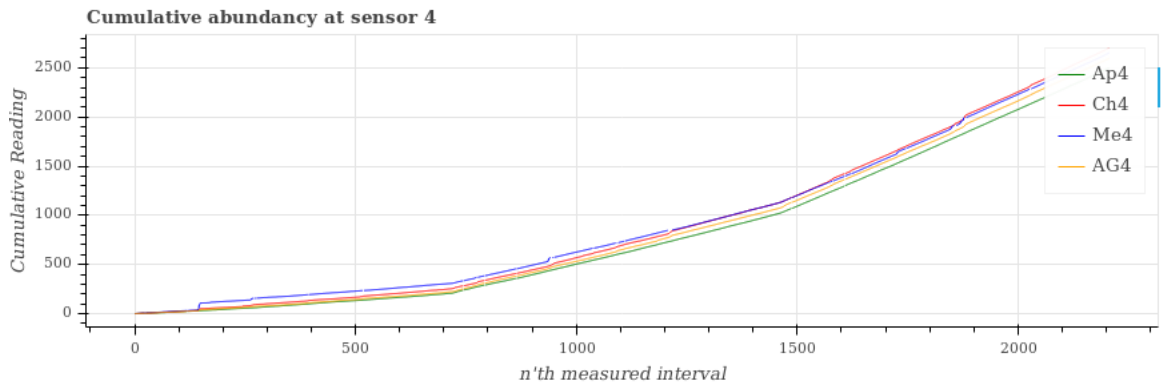

Figure 7: Cumulative plot of chemical abundance at sensor 4

In the cumulative abundance of sensor 4, it can once again be observed that there is a monthly drift when looking at the the seemingly constant increase of the abundance. In this plot, 5 vertical empty lines can be clearly seen between 500 and 2100, these are moments at which none of the chemical readings was recorded. This could be explained by an electric fallout or another event that could have caused widespread damage which required instant maintenance for all sensors.

2.3.2 Chemical readings per hour

In order to potentially catch the culprits red handed it is important to know about the daily interval within which different chemicals are released. The plot below gives the values of all four chemicals in the month of April, with all outliers being still present.

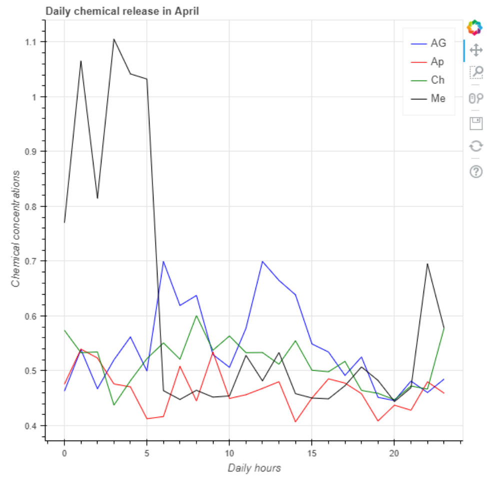

Figure 8: April Chemical release

The figure above shows a peak of Methylosmolene release during the night time. Moreover, the other chemicals seem to follow a random pattern across hours with mean values ranging from 0.4 to 0.7. Only the AGOC-3A graph shows noticable peaks. The following plot gives the corresponding trends of August.

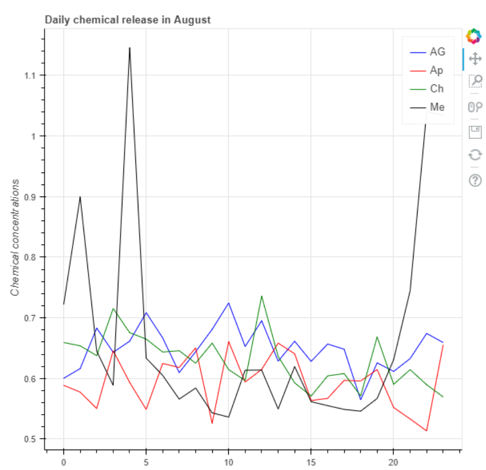

Figure 9: August Chemical release

The graph above shows that the nightly release of Methylosmolene is a consecutive trend in August. Aside from this, thee other chemicals still seem to be randomly emitted across the hours of the day. The range of the remaining three chemicals seems to be the same, with no single chemical showing a substantially different pattern.

