Water Demand Forecasting Using Machine Learning on Weather and Smart Metering Data

Submitted by Maria Xenochristou to the University of Exeter

as a thesis for the degree of

Doctor of Philosophy in Water Informatics Engineering

in September 2019

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature: …………………………………………………………..

Abstract

Water scarcity is a global threat due to lifestyle and climate changes, pollution of water resources, as well as a rapidly growing population. The UK water industry's regulators demand plans from water companies to sustainably manage their water resources, reduce per capita consumption and leakage, and create projections for climate change scenarios. This work addresses critical problems of water demand by expanding the understanding of water use and developing improved forecasting methods.

As part of this effort, the influence of the weather is thoroughly investigated, using a disaggregated, big-data statistical analysis. Results show that the weather effect on water consumption is overall limited, non-linear, and variable over time and households.

Next, a short-term demand forecasting model is developed, based on Random Forests, that predicts household consumption using several socio-economic, customer and temporal characteristics. This model is of significant value due to its accuracy as well as accompanying methodology that allows the interpretation of results.

In order to further improve the forecasting accuracy achieved using Random Forests, a new modelling technique is developed. The new method that uses model stacking and bias correction, outperforms most other forecasting models, especially when past consumption data are not available, as well as for peak consumption days.

Finally, a water demand forecasting model based on Gradient Boosting Machines is trained at different levels of spatial aggregation, for different input configurations. Results show that the spatial scale has a strong influence on the best model predictors and the maximum forecasting accuracy that can be achieved.

The methodology developed here can be used as a guide for researchers, water utilities and network operators to identify the methods, data and models to produce accurate water demand forecasts, based on the characteristics and limitations of the problem.

Table of Contents

Acknowledgments

First and foremost I would like to thank my supervisor, Prof Zoran Kapelan for his continuous support, advice, motivation, faith, patience and assistance, every step along the way. Thank you for accepting me in this PhD program, for enabling me to complete it on time and for always listening and helping me achieve my goals. And of course, thank you for your continuous input in this work, which shaped this thesis.

Next, I would like to thank my industrial collaborator, Dr Chris Hutton, for providing the data to make this project possible, as well as for his guidance, ideas and inspiration during our meetings and his input in this work. Thank you also goes to Prof Jan Hofman for his helpful comments.

I would also like to express my gratitude to Sebastian Gnann and Josh Myrans, who helped improve this thesis. Sebastian, thank you for your extremely thoughtful comments, your time, effort and your ‘simple English’ approach. Josh, thank you for always answering my questions, for your wise words of wisdom and your helpful comments on this thesis. You both make academia the incredibly supportive, intellectually stimulating, friendly and kind place that I experienced it to be.

A big thank you to the Water Research Group at the University of Bristol, where I was hosted over the last 2 years. Even though a PhD can be a lonely path, my one has been quite the opposite, thanks to you. Many thanks to Prof Thorsten Wagener for his generosity, offering me to be part of this group that shaped my PhD experience for the very best. Special thank you to Elisa Bozzolan for being a great friend, colleague and supporter over these years.

Finally, thank you to my parents, for providing me with the education and opportunities in life to achieve my dreams and the upbringing to know it is possible.

List of Tables

List of Figures

Author's declaration

This thesis is presented as a collection of four journal papers that have been published or have been submitted for publication to academic journals. Each one of the papers forms one chapter of this thesis (see chapters 2-5). Modifications have been made to improve consistency throughout the thesis.

All papers have been written by the author and have benefitted from the comments of the co-authors. An explicit statement at the beginning of each chapter states the citation of the paper it corresponds to as well as the contributions of the author and co-authors of the paper.

Definitions

Property Characteristic:	An attribute of the properties in the dataset. It can refer to the garden size, rateable value, council tax band, or metering status of a property.
Customer Characteristic:	An attribute of the customers in the dataset. It can refer to the acorn group, occupancy rate, or consumer behaviour (variations in average monthly consumption).
Household Characteristic:	A property or customer characteristic.
Temporal Characteristic:	An attribute that relates to time. It can refer to the time of day, the type of day (working day or weekend/holiday), the month, or the season.
Weather Characteristics:	An attribute that relates to weather. It can refer to air temperature, soil temperature, humidity, sunshine duration, radiation, rainfall, or number of days without rain.
Variable:	A household, temporal, or weather characteristic can be used as a variable in the analysis. The terms variable and characteristic are often used interchangeably in the text.
Segment:	A homogenous group of consumption or households. All components of a segment share the same household and temporal characteristics (e.g. the same garden size).

Segmentation:	The process of creating consumption or household segments.
Segmentation Category:	A type of consumption or household that has a certain temporal or household characteristic. One segmentation category includes all consumption or household segments that share the same characteristic (e.g. the same garden size).

List of Abbreviations

Acorn	A Classification Of Residential Neighbourhoods
ALE	Accumulated Local Effects
ANN	Artificial Neural Network
BC	Bias Correction
DMA	District Metered Area
DNN	Deep Neural Network
GBM	Gradient Boosting Machine
GLM	Generalised Linear Model
ICE	Individual Conditional Expectations
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MIDAS	Met Office Integrated Data Archive System
MSE	Mean Square Error
Ofwat	Office of Water Services
PCC	Per Capita Consumption
PDP	Partial Dependence Plot
RF	Random Forest
UKWIR	UK Water Industry Research
XGBoost	Extreme Gradient Boosting
XRT	Extremely Randomised Trees

CHAPTER 1

INTRODUCTION

1.1. Motivation

Water is essential for the survival of humans, the preservation of the natural environment, the function of societies, as well as the operation of industry and agriculture. However, water is also a limited resource, threatened by environmental changes and societal reforms, urbanisation, population and business growth, as well as the pollution of water resources.

The UK water industry, privatised in 1989, aims to provide clean water to its customers for four distinguished uses: urban, power generation, industrial and agricultural (Butler and Memon, 2006). Urban water use accounts for the water that is provided to residents (residential demand), businesses (commercial demand) and other organisations within a community or urban area (Billings and Jones, 2008).

The major droughts of 1975/76, as well as the subsequent droughts in the 1990s saw the UK imposing water restrictions and highlighted the vulnerability of the country's water security to weather and climate changes (Parker and Wilby, 2013). Since then, the water industry's regulators have consistently included requirements for assessing potential climate change impacts on the water supply (Beran and Arnell, 1989; Defra, 2003; Downing et al., 2003;

Environment Agency, 2003) and later also for adaptation plans (The UK Government, 2008).

Given the risks to the UK water security, the full extent of the benefits, potentials and limitations of water management need to be well understood and the heterogeneity of water use behaviours taken into account (Parker and Wilby, 2013). Emerging technologies and increases in computing power provide new ways to process large quantities of data in parallel and in reasonable time, which allows extracting values, causes or events from historical data that might have been overlooked in the past (Garcia et al., 2015). This information can be used to develop credible water demand forecasts, as well as pro-active strategies that can assist with optimising network operations and building network resilience.

However, understanding and modelling water demand involves the consideration of a variety of factors such as lifestyle changes, household formation, population growth and weather characteristics, in order to ensure a trustworthy projection for the future. This work uses smart demand metering data, household characteristics and weather variables to gain a better understanding of water demand and its influencing factors, as well as develop an improved water demand forecasting methodology.

The rest of this chapter provides the necessary background information and sets the terms and concepts that are going to be discussed in this thesis. It starts with describing the main aspects of water use and the concept of water demand forecasting, in terms of its characteristics and best-practice approach. Next, the key research questions and objectives of this work are introduced, followed by an outline of the thesis. Finally, a list of available resources, including links to publications and code, are provided at the end of the chapter.

1.2. Background

According to Billings and Jones (2008), urban water demand forecasting is 'the process of making predictions about future water use based on knowledge of historical water use patterns'.

1.2.1. What is water demand?

As part of this work, the first question that needs to be answered is 'what is water demand?'. Some studies (Bellfield, 2001; Merrett, 2004; Rinaudo, 2015) define water demand as the water required by customers for various uses, such as domestic, industrial or agricultural. Another interpretation (Billings and Jones, 2008) defines water demand as 'the total volume of water necessary or needed to supply customers within a certain period of time', including leakage and all other inevitable water losses. In this thesis, the terms water demand, water use and water consumption are used interchangeably to refer to the total amount of water used by customers. This includes water losses on the customer side but excludes the associated water losses within the network (e.g. due to leakage or fraudulent abstractions).

1.2.2. Water demand metering

Traditionally, residential water demand in the UK is not billed based on meter readings. Unmetered customers are charged a fixed amount per year instead, dependent on property characteristics such as the number of bedrooms, type of property, number of occupants or a company average. This is further adjusted according to the property's rateable value, which reflects the rental value of the property and was last updated in the 1970s (Defra, 2008).

Water metering is part of a new, sustainable, environmentally friendly policy that aims to reduce water demand and secure water supply now and in the future. A water meter (similar to a gas or electric meter) is a device that measures how much water is used. Typically, water meters are read twice per year (Ofwat, 2013). Water metering is regarded as the fairest way to charge customers, since it requires them to pay for the volume of water they have used. Historically, most properties in the UK have paid a standard, flat rate for their water use, regardless of actual consumption. However, water companies forecast that more than half of the homes in the UK will be on a meter by 2020 (CIWEM, 2015).

Unlike conventional metering devices, smart meters can record consumption in regular, much more frequent time intervals (e.g. every 15/30 minutes or even a handful of seconds) and are able to communicate that information wirelessly.

Thus, they can provide descriptive statistics (e.g. flow rates) as well as a better understanding of consumption (Pericli and Jenkins, 2015). Potential applications of smart demand data include leak detection and variable water pricing, as well as improved network operations and demand forecasting (McKenna et al., 2014).

1.2.3. Water demand modelling

Water demand modelling can be used for many purposes, such as demand pattern recognition and forecasting, user profiling, as well as identifying the determinants of water consumption.

According to Cominola et al. (2015), the existing literature can be divided into two distinct types, descriptive and predictive studies. Descriptive studies are useful for the analysis of patterns in the data that can improve the understanding of when, where, and why water is used. Predictive studies focus on predicting future demands. Machine learning methods have been employed in the literature for both descriptive and predictive purposes.

More details regarding the types of models and methods that are used in each case are provided in the following.

1.2.3.1. Machine learning models

Machine learning is the process through which machines or computers learn how to perform a task, using data. As machine learning becomes increasingly popular and algorithms become more sophisticated, machine learning based methods have dominated the recent demand forecasting literature. Although they have been so far primarily used for predictions, machine learning methods can find useful applications in descriptive studies. This is facilitated further by the data availability, new techniques, and computing power, which have not been available in the past.

Machine learning techniques can be divided in supervised learning, unsupervised learning, and reinforcement learning. Supervised learning includes prediction tasks where the outcome is known and the algorithm learns to make predictions on new data (Molnar, 2019a). Examples of supervised learning algorithms are Artificial Neural Networks, Random Forests, and Gradient Boosting Machines. In unsupervised learning, for example clustering, the outcome is unknown (Molnar, 2019a). The task in this case is to identify common features and create clusters of data points (Antunes et al., 2018). Finally, in reinforcement learning the machine creates the dataset by running examples and evaluating the results (Antunes et al., 2018), with the aim to maximise a reward.

