Tennet Project: Transformer Noise

Layla Gan - 4760549
Rui Jiang - 4696964
Zhe Min Chia - 4761596
Lara Scavuzzo - 4723635
Rei Yun Tan - 4763777

August 29, 2018

1 Introduction

This report presents the results of our project in collaboration with Tennet, one of the main electricity transmission system operators in Europe. Tennet provides electricity supply to 41 million end-users. In order to do this, they manage an extensive high-voltage grid that connects energy producers with energy consumers.

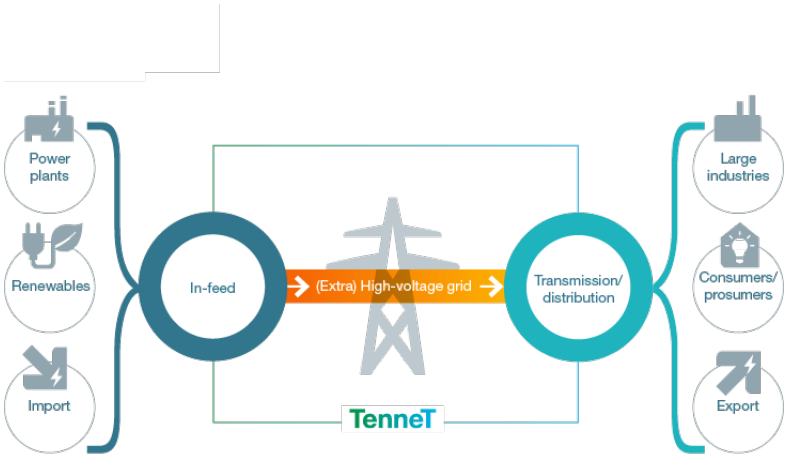

Figure 1: The tennet supply chain. Image credits: [1]

Along Tennet's electric grid, there are several transformer stations. These transformers perform the high to low voltage conversion (and vice versa). The problem lies in the high levels of audible noise that these transformers create. In this project, we study the problem of transformer noise in one particular station, where neighboring house complexes have reported disturbingly high noise originating from Tennet's facility. In particular, we study the correlation of switching operations with the grid signal quality, as we believe that this is closely related with the noise production.

Tennet has addressed the transformer noise by investigating possible alternatives such as active noise cancellation [2] [3]. The long term plan is to replace the transformers, but this will only achieved in a time span of years. In the meantime, we wish to study possible correlations of this noise generation with other factors, so that we can alleviate the disturbances of the surrounding neighborhoods and set a precedent for years to come.

2 Background

The topic of transformer noise has been widely studied. By construction, transformer noise cannot be completely avoided. It can, however, be reduced through proper design, installation and maintenance. It has been shown [4] that the main cause of this noise is magnetostriction, i.e. the expansion and contraction of the iron core due to the magnetic effect of alternation current flowing through the transformer coils. The fundamental frequency of the produced sound is twice the power line operating frequency of the transformer. This is, in this case we expect a 100Hz sound, coming from the 50Hz voltage line.

However, the measured voltage signal from the grid is not a perfect sine. There seems to be a weak correlation between power load in the transformers and noise, but a big relationship between the noise level and the quality of the signal waveform.

For this reason, we centered our study in the correlation between filter switching operations and the power quality. The filters we will study are part of a more complex scheme of power electronics that serves the purpose of transforming from AC current to DC current and viceversa. Tennet suspects that there could be correlations between the filter status and the quality of the sinusoidal waveform of the grid signal. This signal quality could be later correlated to audio noise. Unfortunately, due to several setbacks, we did not have access to the necessary data to do this second analysis in time.

Total Harmonic Distortion

We will measure power quality through a magnitude called total harmonic distortion (THD). In order to define this quantity, let us write the Fourier transform of the voltage in phase k ($k = 1, 2, 3$) as

$$u^{(k)}(t) = \sum_n u_n^{(k)} \cos(n\omega t + \varphi_n)$$

where $u_n^{(k)}$ are the Fourier expansion coefficients, ω is the base frequency and φ_n is a phase term. Then, the THD is defined as

$$\mathrm{THD}_k = \frac{\sqrt{\sum_{n=2}^{\infty} \left(u_n^{(k)}\right)^2}}{u_1^{(k)}}$$

In the following, we will approximate the THD by using information about harmonics up to the 25th, according to our data availability.

We remind the reader that the presence of three phases should not complicate the analysis. The use of a symmetric three-phase power system increases efficiency in power transmission. It consists of three conductors in the grid carrying alternating currents of the same frequency and voltage, but a phase difference of one third of a cycle.

Cross-correlation

In this report we will use the cross-correlation between two signals as a measure of the degree of correlation between them. Therefore, we introduce here the basic concepts regarding the calculation and interpretation of the cross-correlation.

Given two signals $x[n]$ and $y[n]$ with equal sampling frequency, we define the cross-correlation between them,

$$C_{xy}[k] = \sum_{m=-\infty}^{\infty} x[m]y[m-k]$$

If x and y vary in the same way but with some time delay Δ, we expect to see a pronounced peak at $k = \Delta$. Let us consider the following example. Define

$$x(t) = 2cos(13t) + 12t^2 + 4t$$

$$y(t) = 3x(t) - 10$$

These two function vary at the same pace (with no delay). Figure 2 shows these functions (left) and their corresponding cross-correlation (right). As we expected, we find a notable spike in the middle, indicating that the variations happen at the same time. The side peaks correspond to the particular shape of the function, which has some kind of periodicity.

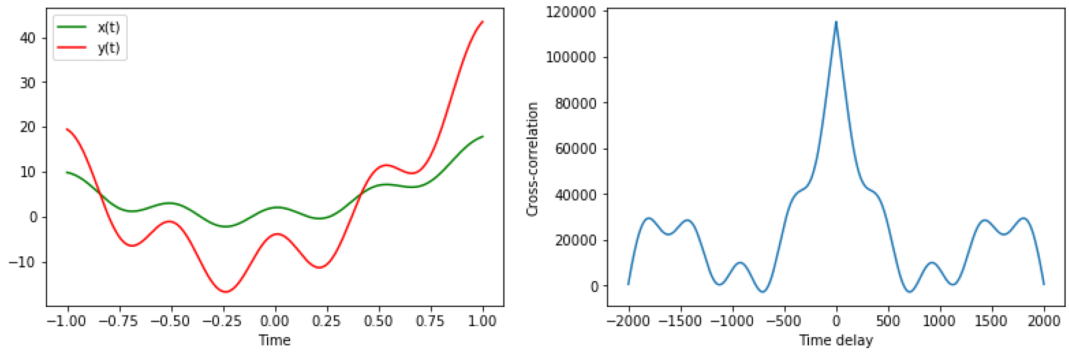

Figure 2: Example: functions $x(t)$ and $y(t)$ (left) and their cross-correlation (right).

3 The dataset

We will analyze data collected over a period of six months (November 2016 to April 2017). The data consists of two separate datasets. The first one contains measurements of the following magnitudes with a sampling time of ten minutes:

- The RMS measure of voltage in each of the phases.
- The harmonic coefficient of each of the phases up to the 25th.
- The THD measure of each of the phases.

All these measurements were taken every ten minutes, in two different points of the grid. We shall refer to these points as point A and point B.

The second dataset is a record of the state of four independent filters over the course of the time period under study. This data is binary, i.e. it is ON/OFF information (see Figure 3). A measurement was recorded every time the state of any of the filters changed. Following the terminology of the dataset, we shall refer to these filters as B11, B12, B21 and B22.

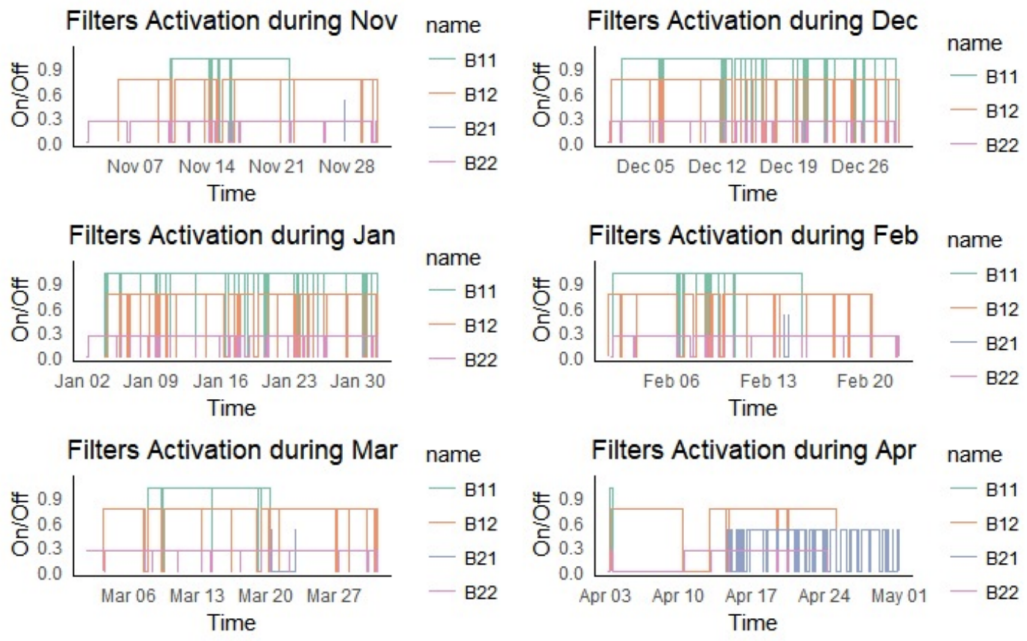

Figure 3: Switching operations over six months (November 2016 to April 2017).

