KTH Architecture and
the Built Environment

Identifying the Influential Factors of the Temporal Variation of Water Consumption

A Case Study using Multiple Linear Regression Analysis

June 2016

Hanna Brandner 2016
Environmental Engineering and Sustainable Infrastructure Program (EESI)
School of Architecture and the Built Environment
Division of Hydraulic Engineering
Royal Institute of Technology (KTH)
SE-100 44 STOCKHOLM, Sweden

Acknowledgements

Firstly I would like to thank my thesis advisors, both at the Royal Institute of Technology (KTH) Anders Wörman and also Hans Hammarlund at Tyréns, a consultancy firm based in Slussen, Stockholm. Both supervisors where available when I needed any help or guidance regarding the thesis work and were open to discussion when it was required. Tyréns kindly allowed me to sit at their offices during my thesis work, which was highly appreciated, and for this I would like to thank Tyréns and the water and wastewater department as a whole. The data provided for this thesis work was organized and collected by Tyréns from the municipalities that were investigated. And for this, I would also like to thank those who provided the data, which includes Göteborg Energi, the water department of the Gothenburg municipality, Kalix municipality Alvesta municipality and also SMHI for its open source of climate data, making this thesis work possible to accomplish. I would also like to thank Svenskt Vatten, as they worked in collaboration with Tyréns on this water development project. I was able to sit in on the meetings organized at Tyréns with Svenskt Vatten, which was a great help with my thesis work and guided me to what questions and issues needed to be addressed regarding the water development project in water consumption.

Contents

Table of Figures

Table of Tables

Definition of Abbreviations

ANN	Artificial neural networks
ARIMA	Auto regressive integrated moving average
DSE	Double sine equation
MLR	Multiple linear regression
OLS	Ordinary least squares
R^2	The coefficient of determination, R squared
R^2_{adj}	Adjusted R squared
SMHI	Swedish Meteorological and Hydrological Institute
SSE	Sum of squared errors
SV	Svenskt Vatten
SWWA	Swedish Water and Wastewater Association
UBW	Urban water demand
VIF	Variance inflation factors

Abstract

This thesis is a part of the water development project conducted by Svenskt Vatten, which is the Swedish Water and Wastewater Association (SWWA) as well as Tyréns, a consultancy company with offices based in Stockholm, Sweden. Prior to this thesis work, a quality assessment was conducted for some of the locations provided by municipalities in Sweden. This thesis builds upon the revised water consumption data, and also continues to work with validating and modifying the water measurement data in order to proceed with the next step of the water development project, which is to identify any trends in the temporal variation of water consumption. The main objective of this thesis work is to investigate the influence of climatic, time-related and categorical factors on water consumption data collected for different regions in Sweden, and includes a number of different sectors such as residential, industrial and agricultural water user sectors. For the analysis of data, spectral analysis and sinusoidal modelling will be applied in order to find the periodicity of the data, and then simulate the fitted sinusoidal equation to the observed water consumption data for the hourly interval period. Multiple linear regression analysis is then used to assess what independent variables such as climate, time-related and categorical variables can explain the variation in water consumption over hourly and daily periods of time.

Spectral analysis identifies high peaks in the spectral density of the data at 12 and 24 hour cycles, for the hourly water consumption data. For the total daily consumption of water, there is a peak at 7 days, which clarifies that there is a weekly pattern occurring throughout the year. The results from the simple linear regression analysis, where the linear relationship between temperature and water consumption was determined, reveals that the water consumption tends to increase within an increasing temperature, where in Lönashult, Alvesta municipality the water demand increased by 5.5% with every 2 ºC rise in temperature, at a threshold of 12 ºC. For Kalix municipality the three areas selected have around 1-2 % increase in water demand with every 2 ºC rise in temperature for the period of May to December. In Gothenburg, areas that were mixed villa areas or areas with summer homes there was a rise of around 2-12 % in water demand, however areas that are situated in the inner city Gothenburg, or that have majority student housing, the water consumption tends to decrease by 2-7% in water demand with every 2 ºC rise in temperature, with a threshold of 12 ºC.

In multiple regression analysis, the hourly water consumption results in adjusted R^2 values were in the range from 0.58 to 0.87 (58-87%) for the best model approach and therefore has a significant relationship between water consumption and the explanatory variables chosen for this study. For the daily water consumption, the adjusted R^2 values were in the range of 0.22-0.83 (22-83%). The adjusted R^2 values are lower for certain areas and can be explained by a number of factors, such as the different variables used for the daily water consumption analysis, as variables that explain more the periodicity of the data such as the sinusoidal fitted variable and hourly or night/day changes in consumption are not included. As well as this, not all independent variables such as the climate variables were available or complete for particular time periods, and also errors in the data can lead to a significantly lower R^2 value.

1 Introduction

1.1 Background into water consumption in Sweden

Half of Sweden's water supply comes from surface waters, which includes lakes as well as surface water runoff from streams and rivers. The other half is supplied by artificial and natural groundwater supplies (Svenskt Vatten, 2016). When the regular water supply is not available for the short or long term, then the water can be supplied from reservoir storages (Svenskt Vatten, 2016). Sweden is a country that has plentiful water, where just 0.5 % of the available water resource in Sweden is extracted for municipality water use (Sydvatten, 2011).

The average water consumption per person is 160 L/day in Sweden, according to Sydvatten (2015), where within a household per day this breaks down to 60 L for personal hygiene, 30 L for toilet flushing, 20 L for washing clothes, 30 L for dishwashing, 10 L for cooking and drinking and 10 L for other purposes. In the summer or drier season the external water use is expected in increase, in areas that have larger green spaces to maintain. Domestic demands is however not the only demand of water. There are also public services such as schools or hospitals, office buildings, restaurant and general stores to consider. As well as this, industries and also agricultural areas will significantly change the overall consumption of water. According to Statistic Sweden (2012), there are three main sectors listed for water uses, which include household, manufacturing industry and agricultural purposes as well as other uses. The manufacturing industry uses the most amount of water at a total of 64% of freshwater volume, then after this it's the domestic sector with 21% use, 11 % for other users and finally agriculture with just 4% of the total freshwater volume, which is shown in Figure 1 (Statistics Sweden, 2010). There is also some percentage that is lost due to leakage from the water distribution system, where according to the European Environment Agency (2003) this was roughly 18% of the overall water supply in urban networks in 2000.

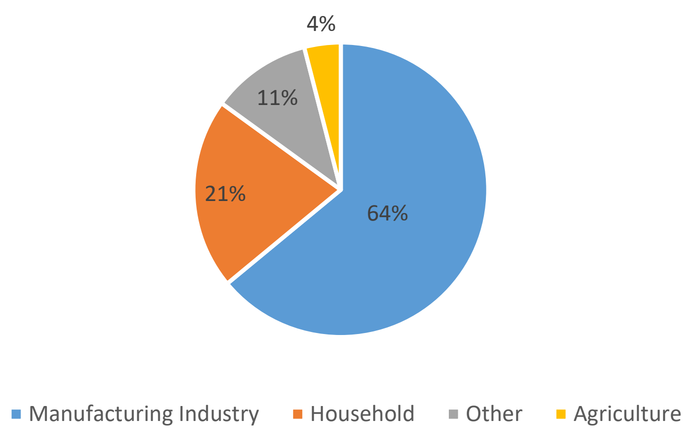


Figure 1: Water use by sector

There are many factors in which the water consumption pattern may be influenced. Throughout an annual period of water consumption, there are cyclic patterns that may affect water consumption such as the weekly, monthly or daily cycle. These periodical cycles can be key to understanding at what times minimum or maximum consumption of water may occur. As well as this, the fluctuations of people during holiday periods can have an impact on water consumption. There are also climatic factors such as temperature and precipitation that can majority influence the consumption of water, however this is more noticeable in more arid countries where external water use may be a higher priority (Gober & Balling Jr, 2007). As well as meteorological influences on the water consumption pattern, anthropogenic climate change can also have an effect on the consumption of water, however this is more key to arid regions that may rely more heavily on rainfall and already have very high temperatures.

Climate change resilience in water consumption is important to understand, as if the water consumption is observed to correlate with climate variables such as temperature and precipitation, then it's important to understand how this will change given future climate scenarios. According to climate scenarios conducted by SMHI the future of Sweden's climate will see a rise in temperatures and as well as the amount of precipitation and higher rainfall intensities. In particular the coldest winter days especially in the northern regions will become warmer and an increase in precipitation is expected mostly in northern Sweden and during the winter season. The average annual precipitation is expected to be 20-60% more than for the period 1961-1990, depending on the climate scenario use. In southern Sweden the precipitation change for the future scenario is expected to stay relatively the same, as the results from models show both an increase and reduction in precipitation, indicating a change near to zero. The temperature is expected to increase by 2-6 °C by the end of the century compared to the period 1961-1990, which depends on the climate scenario used (SMHI, 2015).

1.2 Purpose of the water development project

This thesis is written as a part of the water development project implemented by Stockholm Water and Wastewater Association (SWWA), which has been taken into collaboration with Tyréns, a consultancy firm based in Stockholm, Sweden. The water development project has its key focus on the findings from water consumption data, where it will look into different sectors and water user types such as from residential homes, industries, office buildings, agricultural use and commercial/public water use. The purpose of the water development project is to investigate the current guidelines regarding water consumption which are provided by SWWA publications, namely P83 and P90. Stated in these publications are the minimum and maximum daily and hourly factors, which are used for the design of the water and wastewater supply system. These factors and how they have been identified are to be revised as part of the water development project, as they could lead to misleading quantities of how much water is really required for municipality water use.

The water development project is vital for the design of both pipes for drinking water and the resultant wastewater piping systems. The piping network should be efficiently designed so that it meets the correct demand and supply of water, and it will therefore be necessary to know the actual consumption of water throughout various water demand sectors such as residential, commercial, agricultural and industrial. The findings may also differ depending on the region of interest within the study, as well as other factors such as population size. Furthermore, this study aims to produce a more cost effective approach to pipe design, where improvements to the accuracy of the design values for the total water consumption over given temporal periods could in turn reduce the size of piping.

Initially there will be quality assessment of the data is undertaken, as errors and uncertainties can lead to false or inaccurate measurements of the actual human water consumption, and could be due to leaks in the system, either from the external piping network or within people's homes. This stage of the project both validates the data so that it can be modified and adjusted for further investigation.

A question of what quantifies the measurement of water consumption will also be discussed, where possible defining factors include demand per household, total water consumption per capita, number of employees per office block, or either in terms of location as well as land use in the specific region of interest. It is also important to identify where the peak flows occur on an hourly, daily and annual basis so that when the peak demand is determined for the design of the water distribution system, all expected demands will be met. As well as this, the minimum flows of water consumption will be investigated.

1.4 Aims & objectives

The aim of this thesis work is to investigate the patterns in the temporal variation of water consumption in relation to: periodicity (hourly, daily and monthly), climate variables, and other factors such as population fluctuation, size and location. The water consumption data is provided in hourly units of time, where the total daily consumption of water will also be investigated. The study should increase the accuracy of the water demand and supply water distribution system. Water consumption will be treated as the dependent variable in which independent variables will be tested to see how well they can explain the variance in water consumption.

This will be achieved by proceeding through the following steps:

1) Identify independent variables expected to explain the temporal variation in water consumption
2) Determine the periodicity of the data and fit this to the data by using spectral analysis and sinusoidal modelling
3) Apply multiple linear regression analysis to determine the relationship between the independent variables and the dependent variable (water consumption)

1.5 Statement of the research question

"To identify any factors such as climatic, seasonal or periodic factors that have an influence on the temporal variation of water consumption for different sectors including industrial, commercial, residential and agricultural sectors. "

1.6 Delimitations, limitations and assumptions

When analysis the water measurement data provided, there are still uncertainties that arise in the data such as where there could be leaks or outlying values, which lead to pattern that does not fully represent the actual water demand of say residential homes. From the previous study that looked into the quality of the water measurement data, these errors and uncertainties have been flagged to a certain degree, however it is obvious that some errors still remain in the water measurement data. It is assumed however that the data then provided (containing flagged values) represents the real water consumption, where the values that are marked with a flag are replaced with the average water consumption for that particular hour. This also is limited, as the actual water consumption for the missing values in unknown, however the average hourly value will at least a more accurate pattern rather than an outlier or shift in the data series. As well as this, there are limitations to the lengths of the data provided, where only hourly data for specific years or monthly periods may be accessible, and there are also some data gaps, in particular the Kalix data obtained.

The climate data obtained from SMHI also is also limited, as for the periods that are required, not all the stations have active data, and also the data may contained calculated measurements instead of actual measurement data. There are also large gaps in the datasets for the hourly measurements, which provides difficulties when analyzing the hourly annual water consumption pattern.

Another limit within this study are the accessibility to socio-economic and demographic factors, such as the size of the family per household, income and age. To obtain information about for example the size of the residential home is perhaps easier, however for example income and age of the residents provides more difficult. To analysis further how water consumption should be measurement, such as per household, only rough estimates can be made for areas which do not have an exact number of houses/people within the identified water demand location. To know the exact numbers of people during the course of the year could also provide useful, as the number of persons that are in one area will fluctuate during the year, due to vacation periods for example.

2 Literature Review

The challenges facing todays demand and supply of water is a topic of major concern, as globally there is significant impacts from climate change where in Sweden climate change is expected to intensify rain events leading to increased flooding and contamination of water supplies according to Svenskt Vatten (2014), where in addition to this there is increased pressure on urban environments as populations continue to grow. Although Sweden is a country that has a sufficient availability of water, there are problems concerning the quality of this water as pollution and contaminated water sources are not eligible to meet water quality requirements (SWWA, 2014). It is therefore important to maintain a water distribution system that complies with the expected demand of water, as well as meeting the requirements for peak demand periods, where the pipes need to be design to withhold both an adequate capacity of water volume by allowing for a sufficient size of pipe diameter as well as a sufficient design demand to meet the actual water consumption of the water users (World Health Organization, 2004).

According to the WHO (2004) human consumption of water only accounts for 2% of the actual water supply, where the remaining volume of piped water is used for sanitation, washing, firefighting and irrigation. This is an interesting point to consider when determining what design methods or approaches should be taken, to assume what the variation in total water consumption is for different types of residential homes or common uses of land are for the region of interest.

There are many factors that have been taken into account when investigation what influencing factors there are on water demand patterns where according to research conducted at the Western Sydney University in 2015 there are a range of factors including demographic, socio-economic and climate factors that can influence observed patterns in water demand (Haque, et al., 2015). There are studies that rely purely on socio-economic factors, where Chen *et al* (2012) uses a cross-sectional survey in the Shanghai region which is based on gender, age education, housing conditions, person income, risk perception and personal preferences to assess potential influential factors on residential water consumption. Other studies determine the influential factors by using climate factors such as precipitation and temperature daily variations, where according to Chang (2014) temperature has a positive correlation to water consumption and precipitation has a less significant negative correlation, which was conducted in Portland State, U.S.A. These studies are based globally, and due to huge variations in climate it is important to conduct individual studies specified for the area of interest.

When working with water demand modelling there are some key aspects that should be addressed. As explained by Sarker *et al* (2013) there needs to be a distinction between the base water use and the seasonal water use, where the base water use is assumed to be insensitive to climate conditions, as this is mostly indoor water use, whereas the seasonal water use is more reliant on the changes in climatic conditions, as there is a greater stress on outdoor water demand. The base use is taken in this case as the winter water demand as this is the lowest consumption of water within the annual period, however it should also be tested for any sensitivity to changes in the climate (Sarker, et al., 2013). Other studies such as in Balling Jr *et al* (2008) have discussed the link between the climate conditions and water demand, in particular with houses that have larger lots, pools and higher income-residents the climate variables explain 70% of the monthly variance in water demand for the case in Phoenix, Arizona (Balling Jr, et al., 2008). In Balling et al (2006), which is also a study based in Phoenix, Arizona states the climate variability has particular relevance in Phoenix as the outdoor use of water is a total of 4% of the total water use, which is sensitive to climate conditions such as temperature and rainfall. It is therefore important to also mention the significance of climate conditions on the variation of water demand especially with semi-arid and arid regions, which may be particularly prone to drought seasons. The reliance of rain is also a major issue, as without rain and high temperatures causing increase evapotranspiration, the water demand will be seen to increase as well (European Commission, 2015)

Gutlzer & Nims (2005, p.1778), had an interesting study as the city of Albuquerque's water supply is provided entirely from its groundwater resource, and therefore the consumption of water should not be as sensitive to changes in short term changes in climate conditions. With this stated, it was however found that over 60% of the year to year changes in summer water demand could be explained by the variance in inter annual temperature and precipitation changes, when conducting simple linear regression analysis. On the contrary, there are also studies such as Gegax et al. (1998) and Michelsen et al. (1999) that have found that there was no relationship between precipitation and water demand and only slight correlation with temperature, which was a study conducted for the cities of Oklahoma.

When studying the impact of climate variability on water demand, it is also important to consider the threshold at which water consumption is dependent on temperature, or any other climate variable. In a study by Sarker et al., (2013), which was conducted in Melbourne, Australia the found that there is a temperature threshold of 15.53 °C for Greater Melbourne and a precipitation threshold of 4.98 mm. This indicates that any water consumption that occurs at temperatures higher than 15.53°C is not considered as being directly related to the temperature, and the same logic applied for the precipitation threshold. Other indicators that could therefore explain the variance in water consumption beyond these thresholds are for example changes in the population or other factors (Sarker, et al., 2013).

There is a diverse set of methods used in previous studies relating to water demand analysis, ranging from simple regression based analysis, to cluster analysis when dealing with socio-demographic groups of interest such as in Jorge et al (2015), to sensitivity analysis as well as more complex systems of procedure using artificial neutral networks (Haque, et al., 2015). One technique used in Chang (2014) in the case of Portland, Oregon is the standard regression analysis along with the time series analysis method ARIMA (auto regressive integrated moving average) modelling to test whether the climate variations have stronger determination of the variance when compared to predicting the water demand based on the previous water demand in the time series. It was found that the ARIMA values for the coefficient of determination showed a more significant result than for the ordinary least squares (OLS) regression model, which in this case provides evidence that the memory of the previous days water use is more sufficient as a predictor tool than the climate variation (Chang & Praskievicz, 2014). Other studies conducted by Praskievicz & Chang (2009) in the case of Seoul, Korea uses ordinary least squares (OLS) regression models, where the weather variables used explain up to 39% to 61% of the variance in season water use, so from one-third to two-thirds of the variance explained. ARIMA modelling was also used in this case and explained 66% of the variance in water use. Both the MLR and ARIMA methods take into account seasonal variability of the time series data (Adamowski

& Kaz, 2013).

The cluster analysis performed in Jorge et al (2015) dealt with socio-demographic factors to investigate how they influence the water efficiency within households, where parameters such as number of persons per household, property type, family dimension and age of residents where evaluated where peer group characteristics based on the weekly per capita consumption where determined from cluster analysis. The key findings identified the household efficiency level of a particular cluster, so which type of households would need to make adjustments to their water consumption in order to meet average water consumption values. This method can be used to help identify what socio-demographic factors are more dominant when predicting average water consumption values per capita.

