

PRELIMINARY WORK OF MASTER'S THESIS

---

## Baseline Analysis of Water Supply Data (BAWSD)

---

Jonas Kjeld Kirstein (JKI)

14. december 2015

Technical Documentation: BAWSD Excel Tool

Case Study: Halsnæs Forsyning

Supervisor: Klavs Høgh (KVH@niras.dk)

# 1 Abstract

Collection of monitoring data from water distribution networks can have several benefits. However, often substantial knowledge of the underlying mechanisms are needed, before proper conclusions can be drawn from the analysis of measurement results. It is the aim of this paper to establish a tool that can help to do a simple pre-analysis of the raw data obtained from data loggers in water distribution networks. The aim of the “Baseline Analysis of Water Supply Data Excel Tool” is to give the user the opportunity to generate: 1) raw plots of the obtained data ; 2) average/group the data in a user-chosen interval and plot these new results; and 3) make plots of the weekly water consumption based on a user chosen period. The realization of all three aims occurred with real flow measurements from the water distribution network of Halsnæs Forsyning.

# 2 Preface

This report is part of an internship in the consultancy company NIRAS A/S during fall 2015. The report can be seen as preliminary work of an upcoming master’s thesis in Environmental Engineering. I would like to thank Klavs Høgh for his help throughout the process of creating the BAWSD Excel Tool.

## Table of contents

<b>1 Abstract</b>	<b>2</b>
<b>2 Preface</b>	<b>2</b>
<b>3 Introduction</b>	<b>3</b>
3.1 BAWSD Excel Tool . . . . .	3
<b>4 Method</b>	<b>4</b>
4.1 Manual input prior to application of BAWSD Excel Tool . . . . .	4
4.2 Generic results . . . . .	4
4.2.1 Zone initialization . . . . .	5
4.2.2 Data initialization . . . . .	5
4.2.3 Interpolation of Data . . . . .	7
4.2.4 Compute zone flow . . . . .	8
4.2.5 Raw Data Plot . . . . .	9
4.2.6 Simple time series transformation . . . . .	9
4.2.7 $\Delta T$ plot . . . . .	10
4.2.8 Weekly average plot . . . . .	10
<b>5 Results &amp; Discussion</b>	<b>12</b>
5.1 WDN of Halsnæs Forsyning . . . . .	12
5.2 Raw data plot . . . . .	13
5.3 $\Delta T$ data plot . . . . .	13
5.4 Weekly average plot . . . . .	15
5.5 Additional plots . . . . .	17
<b>6 Conclusion</b>	<b>18</b>
<b>A Manual Input Guide for BAWSD Excel Tool</b>	<b>20</b>
<b>B Halsnæs Water Distribution Network</b>	<b>23</b>

### 3 Introduction

There are several benefits with a collection of long term monitoring data in water distribution networks (WDN). For example, results obtained from analysing flow and pressure measurements are capable of calibrating and optimizing models or finding leakages in the system. However, several problems are encountered when focusing at the individual *raw* results from installed data loggers in the network. Having different start and end points of logging, varying log intervals, missing data and surveying consumers with different behaviours are few of the problems that should be addressed before it is possible to use and analyse raw log results from WDNs correctly.

In this study, the overall aim is to make a simple pre-analysis of the raw data obtained by loggers in a WDN. As a case study, data from Halsnæs Forsyning will be used. The WDN of Halsnæs Forsyning has been sectioned into 27 different district metered areas (DMAs) that mainly consist of domestic households. The majority of these DMAs has a data logger installed at the inlet and/or outlet of the zone, so the water consumption either can be measured or computed. To be able to show and analyse results from the individual zones, a Baseline Analysis of Water Supply Data (BAWSD) Excel Tool has been created that is capable of doing a simple pre-analysis of the obtained data from the water distribution network.

The overall object of this tool can be sectioned into three aims: 1) to generate a simple plot of the (raw) measured data from each logger and zone; (2) to aggregate/extend the logs to a user-defined interval and generate a plot based on this interval; and (3) to compute the mean weekly water consumption over a given period.

#### 3.1 BAWSD Excel Tool

The BAWSD Excel Tool has been programmed in VBA to establish a connection between a SQL-database and Excel. In the SQL database, different parameters, such as flow or pressure, different logging intervals, such as minute-by-minute or hourly logs and different periods of logging have been stored for data loggers in a WDN. The BAWSD Excel Tool has been sectioned into three sheets to make it clear to the user where to generate and where to insert data. The first sheet, *Initialization*, is used to establish a connection between a linked database and Excel. However, the main functionalities of the tool can be found in the second sheet, *Generate\_Data* and a screenshot of this sheet is shown in Figure 1.

CHECK ALL		USER INPUT		VALID TIME PERIOD		COMPUTER OUTPUT BASED ON T1 & T2									
ID	Check	T1	T2	Chosen dT [min]	Tmin (Log Start)	Tmax (Log End)	Biggest Gap Raw Data Time Origin	Biggest Gap - Raw Data[minutes]	From	To	Lowest Data Coverage	Total Data Coverage [%]	Min Raw dT [min]		
MD01	<input type="checkbox"/>	2015-11-17 00:00:00	2015-11-18 00:00:00	10000	03-12-2013	30-11-2015	MAG8000_574802H053_Q_Frem		1	17-11-2015 00:01	17-11-2015 00:02	MAG8000_574802H053_Q_Frem	100.0000		1
MD02	<input type="checkbox"/>	2015-11-17 00:00:00	2015-11-18 00:00:00	60	01-01-2000	30-11-2015	MAG8000_575602H053_Q_Frem		1	17-11-2015 00:01	17-11-2015 00:02	MAG8000_575602H053_Q_Frem	100		1
MD03	<input type="checkbox"/>	2014-01-04 00:00:00	2015-11-18 00:00:00	30	02-01-2014	29-11-2015	MAG8000_575602H053_Q_Frem		47521	29-07-2015 06:21	31-08-2015 06:22	MAG8000_574802H053_Q_Frem	99.9342856		1
MD04	<input type="checkbox"/>	2015-11-17 00:00:00	2015-11-18 00:00:00	60	01-01-2015	29-11-2015	54GK01FT02		1	17-11-2015 00:01	17-11-2015 00:02	54GK01FT02	100		1
MD05	<input type="checkbox"/>	2015-11-17 00:00:00	2015-11-18 00:00:00	60	01-01-2015	30-11-2015	56GK01FT02		1	17-11-2015 00:01	17-11-2015 00:02	56GK01FT02	100		1
MD06	<input type="checkbox"/>	2015-11-17 18:00:00	2015-11-18 00:00:00	60	03-12-2013	30-11-2015	MAG8000_574502H053_Q_Frem		1	17-11-2015 18:01	17-11-2015 18:02	MAG8000_574502H053_Q_Frem	100		1
MD07	<input type="checkbox"/>	2015-11-17 18:00:00	2015-11-18 00:00:00	60	21-03-2013	05-11-2014									
MD08	<input type="checkbox"/>	2015-11-17 18:00:00	2015-11-18 00:00:00	60	01-01-2000	30-11-2015	MAG8000_574902H053_Q_Frem		1	17-11-2015 18:01	17-11-2015 18:02	MAG8000_574902H053_Q_Frem	100		1
MD09	<input type="checkbox"/>	2015-11-17 18:00:00	2015-11-18 00:00:00	60	23-11-2013	30-11-2015	MAG8000_575102H053_Q_Frem		1	17-11-2015 18:01	17-11-2015 18:02	MAG8000_575102H053_Q_Frem	100		1
MD10	<input type="checkbox"/>	2015-11-17 18:00:00	2015-11-18 00:00:00	60	03-12-2013	11-06-2015									
MD11	<input type="checkbox"/>	2015-11-17 18:00:00	2015-11-18 00:00:00	60	06-10-2014	30-11-2015	MAG8000_575802H053_Q_Frem		1	17-11-2015 18:01	17-11-2015 18:02	MAG8000_575802H053_Q_Frem	100		1
MD12	<input type="checkbox"/>	2015-11-17 18:00:00	2015-11-18 00:00:00	60	03-12-2013	02-07-2015									
MD13	<input type="checkbox"/>	2015-11-17 18:00:00	2015-11-18 00:00:00	60	03-03-2015	01-07-2015									
MD14	<input type="checkbox"/>	2015-11-17 18:00:00	2015-11-18 00:00:00	60	01-01-2015	30-11-2015	72GK02FT02		1	17-11-2015 18:01	17-11-2015 18:02	72GK02FT02	100		1
MD15	<input type="checkbox"/>	2015-11-17 18:00:00	2015-11-18 00:00:00	60	03-12-2013	30-11-2015	MAG8000_574302H053_Q_Frem		1	17-11-2015 18:01	17-11-2015 18:02	MAG8000_574302H053_Q_Frem	100		1
MD16	<input type="checkbox"/>	2015-01-01 00:00:00	2015-11-20 00:00:00	60	01-01-2015	30-11-2015	52GK10FT02		1441	17-02-2015 23:59	19-02-2015 00:00	52GK10FT02	95.15489255		0
MD17	<input type="checkbox"/>	2015-11-17 18:00:00	2015-11-18 00:00:00	60	03-12-2013	10-05-2015									
MD18	<input type="checkbox"/>	2015-11-17 18:00:00	2015-11-18 00:00:00	60	17-03-2014	30-11-2015	MAG8000_575902H053_Q_Frem		15	17-11-2015 18:15	17-11-2015 18:30	MAG8000_575902H053_Q_Frem	100		15
MD19	<input type="checkbox"/>	2015-11-17 18:00:00	2015-11-18 00:00:00	60	03-12-2013	30-11-2015	MAG8000_575002H053_Q_Frem		1	17-11-2015 18:01	17-11-2015 18:02	MAG8000_575002H053_Q_Frem	100		1
MD20	<input type="checkbox"/>	2015-11-17 18:00:00	2015-11-18 00:00:00	60	03-12-2013	14-11-2015									
MD21	<input type="checkbox"/>	2015-11-17 18:00:00	2015-11-18 00:00:00	60	03-12-2013	30-11-2015	MAG8000_574102H053_Q_Frem		1	17-11-2015 18:01	17-11-2015 18:02	MAG8000_574102H053_Q_Frem	100		1
MD22	<input type="checkbox"/>	2015-11-17 18:00:00	2015-11-18 00:00:00	60	22-04-2000	02-07-2015									
MD23	<input type="checkbox"/>	2015-11-17 18:00:00	2015-11-18 00:00:00	10	01-01-2000	30-11-2015	MAG8000_574202H053_Q_Frem		1	17-11-2015 18:01	17-11-2015 18:02	MAG8000_574202H053_Q_Frem	100		1
MD24	<input type="checkbox"/>	2015-11-17 18:00:00	2015-11-18 00:00:00	60	03-03-2015	01-07-2015									
MD25	<input type="checkbox"/>	2015-11-17 18:00:00	2015-11-18 00:00:00	60	01-01-2015	30-11-2015	53GK01FT02		1	17-11-2015 18:01	17-11-2015 18:02	53GK01FT02	100		1
MD26	<input type="checkbox"/>	2015-11-17 18:00:00	2015-11-18 00:00:00	60	03-03-2015	02-07-2015									
MD27	<input type="checkbox"/>	2015-11-17 18:00:00	2015-11-18 00:00:00	60	04-12-2013	30-11-2015	MAG8000_575502H053_Q_Frem		1	17-11-2015 18:01	17-11-2015 18:02	MAG8000_575502H053_Q_Frem	100		1