3.1 Data sanity check

In order to verify the validity of the data we were given and to simplify future calculations, we performed some tests before hand.

Test 1: Harmonics

As advised by a Tennet responsible person, we verified that taking into account only the first 25th harmonics is a good approximation of the spectral content of our signal. To prove this, we are using the fact that

$$||u^{(k)}||^2 = \sum_n ||u_n^{(k)}||^2 \approx \sum_{n=1}^{25} ||u_n^{(k)}||^2$$

Results of our analysis are shown in Table 1 where we present the relative error measure, defined as

$$\text{Error}_r = \frac{\sum_{n=1}^{25} ||u_n^{(k)}||^2 - ||u^{(k)}||^2}{||u^{(k)}||^2}$$

Relative error			
Phase	1	2	3
Average	0.06%	0.05%	0.06%
Maximum	0.14%	0.13%	0.13%

Table 1: Relative error when taking into account harmonics up to order twenty-five.

Test 2: Measuring point A vs B

We expect the electrical signal to behave in the same way regardless of the point that was chosen to take the measurements, up to some slack. We are going to test this hypothesis.

First, we performed a 2-sample t-test, obtaining a p-value of zero (within machine precision). Given this result, we have to reject the hypothesis that these two signals come from the same distribution. This is fairly obvious from the look of the plot of the two signals (Figure 4). Clearly, the signals look different but their variations seem to be correlated.

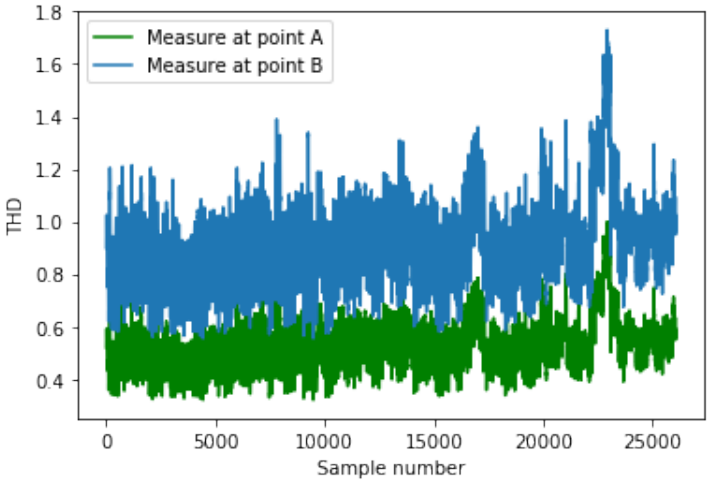

Figure 4: THD measured in point A (green) and B (blue) over time.

In order to prove this, we calculated the cross-correlation of the THD signal at points A and B. Results are shown in Figure 5. We see a clear peak at zero delay, meaning that the signals are highly correlated at every moment in time. Furthermore, we obtained a Pearson correlation coefficient $r = 1$.

Summing up, even though the value of the THD is different, the variations are the same. Given that we will be studying variations in THD, we can use either one of the datasets without loss of generality.

Test 3: Correlations between the three phases

Finally, in this report, we will focus on analyzing the THD signal in phase one. To justify that we can expect to draw the same conclusions on the other two phases, we first run a test to show correlations among them. The results of this test are shown in Table 2.

	Phase 1	Phase 2	Phase 3
Phase 1	1	0.987	0.986
Phase 2	0.987	1	0.984
Phase 3	0.986	0.984	1

Table 2: Pearson correlation coefficient among the THD on different phases.

We can conclude that our treatment of data from only phase one is greatly justified by the high

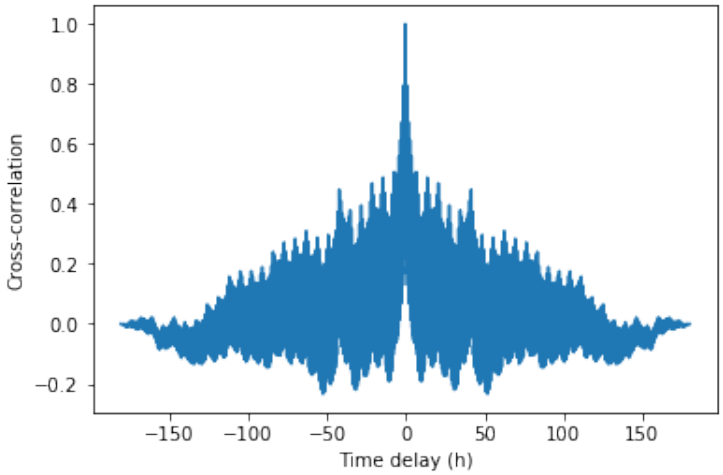

Figure 5: Cross-correlation of THD data at measurement points A and B.

degree of correlation of the three phases, as one would expect given the physical process that is behind the generation of these three signals.

4 Time Series Analysis

We examined the data by graphing it to check for patterns, variability and outliers. We then conduct tests to check stationarity. To stabilise the variance, we proceed to carry out transformation on the data, followed by a decomposition of the time series using STL (Seasonal and Trend decomposition using Loess) to better analyse the trend and seasonal effects. Time series decomposition deconstructs a time series into several components, the trend, seasonal and remainder component. The STL decomposition algorithm would be further explained in Section 4.2. Lastly, we analysed the Autocorrelation (ACF) and Partial Autocorrelation (PACF) for each month to explain the variation observed.

4.1 Exploration Of Data

In this study, the frequency used is 144 observations in one unit of time (day). Note that seasonal effects is observed when there is a fixed and known period and in this case, daily effects is referred to as seasonal effects in the following sections.

4.1.1 Time Plots

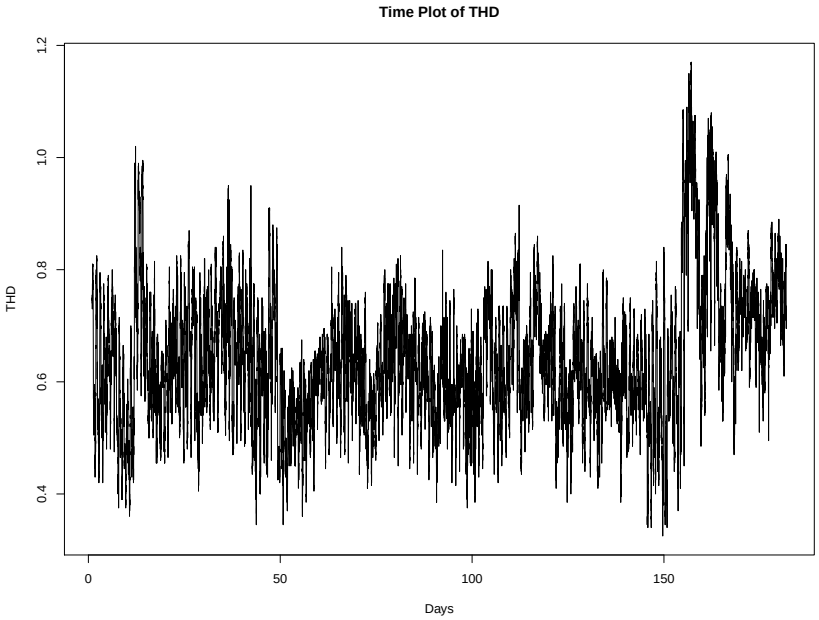

Figure 6: Time Plot of THD

From the time plot of THD throughout the course of 6 months, there appears to be no apparent trend, but we observe a sudden spike in November and April period. There is some indication of cyclic behavior with a period of 25-30 days. THD level does not changes with time. We will further analyze if seasonal effects are present. After conducting Augmented Dickey-Fuller Test to test stationarity of the data, the test yields a p-value of 0.01 which suggest that the time series is stationary. We then split the data and observe the time plots in each month.