Both supervised and unsupervised learning are used within this thesis, although all forecasting models are based on supervised learning methods. Detailed information about the machine learning techniques used in each chapter are provided within the methodology section of the corresponding chapter. A detailed review of the studies that have used these methods for water demand forecasting tasks is also available within the literature review section of each chapter.

A major disadvantage of machine learning methods is their level of interpretability, i.e. understanding how the model makes predictions, as machine learning models are often considered 'black box'. This name implies that information comes inside the box and predictions come out of the box but there is no understanding or knowledge of what is happening inside it. Interpretability should be an important aspect of developing machine learning models, as it is a way to enhance the understanding of a process and ensure the model performs well by sanity checking the results.

Although interpretable machine learning is a relatively new field, few studies developed methods that enable the modeller to peek inside the black box and make conclusions on the role of the input data in making predictions (Goldstein et al., 2015; Apley and Zhu, 2016; Zhao and Hastie, 2018; Fisher et al., 2019; Molnar, 2019). The idea behind many interpretability techniques is to assess how the model predictions change, in terms of accuracy and direction, i.e. whether they increase or decrease, for a change in one or more input variables. A detailed description of the specific interpretability methods used in this study are provided in chapter 3.

1.2.3.2. Descriptive models

The purpose of descriptive studies is to analyse consumption in order to make conclusions regarding the water use of different types of customers, identify the drivers of water demand, as well as explore patterns in the data. The results of this analysis can be used to enhance the understanding of water demand and develop improved demand management strategies.

Typically, descriptive studies (Domene and Sauri, 2005; Babel et al., 2007; Schleich and Hillenbrand, 2008; House-Peters et al., 2010; Chang et al., 2010; Hussien et al., 2016) use simple statistical techniques in order to assess the relationship between consumption and a variety of property, customer, temporal, and weather characteristics. In some cases, machine learning or visual methods have also been employed to identify patterns in water demand or cluster consumption and group households based on their consumption behaviour.

A very common technique used to analyse and gain a better understanding of the dataset is to use descriptive statistics (Domene and Sauri, 2006; House-Peters et al., 2010; Pullinger et al., 2013). These methods are used to provide an overview of the dataset by using measures such as the mean or the variance of a population and demonstrate the frequency of occurrence of a characteristic.

Another very common technique uses econometric and statistical models, such as multiple linear, piecewise, and polynomial regression (Domene and Sauri, 2006; House-Peters et al. 2010; Chang et al., 2010; Hussien et al., 2016) or log-log and semi-log models (Schleich and Hillenbrand, 2008) to investigate the influence of several demographic, behavioural, economic, and environmental factors on water use. These models are popular due to the fact that they are easy to use and interpret.

Other studies estimate the relationship between a variety of influencing factors and water use by assessing the strength of the correlation between them, using the value of a correlation coefficient (Babel et al., 2007; Chang et al., 2010; Hussien et al., 2016). This is a simple approach, although it does not account for the interactions between the variables or the temporal and spatial variation of the effect on water consumption. Methods such as data disaggregation can be useful in accounting for these interactions.

Finally, in some cases, methods such as clustering and data visualisations can offer additional information that would otherwise be very difficult to identify. Clustering methods have been used to find consumption patterns and groups of households with similar consumption behaviour (Pullinger et al., 2013), whereas visual methods can be useful in identifying spatial trends (House-Peters et al., 2010; Chang et al., 2010).

1.2.3.3. Predictive models

There are several water demand forecasting approaches and the most appropriate one needs to be selected with respect to the specific aim, forecasting objective, time horizon, as well as availability and resolution (time and spatial) of the available dataset. One way to group water demand forecasting models is based on their input data and model structure. According to this, they can be classified into micro-component studies, time series analysis, statistical, artificial intelligence, and hybrid models.

In micro-component analysis, ownership level, frequency of use, and volume per use of household appliances, as well as peak use hours, are taken into consideration (Butler and Memon, 2006). Several studies tried to identify patterns and trends using household micro-components (Butler, 1993; Edwards and Martin, 1995; Gurung et al., 2014). However, disaggregating water use requires large amount of data from different sectors, or very high resolution smart demand metering data, that are not typically available. According to the UK Water Industry Research (UKWIR) household consumption forecasting guidance manual, guidance for previous water resources management plans recommended micro-component analysis as the favoured method. However, in the most recent one it was regarded as too data intensive and complex (UKWIR, 2015). In addition, concerns regarding energy spending and carbon emissions (Fidar et al., 2010) also contribute to making micro-component modelling an unattractive option.

Time series models (Froukh, 2001; Kofinas et al., 2014; Brentan et al., 2017; Chen and Boccelli, 2018) are based on the assumption that future trends in water use can be predicted based on historical water use (Billings and Jones, 2008). These models are often used for real-time forecasting and online applications. The Auto-Regressive Integrated Moving Average (ARIMA) method is one of the most important and widely used linear models in time series forecasting, as it has the ability to capture general trends and seasonal variations. The Holt-Winters method is a simple, exponential smoothing method applicable when the time series contain a seasonal component. It is a standard method used for automatic forecasting (Quevedo et al., 2014) and works best when the seasonal variations are roughly constant throughout the series (Kofinas et al., 2014). Although they are quick to train, as well as simple and easy to use, time series models do not typically account for several other variables such as household and customer characteristics that also have an effect on consumption.

Statistical models (Herrington, 1996; Downing et al., 2003; Firat et al., 2009; Haque et al., 2014; Bakker et al., 2014; Fontanazza et al., 2014) consider a variety of variables and estimate statistically historical relationships between dependent and independent variables. This method is very common in the literature, since it integrates the effect of socio-economic and climatic factors, as well as public water policies and strategies. Therefore, it provides water operators with insights regarding the influence of different variables on water use. This is the reason that these models are also frequently used in descriptive studies, where forecasting is not the main goal.

Machine learning algorithms (Froukh, 2001; Cutore et al., 2008; Firat et al., 2009; Bai et al., 2014; Bakker et al., 2014; Romano and Kapelan, 2014; Shabani et al., 2016) have been proven effective to predict short-term, medium-term, and long-term water demand. Artificial Neural Network (ANN) based models are some of the most commonly used machine learning techniques in water demand forecasting and are often suggested as the best in the literature. The downside of these methods is that they are considered 'black-box', hence results obtained this way are harder to interpret. This means that although they can achieve high accuracy, their results cannot be used directly to shape demand management strategies and planning.

Finally, hybrid models (Bakker et al., 2014; Anele et al., 2017) have the advantage of combining different model capabilities, focusing on emphasising positive and reducing negative capabilities of individual models (Kofinas et al., 2014). However, these models can also be hard to interpret as they make predictions by combining the results of individual learners, thus they lack any model structure.

Machine learning and hybrid models are used in this study for their accuracy as well as ability to capture complicated relationships between several predictors. In addition, the use of several interpretability methods allows to use these models not only in order to produce accurate demand forecasts but also in order to gain an improved understanding of the factors that influence water consumption.

1.2.3.4. Model assessment

An essential step of every forecasting methodology is the model assessment, i.e. the process of determining how well the model performed. This is a fairly abstract definition, as it depends on the objective and characteristics of the study. For example, a model might have a very good overall accuracy but perform poorly on peak consumption days, which are of high importance to water utilities. On the other hand, even if the model has a good accuracy for all days, it could be hard to interpret and therefore it might have limited use for operators.

When it comes to water demand forecasting, there is no acceptable level of accuracy pre-defined by the UK water regulators. The cost-benefit of improving forecasts should be considered and the favoured methodology should be determined based on the circumstances. For water scarce areas that are in danger of not being able to fulfil the supply-demand balance, achieving a high accuracy is essential in order to provide guidance and mitigate risks (UKWIR, 2015). However, when potential prediction errors do not threaten the system's capacity to supply water to customers, less costly and sophisticated models can be considered as good alternatives.

In many cases, factors such as the model complexity and training time as well as data requirements might limit the applicability of a model in real-life problems. Thus, the modelling technique needs to be selected based on the appropriate metrics that evaluate its performance with respect to the needs of the case study, while accounting for the requirements and limitations of its application.

Some metrics that are used frequently in the literature appear in the following, where n is the total number of values, O_i and P_i are the i_{th} observed and predicted values, and $\hat{O}$ and $\hat{P}$ are the observed and predicted means, respectively:

- The Root Mean Square Error – RMSE (Dos Santos and Pereira, 2014; Kofinas et al., 2014; Shabani et al., 2016; Tiwari et al., 2016) is the square root of the Mean Square Error - MSE and is expressed as

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(O_i - P_i)^2} = \sqrt{MSE}$$

  The RMSE is a measure of overall performance although it is sensitive to larger errors (Tiwari et al., 2016).

- The coefficient of determination - R^2 (Babel et al., 2007; Bakker et al., 2014; Dos Santos and Pereira, 2014; Haque et al., 2014; Kofinas et al., 2014; Shabani et al., 2016; Tiwari et al., 2016) is expressed as

$$\mathrm{R}^2 = \left[\frac{\sum_{i=1}^{n}(O_i-\hat{O})(P_i-\hat{P})}{\sqrt{\sum_{i=1}^{n}(O_i-\hat{O})^2\sum_{i=1}^{n}(P_i-\hat{P})^2}}\right]^2$$

  The R^2 values vary from 0 to 1 and indicate the degree of correlation between modelled and observed values (Haque et al., 2014).

- The Mean Absolute Percentage Error – MAPE (Bai et al., 2014; Kofinas et al., 2014; Candelieri et al., 2015; Tiwari et al., 2016) is expressed as

$$\mathrm{MAPE} = \frac{100}{n}\sum_{i=1}^{n}\left|\frac{O_i-P_i}{O_i}\right|$$

  The advantage of the MAPE is that it is independent of units and therefore system capacity, which means it can be used to compare results from different studies and utilities (Candelieri et al., 2015).

- The Mean Absolute Error – MAE (Herrera et al., 2010; Dos Santos and Pereira, 2014; Kofinas et al., 2014; Shabani et al., 2016; Antunes et al., 2018) is expressed as

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|O_i - P_i|$$

  The MAE does not assign a higher importance to larger or smaller errors, nor does it take into account the sign of the error. It is merely an indication of the overall agreement between predicted and observed values (Tiwari et al., 2016).

The above performance metrics constitute some commonly cited statistical tests, however, another validation method might be fit for purpose, depending on the respective forecasting aim. Since the selection of the assessment metric could determine the results and conclusions of the study, it is important that this is chosen with respect to the individual aspects of the problem and the research question.

1.2.4. Water demand forecasting

1.2.3.1. Forecast variables

Household water demand can be explored at different temporal and spatial scales, depending on the available data and tools, the selected methodology and the purpose of use. Water demand can be linked to individuals or be aggregated at the household and area level or even across the whole supply zone. It can reflect average annual, monthly, daily or hourly water use, while with the advent of smart demand meters, it can even go down to a few seconds.

Typical end-use studies report per capita consumption (PCC) or per household consumption (PHC) (Gurung et al., 2014). Demand that is analysed at ‘per capita’ or ‘per household’ level is then multiplied by the total population or properties (UKWIR, 2015), in order to determine total demand. PCC can be calculated separately for metered and unmetered customers, as well as for different groups, based on the selected variables or clustering methods. According to Waterwise (2019), the average PCC in England is 150 litres/person/day, although the target is to reduce it to 130 litres/person/day by 2030 (Defra, 2008).