Initialization

PLOT ONLY ZONE

RAW data plot

dT data plot

MIN & MAX

BOUNDARIES

Weekly plot

Figur 1: Screenshot of the buttons that can be pressed in the BAWSD Excel Tool.

Here, the user is capable to generate different plots of the logged data for DMAs. In the third sheet, *SCADA\_Time\_Check*, the user has to insert additional information for the installed loggers of the analysed WDN. The following sections will go into detail how the sheet functions and explain the underlying code.

## 4 Method

In the following, the code/algorithm behind the BAWSD Excel Tool, to generate the pre-analysis of the data, will be described. In other words, the method section could be interpreted as a technical documentation of the Excel Tool. The overall aim of this section is to describe the idea and underlying code of the three different types of plots that can be generated with the Excel Tool. However, before usage of the tool, it is important to consider that some parts of the tool are not generic and require substantial knowledge about the water distribution network and the linked database. Thus, first the manual input will be explained. Afterwards, the generic part of the tool will be listed and explained stepwise. The used examples in this specific section are not in connection to the WDN of Halsnæs and can be interpreted as general examples. On occasion, information can be found in a blue box. The blue box is used to emphasise potential error-sources of the code or certain issues where special attention from the user is needed.

Prior to working with the BAWSD Excel Tool, it has to be mentioned that the underlying code in its current state, during execution, modifies data in the linked database. Under normal circumstances, the algorithm resets the database to its initial state at the end of every execution. However, if an error occurs during the run or the user is forced to stop the execution, it might be that the step of resetting data is not carried out. Hence it is recommended only to run the script on a backup file of the database.

### 4.1 Manual input prior to application of BAWSD Excel Tool

The BAWSD Excel Tool contains three sheets and all of them require specific input from the user. The required manual user input is described in the following:

**Sheet 1 “Initialization”:** Here the specific database properties have to be entered, such as database name, driver, provider, etc. This is essential, because server name, user id and password will likely differ for each user of the Excel Tool. A screenshot illustrating the required user input can be found in Figure 17 in Appendix A.

**Sheet 2 “Generate\_data”:** This is the main sheet of the Excel Tool. However, it has some restrictions: It is only possible to analyse up to 27 DMAs in total and the user has to insert data logger information from the analysed network. For each DMA, the user needs to insert the logger name that measures the inflow to a zone. Moreover, the names of outflow loggers, connected to a DMA, have to be inserted, so a mass balance can be generated. In this case, it is only possible to enter a maximum number of 4 outflow loggers. This means, if a DMA contains more than 4 outgoing data loggers, it is not possible to make a valid water balance. A screenshot illustrating the required user input can be found in Figure 18 in Appendix A.

**Sheet 3 “SCADA\_Time\_Check”:** Each data logger being connected to a DMA can have different log intervals. These have to be entered in this sheet. For example, if a data logger logs in a 60-min interval and later in a 1-min interval, both logging periods have to be entered. Also, if only one logging period exists. A screenshot illustrating the required user input can be found in Figure 19 in Appendix A.

### 4.2 Generic results

The next section describes the code flow step-by-step and how results are obtained by running the Excel Tool. In several cases, user-input is required and if so it is stated. Moreover, for each section, input from the previously described sections is needed. From this point, all results are generated and illustrated in sheet 2: ‘Generate\_data’. The required steps to generate raw data plots are defined in the following sections 4.2.1 to 4.2.4.

### 4.2.1 Zone initialization

A screenshot illustrating the required user input for this section can be found in Figure 20 in Appendix A. The user needs to press the button “Initialization” to generate the described results of this section.

#### User selection:

- One or multiple zones.

#### Output:

- Identifies the overlapping time period, where data is available for all loggers.

First, the user selects one or multiple zones. Afterwards, the code is capable of identifying all loggers that are connected to the selected zone(s). Next, for each logger the start and end time of logging will be determined. With this information, it is possible to identify a time period where data is available for all loggers in one zone. For example, consider a zone with two loggers:

Logger 1 = 01-01-2015 to 31-11-2015	} 01-01-2015 to 27-04-2015
Logger 2 = 01-01-2014 to 27-04-2015	

In this example, data can only be analysed for the period between the 01-01-2015 and the 27-04-2015. All initial steps of the Excel Tool until this section are summarized in the code flow diagram of Figure 2.

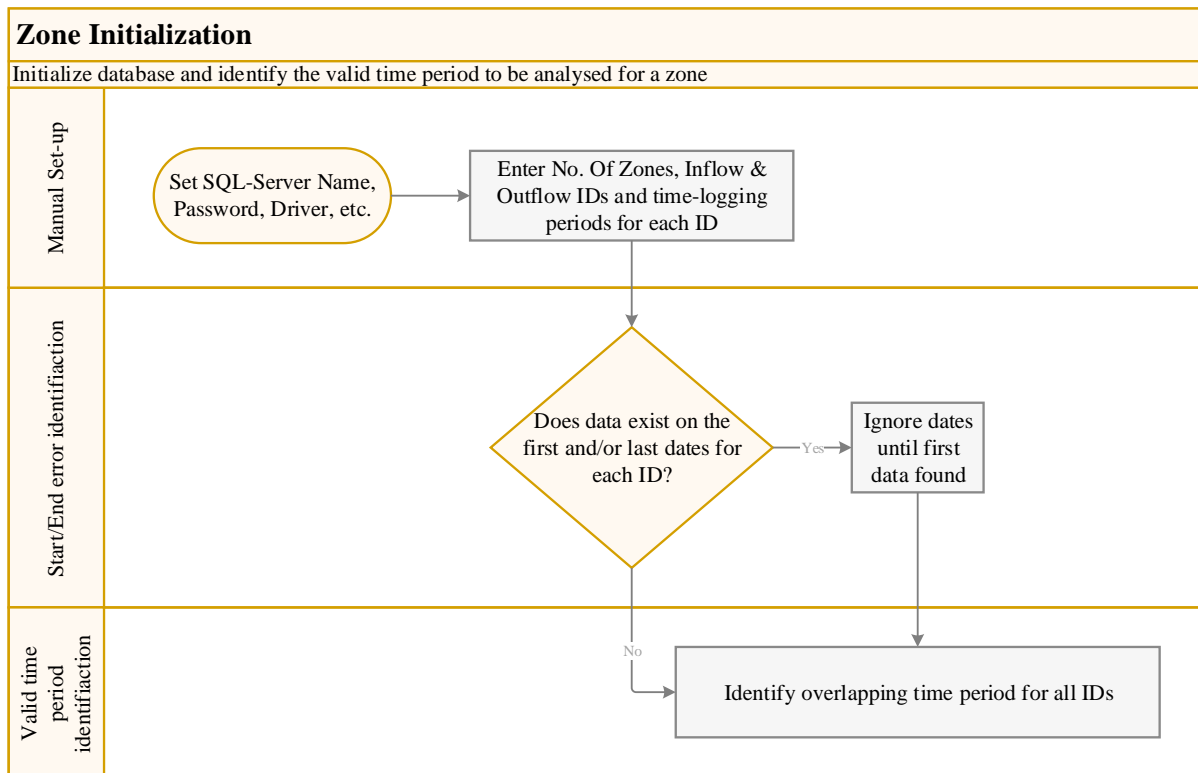


Figure 2: Code flow diagram describing the first step of the data analysis: finding an overlapping time series.