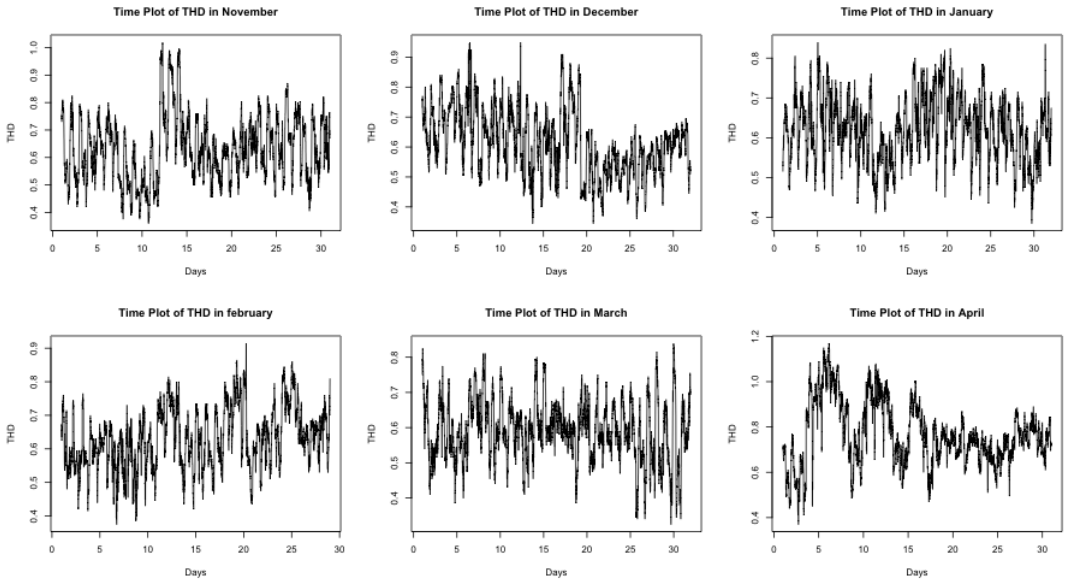

Figure 7: Time Plot of THD in Each Month

From the time plots of THD in each month, there seems to be no apparent trend. There is some indication of seasonal effects. We apply Augmented Dickey-Fuller Test for each month and the test yields a result of p-value$<$0.05 for all months. This suggests that the time series is stationary in each month. A Box-Cox transformation is performed to the data to further stabilise the variance of the data followed by detecting and replacing outliers in the data.

4.2 STL Decomposition

We assume the time series is composed by the 3 components; trend, seasonal and remainder. Remainder component is the residuals from the seasonal and trend fit, and is the portion that cannot be unexplained by the fitting. In our study, as the seasonal variation does not change with the level of THD, we can assume that the three components contribute in an additive manner.[5] We can conduct a decomposition of the time series using STL (Seasonal and Trend decomposition using Loess). "Loess"refers to a locally weighted regression model. We can write:
$y_t = S_t + T_t + R_t$, where S_t refers to the seasonal component, T_t refers to the trend-cycle and R_t refers to the remainder component.
In comparison to the ordinary linear regression model, loess is able to estimate nonlinear relationships.[5] The decomposition algorithm consists of two loops:

- An inner loop smooths the trend and seasonal components. Note that the seasonal component can vary over time.
- An outer loop computes the remainder component, then assigns robustness weights to each observations.

These weights will be used in the inner loop to downplay the importance of outliers in the time series. [6]

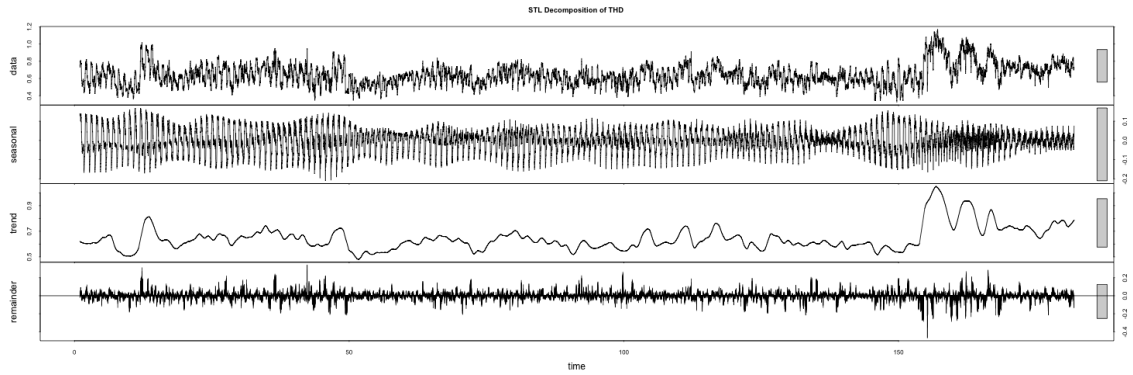

Figure 8: STL Decomposition of THD

This decomposition makes it easier to analyse the trend and seasonal effects as the smoothed series capture the main movement of the time series without all of the minor fluctuations. The plot shows the original data displayed in the first panel, the seasonal variation in the second, the trend variation in the third and the remainder in the fourth panel. Based on the scales of each panels in separate months, roughly 62-75% of series is explained by its seasonal and trend components.
The seasonal variation changes slowly, and it is the least in the last quarter of January and the highest in early November, late December and starting April period. From the trend panel, we can see that, adjusting for seasonality, the level of is remains constant until a spike is observed in April. Also, the remainder component is the largest around the period in April where there are sudden dips and spikes. Hence by analysing the de-trended and seasonality adjusted data, we can observe that the THD level in April behaves differently from the other months. To support the observation, we look into the individual Autocorrelation (ACF) and Partial Autocorrelation (PACF) plots for each month.

4.3 Autocorrelation and Partial Autocorrelation

4.3.1 Autocorrelation In Each Month

Autocorrelation measures the linear relationship between lagged values of a time series. Now we turn to estimating the ACF for data that we have been given. For a time series y_t, we let r_k denote the estimate of the correlation between y_t and y_{t-k} for k $\geq$ 0. Therefore, r_1 measures the relationship between y_t and y_{t-1} while r_2 measures the relationship between y_t and y_{t-2} and so on. Then a ACF plot is a plot of r_1,r_2,.. against lags 1,2,3,.. and so on.

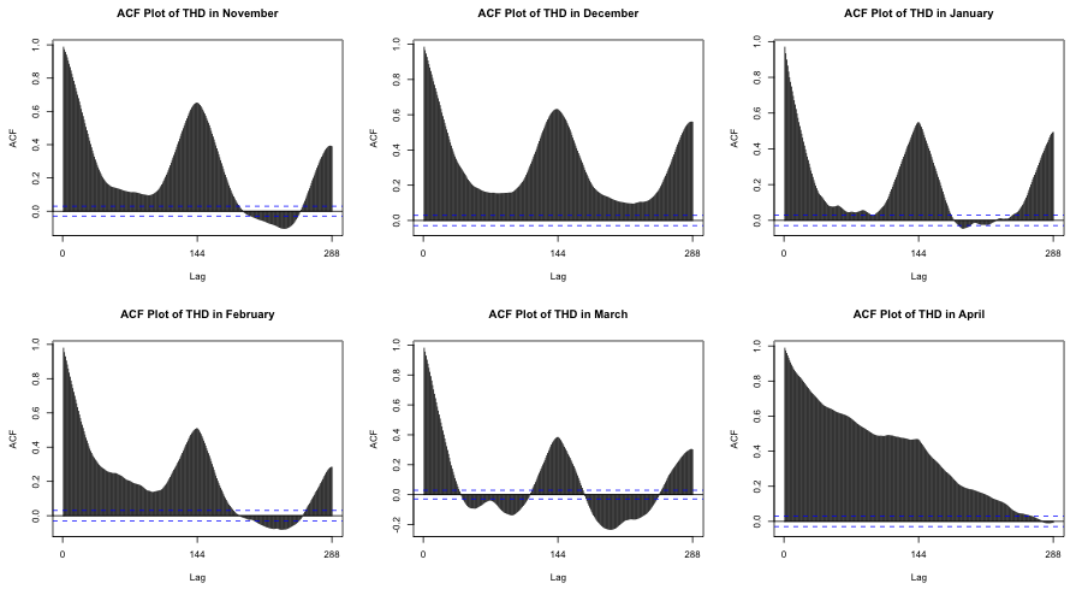

Figure 9: Autocorrelation Plots

From the ACF plot, we notice that all lags are significantly correlated with current series as they all crosses the blue lines which denote the 95% confidence interval. Also, there is a pattern in November to March. We observe the slow decrease in the ACF as the lags increase, this is due to the trend. The regular spikes, which lead to a scalloped shape, are due to the seasonality. The ACF peaks at lags 0, 144, 288 indicate a seasonality of length 144 readings which is in a day. However in April, the decrease in autocorrelation between lags may suggest a weaker seasonal effects compared to the other months.

4.3.2 Partial Autocorrelation

If y_t and y_{t-1} are correlated, then y_{t-1} and y_{t-2} are also correlated. Then y_t and y_{t-2} could be correlated simply because they are both connected to y_{t-1}. To measure the new information in y_{t-2}, we examine the Partial Autocorrelation (PACF) in each months. Partial autocorrelations measure relationship between y_t and y_{t-k}, when the effects of other time lags 1,2,3,...,k$-$1 are removed. Note, the first partial autocorrelation is identical to the first autocorrelation, because there is nothing between them to remove.