An example of the various types of forecast variables, along with their popularity among water utilities, is provided in Table 1.1. The data was obtained from 662 North American water supply systems, on a volunteering basis, and was published in the American Water Works Association (AWWA) water demand survey (Billings and Jones, 2008). Overall, predictions of hourly and peak demands are useful in managing the network and ensuring sufficient water supply, while seasonal and annual predictions are used for planning and development of future strategies (Butler and Memon, 2006). According to Table 1.1, most water utilities are interested in peak-day demands, followed by daily demands.

Table 1.1. Types of urban water demand forecasts reported in the American Water Works Association water demand survey (adapted by Billings and Jones, 2008).

Percentage of US utilities reporting forecast type	Forecast type
73.9%	Peak-day forecasts
65.9%	Daily water-demand forecasts
65.6%	Monthly system water-demand forecasts
65.4%	Annual per capita water-demand forecasts
58.0%	Annual water-demand forecasts by major customer class (e.g. residential, industrial)
57.9%	Revenue forecasts linked with water-demand forecasts

The best forecast variable should be considered when choosing a forecasting method. Here, predictions are made for the daily PCC, at different spatial scales. In addition, predictions over all days as well as peak consumption days are treated separately.

1.2.3.2. Forecast horizons

Depending on the forecast horizon, water demand projections are utilised for different purposes and can be best described by different types of models. Most studies categorise water demand forecasts in short-term, medium-term and long-term. The longer the forecast horizon, the larger the potential forecasting errors (Billings and Jones, 2008). Although there is no defined time-frame that clearly differentiates the forecast types based on their horizon, a general guideline is given in the following.

In most cases, short-term forecasts predict water consumption up to one month ahead and are typically used to optimise the operational and financial management of the system. Specifically, they can assist with reducing energy spending and carbon emissions, as well as avoiding over-abstractions that cause stress to the natural environment. In this work, short-term refers to predictions one to seven days into the future.

Medium-term covers the timeframe between one and ten years. Changes in consumption within this time period are typically influenced by weather changes or changes in the customer base (Billings and Jones, 2008). Medium-term forecasts can assist with planning improvements of the supply system or adjusting water tariffs.

Long-term forecasts look generally ten to thirty years into the future and are used to address future supply needs. They can assist with making long-term capital investments (e.g. major infrastructure costs) or influencing future demand, by promoting or implementing water conservation policies, campaigns and technologies. Since both strategies can become very expensive, it is important to tailor them to the specific needs of the water provider, by considering future needs (Billings and Jones, 2008).

1.2.3.3. Best practice

The UK Water Industry Research institute (UKWIR) published in 2015 a detailed guideline for water companies that outlines a recommended best practice methodology for household water demand forecasting (UKWIR, 2015). The first seven steps of this guide are illustrated in Figure 1.1.

According to this guide, the first step should be reviewing the bigger picture. This means setting out the characteristics of the problem and collectively considering all steps of the process in order to get a general idea of the tools and data that might be used in the study.

The next step focuses on data collection and evaluation. These data could relate to past consumption, weather, occupancy or socio-demographic data, depending on the kind of information the water company is collecting. Aspects such as the vulnerability of the supply area as well as the cost of collecting and processing this data should be taken into account. The choice of the forecasting method as well as the model's accuracy depend on the amount and quality of the available data.

After the data has been collected and processed, their influence on water consumption needs to be determined. According to UKWIR (2015), there are six factors that influence water consumption, the occupancy rate, property type, customer behaviour and socio-demographic characteristics, as well as lifestyle habits and technology. Before considering any of these factors in the demand forecasting model, their influence on water consumption needs to be well understood.

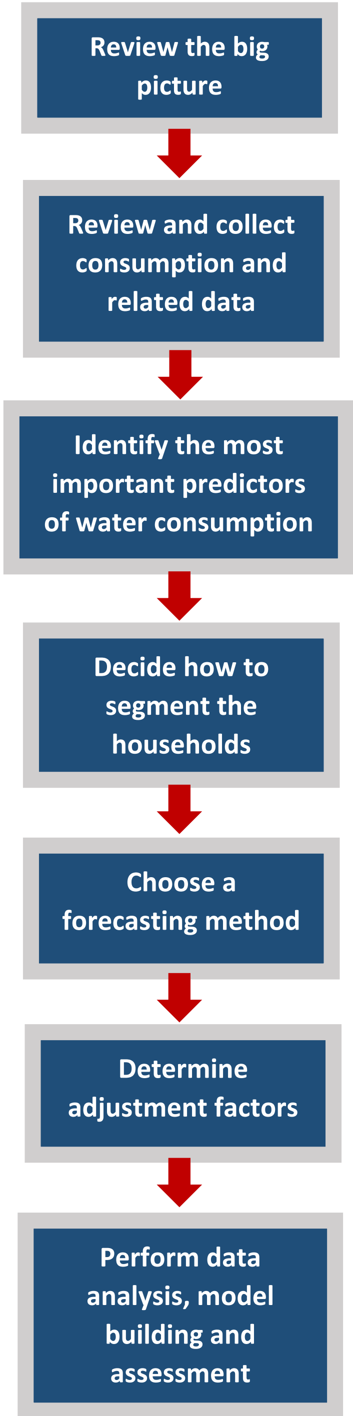

Figure 1.1 Household water demand forecasting best practice (adapted by UKWIR, 2015).

The above factors can be incorporated in the methodology as model predictors or they can be used to segment the households into groups with homogenous characteristics. When segmenting households, a separate forecast is produced for each group. This is often useful if the rate of change in consumption is expected to be different in the future between households with different characteristics. According to the same guide (UKWIR, 2015), further segmenting households will result in more accurate forecasts, since additional information is provided to the model. However, this does not account for the fact that using multiple factors will create smaller household groups, which may also impact the forecasting accuracy.

Based on all of the previous steps, as well as the water availability in the supply area, there are different forecasting options. Each one of them has its unique advantages and shortcomings, which are described in detail in the UKWIR (2015) guideline. Some examples of forecasting approaches are regression models, micro-component analysis, per capita methods or micro-simulation. A combination of two or more of the above methods can also be applied.

The next step is producing a forecast for the maximum consumption of a ‘dry year’. This step assumes that water consumption is influenced by weather conditions and can vary from one year to the next one. Therefore, adjustment factors need to be calculated for the consumption of a ‘dry year’ and a ‘normal year’. The main aim here is to calculate the base water consumption, which covers basic day to day needs, as well as the weather-induced demand, which relates to activities that are triggered by environmental changes.

The last step consists of analysing the data as well as building and assessing the forecasting model. The forecasting model is built using the influencing factors and model structure that were defined during this process and results are assessed by comparing them to real consumption. An uncertainty analysis can also be performed at this stage, by adjusting the values of the uncertain prediction factors within a reasonable range and assessing how this is going to influence results.

The above process describes the suggested best practice for household water demand forecasting in the UKWIR (2015) guide. The methodology developed in this thesis attempts to follow these guidelines, from reviewing the bigger picture until the model assessment. The next three steps that are suggested in the same guide consist of specifically accounting for uncertainty due to model, systematic or data errors; translating all of the above into a final, baseline consumption forecast; and considering potential water efficiency measures, if the supply area is likely to have a negative water supply balance.

1.3. Research questions and aims

The current work explores the topic of residential water demand and specifically the methods, data and influencing factors that are necessary in order to produce accurate forecasts. This section describes the research questions and specific aims of the study.

1.3.1. Research questions

The following key research questions are addressed here:

1. What is the weather influence on water consumption and how does it vary for different household types and time-varying factors?
2. Which are the determinants of water demand and can they be used to make predictions?
3. Can new, sophisticated machine learning techniques and other methods improve the accuracy of current water demand forecasting models?
4. What is the maximum water demand forecasting accuracy that can be achieved at different spatial scales? What are the best predictors at each scale?

1.3.2. Aims and objectives

The overall aim of this work is to develop new methods and knowledge for improved short-term water demand forecasting by using advanced machine learning techniques applied on smart demand metering, weather and other data. More specifically, the objectives of this thesis are as follows:

1. To better understand the link between weather and residential water consumption (addressing research question 1);
2. To identify and analyse the most significant explanatory factors for short-term forecasting of water demand and to understand how these can be used to improve predictions. The possibility of making demand forecasts with limited data (including no past consumption data) will be explored in the process (addressing research question 2);
3. To develop a new demand forecasting methodology that makes use of the latest machine learning techniques, in order to improve the accuracy of existing demand forecasting models. The best performing machine

learning method(s) will be identified in the process (addressing research question 3);

4. To determine the best demand forecasting accuracy that can be achieved at different spatial scales (i.e. for different household groupings), together with the most important explanatory factors at each scale (addressing research question 4).

The main aims and objectives of this thesis and the way these are linked with each other are summarised in Figure 1.2. The first part of this work (Part I, Figure 1.2) is dedicated to understanding the drivers of water demand, as well as how these can be used to make predictions. The second part of the analysis focuses on developing a new, improved methodology that can address several of the main issues in water demand forecasting (e.g. lack of data, peak consumption days) (Part II, Figure 1.2). Finally, the third part combines the knowledge acquired from parts I and II, to explore demand forecasting at different levels of spatial aggregation (Part III, Figure 1.2).

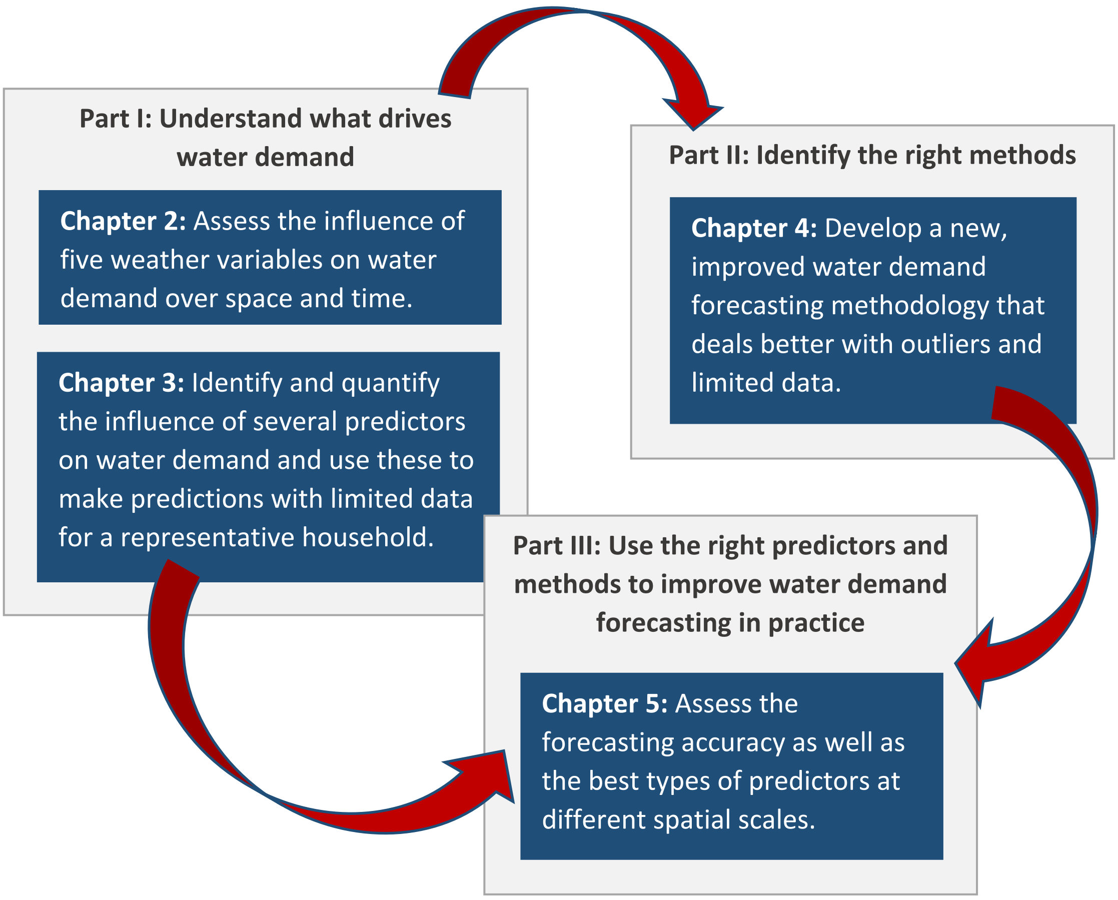

Figure 1.2 Overview of the thesis structure and the main topics that are addressed in each chapter.