Two methods are present in Babel et al (2014) which are sensitivity analysis of explanatory variables as well as artificial neural networks (ANN) in order to predict future trends in water consumption. In this key findings of the study it is revealed that climate variables such as temperature, precipitation, and wind speed had a closer associated with smaller time intervals such as daily or weekly water demands whereas the longer scale annual time periods could be better explained by the socio-economic and demographic variables such as population household income and education level. In reflection to this study it is important to note that in order to properly assess the significant of socio-economic and demographic variables on the trends of water demand a long-term dataset is required.

Another perhaps less represented method is spectral analysis, which has the advantage of treating to observed data depending on its frequency and wavelength, so that specific periods of influences can be noted. In Adamowski & Kaz (2016) the climatological factors are tested for any influence on urban water demand, where temporal patterns can be determined using spectral analysis, and this is then coupled with Fourier and cross-spectral analysis to analyse the significance of the patterns that are detected. It is found by Adamowski & Kaz (2016) that the urban water demand (UWD) is sensitive in the summer months to air temperature, where the temperature is greater to 10-12 °C. The areas with low precipitation show however an inverse relationship to the UWD during the summer months. Finally in the study a 7 day cycle is observed using wavelet transform and Fourier analysis.

A different approach of mathematical modelling has been undertaken for analysis of hourly water consumption data, as described by Lage *et al.* (2012). This looks into how the water distribution system can be improved by the use of models simulating water consumption, where a function containing two periodicities can be applied to fit the observed data. This assumes that the data oscillates at regular intervals, which is the case for hourly flow of water. Over the 24 hour period, Lage *et al* (2012) have combined two sine waves, oscillating around the 12 hour and 24 hour period. By using this double sine equation, which oscillates around a constant value of the averaged water consumption over time, the R^2 is as high as 0.96-0.98 for four areas selected, which are situated in Country Sligo, Ireland and consists of 85% domestic users.

3 Methodology

The diagram below in Figure 2 illustrates the process taken in this thesis work.

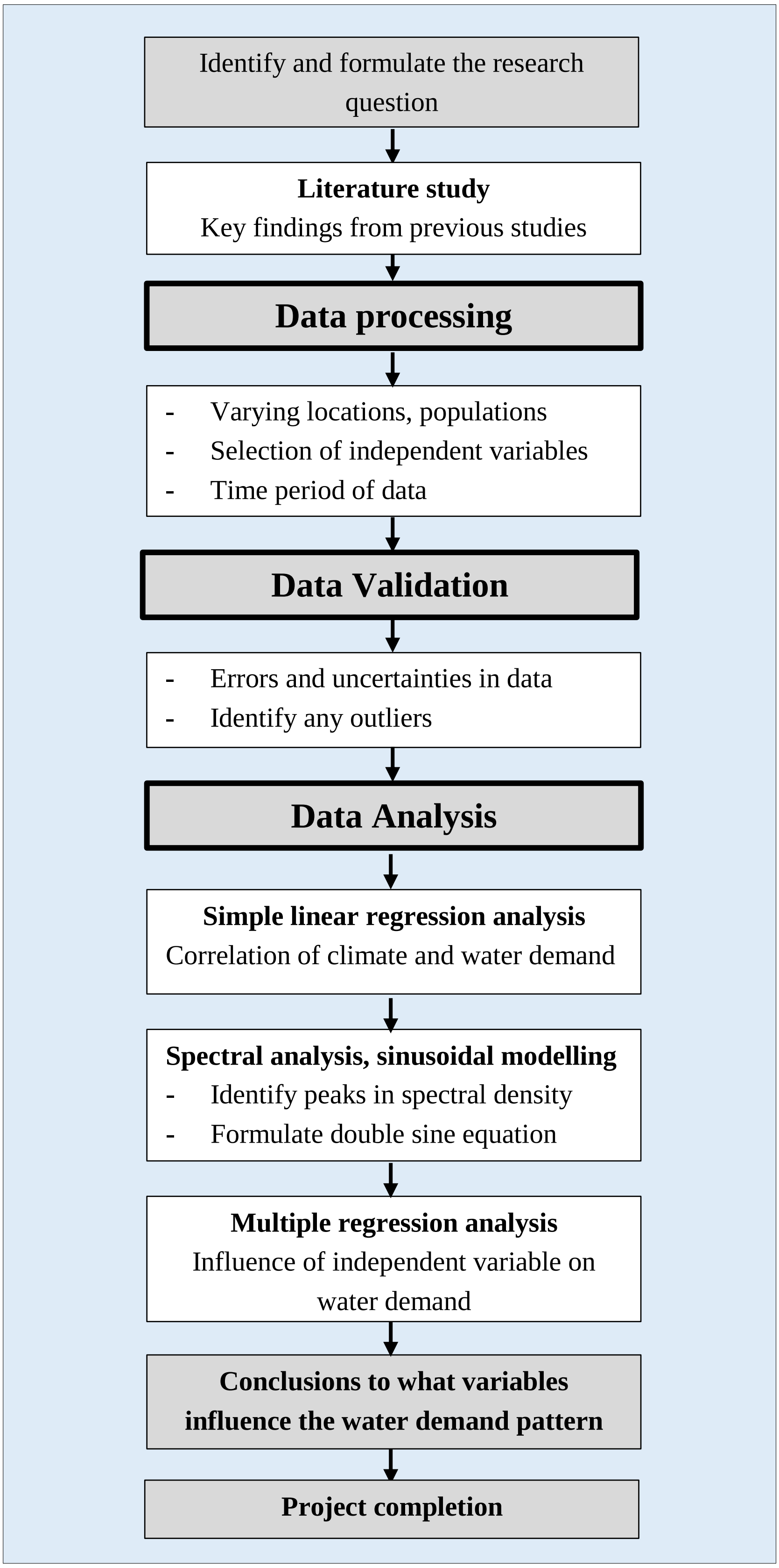


Figure 2: Process of the project

3.1 Data collection

A collection of water measurement datasets were obtained in this study, which included areas within Gothenburg, Kalix, and Alvesta municipalities. The water measurement datasets provided include a range of water-users that include residential, industrial, public/commercial and also agricultural users of water consumption.

3.1.1 City of Gothenburg

The water measurement data obtained from the recycling and water department of the City of Gothenburg (Göteborg Stad) is measured in L/s of water consumption every hour for the year 2014. For individual buildings the data was obtained from Göteborg Energi, an energy company based in Gothenburg and had a longer available time period ranging from years 2011-2015, with units in L/hour.

3.1.2 Kalix municipality

The water measurement data collected for Kalix Municipality includes Kalix municipality, Myrdalen town and Rörbäcks town, however only the data from Myrdalen will be sued for further analysis. The water supply from Myrdalen is suppled from Myrdalens water plant, and the units the water measurement data is provided in is in m3/hour for the months of October, November and December. There is also daily water measurement data for the year 2015, although some data is missing.

3.1.3 Lönashult, Alvesta municipality

The water measurement data collected for Lönashult was provided by the Alvesta municipality, where there are a total number of 46 subscribers and 128 written persons (taken as 174 consumers). The flow of water is measured by Siemens MagFlo, which is a type of electromagnetic flow measurement device and was installed in September 2011. The hourly water measurement data is obtained for 2011-2016, although both 2011 and 2016 contain large periods of missing data, where the data for years 2012 to 2015 are complete datasets. The data is provided in units of m^3/h at hourly intervals.

3.1.5 Climate data

The climate data used for this study is obtained from the Swedish Meteorological and Hydrological Institute (SMHI) form its open data source. The climate data chosen for this study is as follows; precipitation amount (mm/day), average temperature (°C), snow depth (m/day) and sunlight hours (seconds/day).

Hourly and daily data is mostly available for all regions, however data gaps were present especially in the hourly data, so where possible interpolation measures were only one or two points were missing, however with large areas of data missing the data was filled with nearby measuring points. An example of this is the Göteborg temperature data, as the temperature data is taken both from the Centre of Stockholm and also from the measuring collection point at Gothenburg’s airport. The snow depth measurement data was from the Kållered measurement station.

The precipitation and temperature data obtained for Kalix was collected from Haparanda (Temperature), Kalix (Precipitation) and Luleå (Sunlight hours). There is a lack of 'active' data present for Kalix municipality from the SMHI open data source, and therefore these measures were taken of finding the closest measurement stations to Kalix municipality.

3.2 Modifications to the water measurement data

As explained previously, the water measurement data received for Gothenburg contained flagged values for where there are uncertainties or errors in the data. The points of error in the data are derived from outliers which are represented as unusually high peaks in the data, repeated values or where there are shifts in the minimum consumption due to possible leaks in the pipe distribution system.

In order to adjust the values for the Gothenburg dataset that were flagged, they were first removed from the datasets and replaced with the average hourly value for their specific times. This then smooths the original water measurement data as significant outliers are removed and can therefore be used for further analysis. If there is only two or three values missing from the time series, then interpolation is another technique used to fill the missing values.

The water measurement data from Kalix municipality and Lönashult had not been modified or flagged from the original dataset, however as the Kalix data only provided sufficient water measurement data for the daily time period only this was take under consideration. For the water measurement data in Kalix a slightly different approach was used for missing values for the daily water measurement data in 2015. The average of the individual months were calculated and substituted for values that were written in italic (uncertain measurements) or that were already missing. The month of February however is excluded from the data set as there were no reliable measurements in this dataset, and majority of the values were missing. The data from Lönashult was left as the original water measurement data, and seeing as the dataset did not contain any missing data points or shifts within the data that would identify any leaks, therefore the data is assumed to be reliable and of good quality.

In the Lönashult water measurement data was also not flagged previously, however a sufficient 4 year period was obtained per hourly consumption rate. As the data is not normally distributed, due to significant differences between day and night consumption, and seasonal variations in water consumption, a standard test for outliers such as the Grubbs tests could not be taken, as this assumes normality of the data. Instead, the temporal first derivative of the time series data was taken as this is more likely to behave like a Gaussian distribution (Mathematica, 2015). A threshold limit of two or three times the standard deviation is then put in place in order to detect any outliers in the dataset.

3.3 Spectral analysis

Spectral analysis, or analysis of the 'frequency domain' is one method used in time series analysis. Unlike a time domain approach this is a frequency domain approach and therefore analyses the fluctuations of a signal around a stable state, which uses the periodicity to describe the behaviour of the time series (Rust, 2007). The periodicity of a time series can be displayed by using a periodogram, where the periodogram is in fact the sample estimate of the population function, known as spectral density. The spectral density, which as explained is the frequency domain representation of a time series, is directly related to the auto covariance time domain representation of the data. The auto correlation follows the formula below in Eq.1:

$$auto\ correlation = \frac{auto\ covariance}{variance} \quad [1]$$

The spectral density and the auto covariance function contain essentially the same information however they are expressed differently. The function $\gamma(h)$ can be taken as the auto covariance function, with the function $f(\omega)$ taken as the spectral density regarding the same process, where ω denotes the frequency and h denotes the time lag for auto covariance (The Pennsylvania State University , 2016). The auto covariance and spectral density can be explained in the formulas as follows:

Auto covariance: $$\gamma(h) = \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{2\pi i\omega h} f(\omega)d\omega, \quad [2]$$

Spectral density: $$f(\omega) = \sum_{h=\infty}^{h=\infty} \gamma(h)e^{-2\pi i\omega h} \quad [3]$$

The formulas above are known as Fourier transform pairs. The periodogram that is calculated is only a roughly estimate of the spectral density, seeing as it measures only discrete fundamental harmonic frequencies, instead of over a continuum of frequencies, which defines the spectral density. Once the periodogram has been calculated, it can then be smoothed by use of centered moving averages, to improve the estimate of the spectral density. The formula for smoothing a time series can be estimated using the Daniell kernel approach with parameter m, which is a centered moving average for the time series, which creates a smoothing value at time *t* by finding the average of all the values between the times of $t - m$ and $t + m$, which can be shown in the smoothing formula with an $m = 3$ below:

$$\hat{x}_t = \frac{x_{t-3}+x_{t-2}+x_{t-1}+x_t+x_{t+1}+x_{t+2}+x_{t+3}}{7} \quad [4]$$

Weighted coefficients are then assigned so that more weight is given to the center point of the moving average, i.e. for x_t, and minimum weighted coefficients for the outer bounds, x_{t-3} and x_{t+3} (The Pennsylvania State University , 2016).

3.5 Multiple linear regression analysis

The aim of this study is to find both the correlation and regression that certain factors have and how they could explain the temporal and seasonal variance observed in the water measurement data. In order to perform this analysis, a series of independent variables will be selected, which have the potential to explain the temporal variance observed in the dependent variable, which is in this case water consumption. A simple linear regression model is not fit for this purpose, as it is necessary to take into account more than one relevant factor in order to explain the variance in water consumption (Uriel, 2013).

To understand the basis of multiple regression analysis, simple linear regression first has to be explained. Linear regression is a process in which it is possible to make predictions about variable "Y" based on the knowledge you have obtained from variable "X" (Higgins, 2005). So it is essentially a method which looks into the relationship between two variables in order to make predictions based on that relationship. The equation for simple linear regression is shown below:

$$Y = a + bX \qquad [5]$$

Where Y is the dependent variable, a is the Y-intercept, b is the gradient of the line and X is the explanatory variable.

Multiple regression analysis on the other hand is an extension of simple linear regression as it assesses the significance of two or more independent variables with a dependent variable, and can provide more accurate predictions to the trend observed in the dependent variable (Higgins, 2005)

The equation of the model of multiple linear regression, with a given n observations is:

$$Y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_p x_p + \epsilon_i \; for \; \beta_j = 0 \; where \; j = 1,2, \ldots n \qquad [6]$$

The β_0 is the Y-intercept, and the β_j terms are the coefficients of independent variables, otherwise known as the regression coefficients, and are calculated using the least squares estimate of β by minimizing the following equation (Sen & Muni, 1990):

$$S = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{1i} \ldots - \beta_k x_{1i})^2 \qquad [8]$$

The coefficient of determination in multiple regression can be expressed as:

$$R^2 = \left(\frac{SS_{reg}}{SS_{tot}}\right) = 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \hat{y}_i)^2} \qquad [9]$$

Where the $\hat{y}_i$, is the fitted value, $\bar{y}$ is the mean of the dependent variable, and y_i is the observed value of the dependent variable

The total sum of squares, which is proportional to the variance of the data is:

$$SS_{tot} = SS_{reg} + SS_{res} = \sum(y_i - \bar{y})^2 \qquad [10]$$

The regression sum of the squares, which is also known as the explained sum of the squares is:

$$SS_{reg} = \sum(\hat{y}_i - \bar{y})^2 \qquad [11]$$

The residual sum of the squares is:

$$SS_{res} = \sum(y_i - \hat{y}_i)^2 \qquad [12]$$

As mentioned previously, the R^2 value is maximized as the residual sum of squares is minimized, which is optimized when there is the smallest difference between the predicted value and the actual value. The value of R^2 always falls between zero and one, so $0 \leq R \geq 1$. A value nearer to 1 is considered better, as the linear relationship between X and Y is stronger, where a value closer to 0 signifies no or little correlation between the explanatory variable and the dependent variable. So for example, when the R-squared value is 0.5 or below, then only 50% or less of the variance in the data is explained, so the ability of the predictor variable(s) decreases, and is therefore unreliable (Ostertagova, 2012). Any number of variables can be used in the regression equation, however this then means that R^2 value will always increase with the increasing number of explanatory variables used in the regression equation. To protect against any spurious relationships an 'adjusted' R^2 value is used, which is computed by:

$$R_{adj}^2 = 1 - \frac{(1-R^2)(n-1)}{n-k-1} \quad [13]$$

Where n is the total sample size, and k is the number of independent variables used to predict the dependent variable. The adjusted R^2 value will always be smaller than R^2 as it takes into account the number of explanatory variables that are included in the regression equation (Ostertagova, 2012).

Two methods of regression analysis will be used for this investigation, which are the 'best model' regression analysis and stepwise regression analysis. The aim of finding the optimum solution in regression analysis, is to minimize the residual mean square, which in turn maximize the coefficient of determination (multiple correlation value) R^2. The best model method selects all explanatory variables chosen, and computes a regression equation. It identifies the variables that have a p-value of lower 0.01, however all variables are still selected for the model. Because of this, the variables that are not considered statistically significant should be removed by the user from the input variables, and the regression equation should then be computed again, this time only with the independent variables that are statistically significant.

The second technique used is 'stepwise regression'. In this technique, only the explanatory variables that are statistically significant are selected, and then the variables are added one at a time until an optimal solution for R^2 is acquired. If there are two variables that are very similar in terms of characteristics such as sunlight and temperature, and can therefore be described as collinear variables, then one of these will be eliminated from the regression analysis. Another important point to make about the stepwise regression method is that it account for collinearity occurring in the data. Collinearity is when there are two variables that have an excessive correlation between them, and therefore one will be eliminated from the model when finding the optimal solution for R^2. Another approach to identify the collinearity of the data is by the use of variance inflation factors (VIF), where the higher the value for VIF the higher the collinearity between variables, or by the correlation matrix, which identifies the correlation between all explanatory variables, and if the correlation coefficient are larger than 0.8 than they are subject to collinearity (Statistics Solutions, 2016).

3.5.1 Hypothesis

There are a number independent variables that could potentially explain the variance observed in the dependent variable water consumption. In order to test whether they are significant at predicting the outcome of water consumption a null hypothesis will be formulated. The null hypothesis states that:

$$H_0 \rightarrow The\ independent\ variable\ has\ no\ signficant\ effect\ on\ the\ water\ consumption$$

If the null hypothesis is true, then the corresponding regression coefficient will equal to zero. The null hypothesis will be tested against an alternative hypothesis which states that:

$$H_1 \rightarrow The\ independent\ variable\ has\ a\ signficant\ effect\ on\ the\ water\ consumption$$

If the null hypothesis is rejected, then the corresponding regression coefficient will be less than or more than zero, as explained below in Eq. 14 & 15 (Sen & Muni, 1990).

In regression analysis a minimum significance level is set at which the null hypothesis should be rejected. This level is referred to as the p-value, and is applied to test the significance of an independent variable, to see to what extent it affects y, the dependent variable. (Sen & Muni, 1990). The significance of the independent variablex_j, is determined by testing the null hypothesis, as explained below in Equations 14 & 15.The p-value describes whether the null hypothesis should be reject or not, which in this case is based on a probability of below 5% ($p \leq 0.05$) that the null hypothesis is wrongly accepted. The threshold limit, α therefore equals to 0.05 and is known as the significance level, where a Type I error occurs when the true null hypothesis is accepted when it should have been rejected (Lane, 2007).

If the p-value is more than 5% ($p \geq 0.05$) then the null hypothesis should be accepted. The null hypothesis states that the regression coefficient should equal to zero, where:

$$H_0 : \beta_i = 0 \qquad [14]$$

When the null hypothesis is true, the coefficient of regression is equal to zero, which is multiplied by x_j and should thus be removed from the regression equation.

If however the p-value is less than 5% ($p \leq 0.05$) then the null hypothesis should be rejected as there is sufficient evidence that the regression coefficient does not equal to zero, where:

$$H_0 : \beta_i <> 0 \qquad [15]$$

The regression coefficient will received a negative value when it is negatively correlated to the dependent variable, and thus it will receive a positive value when it is positively correlated with the dependent variable.