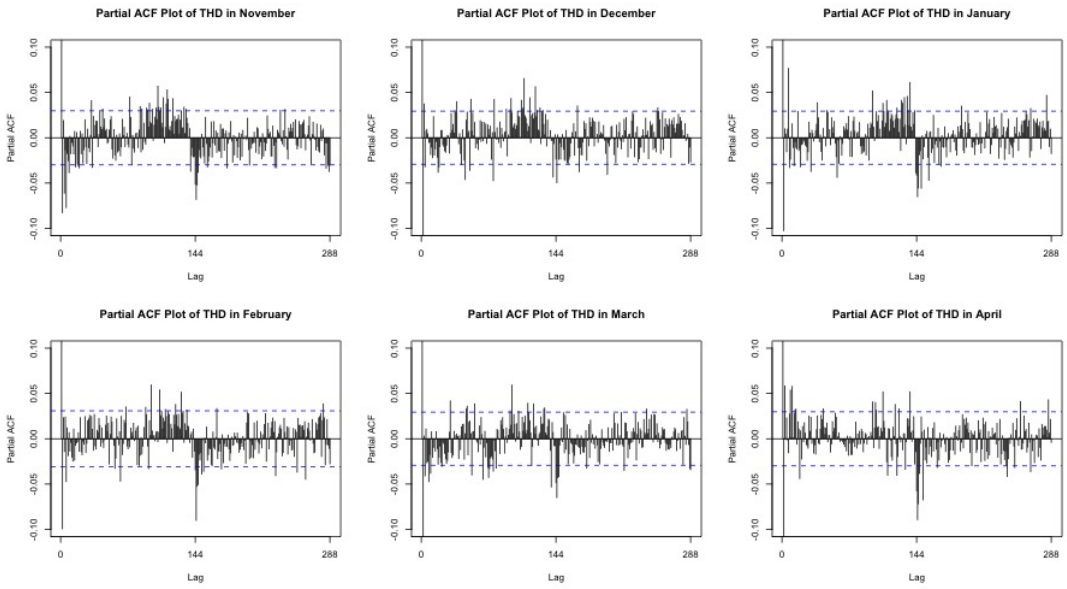

Figure 10: Partial Autocorrelation Plots

Similarly, in the months of November to March, it started with negative autocorrelation between lags followed by a spike just before lag 144. At lag 144, there is a strong correlation between each value and the value occurring a day ago. This is observed throughout all months however in April, it started with positively correlation between points that are 7 to 8 lags apart. Also, there is considerably more significant correlations between lags in April, as seen from the lines crossing 95% confidence interval. Hence, we conclude that THD trend in April is not the same as the other months and we should analyze April separately from the other months, vice versa.

5 Data Analysis

In this section we will study correlations between the filter states and the THD value. First, we make a preliminary analysis of the statistical characteristics of the THD data sequence when each of the filters is on or off (average, median, variance, etc). We also study the effect of having different combined filter states. Then, we visualize these variations to support our findings. These results are backed up by our analysis of the cross-correlations between filter state and THD. Finally we study the possibility of a correlation with filter switching, rather than state.

5.1 ON/OFF data statistics

Our first approach was to study the statistical differences, for each filter, of the THD signal during ON and OFF state. Figure 11 shows the average THD value for ON state (green) and OFF state (red). Overlaid, we can see a box plot, showing the median (orange line), the quartiles (box), the extreme non-outlier values (whiskers) and the outliers (black circles). Furthermore, Table 3 again shows average values of the THD, but this time with confidence bounds. To calculate these values, we considered six sequences of numbers (two per filter, corresponding to ON and OFF). For each of them we present the following numbers

$$\bar{x} \pm 1.96 \cdot \frac{s}{\sqrt{n}}$$

where $\bar{x}$ is the mean of the sequence, s is the estimated standard deviation, n is the number of samples for each case and 1.96 accounts for a 95% confidence bound.

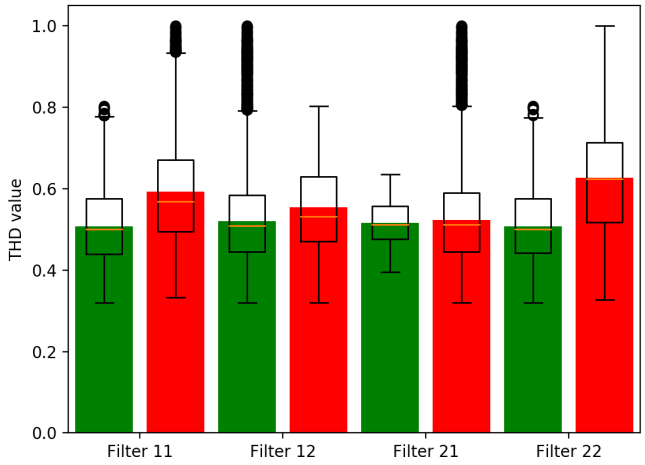

Figure 11: Average value of the THD for ON/OFF states (green/red) of each filter, overlaid with boxplot.

THD average value				
Filter	**B11**	**B12**	**B21**	**B22**
ON	0.5079 ± 0.0012	0.5203 ± 0.0013	0.5167 ± 0.0023	0.5077 ± 0.0011
OFF	0.5933 ± 0.0036	0.5546 ± 0.0046	0.5232 ± 0.0013	0.6275 ± 0.0046

Table 3: Average value of the THD for ON/OFF states (green/red) of each filter, with confidence bounds

What we conclude from this first naïve analysis is the following:

- We can observe clear differences, especially for filters 11, 12 and 22. However, only an analysis of the audio noise data could determine if these differences are relevant. We do not have a bound on the measurement error and the data points given do not have a consistent number

of decimals.Therefore our confidence bound calculations could be corrupted. We need to further analyze the data to draw a final conclusion.

- Filter 21 seems to have a unique behavior and we will therefore take special care of the analysis of this particular filter.

In addition, we performed the same analysis harmonic-component-wise. An example for one of the filters (B11) is shown in Figure 12. We clearly see that increases in THD are due to some specific harmonics, namely number five, seven, eleven and thirteen, with minor contributions of the rest. This analysis could be useful for finding further correlations with audible noise.

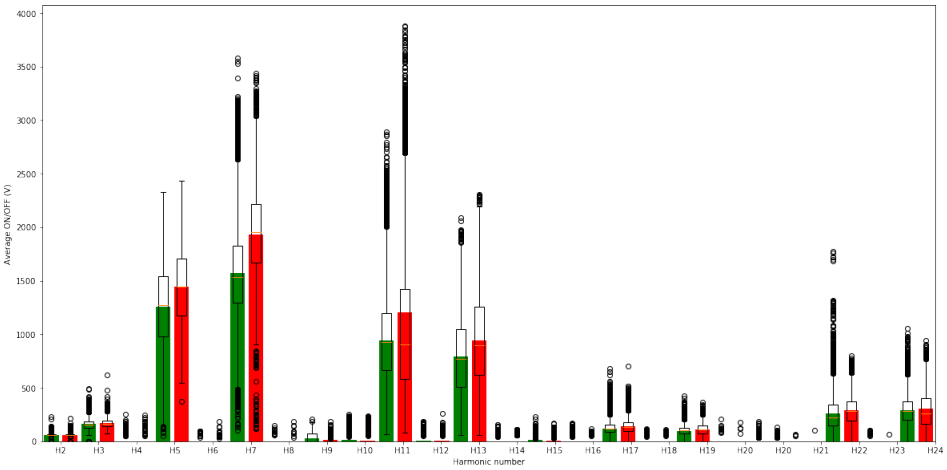

Figure 12: Average value of the THD for ON (green) and OFF (red) state of filter B11, separated by harmonic component, overlaid with boxplot.

5.2 Effect of filter combinations on THD

Our second approach was to consider filter ON/OFF combinations. The idea is that the most common event is many filters being ON at the same time. For this reason, it is not possible to completely isolate the effect of one particular filter. In order to do this analysis, we introduce a binary notation.

The objective of this section is to understand the type of filter combinations used, the proportion of their usage, and the effects of the different filter combinations on THD. We plotted these data across the different months. Thereafter, we realised the filter combinations used in April were uniquely different from the filter combinations during the other five months. Results for Point A are shown in Figure 13 and 14.

Interpreting the plots

The binary code for the filter combinations show the ON/OFF status for each filter, where the leftmost digit indicates the status of filter B11, and the rightmost digit for filter B22 (1 for ON, 0 for OFF).

For each filter combination (x-axis), the box plots indicate the distribution of the THD values. The red dot and the text indicate the mean of the THD values, and the red line shows the percentage occurrence of the filter combination in the time period. The y-axis is shared for both the THD values and the percentage occurrence of filter combinations.

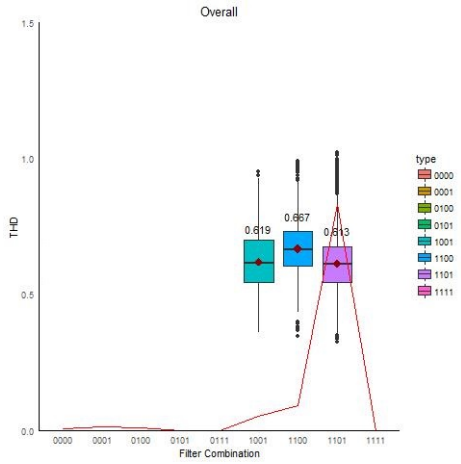

Figure 13: Filter Activation Combination for every month except *April*.

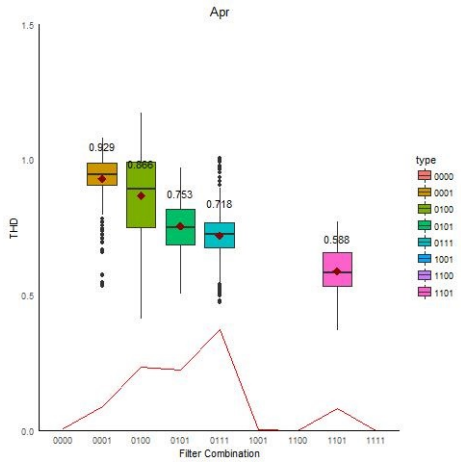

Figure 14: Filter Activation Combination for *April*.