1.4. Thesis overview

The thesis is divided into four methodological chapters (see chapters 2 - 5) and a conclusions chapter (see chapter 6), as well as three appendices (see appendices A-B), containing supporting information for chapters 2, 4 and 5, respectively. Each one of the four methodological chapters corresponds to a research paper (for details see the following section) and addresses one of two aspects that are inherently connected to each other, understanding and modelling water demand. A literature review as well as a description of the data that are used in this study, along with the cleaning and processing of this data are available as part of each chapter. A brief summary of the chapters and appendices is provided in the following:

Chapter 2 (addressing objective 1) focuses on identifying the influence of the weather over space and time. An extensive, big-data analysis is performed that disaggregates consumption into different household types, days and times of the day. The effect of five weather variables, air and soil temperature, humidity, sunshine duration and rainfall is examined for each segmentation of consumption.

Chapter 3 (addressing objective 2) expands on this work by investigating the influence and predictive capability of several household, temporal and weather characteristics on water consumption using a machine learning approach. A Random Forest model is trained on daily consumption records using a variety of explanatory variables, in order to predict daily demand for a representative household. Three interpretable machine learning techniques are also used in order to investigate the influence of these predictors (household, temporal and weather characteristics) on the model's output.

Chapter 4 (addressing objective 3) identifies the tools and methods that can enhance modelling accuracy, for different forecasting aims. As part of this effort, several machine learning models are compared for predictions of daily water consumption one day ahead. The model's performance is assessed for all days in the data as well as peak days, i.e. the 10% of days with the highest consumption. In addition, four bias correction methods are used in order to improve the problem of bias towards the mean, which is a very common, re-occurring problem in the literature that is often overlooked.

Chapter 5 (addressing objective 4) compares the prediction accuracy as well as the best types of variables (e.g. weather, temporal or household characteristics) at different levels of spatial aggregation. For this purpose, several Gradient Boosting Machines are trained on past consumption data, for different household group sizes (from 5 to 600 households) and compared for their accuracy in making predictions with one day lead time. Next, eight model configurations are trained and tested at three levels of spatial aggregation. Predictions are compared for one to seven days into the future, for all days in the data, as well as peak consumption days.

Chapter 6 provides an overview of the work performed, the key results and contributions of the study, as well as recommendations for further research.

Appendix A. Provides supporting information for chapter 2.

Appendix B. Provides supporting information for chapter 5.

1.5. Published work and other resources

The data used in this study is not publicly available and can be requested from different sources. The water consumption and household characteristic data was made available by Wessex Water (www.wessexwater.co.uk) and is protected under a non-disclosure agreement. Interested parties can ask for data access directly from Wessex Water. The weather data was collected and became available by the Meteorological Office of the UK (Met Office) (https://www.metoffice.gov.uk). This data was provided to the author for research purposes only and is available for purchase or under request by the Met Office.

All code for the analysis was developed by the author in R (unless explicitly stated within the thesis) and is available at the following github repository: https://github.com/mariaxen/DemandForecasting.

The work that was carried out during this PhD is summarised in four journal papers that have been published or are currently under review (see chapters 2-5). Part of the work that was carried out during this PhD project is also presented in three conference papers that are available online and are not part of this thesis. A list of all journal and conference publications that were produced as a result of this PhD is available in the following.

1.5.1. PhD candidate’s publications

Journal Papers

Xenochristou, M., Kapelan, Z., and Hutton, C. (2019). Using smart demand-metering data and customer characteristics to investigate the influence of weather on water consumption in the UK. *J. Water Resources Planning and Management*, doi: 10.1061/(ASCE)WR.1943-5452.0001148.

Xenochristou, M., Hutton, C., Hofman, J., and Kapelan, Z. (2019). A new approach to forecasting household water consumption. *J. Water Resources Planning and Management* (under review).

Xenochristou, M., and Kapelan, Z. (2019). An ensemble stacked model with bias correction for improved water demand forecasting. *Urban Water Journal* (under review).

Xenochristou, M., Hutton, C., Hofman, J., and Kapelan, Z. (2019). Water demand forecasting accuracy and influencing factors at different spatial scales using a Gradient Boosting Machine. *Water Resources Research* (under review).

Conference Papers

Xenochristou, M., Kapelan, Z., Hutton, C., and Hofman, J. (2017): CCWi2017: F42 Identifying relationships between weather variables and domestic water consumption using smart metering. Available from: https://figshare.shef.ac.uk/articles/CCWi2017_F42_Identifying_relationships_between_weather_variables_and_domestic_water_consumption_using_smart_metering_/5364565/1.

Xenochristou, M., Kapelan, Z., and Hutton, C. (2018): HIC2018: Smart water demand forecasting: Learning from the data. Available from: https://easychair.org/publications/open/qpH8.

Xenochristou, M., Blokker, M., Vertommen, I., Urbanus, J.F.X., and Kapelan, Z. (2018): CCWi2018: 032 Investigating the Influence of Weather on Water Consumption: a Dutch Case Study. Available from: https://ojs.library.queensu.ca/index.php/wdsa-ccw/article/view/12048/7605.

CHAPTER 2

THE INFLUENCE OF WEATHER ON WATER CONSUMPTION

This chapter was published as a Technical Paper in the Journal of Water Resources, Planning and Management (ISSN: 1943-5452). This publication has been slightly modified in order to improve consistency throughout the thesis. The chapter was written by Maria Xenochristou but has benefited from the comments of the co-authors, Zoran Kapelan and Chris Hutton.

Citation: Xenochristou, M., Kapelan, Z., and Hutton, C. (2019). Using smart demand-metering data and customer characteristics to investigate the influence of weather on water consumption in the UK. *J. Water Resources, Planning and Management*, 36, 3161-3174, doi: 10.1061/(ASCE)WR.1943-5452.0001148.

2.1. Introduction

Water availability is a major concern for water utilities in the UK (Water UK, 2016), because of a growing risk of severe drought impacts, due to changes in the climate and population growth. Accurate projections of demand are an essential part of their short-term forecasting, as well as long-term strategic planning. Managing household water use can lead to a reduction in the requirement for infrastructure investments, help secure water supply in the future, as well as save household energy use and greenhouse emissions (Bello-Dambatta et al., 2014). However, despite the clear benefits, few studies in the literature have focused on water demand forecasting in the UK (Parker and Wilby, 2013).

The advent of smart meters in the late 1990s made water consumption data available at very high temporal (minutes or even seconds) and spatial (household) resolution, enabling a better understanding of the patterns of domestic water consumption (Agthe and Billings, 2002; Schleich and Hillenbrand, 2008; Fox et al., 2009). Such data can be used to model demand at the household (or even micro-component) level and thus maintain the heterogeneity derived from the users' unique characteristics and individual water uses (Parker and Wilby, 2013; Cominola et al., 2015). In addition to household, societal, economic and natural factors, the advance of smart metering allows to account for temporal variations in consumption.

The current chapter proposes a systematic, disaggregated methodology that utilises smart demand metering data in order to identify customer and temporal segments of consumption that are more sensitive to weather changes. It utilises simple statistical methods that could enable the development of improved water demand forecasting models and the implementation of effective demand management strategies. As it can be seen from the next section, a systematic analysis of the weather influence on water consumption by using such data has not been conducted before.

2.2. Water demand influencing variables

Many variables have been investigated in the water demand literature as drivers of water consumption. These can be divided into temporal and household characteristics that are or can be known to water utilities, as well as weather fluctuations that are unpredictable in nature. Since the former follow a relatively stable or periodic behaviour, they are easier to account for and thus it is the influence of the weather that is of high interest to network operators.

2.2.1. Temporal characteristics

Seasonal changes in water consumption, as well as weekly and daily patterns are a widely observed phenomenon (Agthe and Billings, 2002; Cole and Stewart, 2013; Gurung et al., 2014; Parker, 2014; Romano and Kapelan, 2014). Typically, water demand reaches a peak during the summer months, when the water is used for outdoor activities, such as filling water pools or gardening, as well as personal hygiene (Downing et al., 2003; Cole and Stewart, 2013). In a study by Parker (2014) with micro-component data from 100 households in the southeast of the UK, external use showed the highest difference between seasons, followed by shower use. In the same study (Parker, 2014), a weekly cycle was observed for certain water uses, suggesting increased water consumption for washing machines over the weekend. On the other hand, Cole and Stewart (2013) found that water used for irrigation typically occurs between 2 am and 6 am, while water is used for showering between 7 am and 12 pm, as well as 5 pm and 9 pm.

2.2.2. Household characteristics

According to several studies (Khatri and Vairavamoothry, 2009; Mamade et al., 2014; Parker, 2014), socio-demographic variables are the most important for daily consumption patterns. Consumers that live in higher-valued areas tend to have more water-using appliances and larger gardens, therefore an increased water-use (Linaweaver et al., 1967; Chang et al., 2010). This effect of income becomes even more relevant when water is used outdoors (Domene and Sauri, 2006).

In addition, the presence of garden and the property’s metering status have been found to influence the type of end-uses and the share among them, as well as the amount of water a household consumes. Among different household sizes and income groups, the presence of garden is one of the determining factors for increased water use (Domene and Sauri, 2006); households with larger lot sizes and no rainwater tanks tend to use more water for garden irrigation (Loh and Coghlan, 2003). Water use for sprinkling and peak demands is more prominent among metered than unmetered customers (Hanke and Flack, 1968), whereas unmetered households’ external water use is also more responsive to meteorological variables (Parker, 2014).

2.2.3. Weather characteristics

One of the major uncertainties relating to water consumption is the influence of the weather. A number of papers investigated the effect of weather on water demand (Miaou, 1990; Griffin and Chang, 1991; Agthe and Billings, 2002; Gato et al., 2007; Haque et al., 2014; Bakker et al., 2014; Beal and Stewart, 2014; Dos Santos and Pereira, 2014).

Within a variety of weather variables, temperature and rainfall are the ones that are frequently suspected to have an influence on consumption. However, many others such as soil moisture, irradiation, sunshine hours and dry days also appear in the literature (Downing et al., 2003; Goodchild, 2003; Parker, 2014). Most studies found a strong relationship between air temperature and water consumption (Downing et al., 2003; Adamowski, 2008; Cole and Stewart, 2013; Willis et al., 2013; Beal and Stewart, 2014), whereas a much weaker one was identified for rainfall (Downing et al., 2003; Goodchild, 2003; Cole and Stewart, 2013; Beal and Stewart, 2014). Adamowski (2008) concluded that rainfall occurrence rather than amount correlates better with water consumption, whereas the occurrence/non-occurrence of rainfall, five days prior, is an even better predictor of daily water demand.