The plot below shows the development of chemical release in December.

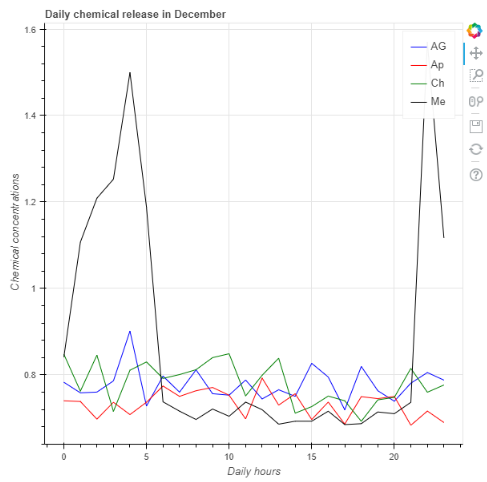

Figure 10: December Chemical release

In December, Methylosmolene still is being released in the night time primarily. Additionally, the remaining three measured substances show comparable internal behaviour as in August, but with the overall range being slightly higher as compared to August. This might hint at elevated emission levels in this month.

All in all and remarkably, Methylosmolene release consistently takes place primarily during the night. Furthermore, we controlled whether the data cleaning process would leave us with another picture. For exemplification we provide the following graph:

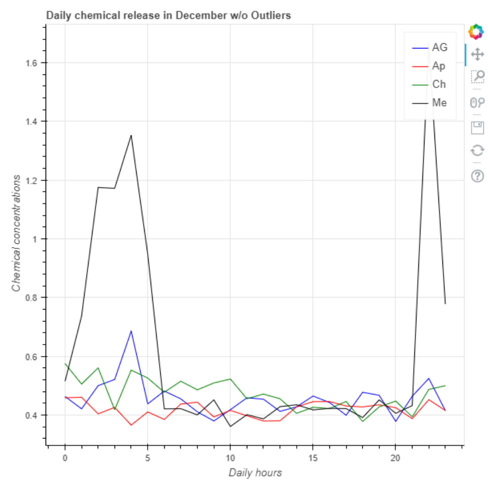

Figure 11: December Chemical release

This graph shows the behaviour of our data in December after Sensors three, four, five and 9 have been removed and all missing values were filled with the mean of the monthly measurement of the corresponding sensor for this chemical. As can be seen, the difference between the clean and raw version of the December graphs is marginal. The same is valid for the months of April and August, thus we can conclude our data cleaning process has not distorted the data in this respect.

2.3.3 Chemical abundance over wind speed

In order to get an idea of the influence that wind speed exerts on the amount of chemicals that reaches the sensors, the Chemical abundance has been scattered over the wind speed in the following plot:

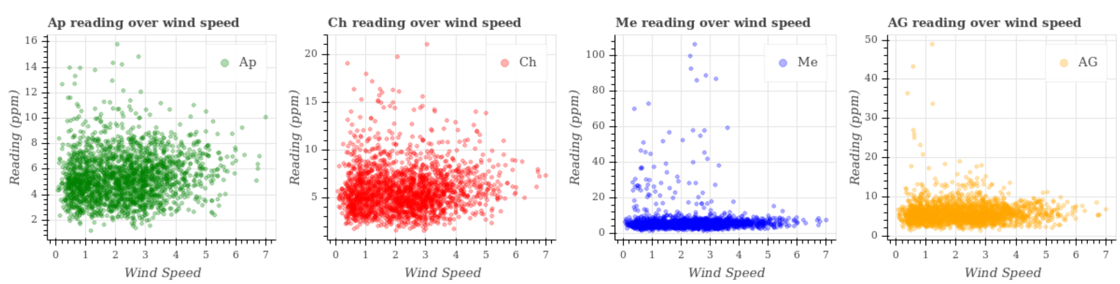

Figure 12: Scatterplot of the chemical abundance over the wind speed

It is noticeable that Methylosmolene and AGOC-3A have relatively high readings in comparison to Chlorodinine and Appluimonia. It also looks as if there is a trend where lower wind speeds seem to produce higher readings. When the y-axis limits are all set to 20, the following scatterplots will be found:

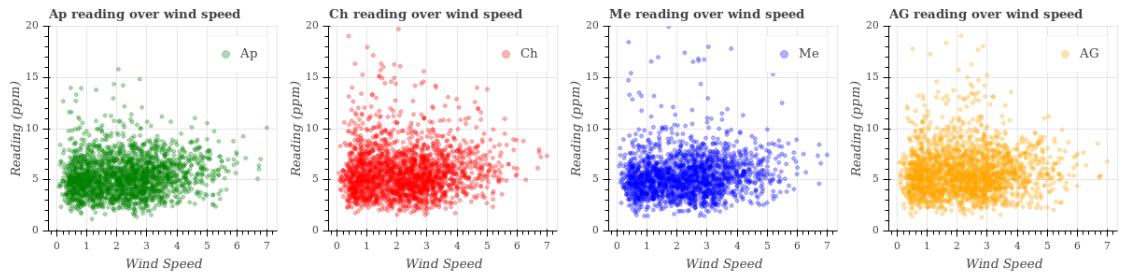

Figure 13: Scatterplot of the chemical abundance over the wind speed

Contrary to what could previously have been observed, now it seems that all chemicals behave similarly within these boundaries: a dense cloud between wind speeds of approx. 0.2 and 4 and the readings range from two to ten. Another shared property is that, once again, the same trend appears where lower wind speeds yield more above-average readings than higher wind speeds.

This trend may be explained through the frequency of the different wind speeds. For this, one could look at the following histogram:

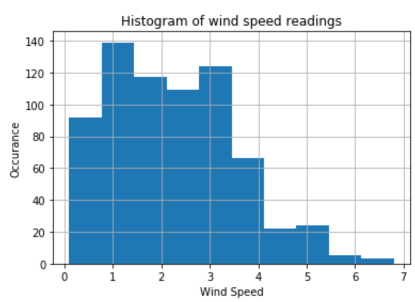

Figure 14: Histogram of wind speed readings

In Figure 8, it becomes apparent that there might not be a connection between the wind speed and the chemical abundance after all, since the amount of readings recorded declines as the wind speed increases.

2.3.4 Pairing chemicals to factories

To answer the third question, where the goal was to determine which factory was responsible for which chemical releases, we used the speed of the wind, the locations of the factories and the sensors, as well as the readings of the different sensors. The search for an answer to his question is based on the assumption that a sensor S will have a high readings for chemical A if the wind is blowing from a factory that emits that chemical A to sensor S.

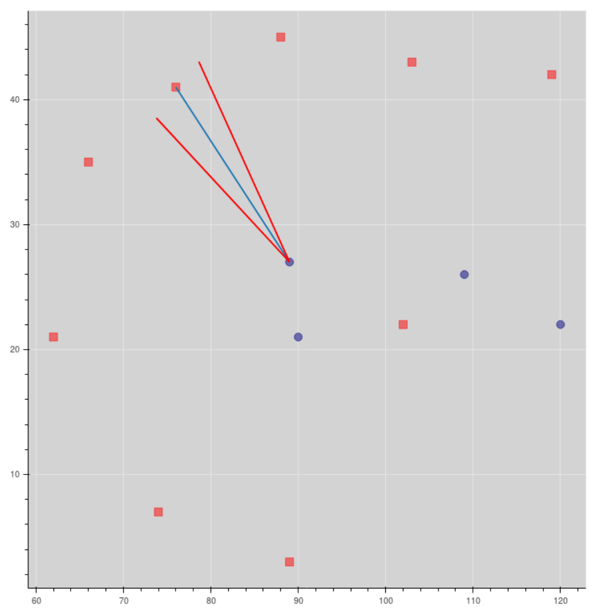

Figure 15: A cone that describes the collection of wind directions under which the factory could possibly influence the measurements of a sensor with a variable range of + or - α degrees

First, the angles between each of the factories and all the different sensors were calculated. It was important to keep in mind that the Meteorological data had North as 0 degrees, but the angles calculated between the factories and sensors had east as 0 degrees. The wind directions in the data were the directions the wind was coming from, not the direction the wind was blowing towards. After correcting for these differences, we assumed that the wind wouldn't blow the particles in a perfect straight line from the factory to the sensor and that there would be some form of diffusion in the direction perpendicular to the direction of the wind. Because there is no way to know exactly how much defusion takes place, we assumed that there would be a range of approximately 10 degrees around the wind direction in which the particles could be found, an example of this can be found in figure 9. Because 10 degrees is just a rough guess, we used 3 ranges of 5, 10 and 15 degrees so we could see how our choise of angle would influence the results.