Analysis

As seen from Figure 13, every month except April predominantly uses filter combinations 1101 and 1100, with the rest of the combinations occurring less than 5% of the time. However, Figure 14 shows that in April, the most commonly occurring combinations are 0111, 0100, followed by 0101. This is highly unusual and different from the other months, all of which follow a pattern similar to Figure 13. These plots were similar to Figure 13 and hence were omitted.

Additionally, April is the month in the dataset with the highest average THD values. In Figure 14 again, we can see that the median and mean THD values for filter combinations 0100, 0101 and 0111 are much higher than those values for 1101, which is the predominant combination for the other months. Hence, the spike in THD values in the month of April could be attributed to the choice of filter combinations which seem to cause more harmonic distortion.

5.3 Visualization of the variations in THD

Until now we have analyzed the statistics of the THD data sequence under different segmentations. To gain more insight, we will devote this section to visualizing the variations that we expect to find given our previous results.

In order to do this, we have created a type of visualization with a line graph for the THD value over the day and a red background color representing the period of the day when each filter is turned OFF. In the following analysis, we focus on the time period of 1st November to 2nd April, as the characteristics and activation of filters in April is very different.

Interpreting the plots

An example plot for a single day is shown below for the day of 12 December 2016. The green line represents THD at point A and the purple line represents THD at point B. In this case, B21 filter was turned OFF for the whole day while filters B11, B12 and B22 were turned OFF during brief periods of the day. The y-axis shows the scale of THD and the x-axis shows the order of the reading through the day. Since readings are taken every 10 minutes, there are a total of 144 readings for each day. In this particular figure, we see that there are certain spikes in the THD in both points A and B, close the the red periods where the filters were turned OFF. In the following section, we will study groups of such plots to analyze the effect of filters on THD.

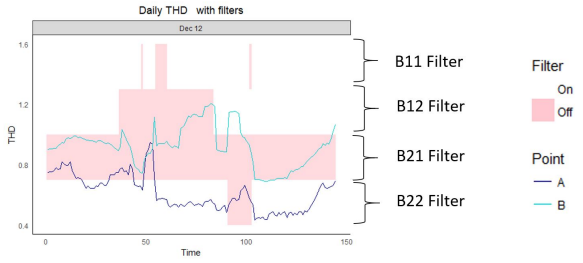

Figure 15: Plot of how THD and filter status changed throughout 12 December 2016.

Variation of THD when B11, B12 and B22 are on

In section 4.2, we saw that the most common filter combination is B11, B12 and B22 ON, while B21 remains OFF. Hence, we use this as a baseline to observe how THD varies typically. We isolated the days in which only this filter combination was present. Figure 16 shows a sample of 50 days. We see that THD generally has the same pattern over the day, with three "humps", firstly increasing then decreasing, then increasing and decreasing again, finally increasing again. We note that THD values do fluctuate without any filter changes, but there are rarely any sudden spikes in the THD values.

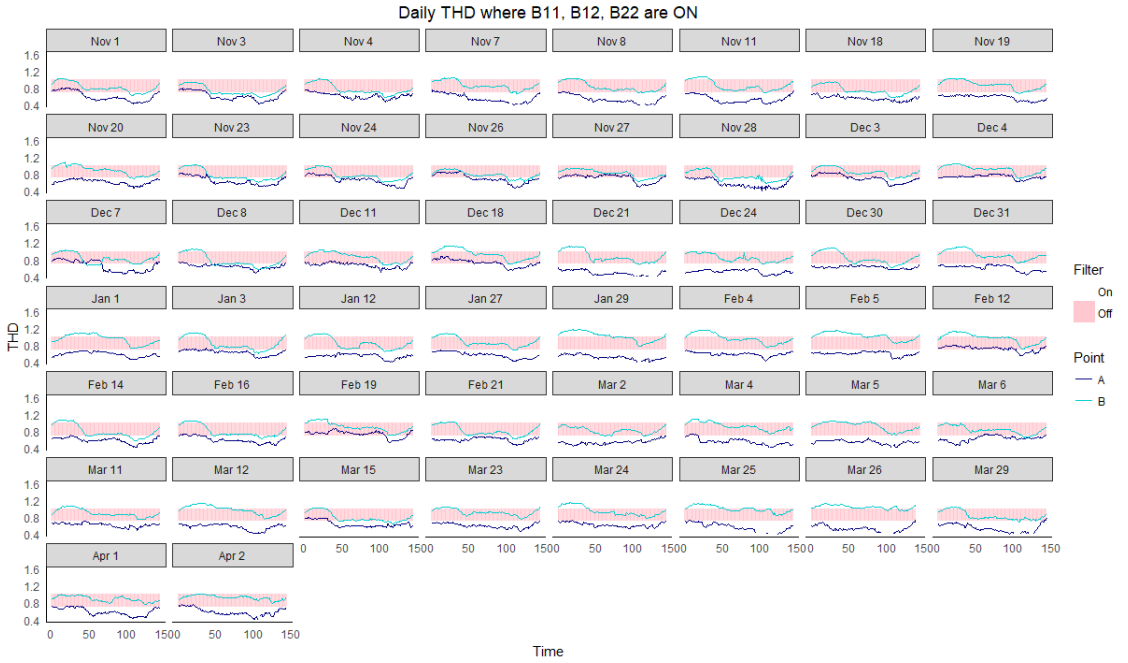

Figure 16: Plot of daily THD variation in point A and B where B11, B12 and B22 filters are turned on.

Variation of THD when B22 is turned off

From our previous analysis we have the suspicion that filter B22 plays an important role in the THD. To study this effect without inference of the other filters, we isolated the following 20 days where only B22 was turned off. We can see during the time periods where the B22 filter was turned off, THD tends to be higher. In particular, when the filter was just turned off, we see a sharp increase in the THD value. This is very unlike the behavior of THD in the previous figure (Figure 16), where no sudden spike was observed. Furthermore, when the filter is turned on, the THD value seems to drop back down. However, we observe that from 22nd February to 1st March where the filter was turned off for eight days, the THD value seems to stabilise.

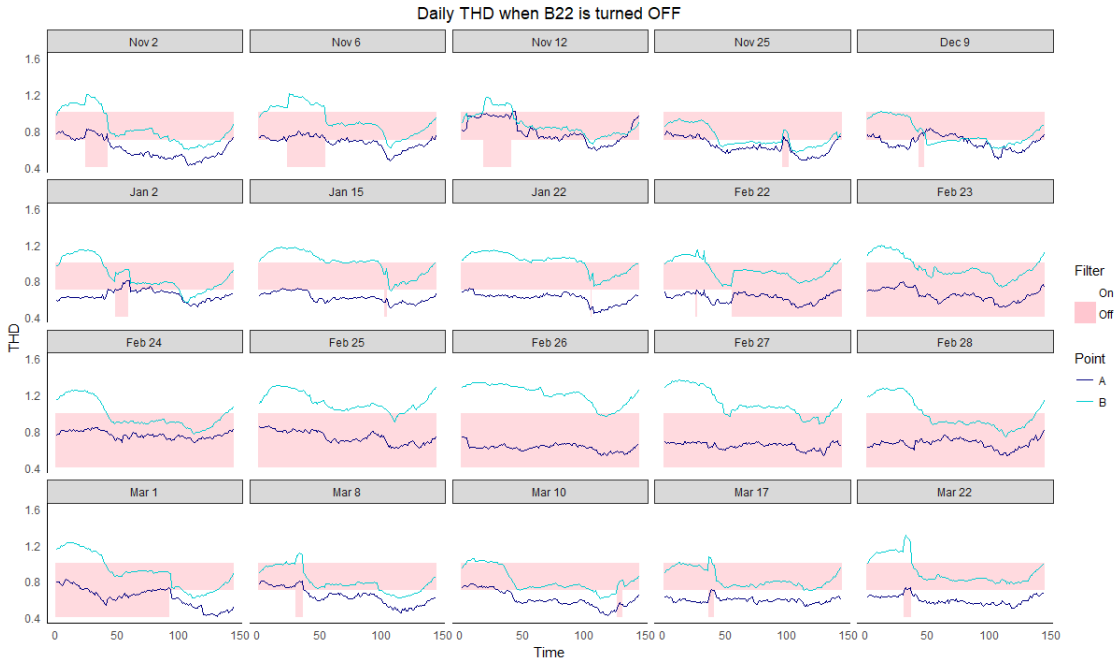

Figure 17: Plot of daily THD variation in point A and B where B22 filter was turned off.

We wanted to make this correlation more precise. In order to do so, we made a smoothed density plot of THD values when B22 was ON and OFF (see Figure 18). Clearly, the distribution of THD has the same shape in the ON and OFF case with a noticeable shift to the right in both points A and B.