Several authors used linear models to quantify the effect of the weather on consumption (Jain et al., 2001; Downing et al., 2003; Goodchild, 2003; Khatri and Vairavamoothry, 2009; Browne et al., 2013; Parker, 2014). Parker (2014) concluded that all indoor micro-components are linearly related to maximum temperature, sunshine hours and amount of rainfall. A non-linear relationship was identified between temperature and external water use, creating the need to identify thresholds of sensitivity to weather variables and piecewise regression techniques (Parker, 2014). Downing et al. (2013) concluded that most of the climate change impact on water use will be due to baths and showers. Parker (2014) on the other hand found that shower use is less sensitive to weather changes compared to external consumption, whereas washing machine use can also be weather dependent. More specifically, Parker (2014) concluded that an increase in temperature and sunshine hours can cause an increase in outdoor and shower use, whereas an increase in sunshine hours and decrease in rainfall can cause an increase in washing machine use.

2.2.4. Summary

Household water use in the UK reflects a variety of time and space dependent variables (Parker and Wilby, 2013). Thus, taking a holistic view of climate effects, as well as temporal and behavioural drivers, is essential in order to forecast demand (Parker, 2014). Although different temporal and social patterns in water use have been widely investigated, the connection between these and the weather has still not been made.

This study performs an in-depth analysis based on a unique water consumption dataset that is based on real and high frequency observations of water use (i.e. smart demand metering data), from a rather large number of houses located in the southwest of the UK. These consumption data are accompanied by equally detailed information on customer and property characteristics, providing a unique opportunity to explore how different days and water users are influenced by weather changes. Further details about the data used in this study can be found in the next section.

2.3. Data

The current study is based in the UK, more specifically in the southwest of England (Dorset, Somerset, Wiltshire and Hampshire). It utilises an extensive dataset that comprises of:

- Smart demand metering data collected from 1,793 properties for a three year period (10/2014 - 09/2017) at 15-30 minute intervals;
- Property characteristics, including garden sizes, rateable values and metering statuses;
- Customer characteristics, comprised of acorn groups and types as well as occupancy rates. Acorn is a geodemographic segmentation of the UK's population based on social factors and population behaviour and it is used to provide an understanding of the different types of customers (CACI, 2014);
- Weather data collected at hourly to daily intervals for the analysed time-period (10/2014 - 09/2017). The weather data was collected from hundreds of stations across the Southwest and acquired as part of the Met Office Integrated Data Archive System (MIDAS) Land and Marine Surface Stations Data (Table 2.1) (Met Office, 2006a; Met Office, 2006b; Met Office, 2006c; Met Office, 2006d; Met Office, 2006e). However, only 56 of them are included in the analysis, based on their proximity to the properties in the dataset.

Since the properties in the dataset are scattered over a relatively large area, daily and hourly information from multiple weather stations is used to calculate one daily value for each weather variable. In order to do this, a weight is assigned to each station, based on the amount of properties that are the closest to it, as opposed to all other weather stations in the area.

The climate in England is characterised by mild temperatures and rainfall well-distributed all year round. Specifically, maximum air temperature averaged from 1981 to 2010 varied from 6.9°C to 20.9°C and sunshine duration from 54.2 to 193.5 hours in total from January to July, respectively (Met Office, 2012). Monthly rainfall varied between 58.4mm and 91.7mm, for May and October, respectively, whereas according to Met Office statistics from 1981 to 2010, it rains on average 132.8 days in a year (Met Office, 2012).

The weather over the analysed time period (10/2014 - 09/2017) in the south and southwest of England was fairly average, with the exception of some hot spells with high temperatures occurring over the summer. The winters were generally warmer than average, whereas all summers were wetter than average. Rainfall and sunshine hours were close to average values overall, with the exception of 2015 that was a rather wet year.

Table 2.1. Summary of the weather variables that are used in this study.

Weather variables	Description	Units	Duration	Dataset
Sunshine duration	total sunshine	hours	00.00-24.00	UK Daily Weather Observation Data
Radiation	total radiation	MJ/m^2	00.00-24.00	Global Radiation Observations
Rainfall	total rainfall	mm	00.00-24.00	UK Daily Rainfall Data
Humidity	mean humidity	%	00.00-24.00	Hourly Weather Observation Data
Soil temperature	mean soil temperature at 10 cm depth	°C	00.00-24.00	UK Soil Temperature Data
Air temperature	max temperature	°C	09.00-21.00	UK Daily Temperature Data

2.4. Methodology

Water demand patterns are best explored and understood through a theoretical framework of coupled human (e.g. societal, economic) and natural (e.g. atmospheric, geological) systems (House-Peters and Chang, 2011; Breyer and

Chang, 2014). The same approach that assumes a two-way, dynamic interaction between the two is adopted here.

2.4.1. Data pre-processing

In order to ensure the credibility of the results, it is necessary to ensure the credibility of the data that is used in the analysis. In the following, the available data are quality and sanity tested for errors and potential interrelations that could influence the results.

2.4.1.1. Water consumption data

Water demand recorded by a water meter at the household level includes supply pipe leakage and internal plumbing losses in the household, alongside genuine domestic consumption. Thus, the water consumption time series are quality controlled in the following, through a series of practical rules that were developed based on thorough analysis of the data. As a result, the following data are removed from the dataset:

- Recordings that correspond to a consumption higher than 450 litres/hour. Considering the average per capita consumption (PCC) in England is 140 litres/person/day (Waterwise, 2019) and swimming pool ownership in the area is very rare, this is considered a safe threshold to exclude leakage without excluding real consumption.
- The days when less than 10% of the total recordings are equal to zero. This rule assumes that at any given day, at least 10% of the time, no residents are using water. In the case of an ongoing leakage, no zero consumption records should be present. This is a generous assumption, in order to ensure that only constant leakages and not real consumption are excluded from the data.
- The months when less than 20% of the total recordings are equal to zero. This rule assumes that at any given month, at least 20% of the time, no residents are using water. Over a month, consumption is expected to be less erratic, as the effect of random daily factors is averaged over many days, therefore the threshold is higher than when looking at the daily scale.

The above rules were tested and found to be effective in excluding leaking properties. After the pre-processing of data, 1,793 properties are included in the final dataset with recordings corresponding to a total duration of 1,019 days.

2.4.1.2. Weather data

The relationship between each pair of weather variables is tested in the following. Table 2.2 demonstrates the Spearman's ρ correlation coefficient, indicating the strength and direction of association between each pair of ranked variables. As it can be seen from this table, by far the strongest correlation is observed between air and soil temperature (ρ = +0.9), followed by radiation and sunshine hours (ρ = + 0.8). Radiation also correlates well with air and soil temperature (ρ = + 0.7), whereas an equally strong but inverse correlation is observed between radiation and humidity (ρ = - 0.7). Finally, a moderate inverse relationship appears between humidity and sunshine hours (ρ = - 0.6). No other significant correlations are identified between the weather variables examined in this study ($\rho < |\pm 0.5|$).

Based on the above and the quality of the data, some recordings are excluded from further analysis. A quality indicator was provided for each weather recording, showing if the data had been quality checked by the Met Office. Weather records that had not been quality checked were excluded from the dataset. In addition, since radiation is strongly correlated with all other weather variables except rainfall and a significantly smaller amount of radiation measurements is available compared to other weather variables (~25%), radiation is removed from further analysis.

Table 2.2. Spearman's ρ correlation coefficient for each pair of weather variables.

Spearman's ρ	Sunshine Duration	Radiation	Rainfall	Humidity	Soil Temperature	Air Temperature
Sunshine Duration	1	0.8	-0.3	-0.6	0.4	0.4
Radiation	0.8	1	-0.3	-0.7	0.7	0.7
Rainfall	-0.3	-0.3	1	0.4	-0.1	-0.2
Humidity	-0.6	-0.7	0.4	1	-0.3	-0.3
Soil Temperature	0.4	0.7	-0.1	-0.3	1	0.9
Air Temperature	0.4	0.7	-0.2	-0.3	0.9	1

2.4.2. Segmentation approach

In order to evaluate the influence of the weather on consumption for different household types and different times, consumption is divided into segments, i.e. groups with homogenous characteristics. Six household variables are used to segment properties, three property (Garden Size, Rateable Value and Metering Status, Table 2.3) and three customer variables (Acorn Group, Occupancy Rate and Monthly Variation, Table 2.4). In addition, to account for temporal variations, three additional variables are used to segment consumption based on the season, the day of the week and the time of day (Table 2.5).

Table 2.3. Property segmentation of analysed consumption data.

Garden Size	Rateable Value	Metering Status
All	All	All
Large (>165 m^2)	High (>190)	Metered
Medium (61-165 m^2)	Medium (135-190)	Unmetered
Small (<60 m^2)	Low (<135)	

Table 2.4. Customer segmentation of analysed consumption data.

Acorn Group	Occupancy Rate	Monthly Variation
All	All	All
Affluent (A-E)	High (>3 occupants)	High (>120 litres/property/day mean monthly difference in consumption)
Comfortable (F-J)	Medium (2-3 occupants)	
Financially Stretched (K-Q)	Low (<2 occupants)	

Table 2.5. Temporal segmentation of analysed consumption data.

Season	Day of the Week	Time of the Day
All	All	All
Summer	Weekends and Bank Holidays	Morning (06.00-12.00)
Spring	Working Days	Afternoon (12.00-18.00)
Autumn		Evening (18.00-24.00)
Winter		Night (24.00-06.00)

Each household and temporal variable divides consumption in two to five segmentation categories (Tables 2.3-2.5). Gardens were divided into small, medium and large by the water company based on their size (Garden Size, Table 2.3). The cutting points for the rateable value (Rateable Value, Table 2.3) that divide one category from the next one are chosen in order to acquire relatively equal groups and therefore remove bias from the grouping. The properties that are classed as unmetered are the ones that are not being charged based on their meter readings but as unmetered properties, as metering can alter the behaviour of the customers (Metering Status, Table 2.3). According to the acorn guide, consumer groups A, B and C are classified as 'Affluent Achievers' and groups D and E as 'Rising Prosperity'. All groups A to E are classified as 'Affluent' in the following. Groups F to J are classified as 'Comfortable Communities' in the same guide, whereas groups K to Q are 'Financially Stretched'. The same grouping is adopted here (Acorn Group, Table 2.4). Occupancy rate groups are created based on the average UK household that consists of two to three occupants (Occupancy Rate, Table 2.4). Therefore, occupancies higher than three are considered high, whereas lower than two are deemed low. Finally, a variation in mean monthly consumption of over 120 litres/property/day is classified as 'High' (Monthly Variation, Table 2.4). The threshold of 120 litres/property/day is chosen in order for this category to include enough households (~600 properties) to create sufficiently large segments but at the same time small enough to distinguish this group from the rest of the properties.

Accounting for all possible combinations of above segmentation categories (34 in total) results in a large number of homogenous consumption segments (115,200) that share the same property, customer and temporal characteristics. The number of segments is calculated as

$$\mathrm{CS}\ (115{,}200) = \mathrm{GS}\ (4) * \mathrm{RV}\ (4) * \mathrm{MS}\ (3) * \mathrm{Acorn}\ (4) * \mathrm{OR}\ (4) * \mathrm{MV}\ (2) * \mathrm{Season}\ (5) * \mathrm{DoW}\ (3) * \mathrm{ToD}\ (5),$$

where CS = Consumption Segments, GS = Garden Size, RV = Rateable Value, MS = Metering Status, OR = Occupancy Rate, MV = Monthly Variations in consumption, DoW = Day Of the Week, and ToD = Time Of the Day.