3.5.2 Sinusoidal fitted variable

The sinusoidal fitted variable assesses the periodicity of the data, and is a function that repeats its values over regular intervals, thus simulating the non-seasonal (base) pattern of the water consumption data. The periodic formula, which contains two sine elements that oscillate around the constant $\bar{w}$ is:

$$W(t) = \bar{w} - A_1 \sin\left(\frac{2\pi}{T_1} t + \varphi_1\right) - A_2 \sin\left(\frac{2\pi}{T_2} t + \varphi_2\right) \qquad [16]$$

$W(t)$ is the volumetric flow rate of water consumption measurement per hour and the $\bar{w}$ is taken as the average water consumption per unit time. The A is the amplitude of the oscillation and φ is the phase, which indicates at which point in the cycle the oscillation begins. The t is the time for the interval of the time series and the two sine elements follow the 12 and 24 hour cycles, which are represented as the T_1 and T_2 respectively, and the equation is fitted for the hourly water consumption data.

3.5.3 Dummy variables

As well as the climate variables and the double sinusoidal equation formulated for the water measurement data, indicator or dummy variables will be used as well. These type of variable consist of only two integers, 0 and 1. For example the daily water measurement data will be split into weekday and weekend consumption, where weekday will be assigned a value of 0, and the weekend, which includes Saturday and Sunday will be assigned a value of 1. See the Table 1 below:

Table 1: Weekend dummy variable

Weekday					Weekend	
Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
0	0	0	0	0	1	1

The reason for this is because there is a rise in the water consumption over the weekend, where for example in residential areas people are more likely to be home, and not at work. The same procedure was implemented for individual week days, so applying 6 variables, for the each week day where Sunday would just include the value 0. This is shown in Table 2 below:

Table 2: Weekday dummy variable

Weekday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Monday	1	0	0	0	0	0
Tuesday	0	1	0	0	0	0
Wednesday	0	0	1	0	0	0
Thursday	0	0	0	1	0	0
Friday	0	0	0	0	1	0
Saturday	0	0	0	0	0	1
Sunday	0	0	0	0	0	0

For the holiday season, the holiday term taken from school holidays were given a value of 1 and the rest of the year were given the value 0. A monthly dummy variable is also created, in order to see a variation over a larger time period, as well as for the holiday seasons as the fluctuations of people can cause the water demand to change. For the hourly water consumption data, a dummy variable is provided for the daytime hours (between 7am-12am), as the consumption generally increases quite significantly during the daytime hours.

4 Study Areas

This study investigates areas which are located in the Gothenburg, Kalix and Alvesta municipalities. Figure 3 below illustrates the three locations of the municipalities, which are situated in Västra Götaland, Norrbotten and Kronoberg counties respectively.

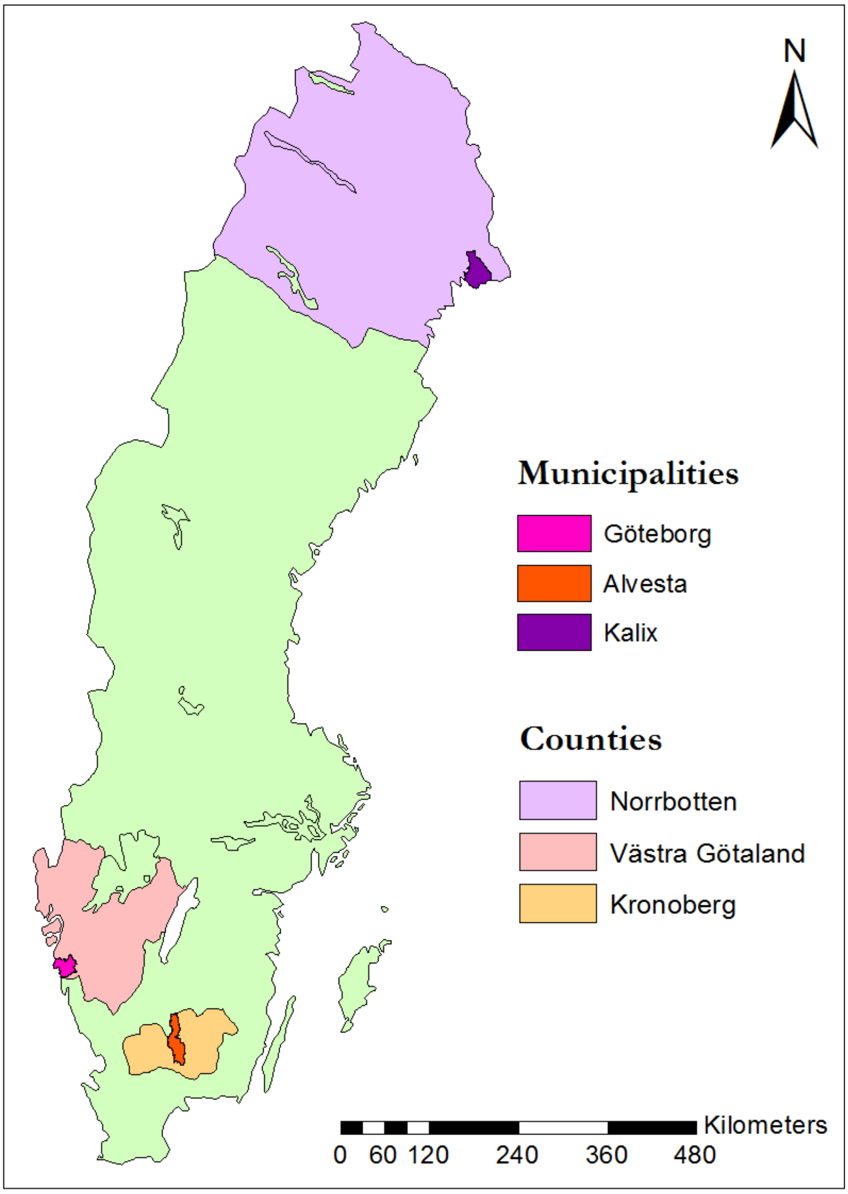


Figure 3: 3 study areas, Sweden

Table 3 that follows, gives the descriptions of the areas within these 3 study locations, where a total of 10 areas have been selected in Gothenburg, due to their varying population sizes (465-17580) and low flagged values. In Kalix, 3 towns were chosen and have a population range of below 50 to 1250 people. Lönashult is selected in Alvesta municipality, which has 128 subscribers, which breaks down into 46 households of 2-3 people per household. Table 3 will be described in more detail in the following sections 4.1, 4.2 and 4.3.

Table 3: Description of all selected areas

Area		Size	Avg. water demand (litres/day/person)	Description	Summer water consumption
Gothenburg (2014)	Amhult	698	**94**	Car repair, pizzeria, general stores, villas	Rise and Fall
	Burmans Gata	1343	150	Warehouses, villas	Rise and Fall
	Änghagsvägen	3706	161	Detached houses, villas, riding school	Rise and Fall
	Slottskogen	17580	178	Apartments, restaurants, schools	Fall
	Tyghusvägen	584	136	Sports facilities, restaurants, student housing, apartments	Fall
	Vattugatan	1707	183	Apartments and other general stores	Fall
	Torshallsvägen	1889	146	Villas, summer homes	Rise
	Näset	14061	318	Villas, summer homes, public facilities.	Rise
	Södra Skärgården	4501	300	Villas, summer homes, other facilities	Rise
	Blomstigen	465	203	Majority villas	Rise
Kalix (2015)	Björkförs	100	**420**	***Agricultural land***, residential homes	Rise
	Tandförs	<50	>197	Residential homes	Rise
	Nyborg	1250	285	Residential, apartment buildings, schools industry	Rise
Alvesta (2012-2015)	Lönashult	128	168	Residential homes	Dependent on year

4.1 Gothenburg

Gothenburg is the second largest city in Sweden and has a population of 491,630 (2007). It is located on the west coast, in the southwest of Sweden. It has a warm climate and moderately heavy precipitation. For the year that the hourly water measurement data was collected, in 2014, the total amount of rainfall for the entire annual period was 971 mm/day, where the wettest month occurred in October and obtained a precipitation amount of 167 mm. The driest month falls in March with a total amount of precipitation of 32 mm. The average, minimum and maximum temperature for the year 2014 (water collection data from 2014) is 10.3 °C, -5.3 °C and 25.9 °C respectively.

A total of 72 areas were provided within the Göteborg dataset, where each dataset is 'flagged' according to the percentage of errors and uncertainties that are contained in the data. 10 areas were chosen to investigate further, which also received percentages of below 5% of flagged values, so that any modification to the original dataset could be limited and accuracy maintained. The following 10 areas shown in Table 3 were used for the study, where this also shows the average water consumption in each of the areas. The highest average water consumption occurs in the locations of Näset with an average water consumption of 318 L/d/p and Södra Skärgården with an average water consumption of 300 L/d/p. Both these areas, in particular Södra Skärgården are areas with summer homes, so the number of people in the summer season will usually increase, therefore leading to an increase in water consumption.

As a way of separating the areas into categories, the summer variance in water consumption is analysed, which can be found in the last column of Table 3. The areas that tend to have summer homes, where people would go to during the summer holiday season, or a large proportion of villa homes have a rise in the water consumption during the summer season. Areas that are in the inner city of Gothenburg, such as Vattugatan and Slottskogen that contain apartments and other facilities such as schools, tend to decrease in water consumption during the summer season. Finally, there are areas containing villa homes but also services such as shops and restaurants that are observed to rise in water consumption during June, but in July it falls again, which has been assumed to be because of the vacation period, where a large amount of the population tend to go on vacation during July. The map of Gothenburg, which displays the 10 selected areas is shown below in Figure 4:

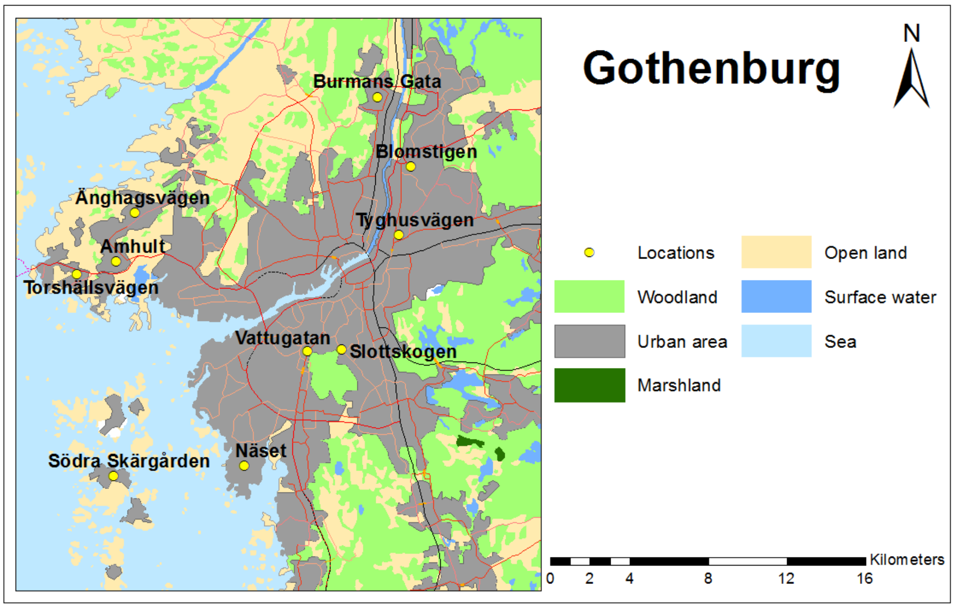


Figure 4: Map of the Gothenburg municipality

4.2 Kalix municipality

Kalix municipality is situated in Norrbotten County, in the northern most part of Sweden, next to the Finnish border. It has a population of 16 248 according to Statistics Sweden (2015). The climate of Kalix is cold and temperate, where even in the driest month there is a great deal of rainfall. The year obtained for daily water measurement data is 2015. For 2015, the total amount of rainfall for the entire annual period is 913 mm. The wettest month occurs in November with 181 mm and the driest month is in April, with 18 mm of total rainfall. When compared to the data for Gothenburg, the wettest and driest months occur a month later. The average, minimum and maximum temperature for 2015 is 4.1 °C, -25 °C and 18.4 °C respectively. The map of Kalix municipality is shown below in Figure 5, with the 3 areas selected for the investigation, as well as Kalix:

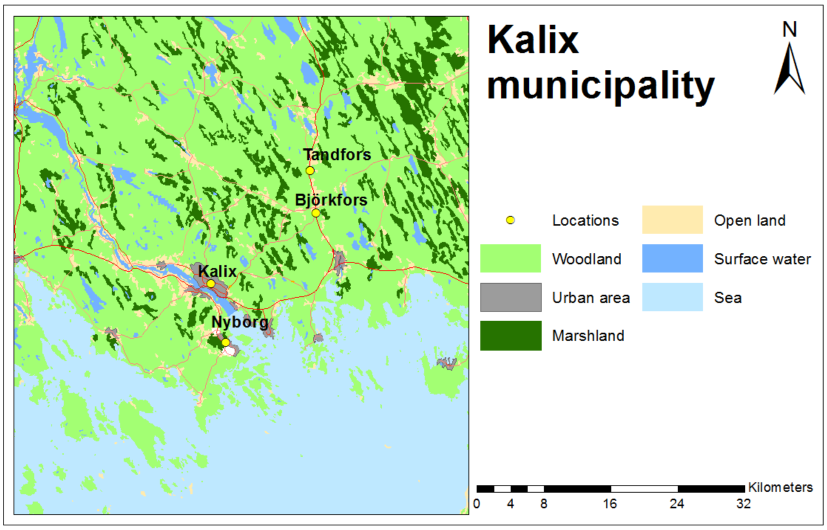


Figure 5: Map of Kalix municipality

What is important to note about the areas in Kalix, is that Björkfors has a high water consumption average of 420 L/day/person. This is assumed to be because of the agricultural land within this area, where cattle are kept on the farms and they have access to the drinking water supply.

4.3 Alvesta municipality

Lönashult is located in the southern part of Sweden, in Kronoberg County, which is located in Alvesta municipality. The data was collected from 2011-2016, however this study focuses on the period between 2012-2015, so for this period (2012-2015) the average yearly precipitation is 648 mm, with the least total amount of annual precipitation occurring in 2013 with 512 mm, and the highest amount of total precipitation in 2014, with 732 mm. The average, minimum and maximum temperature for 2012-2015 is 7.4 °C, -13.2 °C and 24.6 °C respectively. The map of Lönashult, which is located southwest of Växjö, in the Alvesta municipality is shown below in Figure 6:

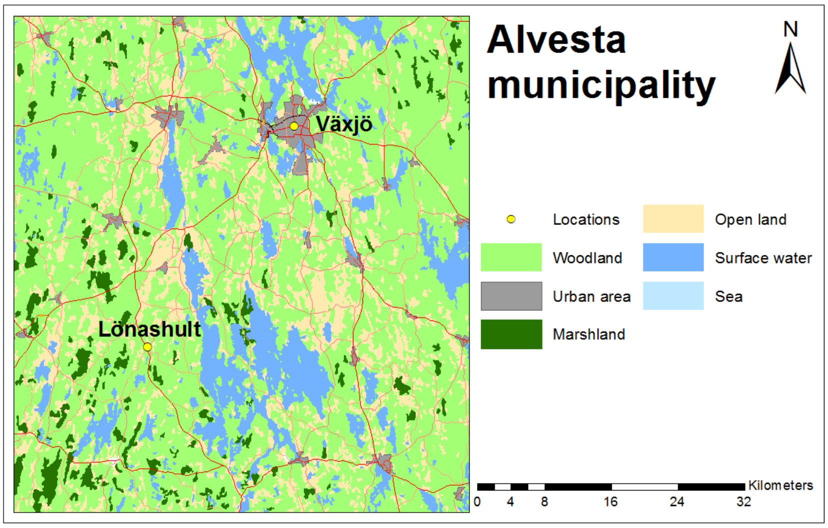


Figure 6: Map of Alvesta municipality

5 Results

5.1 Hourly consumption profiles

As mentioned previously there were 10 areas selected in the Gothenburg municipality. All 10 areas have been separated into similar patterns, which is illustrated in the three graphs in Figure 7. The graphs represent the average hourly consumption pattern over the annual period 2014. Generally there are two peaks in the day and the minimum consumption falls in the night time hours from 12am-6am. The data for Gothenburg is provided for the year 2014, where the average for each hour across the whole year is calculated and is displayed below in Figure 7. The hourly concentration rate, C_h is given in litres per hour per person which is calculated by dividing the population size with the water consumption and converting to litres per day. The term t stands for time in Figure 7 below.

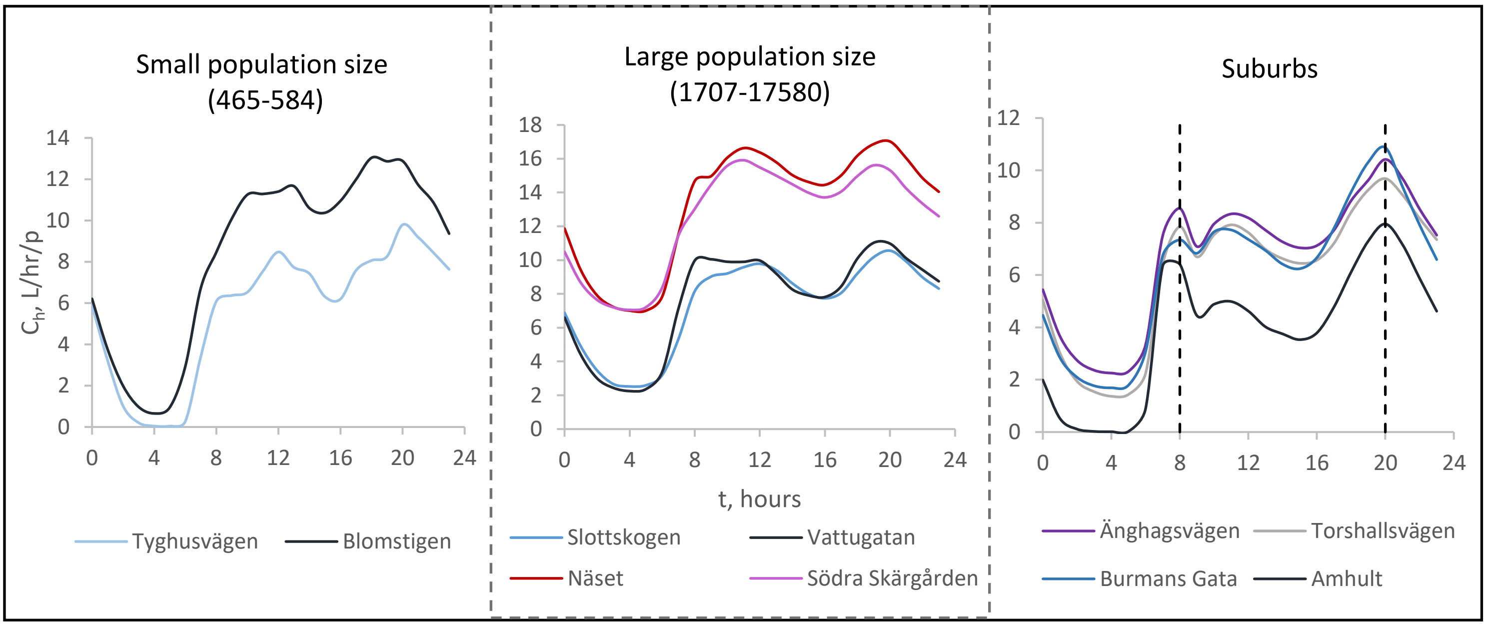


Figure 7: Average hourly consumption patterns for Gothenburg (2014)
a) Small population size b) large population size c) suburbs

Starting from the right, the four areas grouped in Figure 7 c) are all areas situated in the suburbs of Gothenburg, see Figure 4. They all have similar trends in the consumption pattern, where there is a sharp peak in the morning at 8am, following a small peak around 12pm, then another sharp rise and the highest peak occurring at 8pm. What was believe to be the reason for the sharp and early rise in the morning is the fact that these areas are in the suburbs, and therefore are more likely to contain commuters, who require a longer travel time into the city of Gothenburg. The graph in the middle, labelled as Figure 7 b) represents 4 areas that have a relatively large population size, of 1707 to 17580 inhabitants. There are peaks at around midday (12 pm) and later on again at 8pm. What was concluded from these areas, is that both Södra Skärgården and Näset have higher consumption patterns, which is due to the rise seen in the summer periods and contain summer homes, where compared to this Vattugatan and Slottskogen that decrease during the summer period and are inner city areas with apartments.