### 4.2.2 Data initialization

A screenshot illustrating the required user input can be found in Figure 21 in Appendix A.

#### User selection:

- Time period

**Output:**

- Identification of missing dates that are not in compliance with individual logging periods.
- Identification of dates where the logged value is NULL.
- Identification of all duplicate values.
- Identification of largest data gap.
- Identification of total data coverage.

First, for the to be analysed zone, the user has to choose a time period that is in compliance with the results from section 4.2.1. In this period, it is possible that periods for a data logger exist, where data is missing. For example, consider a logger being set to log every hour. The obtained data set might look like the following:

Timestamp	[hh:mm]	12:00	13:00	14:00	15:23	16:23
$\Delta T$	[hh:mm]	-	1:00	1:00	1:23	1:00

It applies for all logs that each time step is valid until the next time step occurs. So in this case, time step 12:00 covers the period between 12:00 to 13:00. In this example, we can see that the difference between the time steps 15:23 and 14:00 is not one hour ( $\Delta T = 1:23$ ). However, the logger was supposed to log every hour and we want every  $\Delta T$  to be of one hour or less. Thus, a new date point is added at 15:00, which makes  $\Delta T$  equal to 1:00 or less at every point in the example data set:

Timestamp	[hh:mm]	12:00	13:00	14:00	15:00	15:23	16:23
$\Delta T$	[hh:mm]	-	1:00	1:00	1:00	0:23	1:00

From the above, we have recognized 15:00 as a missing time step. Next, it is possible that the data logger has logged a date without a valid value (i.e. NULL-value). As for example in this data set:

Timestamp	[hh:mm]	12:00	13:00	14:00	15:23	16:23
Value		1.3	NULL	0.3	5.9	4.5

Easily, we recognize the timestamp of 13:00 to have a missing NULL - value. Finally, there are also cases where duplicate data has been logged. For example in the following data set:

Timestamp	[hh:mm]	12:00	12:00	13:00	14:00	15:23	16:23
Value		1.3	1.3	NULL	0.3	5.9	4.5

We observe that the timestamps of 12:00 and the values of 1.3 occur twice. Consequently, one of these logs is removed. In case multiple duplicates exist, always one is kept only.

The code does not take into account if duplicate timestamps contain different logged values!

If we want to find the largest data gap of a logger and compute the total data coverage for the entire period in a zone, the following example data set is used:

Timestamp	[hh:mm]	12:00	13:00	13:34	14:00	15:23	16:23
$\Delta T$	[hh:mm]	-	1:00	0:34	0:26	1:23	1:00
Value		1.3	NULL	0.8	0.3	5.9	4.5

The *largest data gap* is the step in time, where  $\Delta T$  is highest. In the above example data set, the largest data gap is easily observed to be between 14:00 and 15:23. Such a result can be used to make a simple evaluation on outages and failures of loggers.

The *total data coverage* is a measure for how many data points are missing in the analysed period. Under normal circumstances, a logger is set to log in a pre-defined logging interval. However, in certain periods, e.g. due to outages, data is not logged, even though it should have been. In such a case, the data coverage is calculated by finding the number of logged values vs. the number of values that should have been logged. The measure of the total data coverage can then be used to estimate how robust or valid the logger and/or analysed period is. In the above data set, the logger was set to log every hour and we observe 6 data points. With this information, we can see that only one date value is missing, i.e. where  $\Delta T$  is greater 1. In this case the missing date point is at timestamp 15:00. Thus, we now have 7 data points in total. Moreover, we see that one additional value (at 13:00) of the raw data set is equal NULL. This means that we have 2 out of in total 7 points with missing data. Vice versa, we know 5 out of 7 data points. Thus, the data coverage for this logger becomes:

$$\text{Data Coverage} = \frac{\text{Data points available in analysed period}}{\text{Total data points in analysed period}} = \frac{5}{7} = 71\%$$

The code flow diagram of this and the previous section is summarized in Figure 3.

**Please bear in mind** that the applied Data Coverage algorithm is only a simple estimate. Imagine a logger that is set to take logs every hour for a five hour period. Even though this is a rare case, what if the logger starts to take four successive logs with an individual distribution of 61 minutes? In such a case, the applied algorithm determines that we lack four data points. Consequently, only 5 out of in total 9 data points are available and it should be discussed if 5 out of 9 represent the data coverage adequate, when only four minutes of measurements are missing.

### 4.2.3 Interpolation of Data

**Input:**

- Missing date/data points (Section 4.2.2).

**Output:**

- Interpolation of new data points.

In the previous section 4.2.2, missing dates and data points were identified. To compute values for these missing data, we look at the following example:

Timestamp	[hh:mm]	12:00	13:00	13:34	14:00	15:00	16:00	16:23
$\Delta T$	[hh:mm]	-	1:00	0:34	0:26	1:00	1:00	0:23
Value		1.3	1	0.8	0.3	NULL	NULL	4.5

In this case two values are missing (at 15:00 and 16:00). The missing values will be found by linear interpolation, which between two known points is known as:

$$y = y_0 + (y_1 - y_0) \frac{x - x_0}{x_1 - x_0} \quad (1)$$

where  $x$  and  $y$  are the values between the interval  $x_0$  and  $x_1$ . With equation 1 we can simply compute the missing values of the example data set, as at 15:00:

$$y = 0.3 + (4.5 - 0.3) \frac{15 - 14}{16.38 - 14} = 2.06 \quad (2)$$

Moreover, the value for the timestep at 16:00 can be computed to be 3.83.