The number in brackets represents the number of segmentation categories in which each variable divides consumption. For each segment, consumption is averaged across all properties, for each day of available data. Spatial analysis of the data showed that when aggregating consumption among less than 60 properties, the inherent randomness of water use becomes significant and affects the quality of results. In addition, a sample size smaller than 35 data points is considered insufficient to produce accurate correlation estimates.

Therefore, segments with less than 60 properties or 35 days of consumption recordings are excluded from the analysis.

2.4.3. Assessment of weather-consumption relationship

For each segment (115,200) and weather variable (5), the relationship between consumption and weather is evaluated as follows:

- The Spearman's rank ρ correlation coefficient is used as an indicator of the degree of association between weather and consumption. The Spearman's rank is chosen to assess the degree of monotonic relationship between the variables, since it is better suited to identify non-linear relationships.
- The p-value of the correlation is used to determine the statistical significance of the relationship.
- The gradient of the linear curve that is fitted on the data is used in order to determine the degree of association, i.e. the relative change of consumption for the same change in the weather variable.

In order to filter out segments of consumption for which a weather variable does not have an effect on water demand, correlations with a ρ less than $|\pm0.5|$ or a p-value greater than 0.01 are excluded from the data. The relationship between the weather variable and the consumption in these cases is considered weak or statistically insignificant, respectively.

The gradient of the linear curve that best fits the data is used to filter out results that are statistically significant but not practically significant. This is done by retaining the top 1/3 of the segments with the highest gradient among all significant segments ($\rho > |\pm0.5|$ and $p < 0.01$), for each weather variable. Too often, a relationship between two variables is assessed based on the strength (correlation coefficient) and statistical significance (p-value) of the relationship, without recourse to the effect size, in this case the unit change in consumption for a unit change in the weather variable. In this study, the gradient of the linear curve is deemed acceptable since it is used in relative terms, as a filtering approach, comparing gradients for different segments of consumption. A linear curve (i.e. straight line) still has a higher gradient (for a higher effect) for non-linear relationships – e.g. in the case when a weather variable only becomes significant for demand beyond a certain threshold value.

2.5. Results

2.5.1. Qualitative analysis of weather influence on consumption

The total amount of significant segments ($\rho > |\pm 0.5|$ and $p < 0.01$) that are identified for each weather variable is an indication of the influence this variable has on water consumption, across different customer types and for different times. In this study, 300 significant segments are identified for sunshine hours, followed by humidity and air temperature with 234 and 211, respectively. A less widespread influence is identified for soil temperature, with 125 significant segments, whereas a weak influence is found for rainfall with only 54.

Figure 2.1 shows an example of the distribution between gradients and correlation coefficients for some combinations of weather variables and segmentation categories (e.g. affluent customers, evenings or summers). Each point in Figure 2.1 represents the relationship between weather and consumption for one specific segment, for relationships that are statistically significant ($p < 0.01$). This relationship corresponds to x number of properties and y number of days, where x and y depend on how large the corresponding segment is. The number of properties (x) depends on the six property and customer characteristics (Tables 2.3 and 2.4) and is equal or greater to 60, as mentioned earlier. The number of days (y) depends on the three time-varying characteristics (Table 2.5) and is equal or greater to 35. The total range of correlation coefficient values and gradients for each segmentation category and each weather variable can be found in Appendix A (Figures A1 to A9).

A positive ρ value is usually associated with a positive gradient, whereas a negative correlation coefficient is usually paired with a negative gradient, indicating a direct and inverse, respectively, relationship between weather and consumption (Figure 2.1). However, a few instances in Figure 2.1 have a correlation coefficient and gradient with opposite signs. This is due to the fact that the Spearman’s correlation coefficient is a measure of the monotonic relationship between two variables, which is not always true for the relationship between the weather (especially rainfall) and water consumption, as it becomes apparent from the scatterplots in the next section.

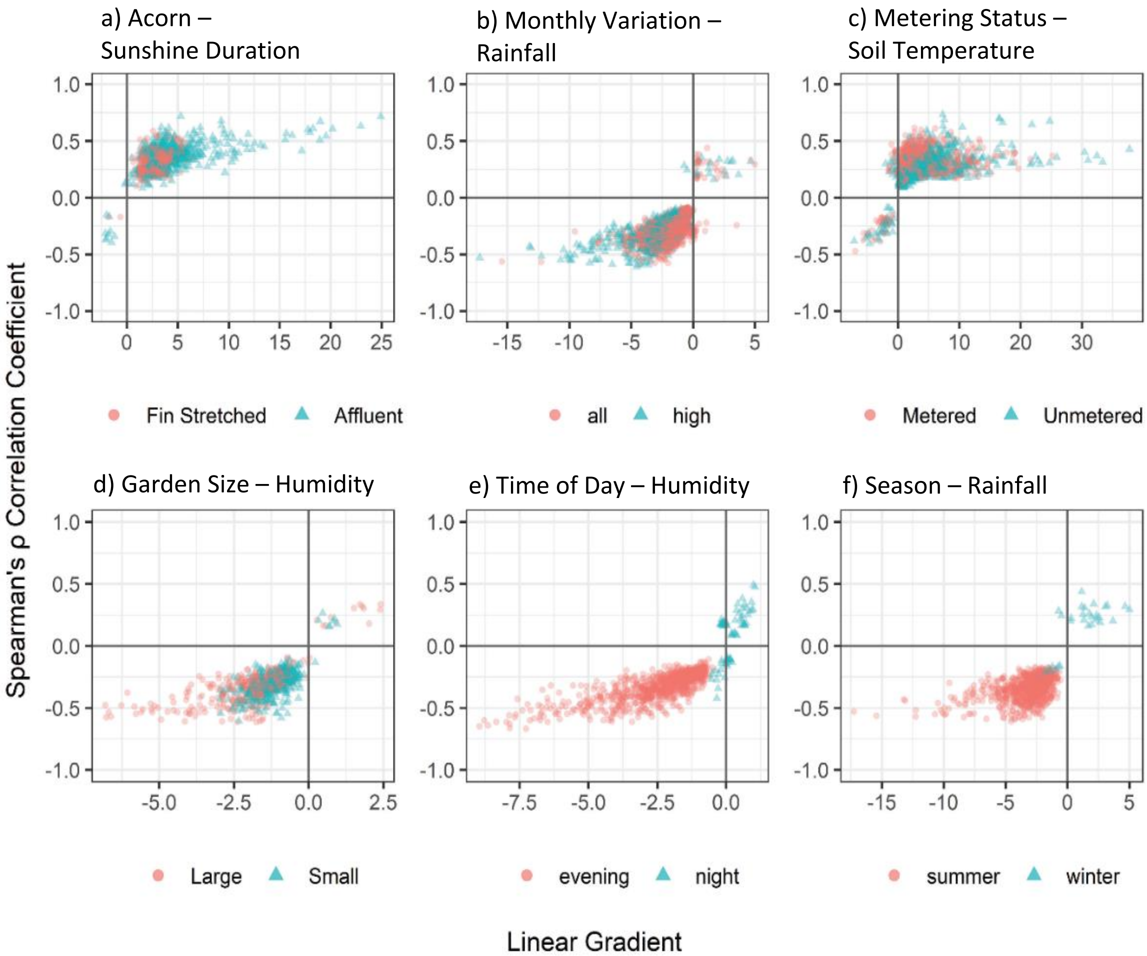

Figure 2.1. Distribution of correlation coefficients and gradients for segments that correspond to various combinations of weather variables and other characteristics (household, customer, temporal). Each point demonstrates the correlation coefficient and gradient for the relationship between consumption and a weather variable, for one segment of consumption.

According to Figure 2.1, consumption falling under certain categories correlates much stronger with the weather. Evening consumption has a significantly stronger negative correlation with humidity, as well as a steeper gradient, compared to night consumption (Figure 2.1(e)). In addition, summer consumption has a much stronger negative correlation with rainfall and steeper gradient, compared to winter consumption (Figure 2.1(f)). The same applies for consumption occurring in properties with affluent residents and large gardens, which appear more sensitive to weather changes (Figure 2.1, (a) and (d)), compared to properties with financially stretched residents and small gardens, respectively. However, results are more difficult to interpret for customers with high variation in their monthly consumption, as well as different metering statuses (Figure 2.1, (b) and (c)).

In order to determine the influence of each weather variable on consumption for each segmentation category, a summary table is created (Table 2.6). This table shows the number of occurrences of each segmentation category among the significant segments, i.e. the ones that are influenced by weather changes. The left column (all gradients - AG) for each weather variable shows the number of significant segments in each category of each characteristic, for all gradients. The right column (top gradients - TG) shows the number of significant segments that also have a gradient among the top 1/3. The category within each characteristic, for each weather variable, which has the highest influence on water consumption (if there is one), is highlighted in bold in Table 2.6.

For example, when looking at the sunshine duration (Sunshine, Table 2.6), the AG column shows that consumption over mornings and evenings is sensitive to changes in sunshine hours. This is because there is a high number of statistically significant relationships with a moderate to high correlation coefficient ($\rho > |\pm 0.5|$) identified between sunshine duration and consumption, for mornings and evenings. However, the same increase in sunshine hours will result in a much higher increase in consumption over evenings (TG column), as 59 of the segments that show the highest sensitivity to sunshine hours (top 1/3 of the gradients) correspond to evening consumption, as opposed to ten segmentations for mornings.

When the ‘All’ segmentation category has the highest occurrence (Rateable Value, Table 2.6), it means that the corresponding characteristic has a weak influence on weather induced demand. Since the ‘All’ segmentation category means that all properties or all days are included in the data, it forms a bigger sample and therefore a positive bias towards it. This simply means that if there is a higher number of segments in this category, it is likely that there are also more significant segments in the ‘All’ category. The reason there are more segments in the ‘All’ category is that segments with less than 60 properties or 35 days are excluded from the analysis. Therefore, including only properties with e.g. high rateable value in a group (instead of all properties) results in smaller groups, thus increasing the probability some of them will not reach the threshold of 60 properties and will be removed from the analysis. The same applies to segments with e.g. just weekends, as this is likely to result in segments with less than 35 days. This means that unless the consumption that belongs in one of the other segmentation categories has a much higher correlation to the weather, the ‘All’ category is going to appear with the highest frequency. In the case of the time of the day, all categories (e.g. morning, afternoon, all) form segments with the same number of days. However, due to potentially missing data for a specific time, the ‘All’ segmentation category is again likely to form more segments.

As it can be seen from Table 2.6, temporal characteristics such as the season, type and time of the day have the highest influence on weather induced demand. The majority of significant segments correspond to summer water use, while soil temperature is the only weather variable that correlates better with consumption during spring (Table 2.6, Season). When the degree of the effect is not taken into account, air temperature has an equally strong effect over spring and summer (Table 2.6, Air Temp - AG). However, almost no significant segments correspond to autumn and winter. Similar results appear for the day of the week and the time of day (Table 2.6, see under corresponding variable name), with the vast majority of the strongest correlations identified during working days and evenings.