Finally, the last graph on the left, in Figure 7 a) represents two areas that have relatively low population sizes of 465-584 inhabitants. Both of the hourly consumption patterns have a more irregular pattern when compared to the other areas, and this may be due to the small population size, as the consumption patterns will vary a lot more, and it is harder to predict how many peaks there will be within the day. With this said, there are still two main peaks occurring, at 12 pm and at 8 pm, which follows the same trend as the middle graph. When compared to the average hourly consumption pattern for Lönashult, situation in Alvesta municipality, the graph for relatively low consumption patterns follows the same trend. In Lönashult where there is a population size of 128 people, there is also more of a sporadic pattern, and a peak occurring at 10 am, another small peak at 2 pm, and again in the evening at 8 pm. What should be noted is that the consumption pattern in Lönashult has its largest peak occurring in the morning, whereas for all areas investigated in Gothenburg the largest peak occurs in the evening. The average hourly consumption rate for Lönashult is illustrated below in Figure 8, where the hourly water consumption rate is given the term C_h:

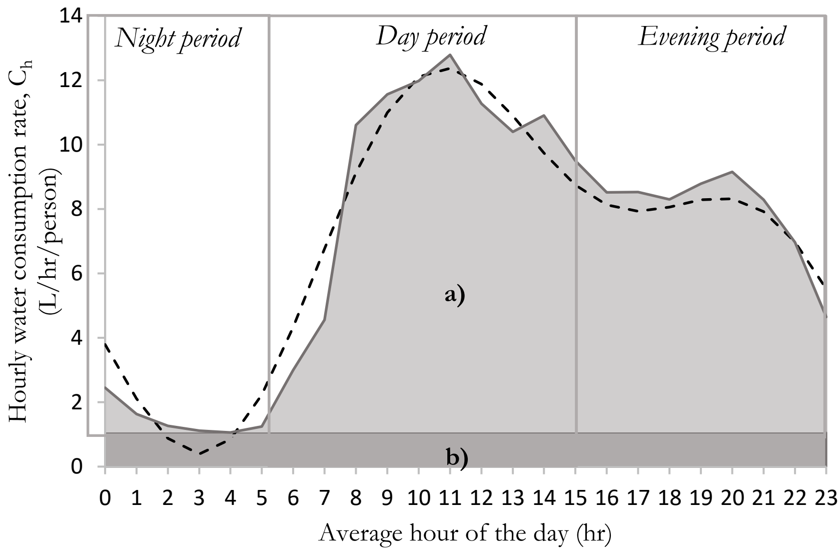


Figure 8: Average hourly consumption for Lönashult, Alvesta municipality (2012-2015)

The average hourly water consumption is calculated for the period 2012-2015, where a double sinusoidal equation was formulated to fit the trend in water consumption. The double sinusoidal equation consists of a constant (average water consumption in m^3/hour), as well as two sine functions that represent the 12 and 24 hour periods, which were identified in the spectral analysis of the data. This is explained further in the Section 5.2 and 5.3. The shaded areas in the graph labelled A) and B), identify different types of water consumption. The shaded area labelled a) is mainly induced by human consumption of water. There are also losses (from the piping network and from the owners) as well as 'random' uses of water, which is labelled as b).

5.2 Spectral analysis

Spectral analysis is used to find any dominant cycles within the annual period and it was found that in the daily total water consumption, had a peak in the spectral density at 7 days, suggesting there is a 7 day cycle in daily water consumption data. The hourly water consumption data shown below in Figure 9 a) reveals a peak in the spectral density at 12 and 24 hours, which is to be expected, as there is a very different pattern of water consumption depending on whether it is night or day. Within the 7 day cycle observed in the spectral analysis, all of the locations excluding that of Tyghusvägen, reveal small peaks in the spectral density that is visible of 2.3 and 3.5 days, suggesting that the 7 day cycle is stepwise, due to the 5 working days and 2 non-working days weekly pattern (Rust, 2007). The 7 day cycle also reveals that this is not perfectly harmonic, as the peak at 7 days is not the only visible peak in the spectral density, suggesting as daily water consumption data this is not as cyclic as the hourly water consumption time period.

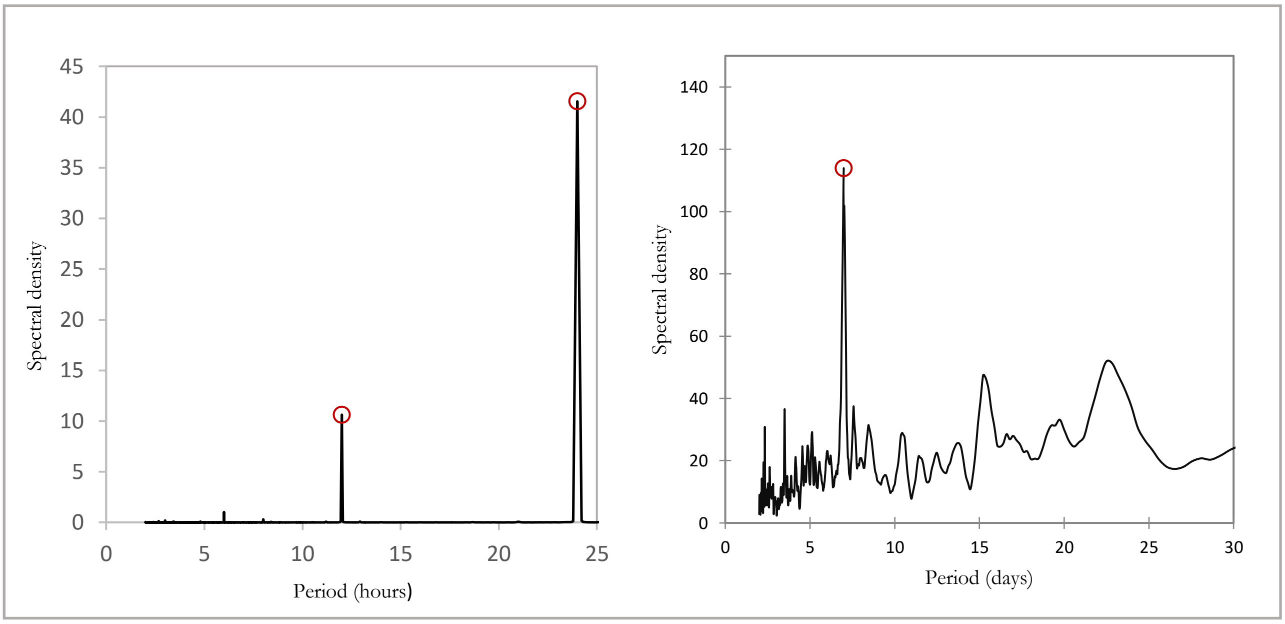


Figure 9: Peaks in the spectral density a) hourly b) daily

Spectral analysis was performed for the data, which shows a small peak in the spectral density at 7 days for the daily water consumption data, and for the hourly water consumption data there is a dominant peak at 24 hours, also a peak at 12 hours and a small peak visible at 6 hours. Figure 9 are the results from Lönashult. Figure 9 b) above illustrates the peak at 7 days for the daily water consumption data. For Kalix, the time series used however is not from the entire year, and only considered the months May through to December. The only complete water measurement data is for Nyborg, where the spectral analysis reveals a small peak at 7 days, 3.5 days and at 2.3 days, where the 3.5 days and 2.3 days could be harmonics of the 7-day cycle, where there is a 5 working days and 2 non-working days (weekend) during the a weekly cycle, as observed as well in the data for Gothenburg. These stepwise peaks are however very small and relatively insignificant.

5.3 Sinusoidal modelling

Sinusoidal modelling was simulated for the hourly consumption data for Gothenburg and Alvesta municipalities, where the hourly data for Kalix was insufficient for further analysis. The fitted equation explains a large proportion of the periodicity observed in the data, however there still remains seasonal variation as well as other variations due to different consumption patterns during weekend as opposed to the week, particularly in the areas located in Gothenburg. The sinusoidal equation was fitting for a period of 10 days, in the beginning of January for all areas included, the month on January is taken as the base consumption, with minimum influence from seasonal variation.

Below, in Figure 10 is an example of a fitted sinusoidal equation for Slottskogen, located in Gothenburg for the annual period 2014. The sinusoidal equation is fitted for the hourly water consumption data and for a 10 day period from 1st January to the 11th January. The equation applied here is take from Eq. 16. And the values used in the equation can be found in Appendix A. This area has a population size of 17 580 which is relatively large in comparison with other areas. It can be seen from the Figure 10 that there is a slight increase in the water consumption pattern, which occurs usually over the weekend period, over both Saturday and Sunday. There are also two significant peaks in the hourly consumption pattern, occurring at around 12-13 pm and another at 8 pm, so during the hours that people will be having lunch and dinner. The minimum consumption as mentioned previously occurs during the night, through the hours of 12 am-6 am.

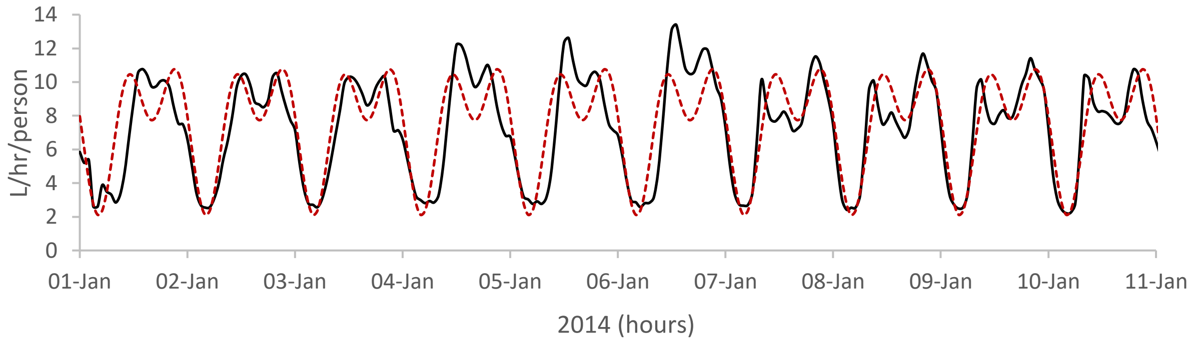


Figure 10: Sinusoidal fitted equation for Slottskogen, Gothenburg

Below in Figure 11, is the double sinusoidal equation fitted for the hourly consumption data of Lönashult, Alvesta municipality. The equation again that is used is Eq. 16 and the values inserted into the sinusoidal equation can be found in Appendix A

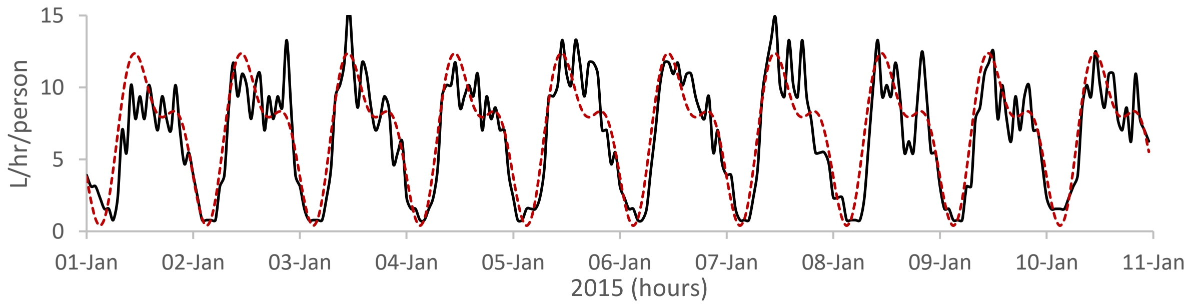


Figure 11: Sinusoidal fitted equation for Lönashult, Alvesta municipality

As mentioned previously the hourly consumption patter is more irregular due to the small population size of 128 people. It may prove harder to simulate a sinusoidal equation that fits with these hourly patterns of smaller population sizes, as the pattern is irregular and the periodicity more complex. In order to reduce the difference between the sinusoidal fitted equation and the actual water consumption, the sum of least squares is minimized to find the optimal solution.

5.2 Climate analysis

The following section includes the results from climate analysis, which looks primarily at the correlation between temperature and water consumption, but also for precipitation and water consumption for the areas situated in Gothenburg.

5.2.1 Gothenburg

In Gothenburg, from roughly mid-May to mid-September the temperature is usually above 15 °C. Temperatures above 20 °C occur in July and the beginning of August, where temperatures above 10 °C occur through the months of May to October in 2014. Higher peaks in the amount of precipitation occur after the summer period, in October 2014, where above 30 mm of rain can occur on maximum days. The annual daily values for temperature and precipitation are illustrated below in Figure 12:

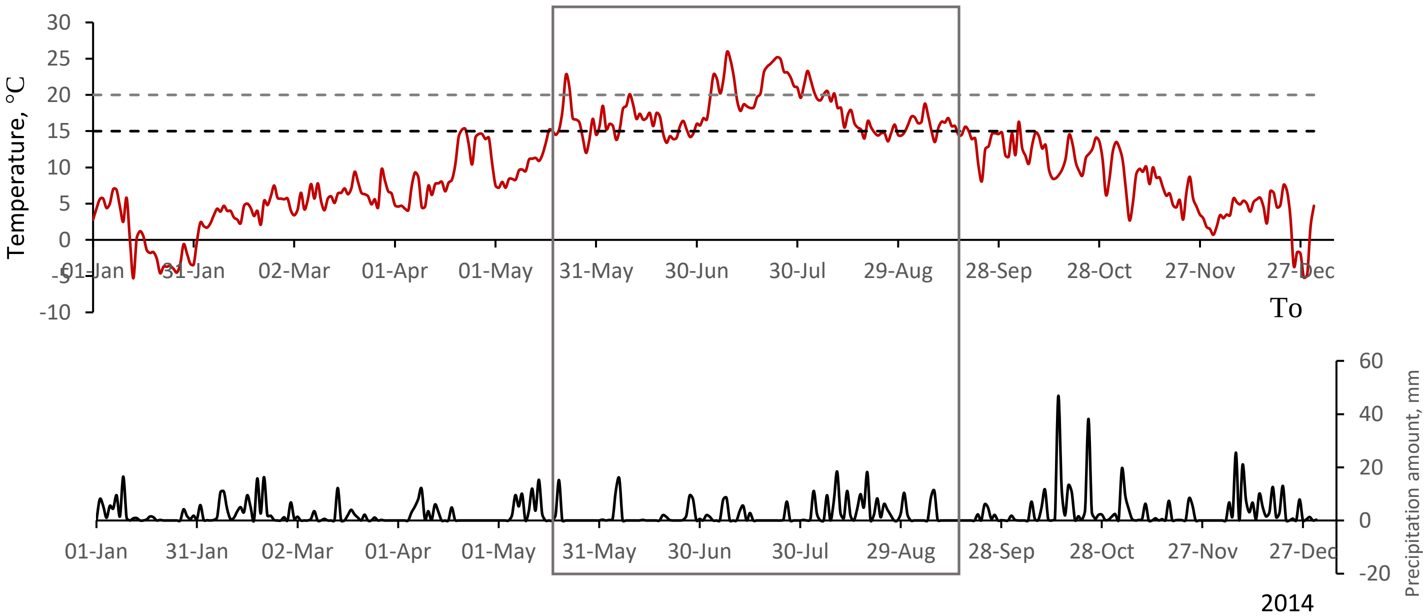


Figure 12: Temperature and precipitation for the period of 2014, Gothenburg (Source: SMHI)

find out how much the water consumption changes in regards to the increase in temperature, scatter diagrams were plotted to evaluate each area and its correlation with temperature, shown in Figure 13. It is observed that in most locations the water consumption pattern shows a change in its pattern with temperatures above 12 °C, as either the consumption will start to increase or decrease at a higher rate than temperatures below 12 °C. Correlation tests with the 10 areas revealed that for temperatures higher than the threshold of 12 °C, the results for Burmans Gata and Amhult were not statistically significant, as they had p-values of <0.01 (1% significance limit). These results are then not shown in Figure 13 below. The units used for the water consumption are in litres per day per person which will be taken as the consumption rate, Cd.

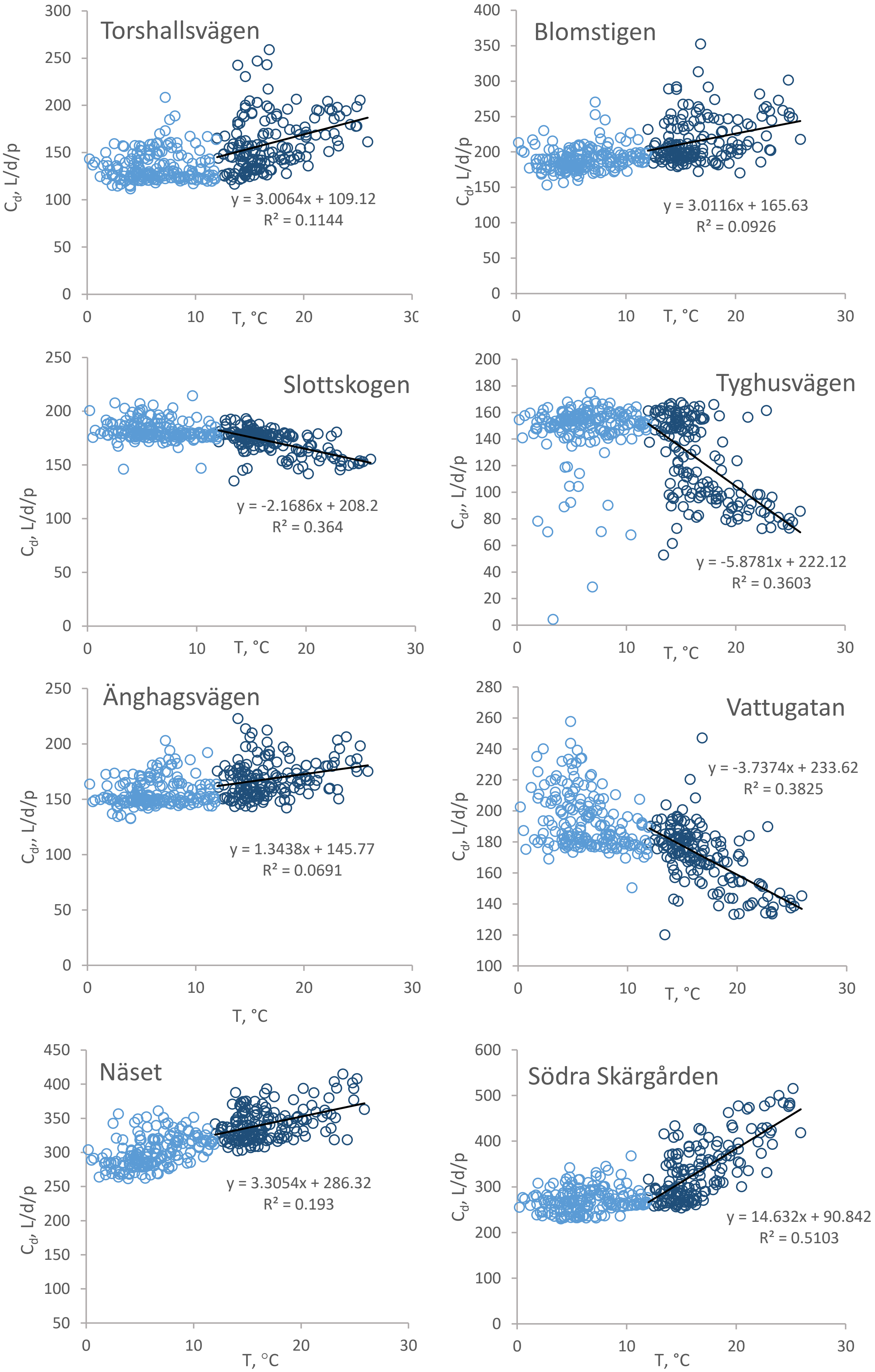


Figure 13: 8 areas in Gothenburg showing the relationship between water consumption and temperature over the annual period of 2014

temperatures above 12 °C. For Tyghusvägen, the water consumption pattern changes very significantly below around 120 L/p/d, and this is considered to be because it contains a large proportion of student housing within this location, and during the holiday seasons majority of students will not be at the accommodation. Therefore, as it will go into in further explanation in the multiple regression analysis, the Tyghusvägen location is likely to be better correlated to the fluctuations of people, i.e. the holiday seasons. The temperatures are above 12 °C in the months from May through to the end of September. Table 4 below identifies the change of water consumption as a percentage as a result of a 2 °C rise in temperature. As predicted, the locations selected for Category 2 show a decrease in water consumption above the temperature threshold of 12 °C, so during the summer season. The locations in Category 1 show a rise in water consumption above the temperature threshold of 12 °C.