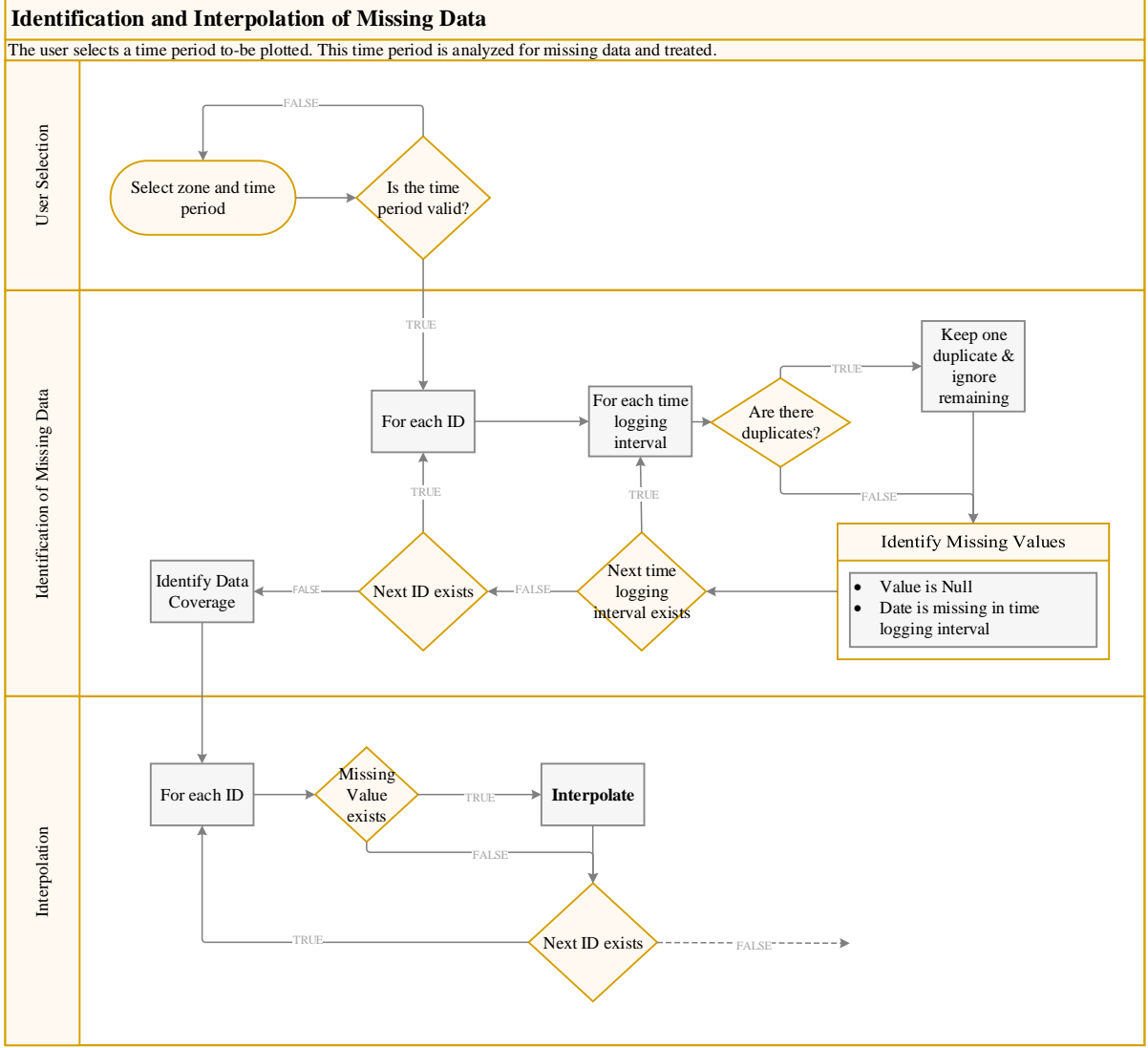


Figure 3: Code flow diagram describing the second step of the data analysis: finding and interpolating missing data.

So far the code interpolates every missing value. This is a simple solution, as far as the distance between known and unknown points is not too long. For example, if you log in an hourly interval and only one value is missing, the interpolated value might be a considerable good estimate. However, if values for a whole week are missing, the interpolated data would be a bad estimate of representing the missing data between the two known points.

The code flow diagram of this section is summarized in Figure 3.

#### 4.2.4 Compute zone flow

As a preliminary step of generating water consumption plots of a zone, the water consumption has to be calculated when a zone has multiple outflow loggers. This is done by a simple mass balance:

$$Q_{Zone} = \Sigma Q_{Log,in} - \Sigma Q_{Log,out} \quad (3)$$

If several loggers are taken into account, it is not only important to look at the individual logging intervals from one logger (as stated in section 4.1), but also at the logging intervals “between” the loggers. Consider



a zone, with one inflow and one outflow logger. Imagine, we know that the inflow logger logs in a 60 min interval whereas the outflow logger logs in a 15 min interval. All logged data represent the average flow value within the logging period, i.e. a 60 min log represents the average flow value within one hour. Thus, for the period of one hour, we have to subtract 4 different outflow values from the same inflow value to compute the water consumption in a zone. If logged values are plotted, they normally should be illustrated in a stair-step plot.

#### 4.2.5 Raw Data Plot

“Raw Data” is in this case not the original data set. It is a data set that might contain interpolated data, as described in the previous sections. To generate a raw data plot, the user needs to insert data as described in section 4.2.1 to 4.2.4 and press the button “Raw data plot” in the Excel Tool.

##### Input:

- Raw data set with interpolated data (section 4.2.1 to 4.2.4).

##### Output:

- Plot of the raw data set.

#### 4.2.6 Simple time series transformation

Before we are capable of generating the next two kind of plots, the obtained time series from the raw data set might need to be transformed into smaller or larger time intervals. For example, if data has been logged in an 1-minute interval, it could be beneficial to get rid of some noise by averaging the minute values into hourly values. In the next example data set, data should have been logged in hourly values:

Timestamp	[hh:mm]	12:00	13:00	13:34	14:00	15:00	16:00	16:23
$\Delta T$	[hh:mm]	-	1:00	0:34	0:26	1:00	1:00	0:23
Value		1.3	1	0.8	0.3	2.06	3.83	4.5

In this case, we want to transform the time series into the average value of every second hour. Thus, a new and transformed data set would have the following attributes:

Timestamp	[hh:mm]	12:00	14:00	16:00
$\Delta T$	[hh:mm]	-	2:00	2:00
Value		-	-	-

We can see that there are several values for each time interval. For the first time period between 12:00 and 14:00, three different values of the timestamps 12:00, 13:00 and 13:34 are found. To compute the average value for two hours, the following equation is applied:

$$Value_{new} = \frac{\Sigma(Value_{old} \cdot \Delta T_{old})}{\Delta T_{new}} \quad (4)$$

From equation 4, we find the new value for the period 12:00 to 14:00:

$$Value_{new} = \frac{60 \cdot 1.3 + 34 \cdot 1 + 26 \cdot 0.8}{120} = 1.11 \quad (5)$$

Likewise the value for the period 14:00 to 16:00 can be computed to be equal 1.18. In this example, the user requested two-hour values for the period between 12:00 and 16:23. As a consequence, the data point from 16:00 will only be representative for 23 minutes. In this interval the value was found to be 3.83. Thus the new and transformed data set would look like:

Timestamp	[hh:mm]	12:00	14:00	16:00
$\Delta T$	[hh:mm]	-	2:00	2:00
Value		1.11	1.18	3.83

#### 4.2.7 $\Delta T$ plot

$\Delta T$  (difference in time) plot describes a plot where data has been extended or averaged in a user-defined  $\Delta T$  interval. To be able to do this, the raw data set from a zone has to be transformed, as mentioned in the previous section 4.2.6. To generate a  $\Delta T$  data plot, the user needs to press the button “dT data plot” and insert information as demanded in the previous sections 4.2.1 to 4.2.4 and 4.2.6 in the Excel Tool. A screenshot illustrating the required user input for this type of plot can be found in Figure 22 in Appendix A.

**Input:**

- Specific  $\Delta T$ .

**Output:**

- Plot of  $\Delta T$  data set.

#### 4.2.8 Weekly average plot

The weekly average plot describes the mean water consumption in one week, based on hourly values over a user-chosen period. The user needs to press the button “Weekly average plot” with input from the previous sections 4.2.1 to 4.2.4 and 4.2.6. A screenshot illustrating the required user input for this type of plot can be found in Figure 23 in Appendix A.

**Input:**

- Time period.
- If a confidence interval should be plotted.
- If min & max should be plotted.

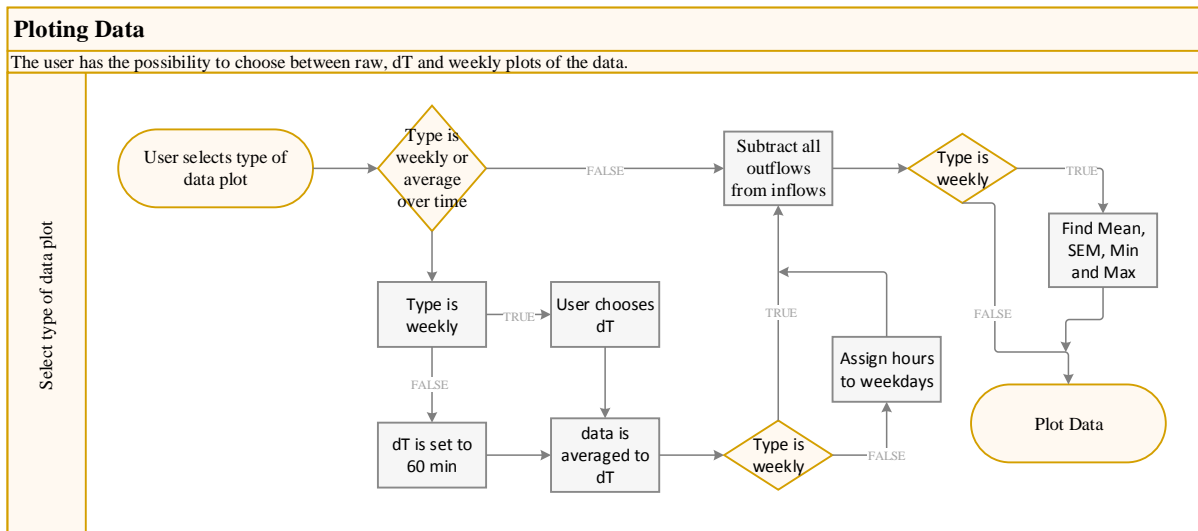
**Output:**

- Plot of mean water consumption over one week.

As shown in section 4.2.6, data is transformed. For this type of plot, all raw data is always transformed to a permanent value of hourly values for the analysed time period. Afterwards, it is possible to sort all hourly values to the individual weekdays. For example, if 12 weeks are analysed, 12 values for each hour of every weekday are available. Next the mean value is found for each hour. Moreover, it is possible to plot the 95% confidence interval (boundaries) and min & max values of a data set. The boundaries are found by:

$$95\% \text{ confidence interval} = \mu \pm 1.96 \cdot \frac{\sigma}{\sqrt{N}} \quad (6)$$

where  $\mu$  is the mean of every hour,  $N$  is the sample size (i.e. number of samples per weekday hour),  $\sigma$  the standard deviation and  $\frac{\sigma}{\sqrt{n}}$  the standard error of the mean. Moreover, the min and max value can be found and they represent the lowest and highest value found for each individual hour over one week. The code flow diagram in Figure 4 summarizes how the three plots are generated.