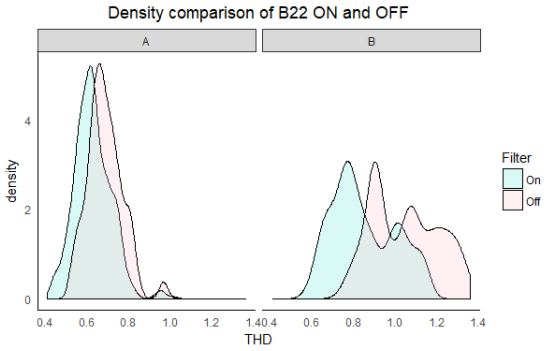

Figure 18: Density plot of THD values in point A and B where B22 filter was turned off.

To test our hypothesis that there the THD value has a different distribution when the filter is on or off, we use a two-sample Kolmogorov-Smirnov test. This is a nonparametric test, with the null hypothesis that both datasets are sampled from the same distribution. Running the test, p-value smaller than $2.2 * 10^{-16}$ was obtained, meaning that we reject the hypothesis that two dataset are sampled from the same underlying distribution. Thus, we conclude that B22 filter actively reduces the THD.

Variation of THD when B12 is turned off

Similarly, to study the effect of B12 on THD without inference of the other filters, we isolated the following 14 days where only B12 was turned off. During the short time periods where the B12 filter was turned off, we observe that THD tends to be higher with a spike at the moment when the filter was turned off, while the THD drops when the filter was turned on. However, unlike B22 filter before, there is a smaller sample size and the effect of B12 filter on THD is not as pronounced.

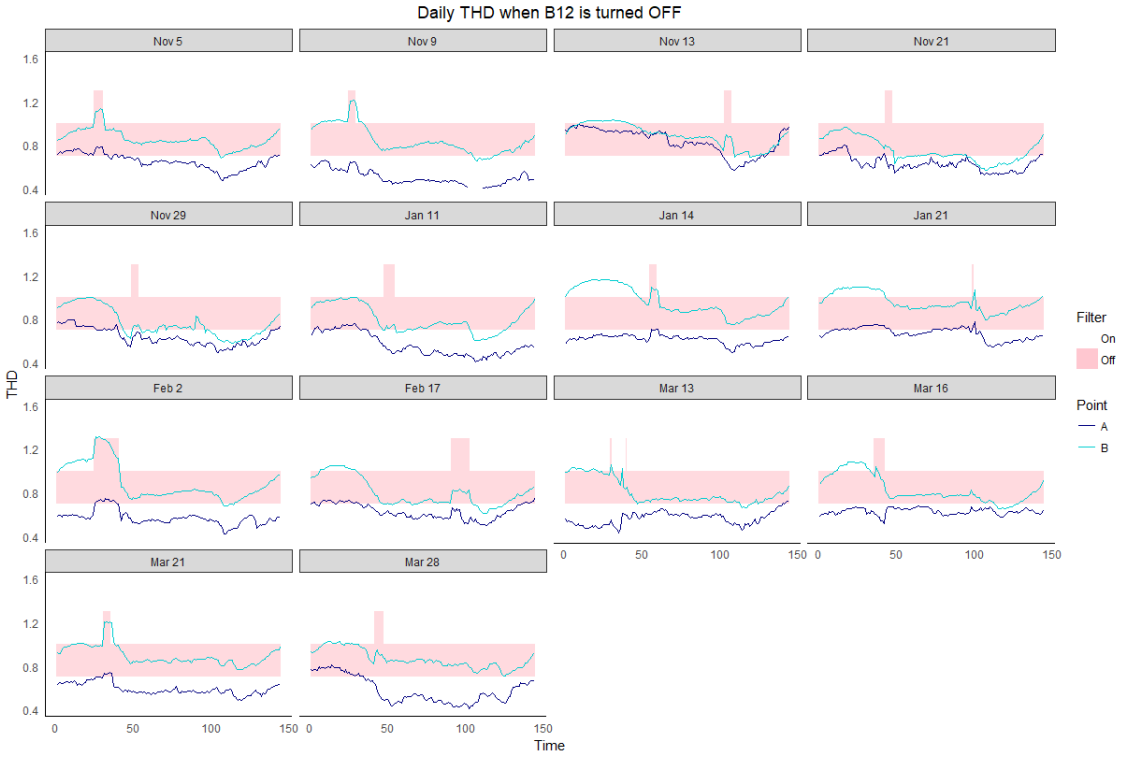

Figure 19: Plot of daily THD variation in point A and B where B12 filter was turned off.

The density plot in Figure 20 shows that, unlike the previous case, the distributions when the filter is on or off look different. In general, when the filter is turned on, there is a higher proportion of smaller THD values. However in point A, we see that there is higher composition for undesirable THD of values higher than 0.9 when the filter is turned on. This could be due to more anomalous values since the sample size is much larger when the filter is turned on.

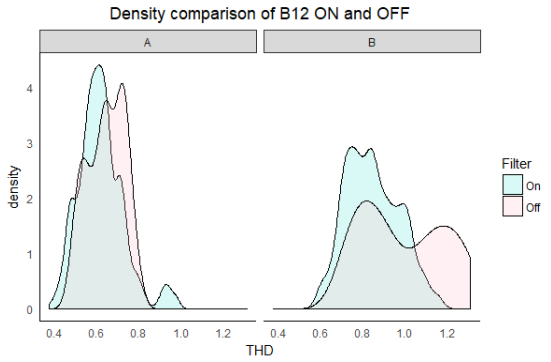

Figure 20: Density plot of THD values in point A and B where B12 filter was turned off.

Using the two-sample Kolmogorov-Smirnov test, a p-value of $2.189 * 10^{-5}$ is obtained. Similarly, we reject the hypothesis that two dataset are sampled from the same underlying distribution. Hence, we may conclude that, while the status of the filter influences THD value, the extent of this effect is not as pronounced, possibly to a lack of data points corresponding to the filter off-state.

Variation of THD when both B12 and B22 are turned off

We can observe a similar trend of spikes in the THD when we consider cases where both filter B12 and B22 are turned off (see Figure 21).

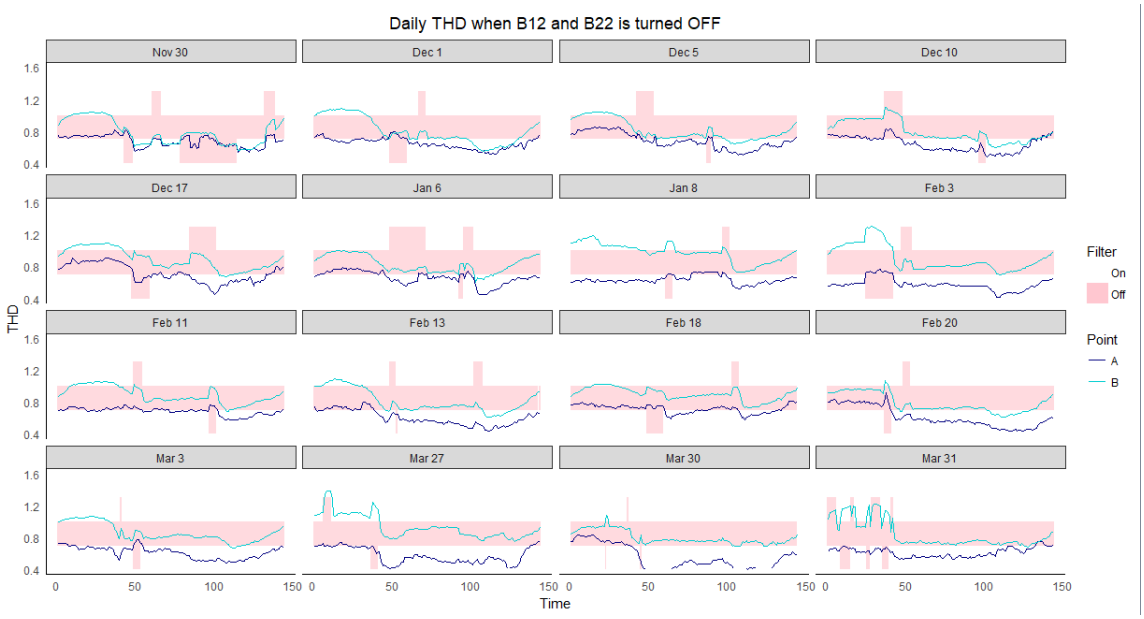

Figure 21: Plot of daily THD variation in point A and B where B12 and B22 filters was turned off.

Variation of THD with multiple filter interaction

In the remaining days, many filters are activated, usually at the same time. Hence, it is hard to study the effect of B11 filter alone. However, we do observe the adverse effect of turning off B12 and B22 filters on THD.

Figure 22: Plot of daily THD variation in point A and B with multiple filters interactions.