Table 4: The percentage of change in water demand due a 2°C increase in temperature

Location	% change in water demand due to 2 °C increase
(1) Änghagsvägen	1.7
(1) Näset	2.1
(1) Blomstigen	3.1
(1) Torshallsvägen	4.3
(1) Södra Skärgården	12.3
(2) Slottskogen	-2.3
(2) Vattugatan	-3.8
(2) Tyghusvägen	-7.2

The highest positive changes in water consumption are in the locations of Södra Skärgården, which mentioned previously is part of the archipelago and will have significant fluctuations of people during the summer season. The highest negative change in water consumption occurs in the location of Tyghusvägen, which drops by 7.2% with an increasing 2 °C. This is as previously stated could occur because of the decrease in students during the summer period. The correlation between the amount of precipitation and water consumption was also considered, where a precipitation threshold of no rainfall to an occurrence of rainfall was chosen, so values of average water consumption below 0 mm/day and above 0 mm/day for the whole year were further analysed. Below Table 5 represents the percentage change of water consumption due to the occurrence of rainfall.

Table 5: Percentage change in average water demand due to occurrence of rainfall (>0 mm/day)

Location	% change in average water demand due to occurrence of rainfall (>0 mm/day)
(1) Änghagsvägen	-5.1
(1) Näset	-3.1
(1) Blomstigen	-9.4
(1) Torshallsvägen	-9.2
(1) Södra Skärgården	-7.7
(2) Slottskogen	2.5
(2) Vattugatan	2.7
(2) Tyghusvägen	3.3

***** (1) is Category 1, (2) is Category 2***

The results in Table 5 reflect the change in average the average water demand for days where rain occurs to days where rain is not occurring. As shown in Table 5, the opposite patterns occurs to that of the rise in temperature, where if the precipitation is considered for the entire year, the average of the water consumption decreases in the Category 1 areas, and increases for the Category 2 areas. This is expected as the precipitation is minimal during the summer season, where the total precipitation during the summer season (May-September) is 351 mm compared to 620 mm in the winter season (October-April), or that the count of no days with no rain is 86 in the summer compared with 78 in the winter season.

5.2.2 Kalix municipality

The three locations for the Kalix measurement data are Björkfors, Tandfors and Nyborg and the daily water measurement data is obtained for the year 2015. Björkfors, which is situated in Myrdalen has a population of 100 and consists of residential homes and agricultural land. The average water consumption is 422 L/day/person, which is a high consumption of water but this is justified due to the fact that the drinking water is also used for the cattle maintained on the farms. The population of Tandfors which is located in Myrdalen is below 50, which taking this as the maximum of 50 persons the average water consumption is 197 L/day/person, and the area consists of only residential homes. The last location is Nyborg, situated in Kalix municipality and has a population of 1250. The area consists of a mixture of building uses such as residential homes, schools as well as industrial use of water consumption. The average consumption is 285 L/day/person.

The temperature reaches above 10°C in the summer season (from beginning of June to the end of September), however the average temperature does not often reach above 15 °C. The temperatures reach below -15°C in the winter season, during January. Below Figure 14 displays the temperature and precipitation for the annual period of 2015, in Kalix municipality as well as the daily water consumption pattern for all three locations:

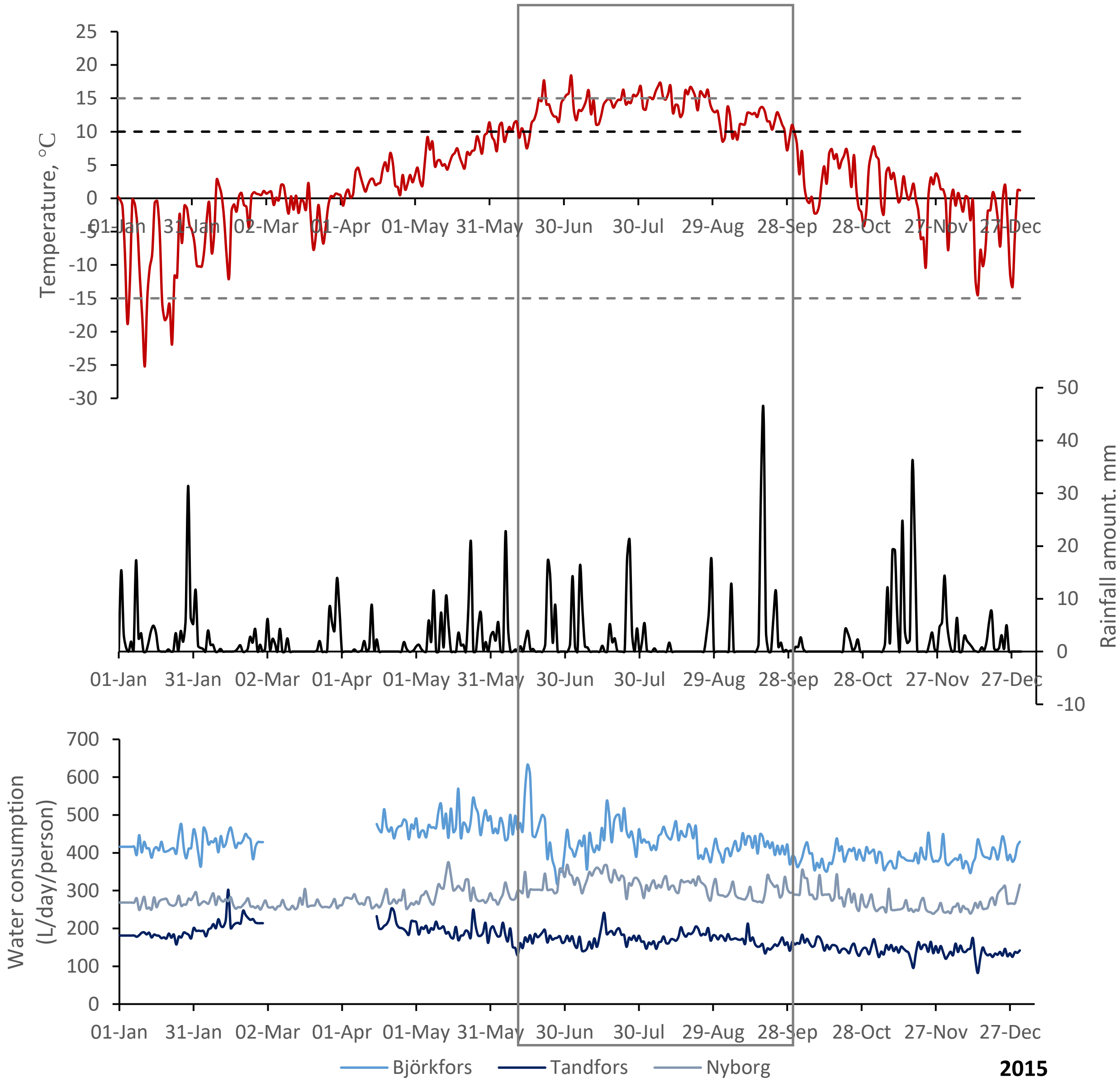


Figure 14: Climate and total daily water consumption for 3 locations in Kalix municipality for 2015

A correlation test was performed for the three locations and temperature, for the period of beginning of May to the end of December, as there is no missing data within this period. All locations had p-value of below 0.0001 so are considered statistically significant, where the coefficient of determinations are 0.12, 0.27 and 0.36 for Björkfors, Tandfors and Nyborg respectively. For a temperature increase of 2°C leads to an increase in water consumption by 1.1%, 2.2% and 1.8% respectively for Björkfors, Tandfors and Nyborg. Table 6 illustrates the p-value and R^2 value, which is shown below:

Table 6: Percentage increase in water demand with a 2°C increase in temperature

Town	p-value	R^2	% increase in water demand with a 2°C increase in T
Björkfors	<0.001	0.12	**1.1** (4.8 L/d/p)
Tandfors	<0.001	0.27	**2.2** (3.7 L/d/p)
Nyborg	<0.001	0.36	**1.8** (5.5 L/d/p)

For both the temperature limits of above 10°C and 15°C the significance testing for the p-value shows that the correlations are not statistically significant, excluding that of Nyborg for above 10°C. This shows a percentage increase of 0.74% with a 2°C rise in temperature, with p-value of 0.001 and an R^2 value of 0.092.

5.2.3 Lönashult

The data for Lönashult was collected for the years 2011 to 2016, however only years 2012 to 2015 will be analyses as they are complete datasets, thus encompassing four complete summer seasons as well as three full winter seasons. From around May to September, where the summer seasons falls in between these months the temperature reached above 10 °C for all years, as illustrated below in Table 7. The water consumption data was first modified using the 1st derivation of the original values, where then a threshold of 2σ or 3σ (2/3 times the standard deviation) is given to identify any outliers, for hourly or daily water consumption respectively. As a result average water consumption for the whole period of 2012-2015 is 168 L/day/person, with a maximum and minimum per capita of 251 L/day and 112 L/day respectively. The average, minimum and maximum of the individual years is shown below:

Table 7: Average, minimum and maximum water consumption for Lönashult, Alvesta municipality (2012-2015)

Year	Water Consumption L/d/p		
	Minimum	**Average**	**Maximum**
2012	137	172	245
2013	132	175	251
2014	122	166	228
2015	112	159	209

It should be noted that the maximum or minimum water consumption values could also be due to leaks or errors still existent in the data. The method used to identify peaks in the data, has only the capacity to single out outliers that are a certain threshold above the standard deviation, however if there is a gradual increase in the values up to an outlier, this will not be identified as the change between the values is relatively small.

As illustrated in Figure 15 the consumption reaches above 200 L/day/person usually during the summer periods, however for the year 2015 the water consumption is relatively low and only contains one peak above 200 L/d/person, however it this could also be an error in the data.

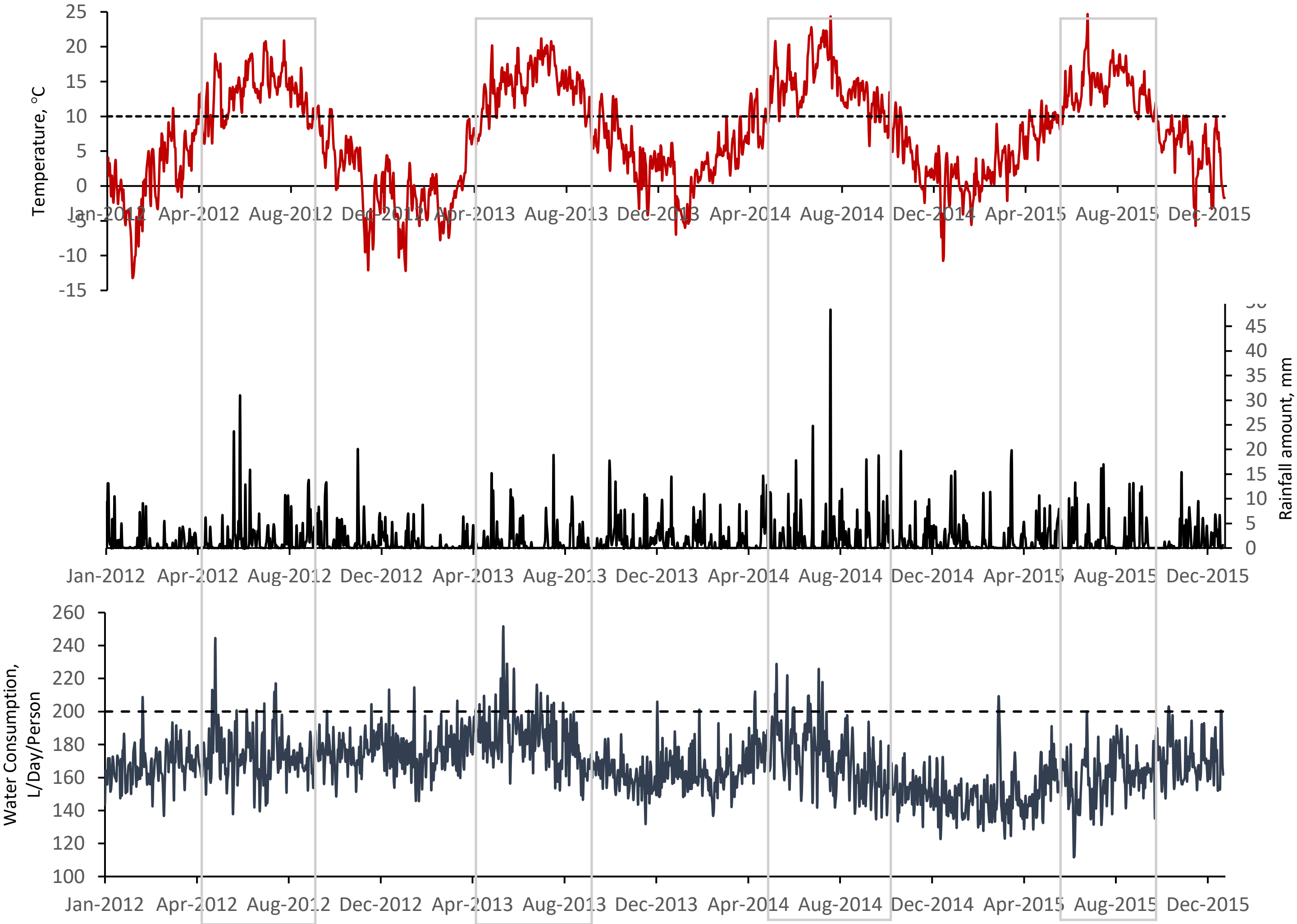


Figure 15: Climate and total daily water consumption for Lönashult, Alvesta municipality over the period from 2012-2015

The year 2014 was investigated further for variance in water consumption and temperature, where the summer months (July and August) were compared to the winter months of (December, January) for any changes in the adjusted R^2. Table 8 below shows the adjusted R squared values for all months:

Table 8: Adjusted R squared values for each month in the year 2014. Lönashult, Alvesta municipality

Month	Adjusted R^2
Jan	0.003
Feb	0.053
March	0.255
April	0.343
May	0.214
June	0.383
July	0.346
August	0.201
September	0.259
October	0.048
November	0.000
December	0.015

From the adjusted R^2 values for hourly water consumption data for each monthly period, the months from March to September are most significant in regards to the link between temperature and water consumption. This also gives good reason as to why a temperature threshold is required, as the correlation between temperature and water consumption can be analysed for the months that are weighted with higher significance, where for the significant months from April to September, the temperature is generally higher than 10 ºC.

When performing a correlation test to see the significance of the temperature in relation to the water consumption, correlations were made for all temperatures across the time period, as well as water consumption at temperature above 10 °C, 12°C and 20°C. The significance of the water consumption observed above 20°C is negligible, where the p-value equaled to 0.938, which is higher than the significance limit of 0.01 (1%). This is possibly due to the lack of data points at temperatures higher than 20°C, as the daily temperatures are taken as an average, which does not represent the maximum temperatures occurring. At temperatures above 15 °C, with every 2 °C increase in temperature the water consumption increased by 2.5 %. At temperatures above 12 °C , the water consumption increased by 5.5 % with every 2°C increase in temperature and when analyzing the complete time period for temperature over the given period, there is a 0.7% increase in water consumption with every 2°C rise in temperature. Table 9 illustrated below, represents the coefficient of determination (R^2), p-value of each of the correlations between different temperature limits.

Table 9: Percentage increase in water demand with a 2°C increase in temperature for Lönashult, Alvesta municipality (2012-2015)

Temperature limit (°C)	p-value	R^2	% water consumption increase with 2°C rise
All	<0.0001	0.060	0.7 % (1 L/d/p)
>12°C	<0.0001	0.105	5.5 % (3.3 L/d/p)
>15°C	0.001	0.052	2.5 % (3.6 L/d/p)
>20°C	0.719	0.005	Not statistically signif.

The coefficient of determination is at a maximum when all temperatures are compared, which shows that this is the limit which explains the most amount of variance between temperature and water consumption. For the precipitation across the given time period, there is no particular trend in the data, where the water consumption is observed to increase or decrease with days that no rain and minimum amount of rain occurs. The pattern in daily precipitation amount does not show any significant correlation with water consumption, thus is assumed to not be an influential variable that will explain any of the temporal variance in the water measurement data.

To understand the difference in water consumption above and below a temperature threshold of 10 °C (this falls in the summer between May-September), the individual yearly periods from 2012-2015 were looked into in Table 10. Figure 15 illustrated the average monthly water consumption for the yearly periods, where 2012 and 2015 both show an increase across the whole year, whereas 2013 and 2014 show an increase during the summer months (May-September). This explains why the consumption pattern for 2012 and 2015 does not change so significantly, with only a 2-3% increase for the average water consumption over 10 °C, whereas for 2013 and 2014 where the water consumption is more closely correlated with the rise in temperature where a 7% increase is observed for a temperature threshold above 10 °C.

Table 10: Average water consumption above and below 10°C for Lönashult, Alvesta municipality (2012-2015)

Year	Water consumption (L/day/person)		% increase
	Average below 10 °C	Average above 10 °C	
2012	125	128	3
2013	126	135	7
2014	118	127	7
2015	118	115	2

5.3 Multiple linear regression analysis

5.3.1 Hourly water consumption

Before inserting all the independent variables for multiple regression analysis, a simple linear regression analysis was applied to test the strength of the linear relationship of the sinusoidal fitted variable, and also temperature, as shown in Table 11 below.