Figur 4: Code flow Diagram describing the generation of three different kind of plots in the BAWSD Excel Tool.

To generate results, the SWCDA Excel Tool will be applied to the WDN of Halsnæs Forsyning.

The WDN of Halsnæs Forsyning is sectioned into 27 DMAs. At each inlet of a DMA, a data logger has been installed, making it possible to monitor the water consumption of a DMA. In some cases, a DMA has several outflow loggers installed. If so, it is necessary to compute the water consumption by subtracting these outflow log data from the values obtained by an inflow data logger. The location of all data loggers and a flow chart overview on how the water moves between the DMAs can be found in Figure 24 in Appendix B.

[illegible]

12

## 5.2 Raw data plot

First of all, it is possible for the user to choose a specific period and a specific zone, and plot the raw data collected by the logging devices connected to the DMA for this period. In the following MD01 (Melby Vest) was chosen as an example zone. The zone consist mainly of summer cabins (78%). Figure 6 and Figure 7 show the water consumption during December 2013 and 2014 respectively for MD01.

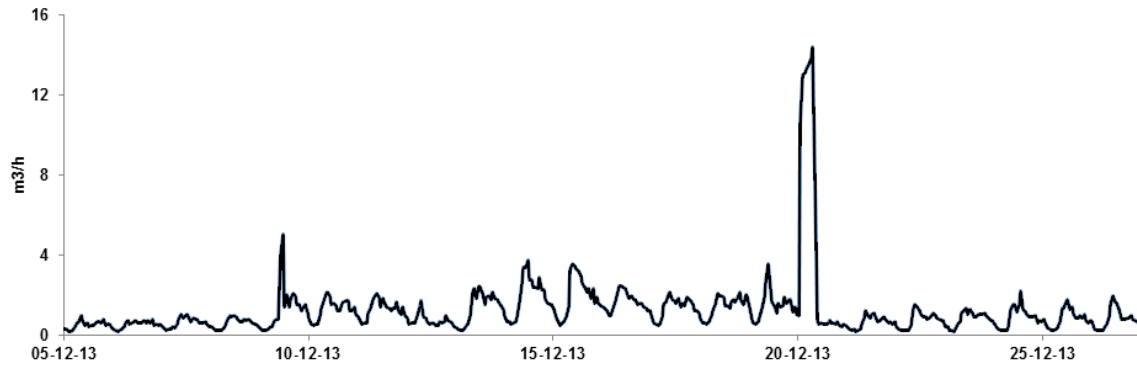


Figure 6: Raw water consumption in zone MD01 during December 2013. Based on logger MAG8000\_574802H053\_Q\_Frem.

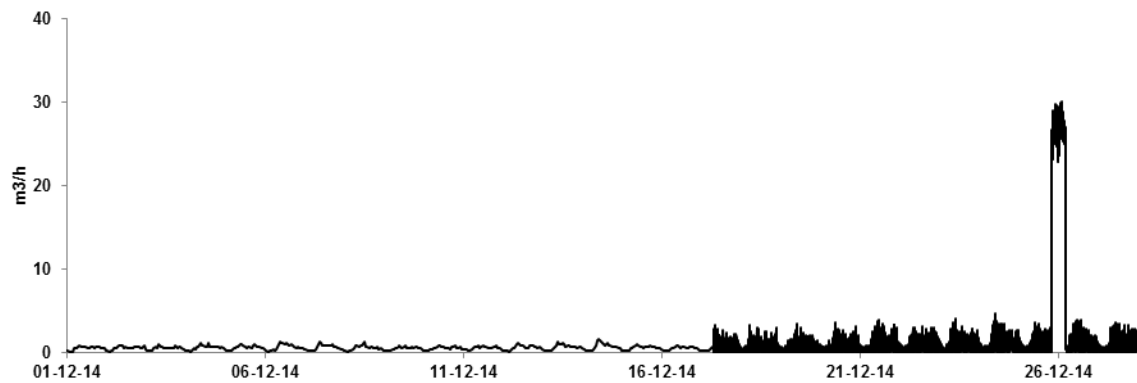


Figure 7: Raw water consumption in zone MD01 during December 2014. Based on logger MAG8000\_574802H053\_Q\_Frem.

In 2013 data collection was conducted in hourly intervals (Figure 6). This changed on the 17th of December 2014 into minute-by-minute logs (Figure 7). Moreover, the plot from 2013 and 2014 show that there has been a great water consumption during the 20th of December 2013 and 26th of December 2014.

## 5.3 $\Delta T$ data plot

Next, a  $\Delta T$  (logging interval) of 60 min was chosen, which averages all data values from the 17th of December 2014 into hourly values. The new result for December 2014 is shown Figure 8. Looking at Figure 7, it could be interpreted that the mean water consumption during the end of December was higher than in the start of the December. Figure 8 disproves this, since the higher resolution of logging values only makes it “look” like there is a higher water consumption.

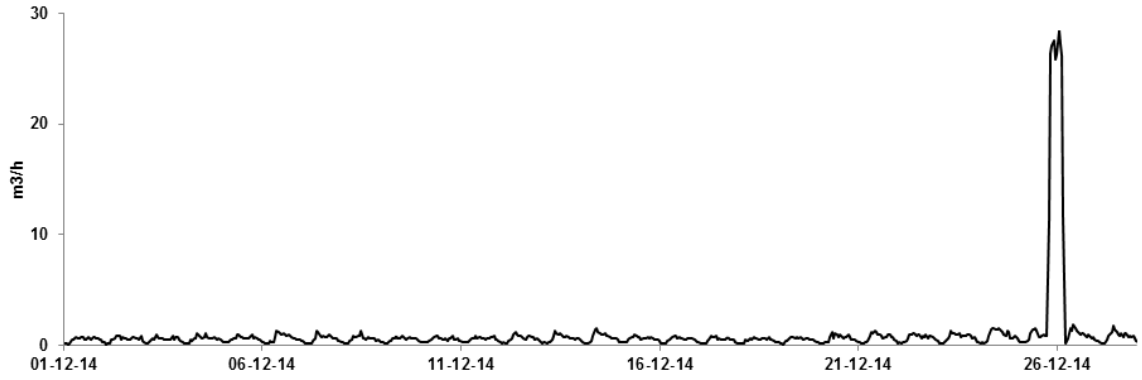


Figure 8:  $\Delta T$  60 min water consumption, zone MD01 during December 2014. Based on logger MAG8000\_574802H053\_Q\_Frem.

As done with Figure 8 in December 2014, it is also possible to show the hourly water consumption in the whole lifetime of the logger (starting from the 12/2013 to 08/2015) (Figure 9).

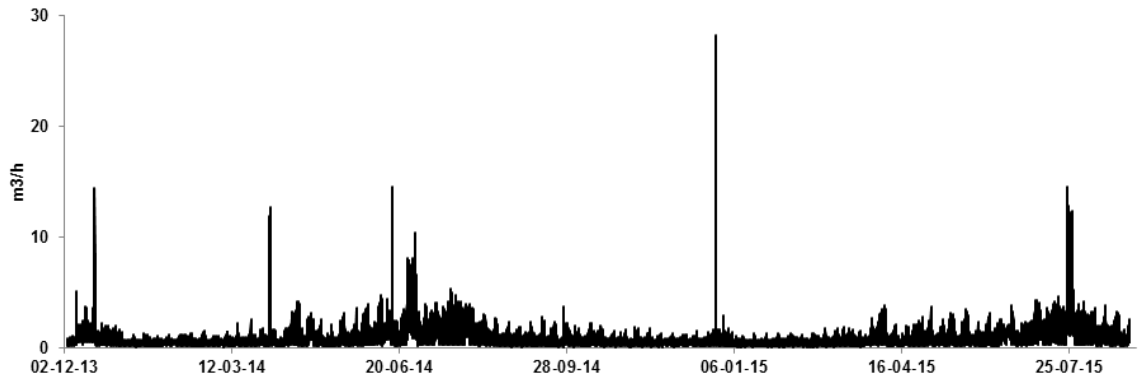


Figure 9:  $\Delta T$  60 min water consumption, zone MD01, December 2013 to August 2015. Based on logger MAG8000\_574802H053\_Q\_Frem.

Even though only one year of full data is available, Figure 9 shows that a seasonal variation with higher water consumption during summer exists (as expected from a zone that mainly consists of summer cabins). The same seasonal trend is also visible when a  $\Delta T$  of 1440 min (i.e. daily values) for the same period is chosen. The result for this  $\Delta T$  is shown in Figure 10.