The next step was to look at all the different measurements of the wind direction. If the angle of the wind direction fell within the range of the angle between a factory and a sensor, the readings for each of the chemicals were stored, with the timestamp, the sensor and the factory as additional information. Table X shows the number of times this accured for each of the angles.

Factory	5 degrees	10 degrees	15 degrees
RFE	190	388	574
KOF	298	613	897
RCT	184	389	589
ISB	274	547	581

Table 5: For each factory, the number of times the wind was blowing from that factory to a sensor within the range of alpha degrees

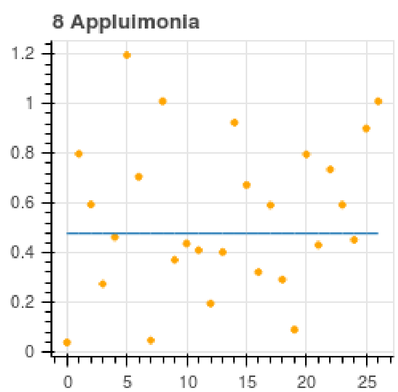

Figure 16: an example of the plots for the RCT factory for sensor 8 from the chemical Appluimonia. The blue line is the average value of the Appluimonia readings of sensor 8 and the yellow dots are the values on the moments where the wind was blowing from the RCT factory to sensor 8 within a range of 10 degrees

To extract the information necessary to answer the question, we first calculated the average value for each of these factory-sensor-chemical combinations (MeanFSC). After this, these were compared to the average value of all the readings of that sensor for that chemical, not just the ones that where recorded when the wind was blowing from that factory to that sensor (MeanFS). The next step was to count the number of times per factory per chemical that a reading was higher than the average value for that chemical for that sensor, if the wind was blowing from the factory to that sensor. These values were quite low and were all around the 50 percent so this wouldn't help us answering the question.

Then, instead of looking at the number of readings, we looked at the values of the readings. We compared the MeanFSC, so the average reading for a specific sensor and chemical, given that the wind was blowing from that specific factory to the sensor, with the MeanFS, the average reading for that sensor and chemical over the entire period.

RFE			
Chemical	5 degrees	10 degrees	15 degrees
Methylosmolene	370	348	483
Chlorodinine	272	126	121
AGOC-3A	517	244	463
Appluimonia	157	101	92

Table 6: the increase in readings as a percentage for factory RFE when the wind blew from the factory to a sensor within the given range compared to the overall average.

KOF			
Chemical	5 degrees	10 degrees	15 degrees
Methylosmolene	638	587	475
Chlorodinine	213	117	124
AGOC-3A	765	734	611
Appluimonia	143	91	93

Table 7: the increase in readings as a percentage for factory KOF when the wind blew from the factory to a sensor within the given range compared to the overall average.

RCT			
Chemical	5 degrees	10 degrees	15 degrees
Methylosmolene	771	200	158
Chlorodinine	183	96	104
AGOC-3A	418	523	570
Appluimonia	189	112	118

Table 8: the increase in readings as a percentage for factory RCT when the wind blew from the factory to a sensor within the given range compared to the overall average.

ISB			
Chemical	5 degrees	10 degrees	15 degrees
Methylosmolene	517	518	412
Chlorodinine	171	125	122
AGOC-3A	369	346	278
Appluimonia	172	107	114

Table 9: the increase in readings as a percentage for factory ISB when the wind blew from the factory to a sensor within the given range compared to the overall average.