With regards to property characteristics (Garden Size, Metering Status, Rateable Value), the influence varies, but it is less prominent than when looking at the temporal ones. Customers with larger garden sizes are overall more influenced by weather changes (Table 2.6, Garden Size), especially humidity as well as air and soil temperature. Although the weather has an effect on both metered and unmetered customers (Table 2.6, Metering Status), the unmetered group shows a higher sensitivity to weather changes, i.e. they will increase their consumption more than the metered group, for the same change in weather conditions (TG column). However, when looking at the results for the rateable value (Table 2.6, Rateable Value), the most significant segments include properties of all rateable values (‘All’ segmentation category) and not a specific type (e.g. ‘High’, ‘Medium’ or ‘Low’). Therefore, no rateable value category seems to be particularly influenced by the weather.

Table 2.6. Number of significant segments, i.e. the ones that have an absolute Spearman's ρ correlation coefficient higher than |±0.5| at 99% confidence interval, for each category and weather variable (Sunshine duration, Rainfall, Humidity, Soil Temperature, Air Temperature), for all gradients (AG), as well as a gradient among the top 1/3 (TG).

Characteristic	Segmentation category	Sunshine		Rainfall		Humidity		Soil Temp		Air Temp	
		AG	TG	AG	TG	AG	TG	AG	TG	AG	TG
Season	All	17	8	0	0	12	6	61	20	65	20
	Summer	**214**	**85**	**52**	**16**	**216**	**71**	4	2	63	**33**
	Spring	61	6	2	0	6	0	**57**	**19**	**74**	17
	Autumn	8	1	0	0	0	0	2	0	6	0
	Winter	0	0	0	0	0	0	1	0	2	0
Day of the Week	All	47	17	9	5	42	23	25	11	31	17
	Weekends	34	10	16	3	27	7	20	9	20	10
	Work days	**219**	**73**	**29**	**8**	**165**	**47**	**80**	**21**	**159**	**43**
Time of the Day	All	140	31	35	1	106	8	50	4	119	23
	Morning	**94**	10	3	0	48	4	20	1	24	3
	Afternoon	0	0	3	2	0	0	2	2	1	1
	Evening	66	**59**	**13**	**13**	**80**	**65**	**51**	**34**	**64**	**43**
	Night	0	0	0	0	0	0	2	0	2	0
Garden Size	All	213	69	48	15	156	50	81	24	137	44
	Large	**39**	**21**	**3**	0	**44**	**22**	**38**	**14**	**53**	**19**
	Medium	36	10	**3**	**1**	20	5	4	3	18	7
	Small	12	0	0	0	14	0	2	0	2	0
Metering Status	All	162	50	28	6	123	41	73	20	120	43
	Metered	**72**	15	3	0	53	12	**30**	7	**49**	5
	Unmetered	66	**35**	**23**	**10**	**58**	**24**	22	14	41	**22**
Rateable Value	All	251	79	42	14	199	70	112	37	184	64
	High	17	11	7	2	15	6	9	2	13	2
	Medium	26	6	5	0	17	1	1	0	5	0
	Low	6	4	0	0	3	0	3	2	8	4
Acorn Group	All	187	56	29	9	136	40	52	12	104	29
	Affluent	**55**	**35**	**23**	**6**	**60**	**31**	**59**	**27**	**82**	**40**
	Comfortable	43	8	2	1	28	6	13	2	23	1
	Fin Stretched	15	1	0	0	10	0	1	0	1	0
Monthly Variation	All	222	44	21	2	147	32	56	8	113	18
	High	78	**56**	**33**	**14**	87	**45**	**69**	**33**	97	**52**
Occupancy Rate	All	178	57	21	5	127	36	64	25	100	30
	High	6	3	0	0	9	3	1	0	2	0
	Medium	107	40	**33**	**11**	94	**38**	45	**16**	94	**40**
	Low	9	0	0	0	4	0	15	0	14	0

A clearer distinction appears between different customer characteristics (Acorn Group, Monthly Variation, Occupancy Rate). Residents of higher socio-economic status are more likely to alter their consumption due to weather changes (Table 2.6, Acorn Group), as more than half of the strongest correlations between consumption and air/soil temperature are identified for segments with affluent residents. Customers with high variation in their monthly consumption also dominate the most sensitive segments for air and soil temperature, as well as rainfall (Table 2.6, Monthly Variation). Similar results, although a bit weaker, appear for properties with medium occupancy rate (Table 2.6, Occupancy Rate).

2.5.2. Quantitative analysis of weather influence on consumption

In order to further explore the above results, five figures are created, one for each weather variable (Figures 2.2-2.6). Each figure demonstrates how consumption correlates to a weather variable, for two different segments, i.e. across different properties and days in the data. Each point in Figures 2.2-2.6 corresponds to a single day for which data is available and shows the mean water consumption (averaged across all properties in the corresponding segmentation) for that day. The red line represents the linear curve that best fits the data and gives a visual representation of the degree of the effect a weather variable has on consumption. In order to visualise the simultaneous effect of different weather variables, three of them are incorporated in each figure. One is represented on the x axis, as the independent variable, while the other two are represented using point size and colour ranges.

Figure 2.2 shows the correlation between total sunshine duration (hours/day) and average daily consumption, for summer evenings and affluent residents, with high variation in their monthly consumption, in unmetered properties, during weekends (Figure 2.2, plot 1), as opposed to working days (Figure 2.2, plot 2). According to Figure 2.2, an increase of 1 hour in sunshine duration could lead to an increase of up to 6 litres/property/day in water consumption for certain customer and temporal characteristics (Figure 2.2, plot 2). For plot 1, which corresponds to weekend consumption, for otherwise the same characteristics, there is very high variability in consumption and no clear trends. On working days on the other hand (Figure 2.2, plot 2), there is remarkably less uncertainty and consumption shows a steady increase for an increase in sunshine hours, which becomes clearer after sunshine exceeds five hours/day.

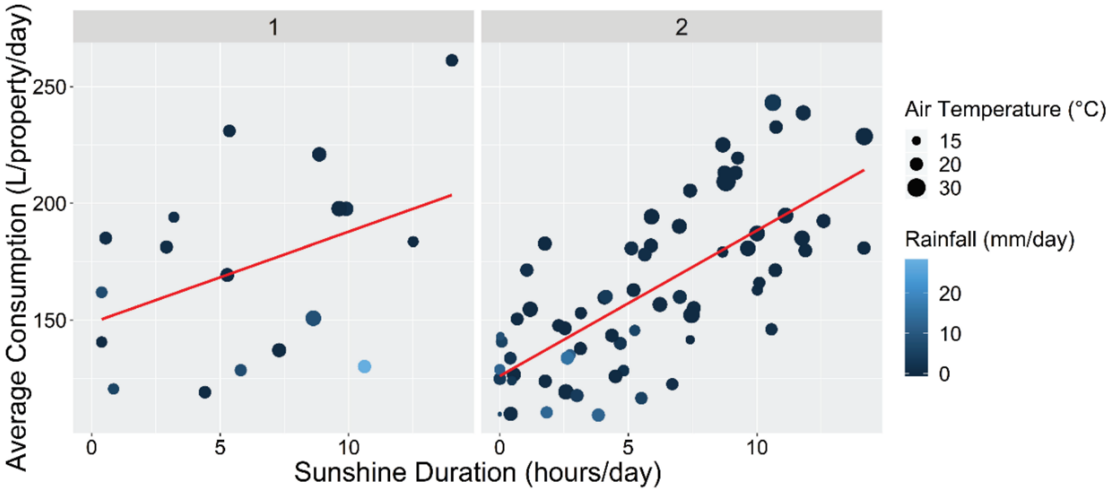

Figure 2.2. Correlation between total sunshine hours (hours/day) and average daily consumption (averaged across all properties), for summer evenings and affluent residents with high variation in their monthly consumption, in unmetered properties, during (1) weekends and (2) working days.

Figure 2.3 shows the correlation between rainfall (mm/day) and average daily consumption for all properties and days in the data (Figure 2.3, plot 1), as opposed to consumption occurring during summer working days, for households with affluent residents and high variation in their monthly consumption, in unmetered properties (Figure 2.3, plot 2). Although there is not a high correlation between amount of rainfall and amount of consumption, high values of water consumption always occur when rainfall amount is zero or close to zero. No rainfall does not necessarily mean that consumption is high, but unusually high consumption always indicates no rainfall (or close to none) (Figure 2.3). It is also remarkable that although rainfall values are fairly similar between plots 1 and 2, higher soil temperature (light blue points) and higher sunshine duration (larger points) correlate with higher values of consumption, for the same rainfall amount.

Figure 2.4 shows the relationship between humidity (%) and average daily consumption for working days and customers with high variation in their monthly consumption, in unmetered properties with high rateable value during the winter (Figure 2.4, plot 1) and summer months (Figure 2.4, plot 2). According to Figure 2.4, humidity is also inversely related to consumption. An increase of 1% in humidity could cause a decrease of 2.5 litres/property/day in consumption for certain segmentations over the summer months (Figure 2.4, plot 2), whereas no such effect is observed for the same properties over the winter (Figure 2.4, plot 1).

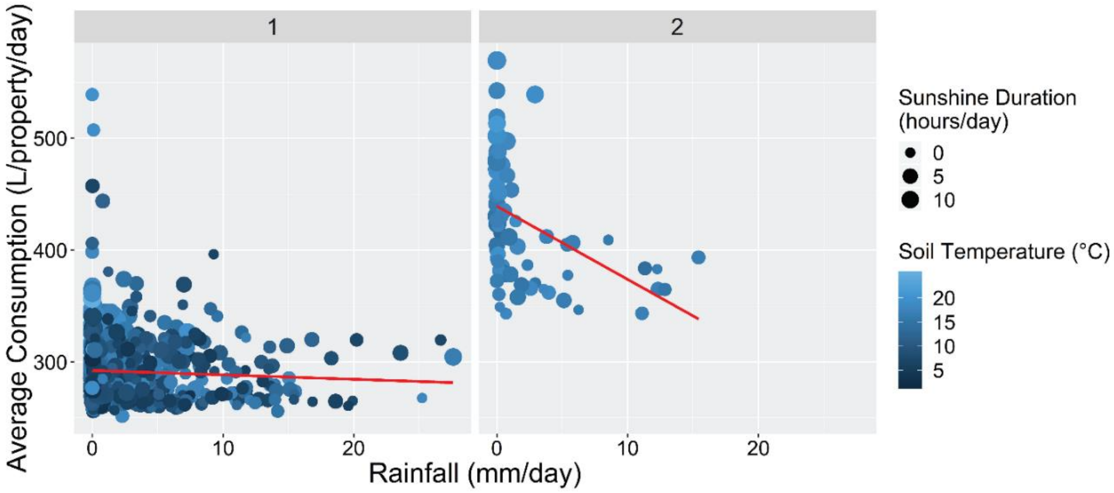

Figure 2.3. Correlation between total rainfall (mm/day) and average daily consumption (averaged across all properties), for (1) all properties and days, and (2) properties with affluent residents with high variation in their monthly consumption, in unmetered properties, during summer, working days.