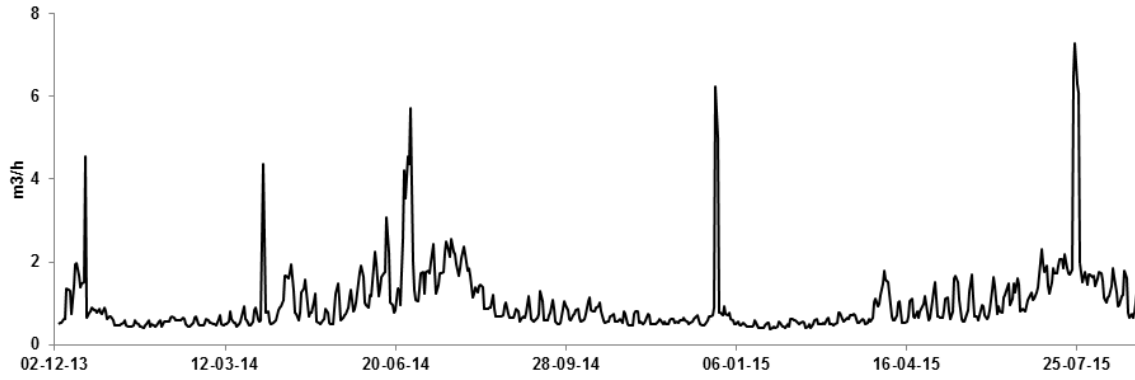


Figure 10:  $\Delta T$  1440 min water consumption, zone MD01, December 2013 to August 2015. Based on logger MAG8000\_574802H053\_Q\_Frem.

There are several explanations for the observed peaks in the above figures, but these are outside of the scope of this paper. Possible explanations include tapping of water from fire hydrants, pipe bursts, filling of a swimming pools, outages of data loggers, high public holiday consumption, etc.

#### 5.4 Weekly average plot

The last objective of this tool was to generate average weekly plots. As with the Raw Data Plot and  $\Delta T$  Data Plot, the user can choose a specific period to be analysed. In Figure 10, a lower consumption during the period December 2013 to March 2014. However, from May 2014 to August 2014, it looks like that the water consumption has increased. To prove this, a weekly plot can be generated for both periods. The result for the first period between December 2013 and March 2014 is shown in Figure 11.

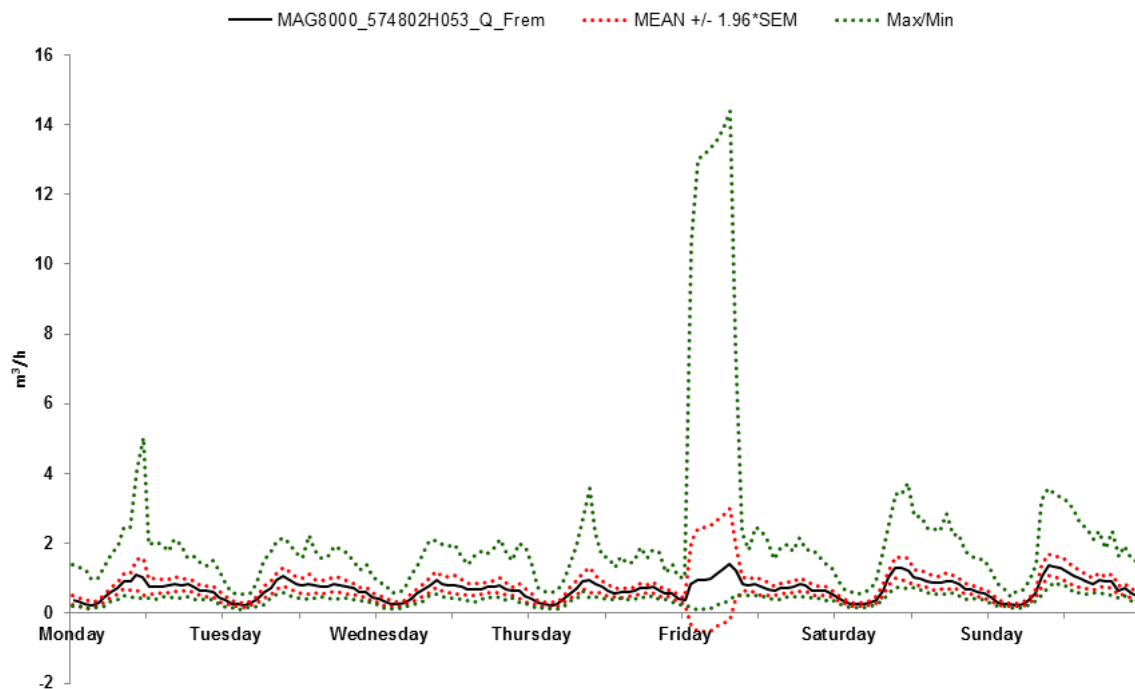


Figure 11: Average weekly water consumption in the district metered area MD01, December 2013 to March 2014.

The water consumption swings around  $0.5 \text{ m}^3$  per hour [black line]. The green line shows the observed min and max values for all hours. The red lines show the 95% confidence intervals of the mean. From the “max” line it can be observed, that for all days, there has been a period with a substantially higher water consumption, peaking on a Friday. As mentioned earlier, it is outside the scope of this paper to analyse the possible reasons for this. The next period between May 2014 and August 2014 clearly shows that the average water consumption has increased during summer (see Figure 12). Whereas the average consumption in colder months peaked around  $1.5 \text{ m}^3$  per hour (Figure 11), it is above  $2 \text{ m}^3$  per hour during summer months (Figure 12).

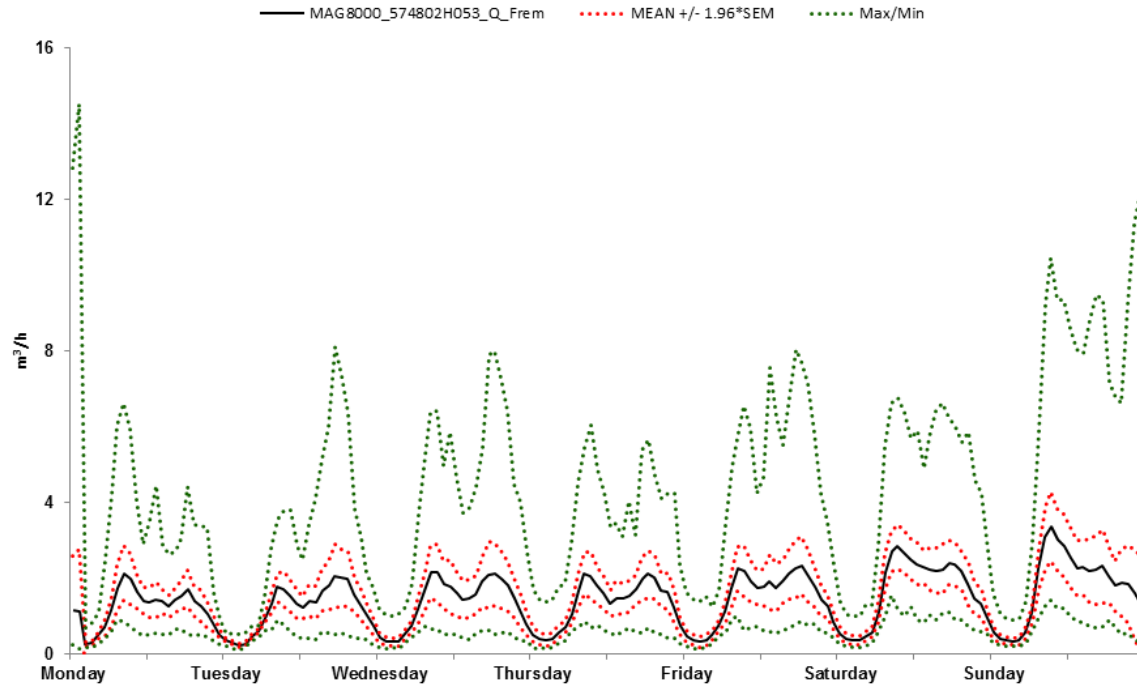


Figure 12: Average weekly water consumption in the district metered area MD01, May 2014 to August 2014.

Figure 12 shows that the water consumption during the weekend is greater than during weekdays. The next figure shows the average water consumption based on the whole logging period of the data logger installed in zone MD01 (Figure 13). It could be argued that the figure represents the daily water consumption for a “normal” week in zone MD01 better than the previous two figures because a longer period with more data has been taken into account. The maximum flow for each hour has been removed. Water consumption is highest during Sundays and lowest from Monday to Wednesday. Additionally, most variances in the mean water consumption are observed on a Thursday and on a Friday.