Table 11: Adjusted R squared comparing the hourly consumption patterns and the sinusoidal fitted variable and temperature

Location		R^2_{adj}	
		Sine	Temp
Gothenburg (2014)	Amhult	49	<5
	Södra Skärgården	50	**26**
	Torshällsvägen	55	<10
	Tygshusvägen	58	<5
	Änghagsvägen	60	<5
	Burmans Gata	60	<5
	Vattugatan	66	<5
	Slottskogen	72	<5
	Näset	71	**14**
	Blomstigen	75	<5
Lönashult, Alvesta municipality	2012	70	4
	2013	70	6
	2014	69	7
	2015	71	-

When observing the correlation between the sinusoidal fitted variable and water consumption, the adjusted R squared value ranges from roughly 50 % to 75 %, which is considered reasonable good as a deterministic variable for the trend observed in water consumption. Generally temperature is not so significant by itself to predict the outcome of water consumption, however for Södra Skärgården and Näset the adjusted R squared values are higher, from 14 to 26 % variance in water consumption explained by temperature. Both these areas as mentioned previously have high water consumption patterns during the summer, as they tend to have more summer homes located in the area. After this analysis, the multiple regression analysis is perform where the following independent variables were inputs into the model:

I. Temperature
II. Double sinusoidal equation
III. Dummy variables
- Night/day (night is 12am-6am assigned value 0)
- Weekend (Sat-Sun)
- Vacation days (based on school holidays)
- Day of the week
- Hour of the day

Both a best model and stepwise approach were applied, where for the best model, all input variables are initially used in the regression equation, but through a process of elimination depending on the p-value of the individual variables, variables that are considered to not be statistically significant are removed. In Table 12 below, the results of the adjusted R^2 for the hourly consumption of water are illustrated, for both the Gothenburg and Alvesta municipalities.

Table 12: Results from best model and stepwise regression analysis, for the hourly consumption data

Area		Best model		Stepwise	
		R^2_{adj}	*Additional variables*	R^2_{adj}	*Variables*
Gothenburg (60-80%)	**Amhult**	57	Sine, temp, weekend	55	Night, hourly
	Torshällsvägen	66	All*	60	Night, hourly
	Tygshusvägen	73	Weekday	73	Night, hourly, holiday, sine
	Änghagsvägen	67	Sine, temp, weekend	65	Night, hourly
	Burmans Gata	71	Temp, weekday, weekend, hour	65	Night, sine
	Slottskogen	78	Temp, weekday, weekend, hour, holiday	74	Night, sine
	Vattugatan	73	All	66	Night, sine
	Södra Skärgården	77	All	57	Night, holiday
	Näset	80	Temp, weekday, hour, month	72	Night, sine
	Blomstigen	81	Sine, temp, weekend, night	80	Hourly, monthly
Alvesta, Lönashult	**2012**	87	All**	87	Sine, hour
	2013	85	All	84	Sine, hour
	2014	83	Weekend, night, temp	82	Sine, hour
	2015	86	Weekend, night, month	84	Sine, hour

***All includes Sine, temp, weekend, weekday, month, hour, night/day, and holiday*

*** Excluding holiday*

**** No complete temperature data available for this year*

When comparing the adjusted R^2 values in the best model approach, they are generally higher than the stepwise adjusted R^2 results. This is to be expected as more variables are used, and therefore a more complex model is produced. The variables that tend to remain in the regression model in the stepwise regression model are the sine, hourly and night/day variables. This signifies the strong periodicity present in the hourly consumption data. For the results in Lönashult the adjusted R^2 is significantly high at 82-87%, where the results from the stepwise regression analysis are only 1-2% lower, and use just two variables in the regression equation. This reveals that even though there are variables that are statistically significant and are therefore used in the best model approach, they do not improve the model so significantly. The hourly water consumption data for Lönashult has a more constant pattern when compared to that of Gothenburg, which is illustrated in Figure 16 below. In Figure

16, the blue shaded pattern is what has been simulated by the regression model given the input independent variables, and the red shaded pattern indicates the actual hourly water consumption data for Lönashult, Alvest municipality, here shown for the year 2015. It is clear to see that the main bulk of the water consumption pattern has been identified by the model, however there are some variations in the maximum consumption values in the actual data that has not been simulated by the regression model.

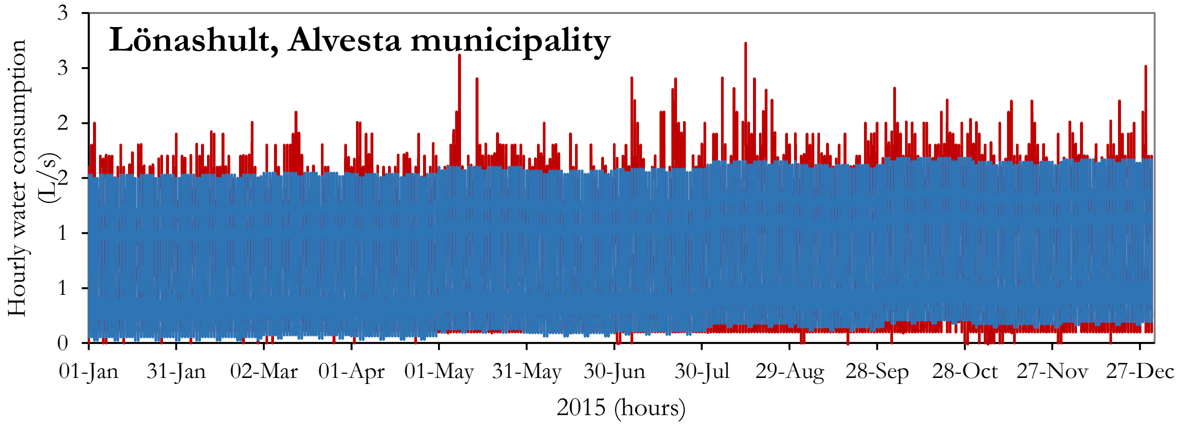


Figure 16: Hourly water consumption, multiple regression analysis for Lönashult, Alvesta municipality

In Gothenburg, there tends to be more variation in the summer season where for areas such as Södra Skärgården that contain summer homes and is situated on the Gothenburg Archipelago, the water consumption increases quite significantly during the summer period, as shown below in Figure 17. Again the main bulk of the data is identified, however there is still some variations in the higher values observed in the actual data. What is clear for the results of Södra Skärgården is that there is a significant amount of leakage or unknown water consumption, where the minimal consumption starts at around 5 L/day/person. This is likely to be a result of the relatively large population size, where both Näset and Slottskogen also have a shifted minimum consumption, see Appendix C.

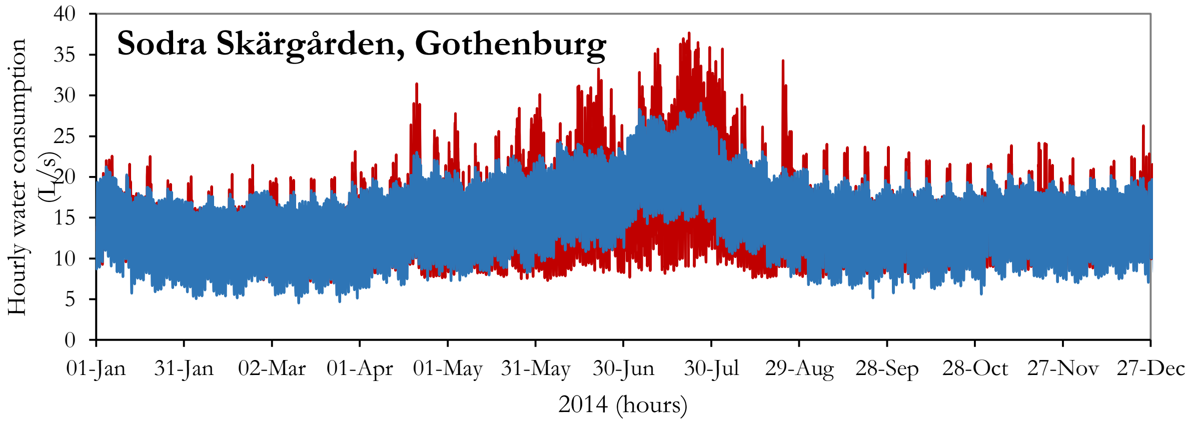


Figure 17: Hourly water consumption, multiple regression analysis for Slottskogen, Gothenburg

5.3.2 Daily water consumption

Table 13 below are the results for both the best model and stepwise regression methods, for the daily water consumption series for Gothenburg, Kalix and Alvesta municipality.

Table 13: Results from best model and stepwise regression analysis for the daily water consumption data

Areas		Best model		Stepwise regression	
		R^2_{adj}	*Additional variables*	R^2_{adj}	*Variables*
Gothenburg	**Änghagsvägen**	68	Temp, weekend, month	58	Sunlight, day
	Amhult	67	Temp, sunlight, weekend	59	Month, day
	Blomstigen	-	***	61	Sunlight, temp, snow depth, weekend, month
	Slottskogen	-	Weekend	65	Holiday, month, day
	Vattugatan	70	Holiday	66	Weekend, month
	Näset	73	Temp, snow depth	68	Sunlight, month
	Torshällsvägen	-	Temp, holiday, day, month	72	Sunlight, weekend
	Burmans Gata	-	Temp, weekend	72	Sunlight, month, day
	Tygshusvägen	76	Temp	75	Holiday, month
	Södra Skärgården	83	Temp, weekend, holiday	75	Sunlight, month
Kalix	**Björkfors**	-		47	Sunlight, month
	Tandfors	-		63	Sunlight, month
	**Nyborg ** **	-	Sunlight	66	Holiday, day, month
Lönashult	**2012**	-		22	Temp, month, day
	2013	42	Temp	40	Sunlight, month
	2014-2015	-		42	Sunlight, month, weekend

** All variables include temp, sunlight, precipitation, snow depth, weekend, weekday, month, holiday*
***Nyborg- Sundays show highest positive coefficient in the regression equation. See Appendix B*
**** - signifies that this is the same result as he stepwise regression method*

From Table 13 above it is evident that the results for adjusted R^2 values are higher, as previously mentioned in the previous section on hourly water consumption results. The range of R^2 values are higher for Gothenburg, when compared to both the hourly results and to Kalix and Lönashult, with 68-83 % of the variance in water consumption explained by the independent variables chosen. This could be relatively high as firstly the data for Kalix is a smaller dataset and more prone to errors, and also has a more sporadic pattern in water consumption, whereas Lönashult, seeing as it had very high correlation with variables such as the sinusoidal fitted variable, it has a more constant pattern with higher periodicity, and therefore won't be as strongly correlated to variables such as holiday season or temperature. The results for Björkfors is relatively low with an adjusted R^2 value of 47%. This could be explained by the large spike in the water consumption data, which is likely to be an error rather than the actual consumption of water. This tends to lower the predictability of the independent variables, and thus lowering the R^2 value.

In the following example, the results for Vattugatan will be displayed, which is located in Gothenburg. The R^2 value is 66%, which is a good result and shows that the explanatory variables prove to be good predictors of the trend in water consumption. The R^2 value could however have the potential to be increased if spikes did not appear in the data, where in Figure 18 below it is evident that there are some outliers contained in the water consumption pattern that may not be caused by the actual human water consumption:

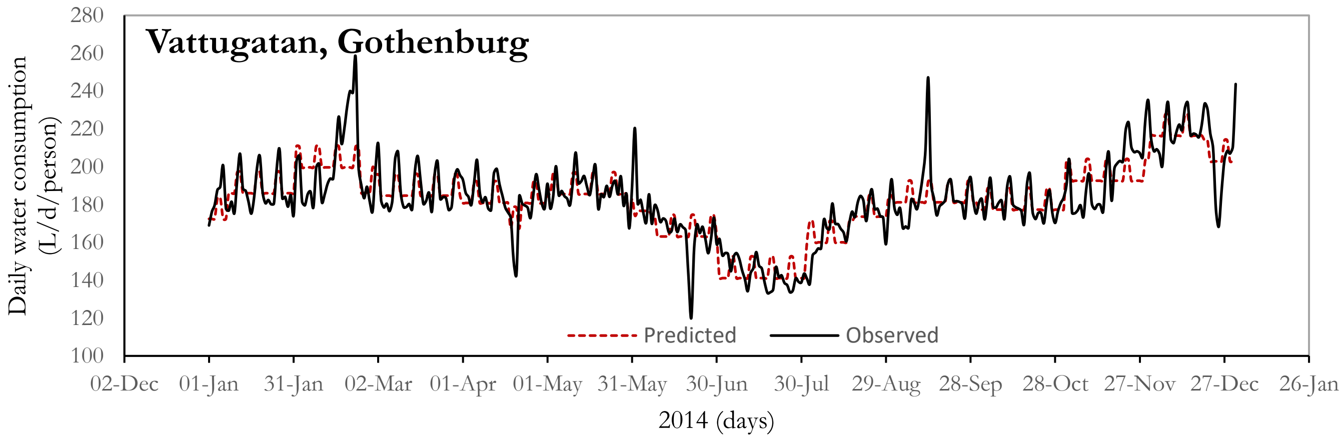


Figure 18: Daily water consumption for Vattugatan. The results from multiple regression analysis are displayed with the observed values

In Gothenburg, there tends to be more variation in the summer season, where areas such as Tyghusvägen, which contain majority student housing and tends to see a decrease in the water consumption during the summer and the Christmas period as there will be a fluctuation of people out of this area during the vacation period. Other areas that have similar patterns to Tyghusvägen are Vattugatan and Slottskogen, which are located in the inner city of Gothenburg and contain apartments, general stores and schools and tend to decrease in water consumption during the summer, where the graphs for the multiple linear regression results for the hourly water consumption can be found in Appendix C. The results from the multiple regression analysis for the daily consumption pattern for Tyghusvägen in shown below in Figure 19:

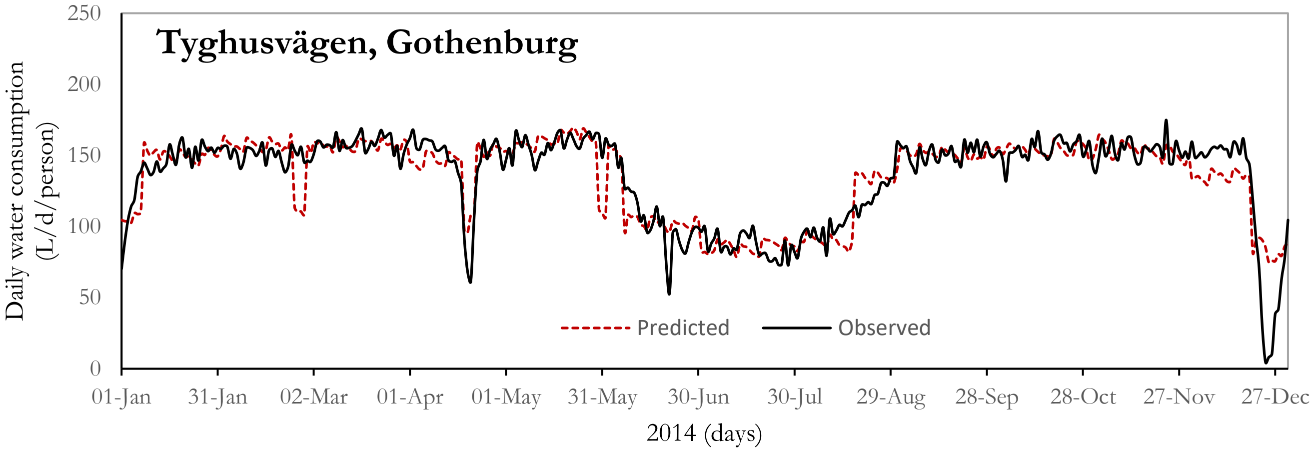


Figure 19: Results from the multiple regression analysis for the daily water consumption pattern for Tyghusvägen, Gothenburg

Below in Figure 20 the results from the multiple regression analysis model for Nyborg, Kalix is displayed. There is an evident rise in the water consumption during the summer, however the water consumption tends to stay higher for a longer period, and even toward the end of September the water consumption in relatively high. Unlike Björkfors and Tandfors, Nyborg is relatively large (population of 1250) and contains public services such as schools, so has a stronger correlation with variables such as the holiday season, and a more define weekly pattern in consumption.

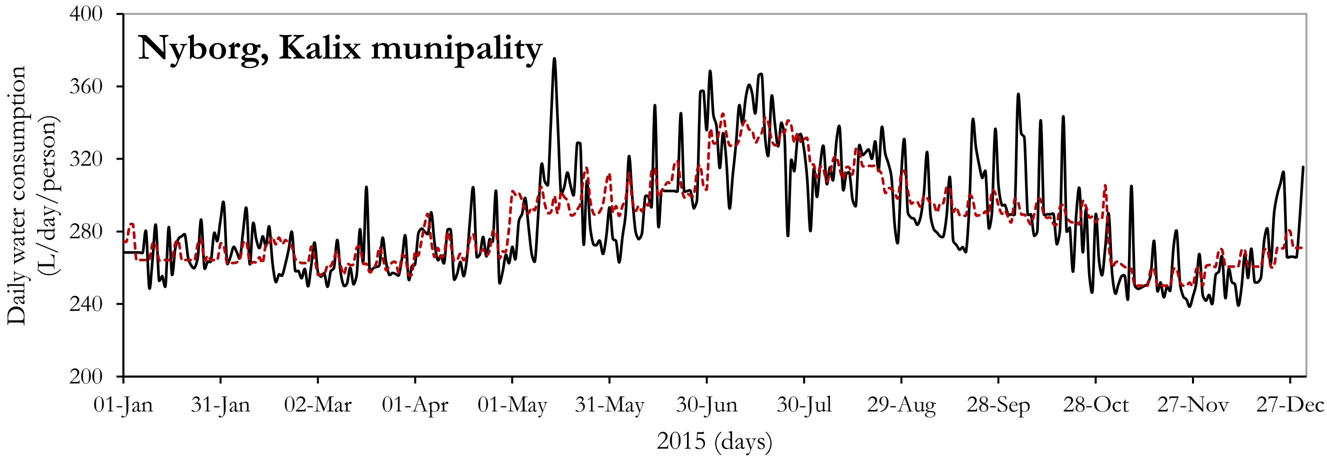


Figure 20: Results for multiple regression analysis for Nyborg, Kalix municipality (2015)

8 Discussions & Recommendations

It is essential for this research into water consumption to validate the data before beginning with the analysis, as errors and uncertainties in the data such as huge spikes in the water consumption that do not match the actual water demand, such as the one observed in Björkfors can lead to misleading values of R^2 that tend to be lower than expected. With this said, the process of validating the data can in itself lead to errors, so validation and modification methods have to be carefully selected, so that they obeyed by the nature of the data. To identify whether there are leaks in the data, it would be possible to do a mass balance analysis, to investigate whether the total water supplied by the municipality is equal to what is observed in the water consumption time series.

The sinusoidal equation should be adjusted to fit weekend periods and also fitted for the summer months in addition to the 10 days in January, which is assumed to be the month with a minimum consumption of water and therefore represents the base values of the data, that does not differ to seasonality or other factors inducing a variation in water demand from the regular cyclic pattern.