3 Results

When looking at the tables six, seven, eight and nine, it becomes clear that the readings within the ten degrees range are near the middle of the readings of the five and fifteen degrees angles. This isn't true for all the factory-chemical combinations (RFE vs AGOC-3A for example, see table 6), but because in most cases the ten degrees looks like a good approximation that is the angle that we will be using in the rest of the analysis.

factory	chemicals	# overlaps	% correct	% increase	z-score
RFE	Methylosmolene	1166	44	392	0.32
RFE	Chlorodinine	1246	50	250	-0.44
RFE	AGOC-3A	1326	52	453	0.64
RFE	Appluimonia	1246	44	169	-0.86
KOF	Methylosmolene	1350	44	616	1.50
KOF	Chlorodinine	1454	48	224	-0.57
KOF	AGOC-3A	1558	51	795	2.45
KOF	Appluimonia	1454	44	158	-0.92
RCT	Methylosmolene	1419	42	249	-0.44
RCT	Chlorodinine	1432	44	167	-0.87
RCT	AGOC-3A	1448	50	493	0.85
RCT	Appluimonia	1432	45	175	-0.83
ISB	Methylosmolene	1732	44	478	0.77
ISB	Chlorodinine	1752	42	187	-0.77
ISB	AGOC-3A	1775	47	324	-0.04
ISB	Appluimonia	1753	45	187	-0.77

Table 10: Analysis of possible factory-chemical relations

Table 10 shows the data collected by all the sensor. '# Overlaps' show's number of times the wind blows from that factory to a sensor. '% correct ' shows the average percentage for all the sensors of the times that the reading for that chemical was higher than the average for that sensor-chemical combination. '% increase' shows how the average reading when the wind blew from that factory to a sensor, relative to the overall average. 'z-score' displays how many standard deviations apart the increased value is from the 'increase' average. Some rows have been marked in green, this indicates that there is a likelyhood that these factories are connected to the chemical in the same row, this is based on the percentual increase.

factory	chemicals	# overlaps	% correct	% increase	z-score
RFE	Methylosmolene	382	39	348	0.34
RFE	Chlorodinine	387	45	127	-0.69
RFE	AGOC-3A	392	46	247	-0.13
RFE	Appluimonia	387	46	101	-0.81
KOF	Methylosmolene	586	43	587	1.47
KOF	Chlorodinine	608	43	118	-0.73
KOF	AGOC-3A	630	46	742	2.20
KOF	Appluimonia	608	46	91	-0.86
RCT	Methylosmolene	385	37	200	-0.35
RCT	Chlorodinine	386	44	97	-0.83
RCT	AGOC-3A	390	50	528	1.19
RCT	Appluimonia	386	45	112	-0.76
ISB	Methylosmolene	522	43	518	1.14
ISB	Chlorodinine	540	43	126	-0.70
ISB	AGOC-3A	561	52	350	0.35
ISB	Appluimonia	541	44	107	-0.79

Table 11: Analysis of possible factory-chemical relations (excluding the possibly defective sensors 3, 4, 5 and 9)

In Table 11, a (likely) more reliable analysis is being portrayed. It is remarkable that this dataset shows different pairs as likely connections. This is possibly caused by either the excluded sensors causing fuzzyness in the data or by the considerably lower sample size of the overlaps.

4 Conclusion

The three questions that were posed for this project were as following:

1. Are all the sensors working properly at all times? Are there any unexpected behaviors of the sensors through analyzing the readings they capture?
2. Which chemicals are being detected by the sensor group? What patterns of chemical releases can be seen?
3. Which factories are responsible for which chemical releases?

After cleaning and analyzing the data, observations have led to the following conclusions:

1. Sensor 3, 4, 5, and 9 appear not to function appropriately. In Table 3, it can be seen that Sensor 3 has its readings spread out more widely than the other sensors. In figure 3 the behaviour of sensor 4 is displayed as it has its readings linearly drifting upwards, sensor 5 has its readings spread

out over time (more than the other sensors), and sensor 9 shows a sudden spread in readings midway through the second test period.

2. As the division of the original data into different months and the hourly performance of the sensors within these monthly blocks has shown, Methylosmolene is being emitted almost exclusively during night-time. This shows that the people responsible for this in all likelihood tried to disguise their actions. Aside from this datachunking based upon time intervals, we have tested for a causal relation between the degrees of the wind and factory/sensor paths. This lead us to the main conclusion presented below.