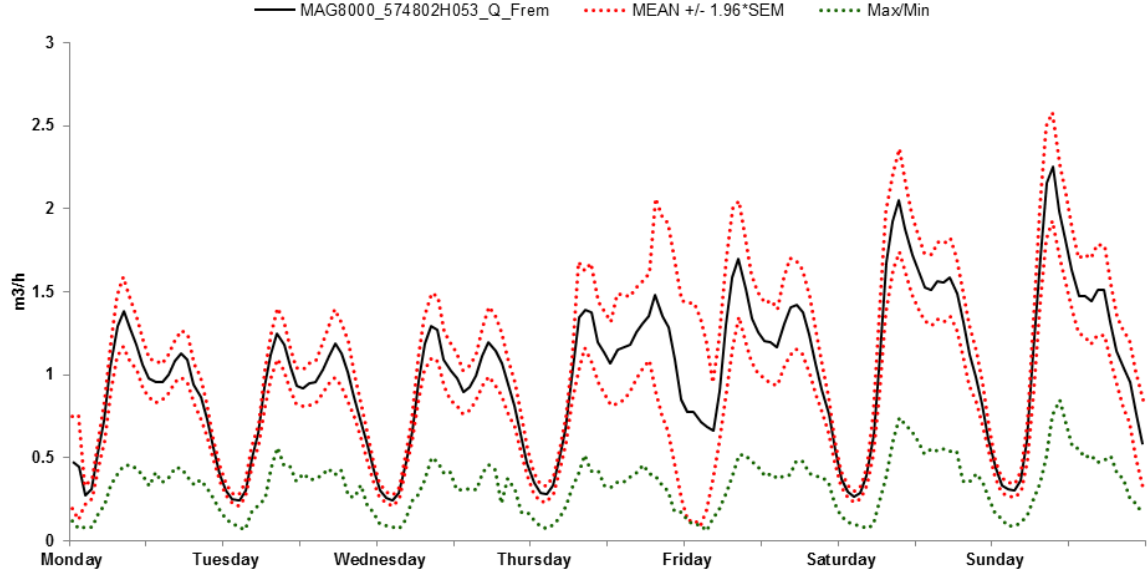


Figure 13: Average weekly water consumption, MD01, December 2013 to August 2015. Max removed.

## 5.5 Additional plots

Finally, additional plots for other sections could be generated. Figure 14 shows the weekly water consumption in zone MD03, mainly consisting of domestic households.

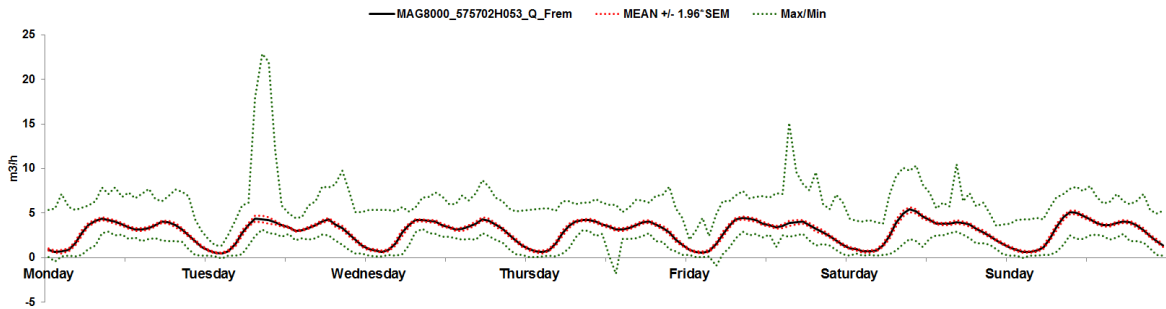
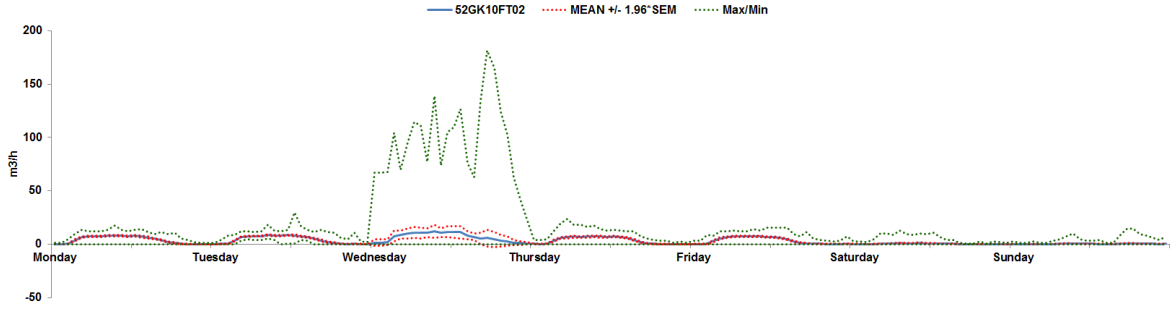


Figure 14: Average weekly water consumption, MD03, January 2014 to November 2015.

Here, it seems that the water consumption is only slightly higher during weekends, than it was the case with the summer cabin zone MD01 (e.g. Figure 13). The next Figure 15 shows the weekly water consumption of zone MD16. This zone consists 100% of domestic households.



Figur 15: Average weekly water consumption, MD16, January to November 2015.

The max value from Wednesday to Thursday in the figure shows that there has been an extraordinary demand of water consumption. Removing this “max” curve, leads to Figure 16:

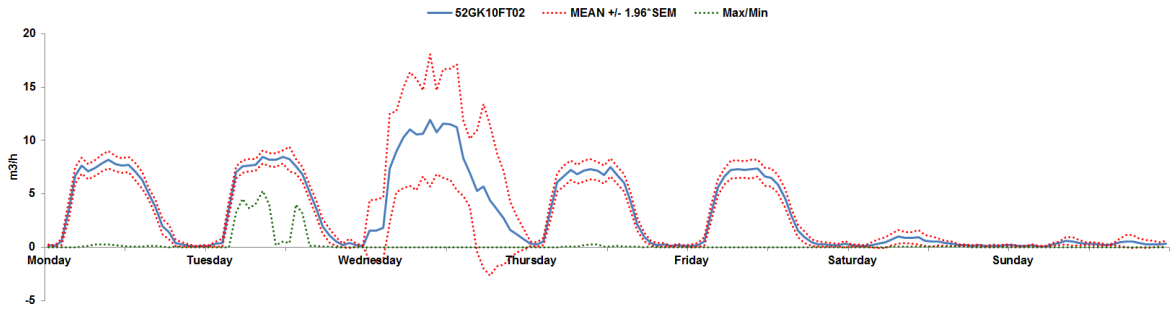


Figure 16: Average weekly water consumption, MD16, January to November 2015. Max removed

Figure 16 shows that additional knowledge and analysis of the obtained logging data is required to be able to explain such an “odd” water consumption curve. Possible reasons for the odd curve might be consumers with unknown consumption patterns, frequent and reoccurring failures/outages of the logger, etc.

## 6 Conclusion

The Baseline Analysis of Water Supply Data Excel Tool is capable of conducting a simple analysis on raw monitoring data from water distribution networks. The tool can be used to show the raw data set, i.e. the water consumption within a district metered area. Moreover, it is possible to group the data into a user chosen interval and plot these as a result. Finally, the tool is capable of generating plots that represent the average weekly consumption over a user-chosen period. Also, some of the tool’s limitations were discovered. For example, the generation and understanding of extraordinary water consumption patterns require deeper knowledge of the analysed area and a more detailed sorting/filtering of the obtained data than the Excel Tool can provide.

# APPENDIX

## A Manual Input Guide for BAWSD Excel Tool

In the following, the manual input required to generate data plots will be listed in the form of screenshots (Figure 17 to 23).

SETTING	INPUT
UserID	
Password	
Provider	MSDASQL.1
Driver	SQL Server
Server	PC119862\SQLEXPRESS
Database	Halsnaes_Forsyning_RawData

In this Excel sheet, it is possible to generate various plots of "raw logging data" from this is Excel Sheet is intended to work:

Figure 17: Screenshot of required User-Input on sheet 1 “Initialization”. In the column **Input**, the user has to insert known properties of the to be analysed database and water distribution network.

2	A	B	C	D	E	F	G
1	CHECK ALL <input type="checkbox"/>						
2							
3	ID	Check	Q in	Qout1	Qout2	Qout3	Qout4
4	MD01	<input checked="" type="checkbox"/>	MAG8000_574802H053_Q_Frem				
5	MD02	<input type="checkbox"/>	MAG8000_575602H053_Q_Frem				
6	MD03	<input type="checkbox"/>	MAG8000_575702H053_Q_Frem	MAG8000_575602H053_Q_Frem	MAG8000_574802H053_Q_Frem		
7	MD04	<input type="checkbox"/>	54GK01FT02	MAG8000_575702H053_Q_Frem			
8	MD05	<input type="checkbox"/>	86GK01FT02				
9	MD06	<input type="checkbox"/>	MAG8000_574502H053_Q_Frem				
10	MD07	<input type="checkbox"/>	MAG8000_575402H053_Q_Frem				
11	MD08	<input type="checkbox"/>	MAG8000_574902H053_Q_Frem				
12	MD09	<input type="checkbox"/>	MAG8000_575102H053_Q_Frem				
13	MD10	<input type="checkbox"/>	MAG8000_506202H142_Q_Frem				
14	MD11	<input type="checkbox"/>	MAG8000_575802H053_Q_Frem				
15	MD12	<input type="checkbox"/>	MAG8000_574602H053_Q_Frem				
16	MD13	<input type="checkbox"/>	52GK01FT02	MAG8000_583802H025_Q_frem	MAG8000_582702H123_Q_frem	MAG8000_574302H053_Q_frem	72GK02FT02
17	MD14	<input type="checkbox"/>	72GK02FT02				
18	MD15	<input type="checkbox"/>	MAG8000_574802H053_Q_Frem				

Figure 18: Screenshot of required User-Input on sheet 2 “Generate\_data”. In the column **C** (Q\_in), the logger name that measures the inflow for a zone/DMA is inserted. In **D** (Q\_out1) to **G** (Q\_out4) the name of outflow loggers can be listed.