Summary of trend of Harmonics

Since THD is heavily influenced by the harmonics, we performed the same visualisation of the harmonics over the period of 6 months to see which harmonic contributes most to the variation of THD. To be concise, we summarised the general characteristics of each harmonic in the following table:

Harmonic	Min.	1st Quantile	Median	3rd Quantile	Max	General trend
2nd	0.00	54.42	59.59	63.93	595.06	Amplitude of the harmonic remains fairly stable around 60, with brief periods where it dips to 0. There are random, occasional short spikes upwards.
3rd	0.0	117.6	154.0	194.5	1142.4	Trend of harmonic seems fairly constant across all the days. There are random spikes upwards in amplitude that does not seem to coincide with switching of filters OFF.
4th	0.000	0.000	0.000	0.000	793.203	Harmonic amplitude is mostly 0.
5th	51.39	1296.96	1757.74	2354.97	4521.30	The shape of the line graphs of the fifth harmonic closely resembles the THD. Similarly, we observe an increase in amplitude when certain filters are switched OFF.
6th	0.0000	0.0000	0.0000	0.0000	440.0683	Harmonic amplitude is mostly 0.
7th	103.3	1395.5	1692.5	2012.8	3726.3	Trend of harmonic seems fairly constant across all the days. There are random spikes upwards in amplitude that sometimes seem to coincide with switching of filters.
8th	0.0000	0.0000	0.0000	0.0000	300.8699	Harmonic amplitude is mostly 0.
9th	0.00	0.00	0.00	0.00	234.72	Harmonic amplitude is mostly 0.
10th	0.000	0.000	0.000	0.000	256.101	Harmonic amplitude is mostly 0.
11th	0.0	892.7	1099.7	1354.7	3879.9	Trend of harmonic seems fairly constant across all the days. There are random spikes upwards and downwards in amplitude that sometimes seem to coincide with switching of filters OFF.
12th	0.000	0.000	0.000	0.000	257.327	Harmonic amplitude is mostly 0.
13th	0.0	695.8	895.4	1198.5	4021.3	Harmonic seems display different trends on different days. There are random spikes upwards and downwards in amplitude that sometimes seem to coincide with switching of filters OFF.
14th	0.0000	0.0000	0.0000	0.0000	159.5601	Harmonic amplitude is mostly 0.
15th	0.000	0.000	0.000	0.000	229.715	Harmonic amplitude is mostly 0.
16th	0.0000	0.0000	0.0000	0.0000	182.8592	Harmonic amplitude is mostly 0.
17th	0.00	84.09	118.47	157.52	702.74	Harmonic amplitude seems to fluctuate erratically, dropping to 0 for certain time intervals.
18th	0.0000	0.0000	0.0000	0.0000	117.7291	Harmonic amplitude is mostly 0.
19th	0.00	0.00	0.00	101.37	425.99	Harmonic amplitude differs for point A and B. At point A, harmonic amplitude seems to fluctuate erratically, dropping to 0 for certain short time intervals. At point B, harmonic amplitude is 0 most of the time, only fluctuating erratically for certain short time intervals.

Table 4: Summary of characteristics of 2nd to 19th harmonic

Harmonic	Min.	1st Quantile	Median	3rd Quantile	Max	General trend
20th	0.00000	0.00000	0.00000	0.00000	211.12016	Harmonic amplitude is mostly 0.
21st	0.0000	0.0000	0.0000	0.0000	187.7155	Harmonic amplitude is mostly 0.
22nd	0.00000	0.00000	0.00000	0.00000	101.31105	Harmonic amplitude is mostly 0.
23rd	0.0	225.8	397.2	581.7	1776.3	Harmonic seems display different trends on different days. There are "plateau" and "bathtub" shaped patterns in the line graph that do not seem to correlate with switching of filters.
24th	0.00000	0.00000	0.00000	0.00000	103.48451	Harmonic amplitude is mostly 0.
25th	0.0	255.5	386.9	675.1	1827.0	Trend of harmonic seems fairly constant across all the days. There are random spikes upwards and downwards, "plateaus" and "bathtub" patterns in amplitude that sometimes seem to coincide with switching of filters OFF.

Table 5: Summary of characteristics of 20th to 25th harmonic

Analysis of trend of Harmonics

Based on the magnitude of the amplitude of the Harmonics, we see that the fifth harmonic contributes most significantly to the variation in the THD. Other harmonics that may contribute significantly to the trend of THD include 7th, 11th and 13th harmonic. An interesting trend is that most of the even numbered harmonics, with the exception of the second harmonic has zero amplitude.

5.4 Cross-correlation of filter status with THD

To support our findings, we also calculated the cross-correlation of the THD signal with each of the filter-status signals (see Figure 23).
These results back up our findings.

- First of all, the effect of filter B11 is masked by the constant switching of other filters. Since there is no time period where only this filter is being switched, its correlation with the THD is noisy. However, there seems to be, if any, a negative correlation, i.e. the filter being on, improves the THD.
- The (negative) correlations with filters B12 and B22 are very apparent.
- The noisy nature of the cross-correlation with B21 could be due to the small amount of samples. The reader must bear in mind that the filter-status signal we are using here is mostly zero. That is why if we delay the filter switching, the THD only multiplied by zeros (hence the zero cross-correlation for positive delay). On the other hand, if we instead delay the THD, we are multiplying different parts of the signal by a small quasi-step function, and therefore the cross-correlation mimics the variations of the THD. The positive peak at zero delay could be related to the different statistics of the month of April. We could not find clear evidence to blame this different statistics to filter B21, given the small amount of data.

It is interesting to see that there are some side spikes in the cross-correlation of B22. This suggests that there could be some periodicity in the switching of this filter. Indeed, if one plots the autocorrelation of filter B22 on top of its cross-correlation with the THD, it is apparent that the cause of the spikes is some kind of repeated pattern in the status of filter B22 (Figure 24).

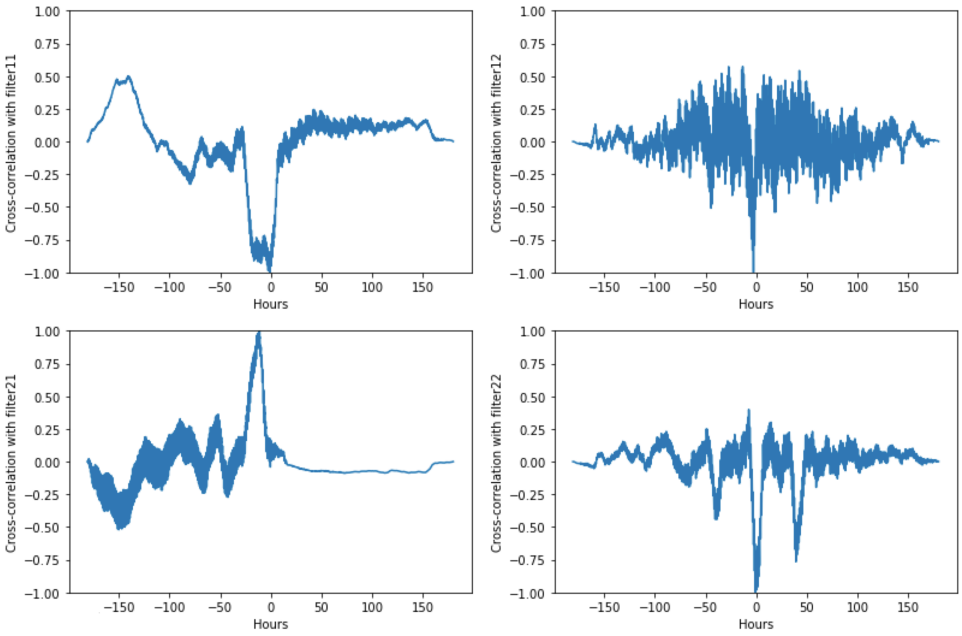

Figure 23: Cross correlation of the THD signal with each of the filter-status signals

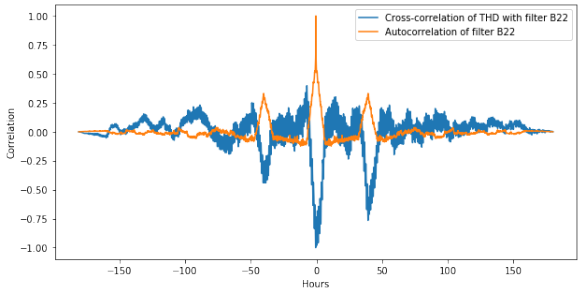

Figure 24: Cross correlation of the THD signal with each of the filter-status signals

5.5 Effect of switching filters ON/OFF over time

To further analyze the effects of switching ON/OFF individual filters on the THD values over time, two dimensional density plots were made for each filter, where the x-axis shows the time elapsed in hours since the last time that particular filter was switched ON or OFF, and the y-axis shows the normalized THD value at time t at point B. The plots are shown below, along with the formula used to normalize the THD values.

$$THD(t)_{Normalized} = \frac{THD(t) - THD(t)_{On/Off}}{THD(t)_{Off}}$$

$THD(t)_{On/Off}$ refers to the THD value when the filter was last switched on/off.