For further analysis, different approaches to regression analysis should also be investigated, such as non-parametric regression analysis that deals with non-linear data. For the temperature correlations is was observed as well that some trends fitted better with a polynomial trend line, which should be investigated further, as when analysis the correlation between temperature and water consumption, below a certain temperature the water consumption will not continue to decrease, but rather stay constant or even increase again.

Other factors that could be applied to the model include socio economic factors, where more detailed findings on individual houses, i.e. income, number of persons per household, area of outdoor space could result I further explanation to the trends observed in the temporal variation of water consumption. A more specific and detailed description of the areas and what they contain, in terms of percentage of each type of building, in order to get a more exact reason for why the water consumption differs depending on what type of area it is.

Finally for further improvements more datasets should be investigated, in order to verify the statements made in this thesis work, such as the distinctive hourly pattern observed for different types of areas in Gothenburg. Longer time periods are also necessary to see if there are any changes in the consumption patterns over the annual periods, as observed in Lönashult where even though the water consumption stays relatively constant, different patterns emerge for different years.

9 Conclusions

Firstly, to discuss the validation and modification of the data from Lönashult, Alvesta municipality using the 1st derivative approach, and assigning a threshold limit determined by the standard deviation of the data. This method is a good method for identifying single outliers in the data, however, it is quite common for an accumulative increase to occur in the water consumption, resulting in more than one value that has shifted from the standard water consumption pattern either due to a leak or other unknown causes. These types of errors or uncertainties in the data tend to go unnoticed by the 1st derivative analysis, as the change between the values is too small to be identified, however overall it may be an incorrect measurement of the water consumption. With this said, it is unlikely to always find each and every error contained in the datasets, and it is important not to remove any outliers or 'errors' just to smooth out the data and allow it to fit better with the regression model, as these assumed errors could also be a part of the actual consumption of water.

The results from spectral analysis agreed with what was expected from the data, that there would be a dominant cycle in the 12 and 24 hour periods, as well as the 7 day period for the daily water consumption data. As for the sinusoidal fitted equation, this was a good variable to use in predicting the outcome of the hourly consumption of water. The double sine equation by itself explained 50-75% of the variance observed in the water consumption data, which shows a strong linear relationship between the periodical variable that is fitted with the sinusoidal equation and the dependent variable, water consumption. This leaves 25-50% unexplained variance, which could be due to seasonality or unexplained variations in the data. Areas with higher seasonality in the data tend to have lower strength in the linear relationship between the fitted sinusoidal equation and water consumption, because the DSE is constant through time and is fitted for 10 days at the beginning of January, thus it is only formulated for the lowest base water consumption pattern.

In order to understand the different types of hourly consumption patterns that exist, the average across the year(s) was investigated, for the areas within Gothenburg and Lönashult, Alvesta municipality. From the results of the average hourly water consumption in Gothenburg, it is clear to see that the pattern differs depending on what type of area it is, for example into relatively large population sizes to small, and also for suburb areas of cities where people tend to commute. The hourly consumption pattern for Lönashult also follows the same pattern for relatively small population size, however what should be noted is that the highest peak is at 8pm and not in the morning period. Generally it can be concluded that there a two noticeable peaks within the day, one during morning/lunch time and another usually at 8pm in the evening, at dinner time. The minimal water consumption occur at night between the hours of 12-6am, where human consumption of water is at its lowest.

For the temperature correlations, generally speaking there is an increase in the water consumption with rise in temperature, excluding 3 areas located in Gothenburg. It is important to note that the type of area under investigation is vital for understanding the relationship between temperature and water consumption as other factors such as population fluctuation plays an important role. If the area is assumed to stay consistent in terms of number of water consumers during the summer season, then the water consumption tends to rise. However, if this area is an area where either there will be less people here during the summer season, of for example if shops, schools and other public services may be closed during parts or all of the summer period, then the water consumption will decrease regardless of the increase in temperature, and especially with areas they will not have any if all at any external use of water. For Gothenburg, the areas that contain villa homes as well as mixed use sectors, and also summer vacation areas have around 2-12 % increase in water demand with a 2 ºC rise in temperature, with Södra Skärgården at 12.3 %. Areas such as Vattugatan, Slottskogen and Tyghusvägen that contain apartments and student housing tend to decrease in the water demand by 2-7% with every 2 ºC rise in temperature, taken at a threshold of 12 ºC. The 3 areas selected in Kalix municipality have an increase of around 1-2 % in water demand with every 2 ºC rise in temperature, for the period of May to December. In Lönashult there is a 5.5% rise in water demand with every 2 ºC rise in temperature, set at a threshold of 12 ºC.

Snow depth and precipitation have relatively small or no correlation with water consumption, however in certain areas it may be falsely correlated if there is a leak in the period where the snow depth is high, as then it assumes this leak is directly related to snow depth, however it is more likely to be due to water inside the pipes expanding when it gets closer to the freezing point during the winter season,, which then leads to bursts or leaks in pipes when the pressure within the pipes becomes too high for the capacity of the pipe to contain, therefore rupturing.. The independent variable sunlight hours tends to be chosen over temperature when using stepwise regression analysis, and overall will have be more statistically significant then temperature when comparing the linear relationship with water consumption, which could be due to the fact that there are negative values in temperature, which will not correlate as well with water consumption. When investigating Lönashult in more detail, the R^2 tends to increase during the summer months in contrast with the winter months when compared to temperature. June has the highest R^2 value with 38%, for the annual period of 2014.

The hourly consumption data had significantly high adjusted R squared values of 58-87% for the best model results, which shows that the periodicity of the data is extremely important when analysis the trend or variance in hourly water consumption data. The results from Lönashult alone has an adjusted R squared range of 83-87%, as there is no significant change in the hourly water consumption pattern over the annual period. When compared to the results from the daily consumption pattern, the adjusted R^2 for the years investigated for Lönashult is much lower, at 22-45%. 2012 is particularly lower, but this is possibly due to there being no complete sunlight hour data available for this period, this this variable was excluded in the regression equation. Errors in the data can result in a significant decrease in the R^2 value, as there are significant spikes in the dataset that may not be due to the actual human water consumption pattern, but due to leaks, for example, which is evident for Björkfors, and this contains a large spike in the data and also has a lower adjusted R squared value of 47%. The adjusted R squared stays relatively the same for both the hourly and daily analysis, where the range of adjusted R squared is 58-81% and 68-83% for the hourly and daily data respectively. The adjusted is higher for the daily data, showing the significance of variables such as the holiday season or climate variables such as temperature or sunlight hours for predicting the outcome of water demand.

When choosing which regression analysis technique to use, the time constraints of the project comes into play, as the best model technique requires more time to choose and select appropriate variables, whereas in stepwise regression this stage of the procedure is applied in the model. The R^2 value tends to be higher in the best model approach, as more variables are used in the regression equation. The best model approach is recommended, when taking into consideration that only significantly significant variables are included and collinearity between variables is eliminated.

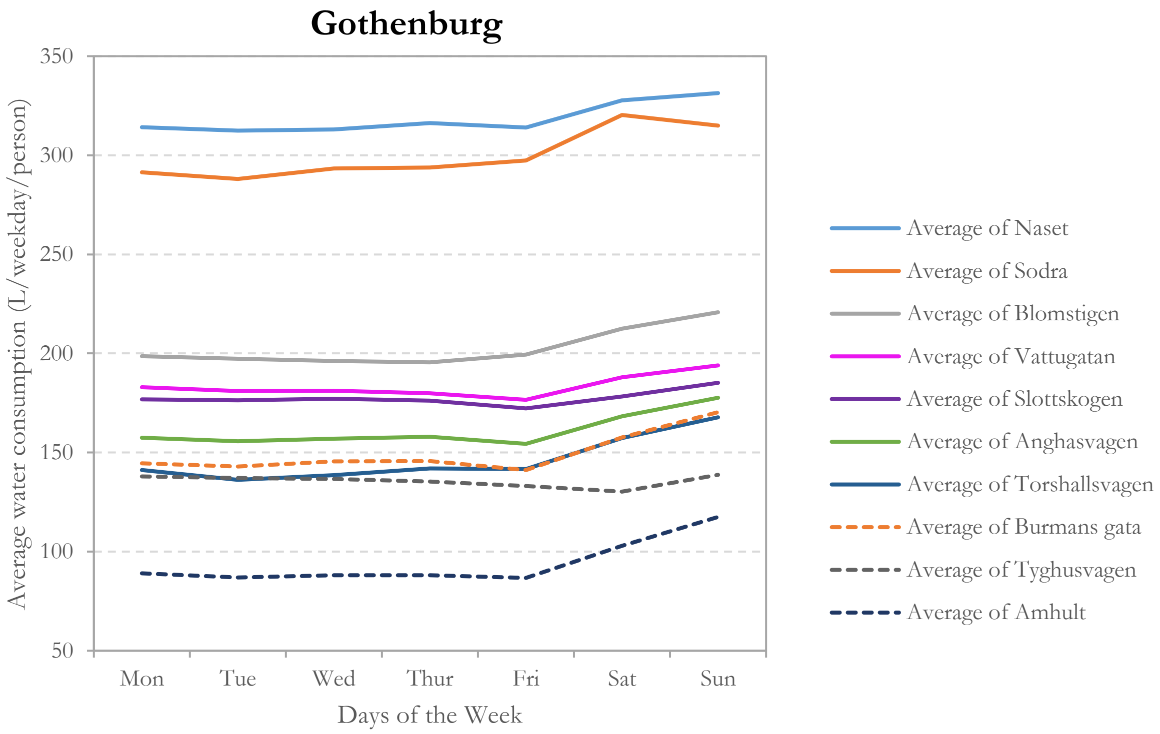
10 Appendices

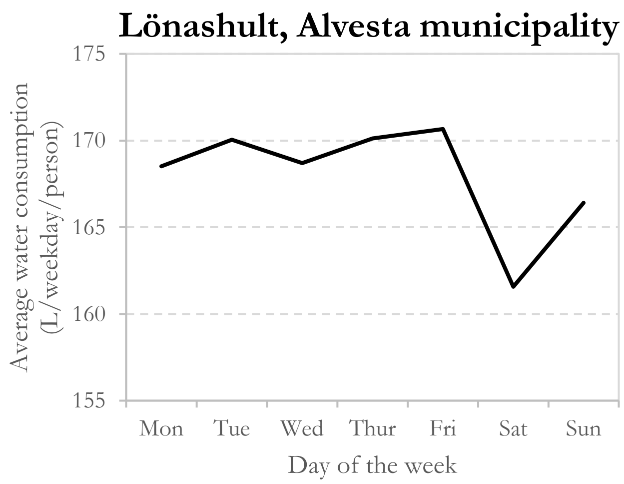
Appendix A Sinusoidal modelling

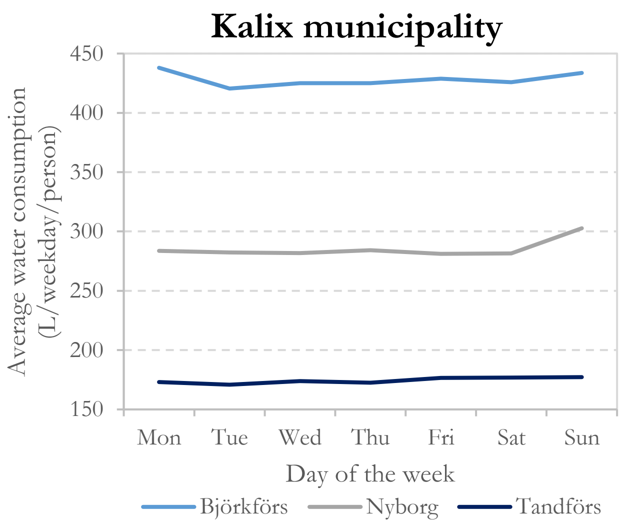
The double sinusoidal equation was fitted for the original units of the hourly water consumption data, where the Gothenburg data was given in L/s at hourly intervals, and for Lönashult, Alvesta municipality this was given in m3/hr every hour.

Location		$\bar{w}$	A_1	A_2	φ_1	φ_2
Gothenburg	Amhult	0.67	0.39	0.41	3.9	3.9
	Blomstigen	1.09	0.39	0.56	3.7	3.8
	Burmans Gata	2.30	1.2	0.8	3.9	3.7
	Näset	51	11	15.5	3.4	3.8
	Södra Skärgården	15.6	3.1	4.9	3.3	4
	Slottskogen	37	13	13.8	2	3.3
	Torshallsvägen	3.20	1.30	1	3.7	3.9
	Tyghusvägen	0.91	0.50	0.52	3.6	3.7
	Vattugatan	3.60	1.4	1.5	3.5	4.1
	Änghagsvägen	6.9	2.4	2.5	3.5	4
Alvesta	Lönashult	0.9	0.33	0.57	0.2	1.4

Appendix B Weekly water consumption patterns

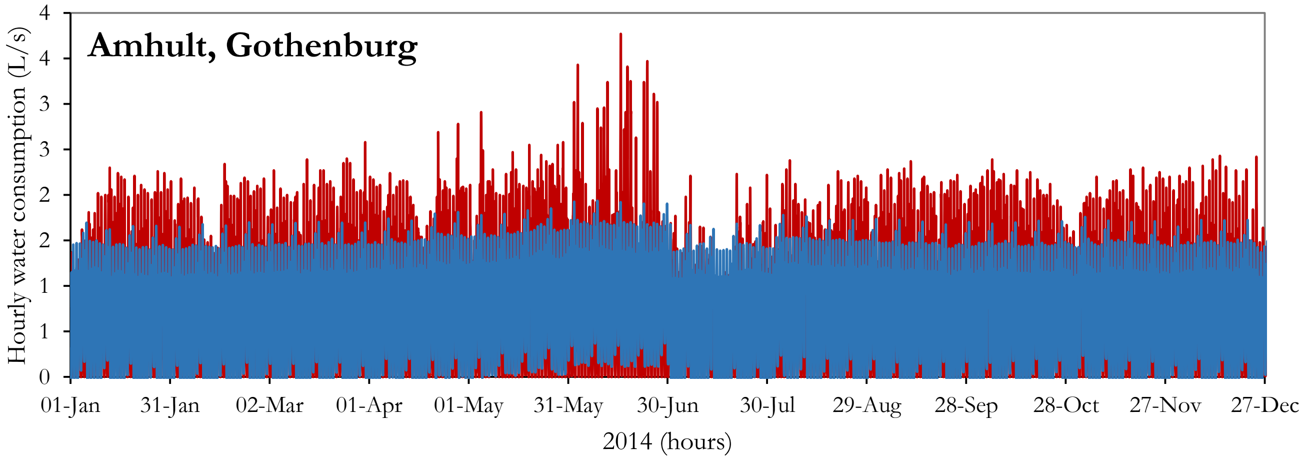


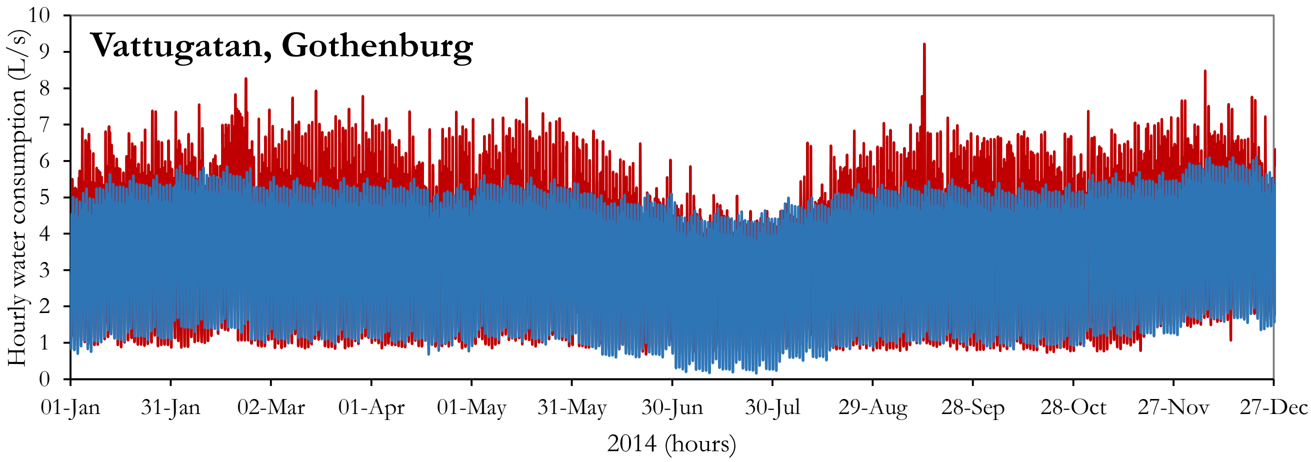


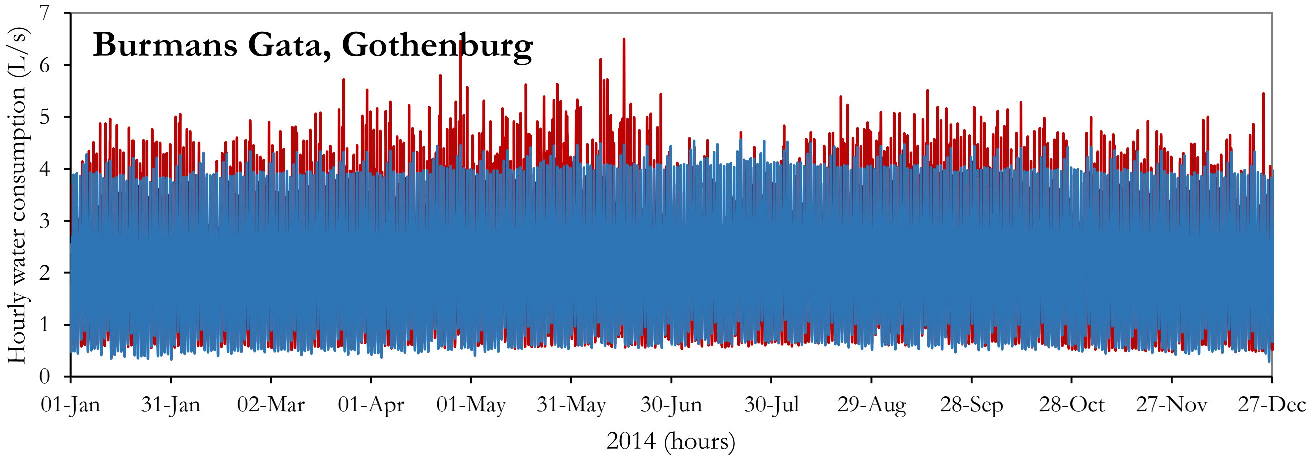


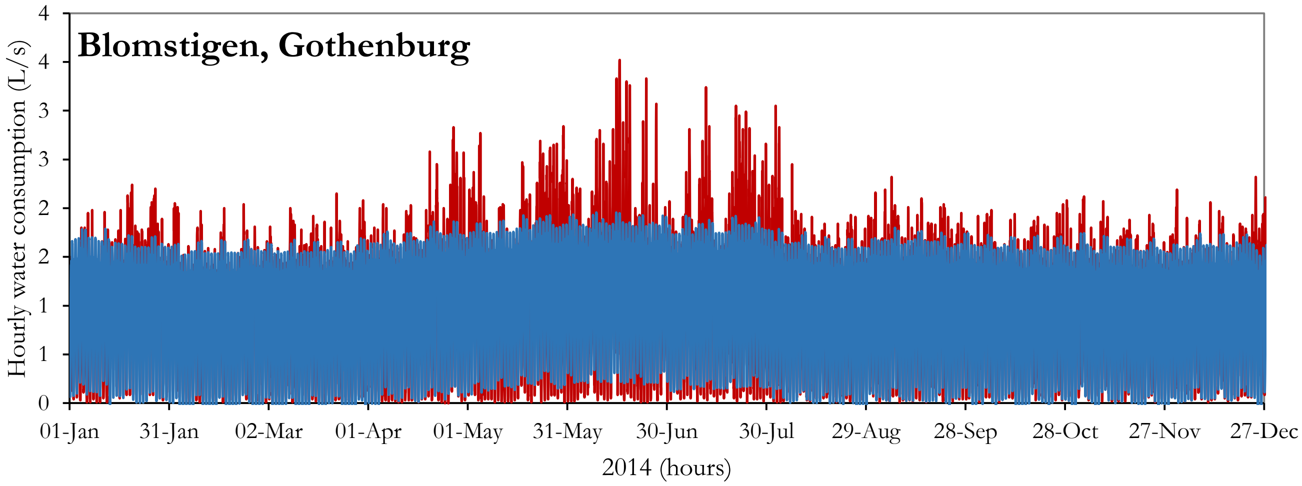
Appendix C Hourly water consumption data