   Furthermore, in the '% succes' column of Table 12, it becomes apparent that only between 37 and 52 percent of the time at which an increased reading would be expected at a sensor, this also is the case. This further supports the conclusion that it is likely factories only produce the chemicals about 40% of the time with a margin of 10%.

3. Table 12 shows the factory-chemical combinations with a positive z-score in green. A positive z-score means that the average reading when the wind blew from that factory to a sensor was substantively higher than the overall average.
   For the RFE, Methylosmolene is the only chemical with a positive z-score but it is a very low one. This means that based on our findings it looks like RFE is not one of the factories that emits pollution. The same goes for ISB and the emition of AGOC-3A, a small z-score is measured but it is not high enough to conclude that the ISB emits AGOC-3A.

   Based on our findings these are the most likely relations between factories and the chemicals they emit:

   RFE: -
   KOF: Methylosmolene and AGOC-3A
   RCT: AGOC-3A
   ISB: Methylosmolene

   There was no substantial increase for Chlorodinine or Appluimonia for any of the factories.

5 Discussion

5.1 Unused data

In the delivered dataset, Elevation was also included. Since this contained only a single entry, it has been disregarded in the analyses as it could not lead to any discoveries that would uncover relevant relations between variables.

5.2 Sensor malfunctions

The sensors around the city all together cover around 200 degrees of the surrounding area. From the east to the south there are no sensors around the city. Sensor 6 covers this part for the Roadrunner Fitness Electonics and the Kasios Office Furniture factories but this is not the case for the Radiance ColourTek and the Indigo Sol Boards factories. When looking at the sensors north of the city, these turn out to be sensors 3, 4, 5 and 9, the same sensors that didn't function properly. This means that for the Radiance ColourTek and Indigo Sol Boards the sensors only cover 90 degrees of the surrounding city, all the air. These are the same factories where the pollution doesn't increase that much when the wind blows over them towards a sensor, looking at the standard deviations away from the average reading. When we leave the dis functioning sensors out of the analysis the standard deviations increases a lot for these factories. For the other to factories only the AGOC-3a reading for the Roadrunner Firness Electonics changes drastically.

In a future research it would be good to place the sensors in a way that the surround the entire city. In the current setting, there are a lot of readings where the wind is blowing away from the sensors and no reading is measured. Especially because the factories are so close together it could be important to be able to get the readings with different wind directions. The way our research works is that we look at the increase at a sensor when the wind is blowing from a factory to that sensor, but when the wind is blowing towards the west for example, sensors one would pick up the chemicals released by both the Kasios Office Furniture and the Indigo Sol Boards factories and on that moment there is now way to distinguish which factory is responsible for the reading. Being able to have both the readings when the wind blows from several factories to the same sensor, and when the wind blows from only one factory to a sensor could help narrow down which factory is responsible for which pollution.

The fact that there were no substantial increases for Chlorodinine and Appluimonia doesn't mean that these chemicals are not being emitted by one of the factories, maybe the quantities are so small the differences are not noticeable. The chemicals could be emitted on the times the wind blew in the other direction or maybe these were the chemicals measured by sensors 3, 4, 5 and 9, which were the ones that broke down.

The fact that the sensors which are suspected of dysfunctional behaviour are all next to each other is rather suspicious. It is likely that there is an external reason why these sensors are failing to function the same as the others. A possibility is that one or more of the factories could be emitting a chemical that is harmful enough to cause the sensors that capture it to break down, another explanation is that the city could have built something between the factories and these sensors but there is no way to be certain about this. In a future research project, it might be a progressive idea to place multiple sensors at the same place in case one of them breaks down, or to check the functioning of the sensors throughout the year, so that if odd measurements are being recorded, the problem could be solved, resulting in more reliable data.

6 References

- 1. Nickling, W. G. (1988). The initiation of particle movement by wind. Sedimentology, 35(3), 499-511.
- 2. VAST. (2017). VAST challenge 2017. issued on 28-6-2018, from http://www.vacommunity.org/VAST+Challenge+2017+MC2