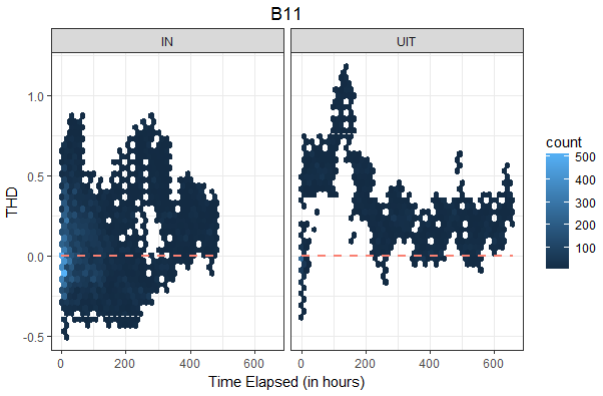

Figure 25: Density plot of normalized THD values over time after B11 filter is switched ON (left) and OFF (right)

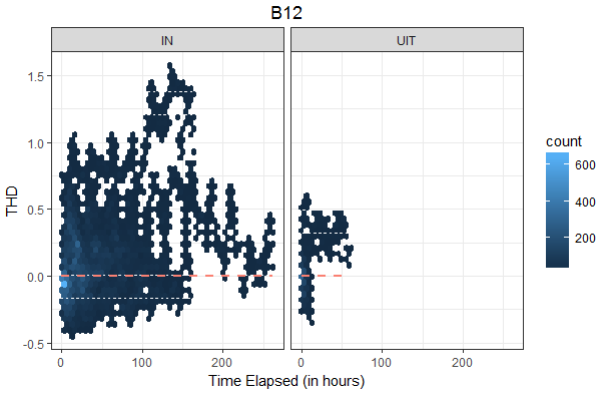

Figure 26: Density plot of normalized THD values over time after B12 filter is switched ON (left) and OFF (right)

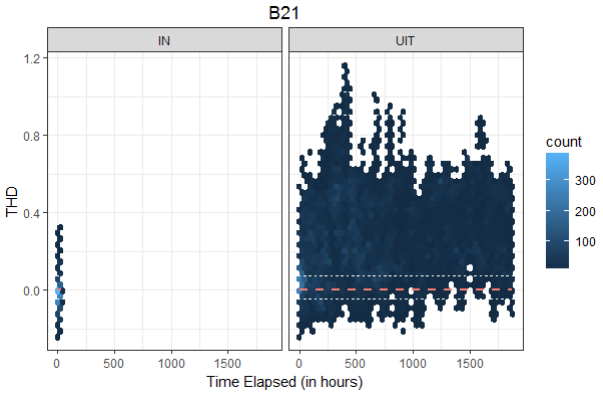

Figure 27: Density plot of normalized THD values over time after B21 filter is switched ON (left) and OFF (right)

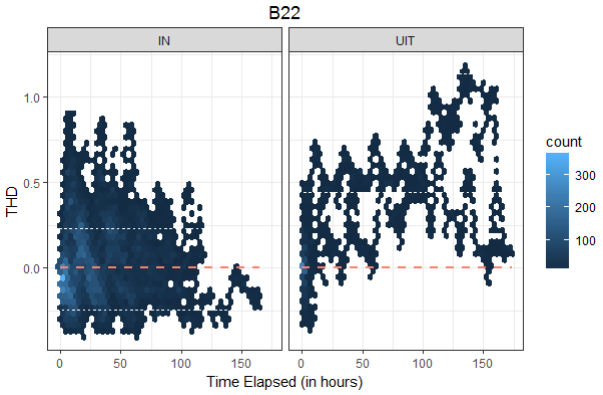

Figure 28: Density plot of normalized THD values over time after B22 filter is switched ON (left) and OFF (right)

Overall Trend

Most of the density plots show that over time, the THD values appear to fluctuate with a sinusoidal trend. The sinusoidal trends appear to be overlapping; some with of 24 hours (daily) and some with significantly longer periods.

From Figure 28, we can see that after filter B22 is switched off, the THD values tend to increase over time. Furthermore, there seems to be downward trend on the THD value in the long run after the filter is switched on. These 2 plots thus suggest that switching on filter B22 helps to lower the THD value in the long run, while switching it off worsens it.

Unlike for filter B22, there seems to be no clear trend for the effect of switching on/off these filters on the THD. The density plots for these filters are also too dense to be able to interpret any trends.

6 Conclusions

We have employed various data visualization and statistical analysis techniques to study the correlation between switching of filters and THD. First of all, we performed different studies on the THD as a time series (section 4), to conclude that the behavior of the signal in April showed some clear statistical differences.

A first look into the ON/OFF statistics (section 5.1), gave us a clue that filters B11, B12 and B22 have a positive impact in the THD (they reduce it). In section 5.2 we made an attempt to generalize this analysis to different filter combinations. However, given that the data is not close to being uniformly distributed over different combinations, we were not able to draw major conclusions.

On the other hand, with our data visualization techniques, density calculations and cross-correlation, we were able to see the following:

- We could not isolate the effect of filter B11 in order to have an irrefutable conclusion. However, we see in Section 4.4 that THD tends to get worse the moment the filter B11 is switched OFF.
- Turing ON filters B12 and B22, actively helps to reduce the THD, although the correlation with filter B12 is smaller
- Due to the differences in characteristics of THD in April, and the fact that Filter B21 was mainly only turned on in April, we could not perform a sound analysis on the effects on Filter B21.

Nevertheless, we see some promising evidence that these Filters do aid in reducing THD values overall.

With only half a year worth of data from November to April, we cannot isolate the causes of the erratic behavior of THD, whether it is caused by Filter B21 or it is a seasonal effect. Thus, further study on a full year of data is still required to understand the effect of filters on THD. More analysis could be done on the harmonics to understand their full effect on THD and audible noise, as we observed that certain harmonics are more prominent than others. Due to legal issues with audible noise data produced by the transformers, we were regrettably unable to get the dataset on time so more work regarding the effect of THD on audible noise produced from the transformer needs to be done.

References

[1] https://www.tennet.eu/company/profile/about tennet/.

[2] C. Ross, "Experiments on the active control of transformer noise," *Journal of Sound and Vibration*, vol. 61, no. 4, pp. 473–480, 1978.

[3] G. E. Warnaka, "Active attenuation of noise-the state of the art," *Noise Control Engineering Journal*, vol. 18, pp. 100–110, 1982.

[4] A. J. Moses, P. I. Anderson, T. Phophongviwat, and S. Tabrizi, "Contribution of magnetostriction to transformer noise," in *Universities Power Engineering Conference (UPEC), 2010 45th International*, pp. 1–5, IEEE, 2010.

[5] R. B. Cleveland, W. S. Cleveland, and I. Terpenning, "Stl: A seasonal-trend decomposition procedure based on loess," *Journal of Official Statistics*, vol. 6, no. 1, p. 3, 1990.

[6] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*. OTexts, 2014.

Appendix I: Work distribution

Name	Tasks
Layla	(4) Time Series Analysis
Rui Jiang	Appendix II
Zhe Min	(5.2) Effect of filter combinations on THD, (5.5) Effect of switching filters ON/OFF over time
Lara	(1) Introduction, (3) The dataset, (5.1) ON/OFF data statistics, (5.4) Cross-correlation
Rei Yun	(5.3) Visualisation of variations in THD

Appendix II: Noise data

Note: We obtained the noise data only five days before the deadline. In this appendix we present a preliminary study that is as complete as possible given the time limitations.

Bewoner is the noise data collected from the resident, Loc1 and Loc2 are data collected from the station. It is reasonable to ignore the delay caused by the distance between Loc1 and Loc2. The delay caused by the distance between Bewoner and stations can be set as 1 sec, and this can be seen as a constant. We check the pearson correlation coefficient among those to see which data works better as the generator of the noise.
Since the data is collected by second, we split it into two parts to save some computing time.

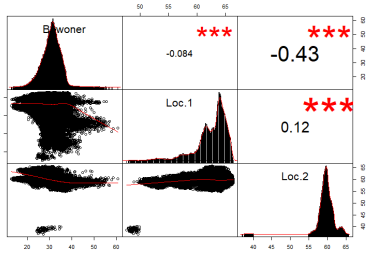

Figure 29: 100Hz Correlation for week 1-8

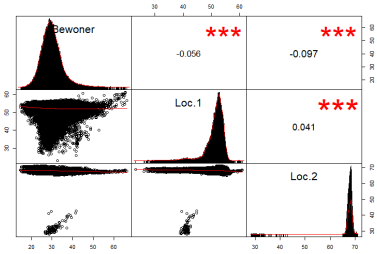

Figure 30: 200Hz Correlation for week 1-8

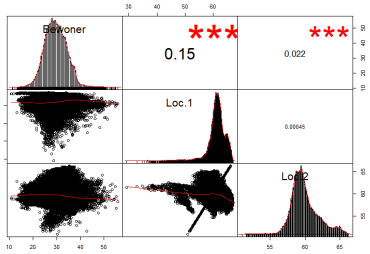

Figure 31: 100Hz Correlation for week 9-16

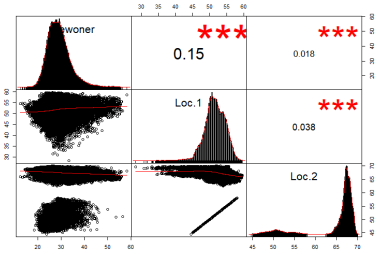

Figure 32: 200Hz Correlation for week 9-16

It is strange to see that the correlation between Loc1 and Loc2 is so weak, and that between Locs and Bewoner can even be negative at some points. This may be caused by some physical phenomena which needs to be further checked.
Though we can't use the original data as time series because of the lack of stationarity and the strong trend inside it, we can use the difference of the data as time series. However, even these data were detrended and fitted with MA(2) models, the correlation between them are still not what we expect.
In conclusion, we think the noise data and the equipment which collected these data need to be checked again. It would be nice to conduct several experiments in the station to calibrate the standard.