	A	B	C	D	E	F	G	H
1		Intervals	Interval [ Min]	Fra	Til	Interval [Min]	Fra	Til
2	MAG8000_574802H053_Q_Frem	2	60	2013-12-03 11:25:00	2014-12-17 06:44:00	1	2014-12-17 06:44:00	2015-08-31 06:46:00
3	MAG8000_575602H053_Q_Frem	3	15	2013-12-03 20:06:00	2014-03-05 19:51:00	60	2014-03-05 19:51:00	2014-12-17 06:19:00
4	MAG8000_575702H053_Q_Frem	4	15	2014-01-02 10:06:00	2014-01-02 20:51:00	60	2014-01-02 20:51:00	2014-10-06 07:02:00
5	54GK01FT02	1	1	2015-01-01 00:00:00	2015-08-31 23:59:00			
6	86GK01FT02	1	1	2015-01-01 00:00:00	2015-08-31 23:59:00			
7	MAG8000_574502H053_Q_Frem	3	60	2013-12-03 14:01:00	2014-09-30 09:50:00	15	2014-09-30 09:50:00	2014-12-16 18:46:00
8	MAG8000_575402H053_Q_Frem	7	60	2013-03-21 11:42:00	2013-11-11 10:40:00	15	2013-11-11 10:40:00	2013-11-12 11:29:00
9	MAG8000_574902H053_Q_Frem	2	60	2013-12-03 11:09:00	2014-12-17 08:44:00	1	2014-12-17 08:44:00	2015-08-31 08:46:00
10	MAG8000_575102H053_Q_Frem	2	60	2013-11-23 13:17:00	2014-12-16 11:08:00	1	2014-12-16 11:08:00	2015-08-31 11:10:00
11	MAG8000_506202H142_Q_Frem	2	60	2013-12-03 21:48:00	2014-12-16 11:20:00	1	2014-12-16 11:20:00	2015-06-11 11:22:00
12	MAG8000_575802H053_Q_Frem	2	60	2014-10-06 13:18:00	2014-12-16 12:00:00	1	2014-12-16 12:00:00	2015-08-12 07:17:00
13	MAG8000_574602H053_Q_Frem	2	60	2013-12-03 12:55:00	2014-12-16 10:44:00	3	2014-12-16 10:44:00	2015-07-02 10:46:00

Figure 19: Screenshot of required User-Input on sheet 3 “SCADA\_Time\_Check”. In the column **A**, the logger name is required. In column **B** (Intervals) the number of different log intervals have to be entered for a logger. For each set of log interval, the log interval has to be inserted as well as the start and end point of logging in this interval.

A	B	H	I	J	K	L	M
CHECK ALL	<input type="checkbox"/>	USER INPUT			VALID TIME PERIOD		
ID	Check	T1	T2	Chosen dT [min]	Tmin (Log Start)	Tmax (Log End)	Biggest Gap Raw Data Time Origin
MD01	<input checked="" type="checkbox"/>	2013-12-04 00:00:00	2015-08-30 00:00:00	10080	03-12-2013	31-08-2015	MAG8000_574802H053_Q_Frem
MD02	<input type="checkbox"/>	2015-04-05 00:00:00	2015-06-06 00:00:00	60	01-01-2000	29-07-2015	MAG8000_575602H053_Q_Frem
MD03	<input type="checkbox"/>	2015-01-05 00:00:00	2015-07-27 00:00:00	30			

Figure 20: Screenshot of required User-Input on sheet 2 “Generate\_data”. In the column **Check** (B), the user can check (or uncheck) if a zone has to be analysed. The valid time period for this zone will be listed in column **K** (Tmin (Log Start)) and **L** (Tmax (Log End)). To generate the results it is necessary to press the button “Initialization”.

A		B	H		I	J	K	L	M		
CHECK ALL		<input type="checkbox"/>	USER INPUT				VALID TIME PERIOD				
ID	Check	T1	T2	Chosen dT [min]	Tmin (Log Start)	Tmax (Log End)	Biggest Gap Raw Data Time Origin				
MD01	<input checked="" type="checkbox"/>	2013-12-04 00:00:00	2015-08-30 00:00:00	10080	03-12-2013	31-08-2015	MAG8000_574802H053_Q_Frem				
MD02	<input type="checkbox"/>	2015-04-05 00:00:00	2015-06-06 00:00:00	60	01-01-2000	29-07-2015	MAG8000_575602H053_Q_Frem				
N			O		P	Q		R		S	
COMPUTER OUTPUT BASED ON T1 & T2											
Biggest Gap - Raw Data[minutes]		From		To	Lowest Data Coverage		Total Data Coverage [%]		Min Raw dT [min]		
79		07-10-2014 06:25		07-10-2014 07:44	MAG8000_574802H053_Q_Frem		99.9992		1		
1		05-04-2015 00:01		05-04-2015 00:02	MAG8000_575602H053_Q_Frem		100		1		

Figure 21: Screenshot of required User-Input on sheet 2 “Generate\_data”. In column **H** (T1) and **I** (T2), the user inserts a valid time period to be analysed. The logger name of the logger with the biggest gap in the analysed time period will show up in column **M** (Biggest Gap Raw Data Time Origin). Moreover, the length of the gap, and start and end time of this gap will be listed in column **N** (Biggest Gap Raw Data) to **P** (To). Finally, the Data Coverage will be shown in column **R**. The lowest log interval is shown in column **S** (Min Raw dT [min]).

I		J	
USER INPUT			V
T2	Chosen dT [min]	T1	
2015-08-30 00:00:00	10080		
2015-06-06 00:00:00	60		
2015-07-27 00:00:00	30		
2015-08-24 00:00:00	60		

Figur 22: Screenshot of required User-Input on sheet 2 “Generate\_data”. In column **J** (chosen dT [min]), the user can enter a specific dT value. By clicking on “dT data plot” afterwards, a dT plot is generated.

<input checked="" type="checkbox"/> MIN & MAX	<b>Weekly plot</b>
<input checked="" type="checkbox"/> BOUNDARIES	

Figur 23: Screenshot of required User-Input on sheet 2 “Generate\_data”. Here the user can check boundaries and min/max if 95% confidence intervals and the minimum and maximum values should be plotted, respectively. By clicking on “Weekly plot” afterwards, a weekly average water consumption plot is generated.



Tabel 1: Share of the total water consumption divided into domestic households, summer cabins and industries within district metered areas in 2014.

<b>Zone</b>	<b>Type</b>	<b>Households[%] Consump.</b>	<b>Summer cabins [%] Consump.</b>	<b>Industry[%] Consumption</b>
MD01	Evetofte zone	22.53	77.45	0.03
MD02	Evetofte zone	13.60	86.4	0
MD03	Evetofte zone	84.51	0.33	15.17
MD04	Evetofte zone	13.64	86.2	0.16
MD05	Kappelhøj VV zone	11.99	86.68	1.33
MD06	Kappelhøj VV zone	98.13	0	1.87
MD07	Kappelhøj VV lavzone	96.23	3.55	0.22
MD08	Kappelhøj VV lavzone	81.51	0	18.49
MD09	Kappelhøj VV lavzone	97.7	0	2.3
MD10	Kappelhøj VV lavzone	78.21	0	21.79
MD11	Kappelhøj VV zone	85.86	0	14.14
MD12	Kappelhøj VV zone	85.99	14.01	0
MD13	Kappelhøj VV zone	100	0	0
MD14	TF1 højzone	98.83	0	1.17
MD15	Kappelhøj VV zone	82.21	9.84	7.95
MD16	Ølsted zone	100	0	0
MD17	Ølsted zone	99.4	0	0.6
MD18	Ølsted zone	100	0	0
MD19	Ølsted zone	98.52	0	1.48
MD20	Ølsted zone	99.39	0	0.61
MD21	Ølsted zone	73.25	0	26.75
MD22	Ølsted zone	77.26	0	22.74
MD23	Ølsted zone	73.36	0	26.64
MD24	Kappelhøj VV lavzone	31.51	0	68.49
MD25	Ølsted zone	97.6	0	2.40
MD26	Kappelhøj VV lavzone	96.22	0	3.78
MD27	Kappelhøj VV lavzone	78.24	0	21.76