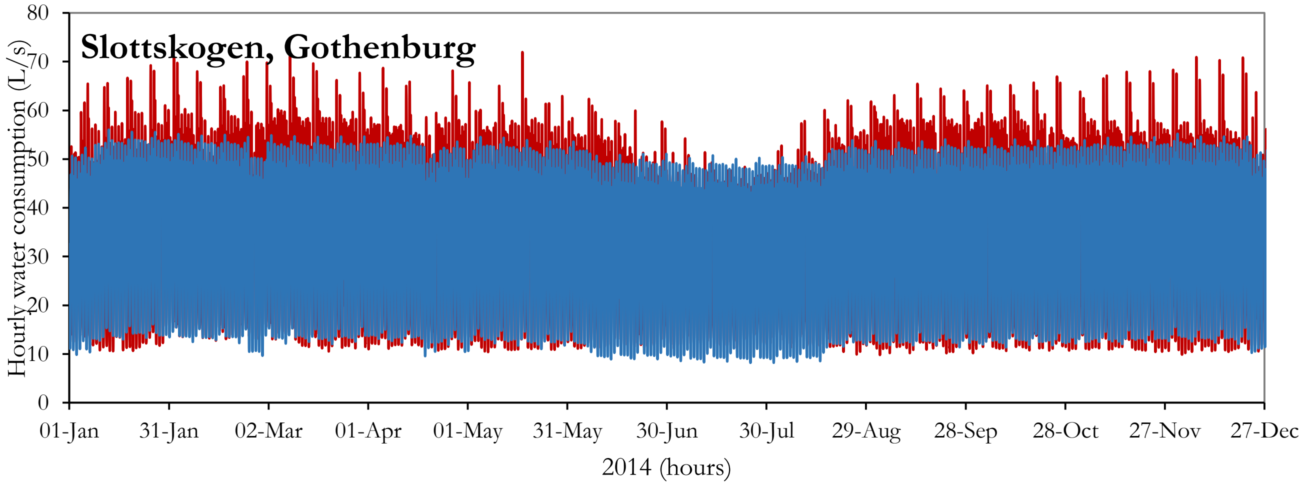
These are the results for multiple regression analysis for the best model approach, using only explanatory variables that have a significant linear relationship with the dependent variable, hourly water consumption. Six areas have been selected from Gothenburg, in order to illustrate different types of consumption patterns that exist. The red hourly water consumption represent the observed water consumption, where the blue represent the prediction from the regression model.

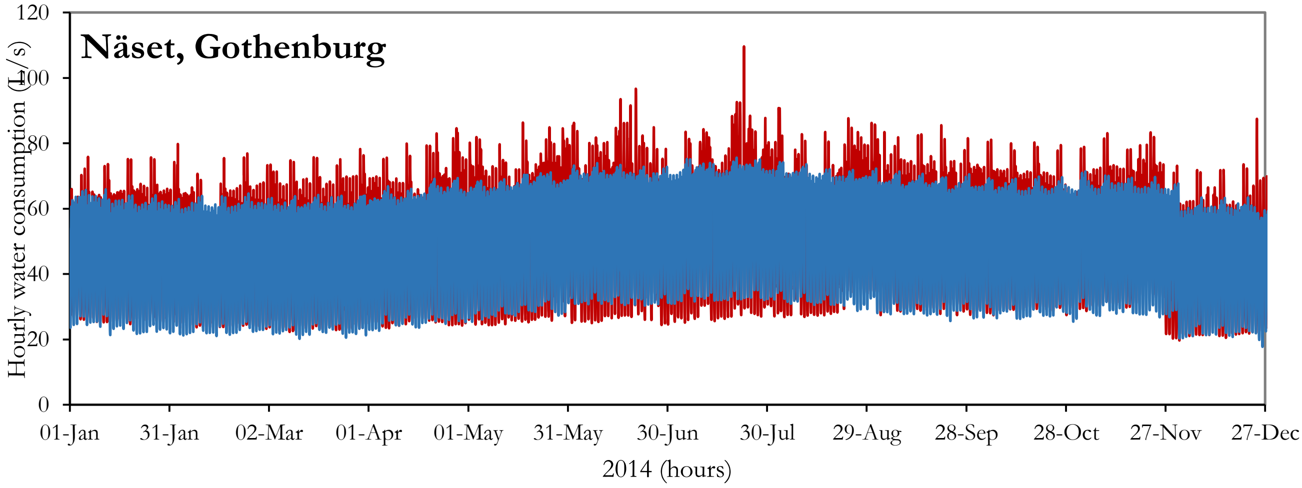




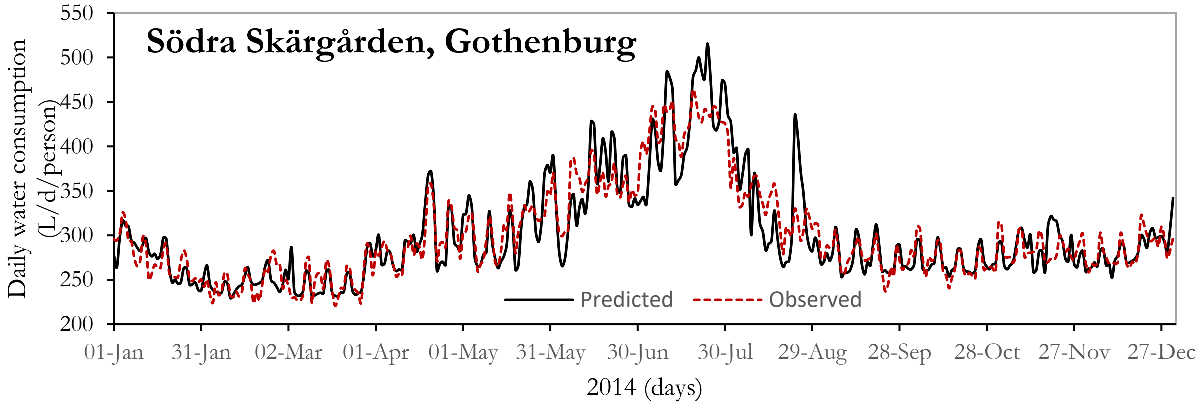


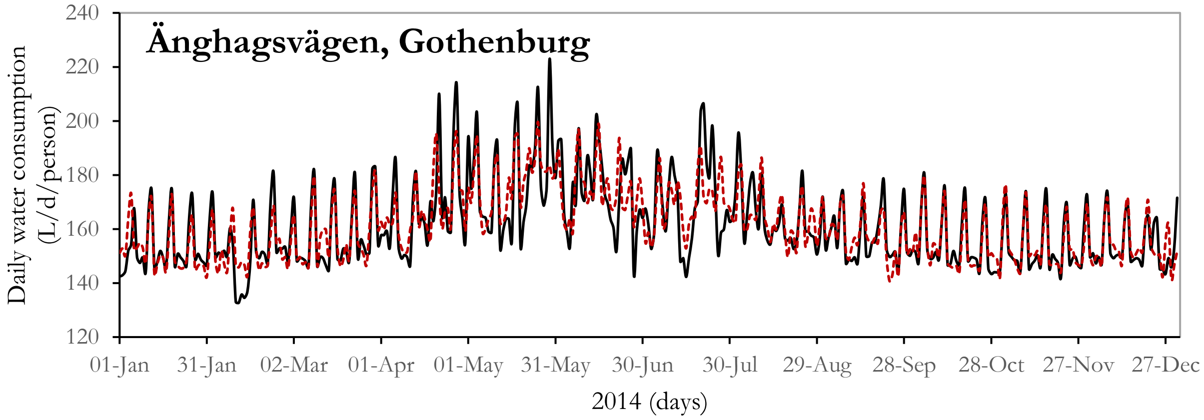

Blomstigen, Gothenburg
Hourly water consumption (L/s)
2014 (hours)
01-Jan
31-Jan
02-Mar
01-Apr
01-May
31-May
30-Jun
30-Jul
29-Aug
28-Sep
28-Oct
27-Nov
27-Dec

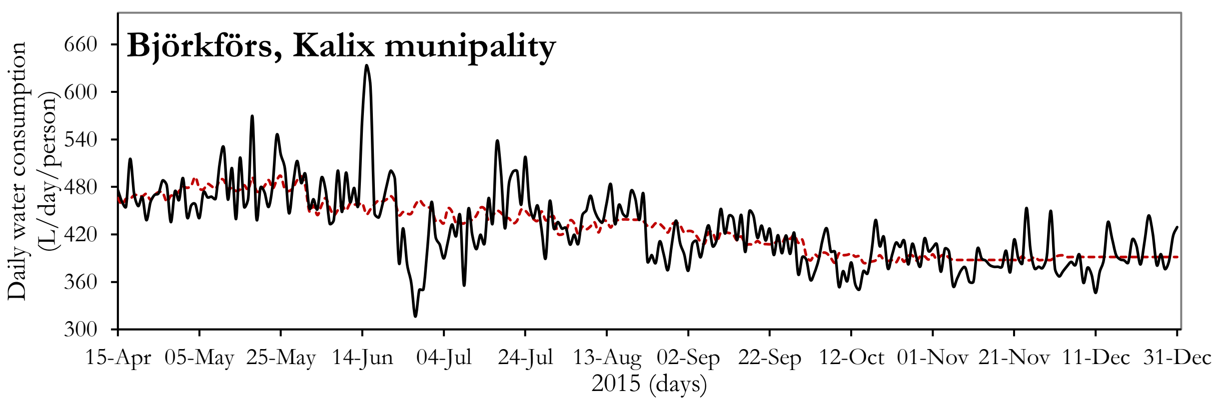

Slottskogen, Gothenburg
Hourly water consumption (L/s)
2014 (hours)
01-Jan
31-Jan
02-Mar
01-Apr
01-May
31-May
30-Jun
30-Jul
29-Aug
28-Sep
28-Oct
27-Nov
27-Dec

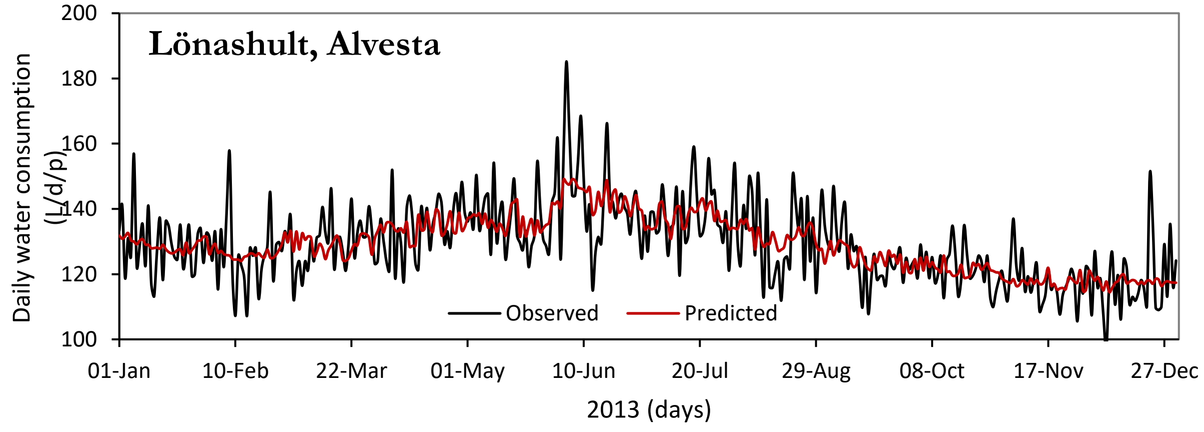

Näset, Gothenburg
Hourly water consumption (L/s)
2014 (hours)
01-Jan
31-Jan
02-Mar
01-Apr
01-May
31-May
30-Jun
30-Jul
29-Aug
28-Sep
28-Oct
27-Nov
27-Dec

Appendix D Daily water consumption data









References

Adamowski, J. et al., 2012. Comparison of multiple linear and nonlinear regression, autoregressive intgrated moving average, artifical neural netwrok, and wavelet artifical neural netwrok methods for urban water demand forecasting in Montreal, Canada. *Water Resources Research* , 48(1).

Adamowski, J. & Kaz, A., 2013. A Spectral Analysis Based Methodology to Detect Climatological Influences on Daily Urban Water Demand. *Internation Association for Mathematical Geosciences* , Volume 45, pp. 49-68.

Anon., 1998. Effectiveness of Conservation Policies on New Mexico Residential Water Demand. *New Mexico Journal of Science* , Volume 38, pp. 104-126.

Babel, M. S., Maporn, N. & Shinde, V. R., 2014. Incorporating Future Climate and Socioeconomic Variables in Water Demand Forecasting: A Case Study in Bangkok. *Water Resource Management,* 28(7), pp. 2049-2062.

Balling Jr, R. C., Gober, P. & Jones, N., 2008. Sensitivity of residential water consumption to variations in climate: An intraurban analysis of Phoenix, Arizona. *Water Resource Research,* Volume 44, pp. 1-11.

BMC Public Health , 2012. *Change of water consumption and its potential influential factors in Shanghai: a cross-sectional study.* [Online]
Available at: http://www.ncbi.nlm.nih.gov/pubmed/22708830#

Chang, H. & Praskievicz, S., 2014. Sensititivity of Urban Water Consumption to Weather and Climate Variability at Multiple Temporal Scales: The Case of Portland, Oregon. *Internation Journal of Geospatial and Environmental Research,* Issue 1.

European Commission, 2015. *Science for Environmental Policy.* [Online]
Available at:
http://ec.europa.eu/environment/integration/research/newsalert/pdf/sustainable_management_of_water_in_an_arid_region_413na5_en.pdf
[Accessed 8 May 2015].

European Environment Agency, 2003. *Water use efficiency (in cities); leakage,* s.l.: EEA.

Gober, P. & Balling Jr, R. C., 2007. Climate Variability and Residential Water Use in the City of Phoenix, Arizona. *Journal of Applied Meterology and Climatology* , Volume 46, pp. 1130-1137.

Gutzler, D. S. & Nims, J. S., 2005. Interannual Variability of Water Demand and Summer Climate New Mexico. *Journal of Applied Meteorology,* Volume 44, pp. 1777-1787.

Haque, M. M., Egodawatta, P., Rahman, A. & Goonetilleke, A., 2015. Assessing the significance of climate and community factors on urban water demand. *Internation Journal of Sustainable Built Environment,* 4(2), pp. 222-230.

Higgins, J., 2005. *Chapter 4: Introduction to Multiple Regression.* [Online]
Available at: http://www.biddle.com/documents/bcg_comp_chapter4.pdf

Jorge, C., Vieira, P., Rebelo, M. & Covas, D., 2015. Assessment of water use efficicieny in the household using cluster analysis. *Procedia Engineering,* Volume 119, pp. 820-827.

Lage, P. et al., 2012. *Mathematical modelling of domestic water flow,* Dublin: University College.
Lane, D. M., 2007. *Type I and Type II Errors.* [Online]
Available at: http://onlinestatbook.com/2/logic_of_hypothesis_testing/errors.html

Mathematica, 2015. *How to remove outliers for the data.* [Online]
Available at: http://mathematica.stackexchange.com/questions/81121/how-to-remove-outliers-from-data

Michelson, A. M., McGuckin, J. T. & Stumpf, D., 1999. Non price water conservation programs as a demand management tool. *The American Water Resources Association (JAWRA),* 35(3), pp. 593-602.
Ostertagova, E., 2012. Modelling using polynomial regression. *Procedia Engineering,* Volume 48, pp. 500-506.

Praskievicz, S. & Chang, H., 2009. Identifying he Relationship Between Urban Water Consumption and Weather Variables in Seoul, Korea. *Physical Geography,* 30(4).

Rust, H., 2007. *Spectral Analysis of Stochastic Processes,* Potsdam: Nonlineat Dynamics Group, University of Potsdam.

Sarker, R. C., Trinidad, S. G.-. & Imteaz, M., 2013. *20th International Congress on Modelling and Simulation.* [Online]
Available at: http://www.mssanz.org.au/modsim2013/L7/sarker.pdf

Sen, A. & Muni, S., 1990. Multiple Regression. In: *Regression Analysis: Theory, Methods, and Applications.* New York: Springer, pp. 28-49.

SMHI, 2015. *Climate indicators- temperature.* [Online]
Available at: http://www.smhi.se/en/climate/climate-indicators/climate-indicators-temperature-1.91472
[Accessed 09 05 2016].

Souza, A. d. et al., 2015. Climatic Variations and Consumption of Urban Water. *Scientific Research Publishing,* Volume 5, pp. 292-301.

Statistics Solutions, 2016. *Statistics Solutions: Advancement Through Clarity.* [Online]
Available at: http://www.statisticssolutions.com/assumptions-of-multiple-linear-regression/

Statistics Sweden, 2010. *Water by by sectors, 2010.* [Online]
Available at: http://www.scb.se/en_/Finding-statistics/Statistics-by-subject-area/Environment/Water-use/Water-withdrawal-and-water-use-in-Sweden/Aktuell-Pong/2010A01/Water-use-by-sectors-2010/

Svenskt Vatten, 2016. *Svenskt Vatten- Vattentjänster.* [Online]
Available at: http://www.svensktvatten.se/vattentjanster/dricksvatten/ravatten/

SWWA, 2014. *A Vision for Water,* Stockholm: Swedish Water and Wastewater Association.

Sydvatten, 2014. *Sydvatten.* [Online]
Available at: http://sydvatten.se/virtuellt-vatten/

Sydvatten, 2015. *Vattenförbrukning.* [Online]
Available at: http://sydvatten.se/vattenforbrukning/

The Pennsylvania State University , 2016. *12.1: Estimating the Spectral Density.* [Online]
Available at: https://onlinecourses.science.psu.edu/stat510/node/80

Uriel, E., 2013. Multiple linear regression> estimation and properties. In: Valencia: University of Valencia.

World Health Organization, 2004. Design and operation of distribution networks. In: R. Ainsworth, ed. *Safe Piped Wter:Managing Microbial Water Quality in Piped Distribution Systems.* London, UK: IWA Publishing, pp. 38-68.