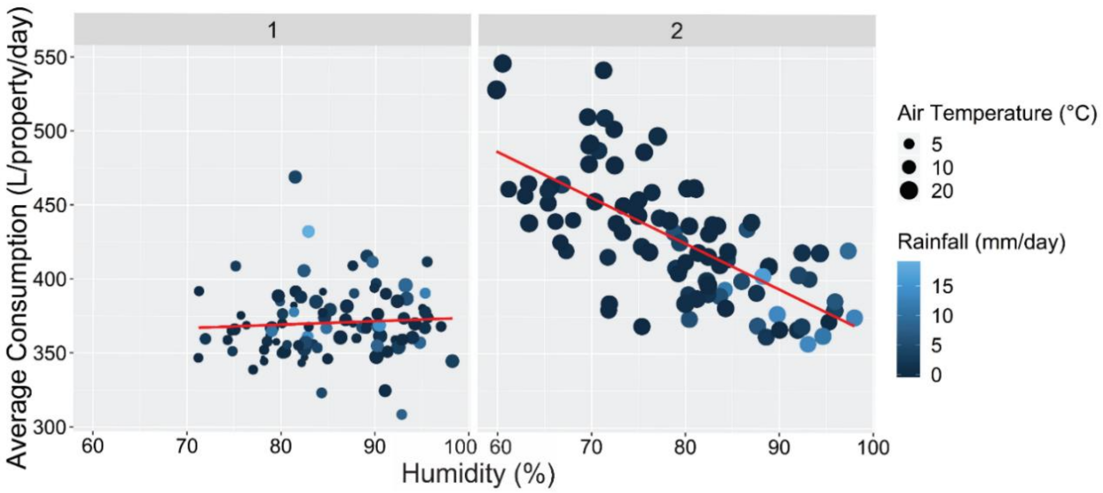

Figure 2.4. Correlation between humidity (%) and average daily consumption (averaged across all the properties), for working days and customers with high variation in their monthly consumption, in unmetered properties with high rateable value, during (1) winter and (2) summer months.

Figure 2.5 shows the correlation between soil temperature (°C) and average daily consumption for working days and affluent residents with high variation in their monthly consumption, in metered (Figure 2.5, plot 1) and unmetered (Figure 2.5, plot 2) properties. An increase of 1°C in soil temperature could cause on average an increase of ~7.5 litres/property/day in consumption for certain customers and days, in unmetered properties (plot 2), whereas a much lower increase (~3 litres/property/day) is observed for metered properties (plot 1). It is worth noting that the effect of soil temperature on consumption shows a clear non-linear trend, as it only starts to become noticeable when soil temperature exceeds 15°C for metered, as opposed to 10°C for unmetered properties. For temperatures higher than 20°C, consumption rises near-exponentially for a further increase in soil temperature, in the unmetered group. For these higher temperatures (>20°C), higher sunshine hours and lower humidity are associated with higher consumption in unmetered properties, as the smaller, light blue points can be found at the upper part of plot 2, for the same soil temperature.

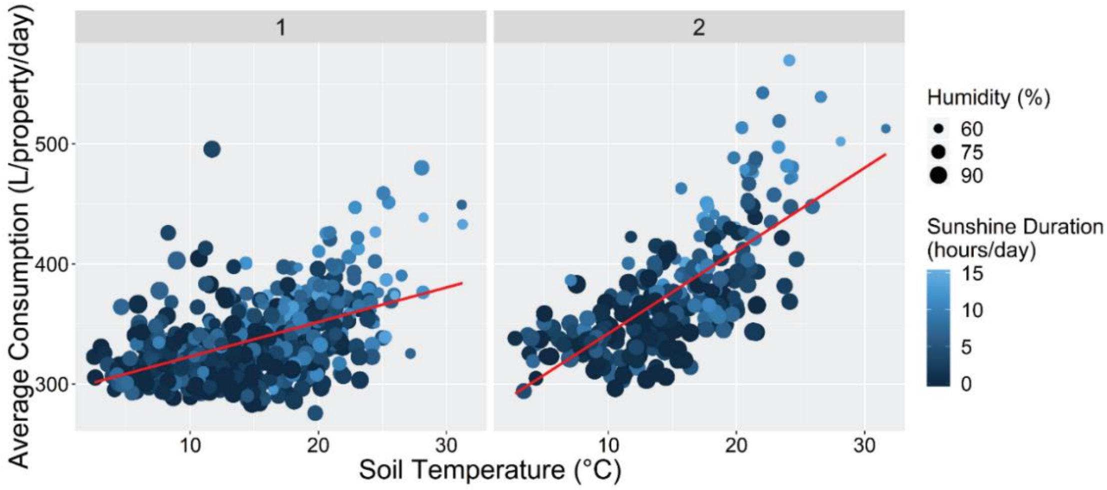

Figure 2.5. Correlation between soil temperature (°*C*) and average daily consumption (averaged across all properties) for working days and affluent residents with high variation in their monthly consumption, in (1) metered and (2) unmetered properties.

Figure 2.6 shows the correlation between air temperature (°C) and average daily consumption for working days and customers with high variation in their monthly consumption, in unmetered properties, with financially stretched (Figure 2.6, plot 1) as opposed to affluent (Figure 2.6, plot 2) residents. The trend for air temperature is very similar to the one observed for soil temperature, demonstrating a non-linear relationship. Higher consumption is associated with higher air temperature and lower humidity, as the darker blue points can be found at the upper part of both plots, although the association with rainfall is less clear; no rainfall (smallest points) does not necessarily imply high consumption, as the small points are scattered throughout both plots, but high rainfall (larger sized points) is always associated with decreased consumption (bottom part of both plots). An increase in air temperature of 1°C could lead to an increase in consumption of ~7.5 litres/property/day, for segments with affluent residents. A much smaller influence is observed for financially stretched customers, with an average increase of ~2.5 litres/property/day. Similarly to soil temperature, only when air temperature exceeds ~15°C for financially stretched or ~10°C for affluent customers, the influence on water consumption becomes significant.

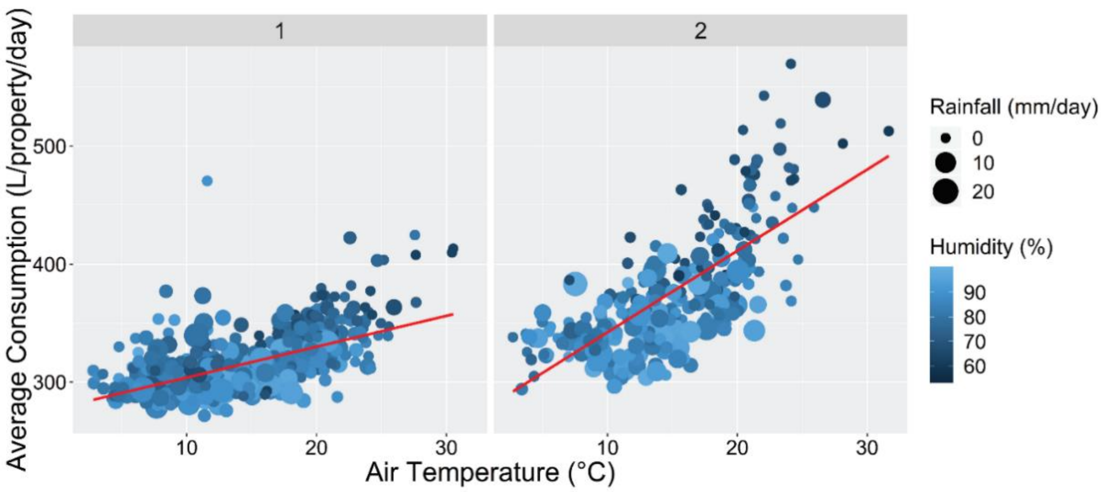

Figure 2.6. Correlation between air temperature (℃) and average daily consumption (averaged across all properties), for working days and customers with high variation in their monthly consumption, in unmetered properties with (1) financially stretched and (2) affluent residents.

The above results confirm what was observed before, that certain types of customers during certain times of the year, the week or the day are more sensitive to weather fluctuations than others. The results and observations from this study are analysed further and compared with findings from the literature in the next section.

2.6. Discussion

When looking at weather induced demand, it is important to identify the primary water uses that drive it, outdoor use as well as baths and showers (Downing et al., 2003; Parker, 2014). As previous studies found out, these water uses are more likely to occur during certain times, as well as for certain households and customer types. Outdoor use is more likely to occur over the summer (Downing et al., 2003; Waterwise, 2009; Cole and Stewart, 2013; Parker, 2014), as well as night hours (Cole and Stewart, 2013), for households with larger gardens (Loh and Coghlan, 2003; Domene and Sauri, 2006), as well as customers that are unmetered (Hanke and Flack, 1968) and have a higher socio-economic status (Linaweaver et al., 1967; Domene and Sauri, 2006; Chang et al., 2010;). Water use for personal hygiene occurs more frequently over the summer, as well as morning and evening hours (Cole and Stewart, 2013). It is therefore expected that for these times and households, the effect of weather on water consumption is more prominent. The results of this chapter are in general agreement with above studies.

The strongest relationships between weather and demand are identified for evenings and working days, primarily over the summer, whereas air and soil temperature also have an effect in spring. As pointed out before, the effect of temperature on consumption becomes noticeable after it exceeds a certain threshold (~10°C-15°C), which in the UK is reached around spring. In addition, the summers over the three years in the dataset were wetter than average, therefore soil temperature would probably decrease due to the evaporation of rainwater from the ground. This could lead to a weaker correlation between soil temperature and water consumption over the summer, even among the segmentations that show the highest sensitivity to soil temperature (i.e. the ones with the highest gradients). Regarding weekends and holidays, this is when people tend to have less constrained schedules and/or are frequently away from home, therefore their behaviour is less likely to be consistently influenced by the weather. Out of the weather variables, rainfall is the only one that has an effect on consumption (although weaker) during weekends. Parker (2014) identified an inverse correlation between washing machine use and rainfall, as well as a weekly pattern indicating that people are more likely to wash over the weekend, which could explain this mild effect.

The customer type also contributes in explaining sensitivity to weather, as affluent customers with high seasonal variations in consumption, in medium occupancy households, are more prone to change their water use due to weather changes. As pointed out by Allon and Sofoulis (2006), understanding the social standards, expectations and habits that relate to water use is just as important as the practical activities that constitute water consumption, if not more so. For example, water used for irrigation might be more related to expectations and care for garden aesthetics, which relate to higher socio-economic status, rather than the size of the garden itself. Furthermore, since households with medium occupancy are occupied by two to three residents, their behaviour is more consistent and easier to correlate to weather changes. Consumption in households with one resident is probably too erratic to form a statistically significant correlation with the weather, whereas in households with more than three residents, weather induced demand (e.g. garden watering) is probably a small percentage of the overall consumption and thus this increase is overlooked. Finally, assuming that water demand is made up of base consumption, seasonal consumption and weather-dependent consumption (Bakker et al., 2014), the high fluctuations in monthly water use observed for certain households are likely due to seasonal and weather-related activities. Thus, it is reasonable that customers with high variation in their monthly water consumption show a higher sensitivity to weather changes.

A more modest influence is identified for household characteristics. Unmetered households with larger gardens are more sensitive to most weather variables, although the rateable value makes little to no difference. Garden size has a rather weak effect on weather related consumption, which becomes stronger for air and soil temperature, as well as humidity. This implies that customers with larger gardens likely increase their consumption in warmer and less humid weather, in order to satisfy garden watering requirements. The same applies to unmetered customers, as prior research concluded that their outdoor use is more sensitive to climatic conditions compared to the metered group (Parker, 2014). Finally, the rateable value of the properties is the factor with the least significance. Although this was originally used as a proxy of the housing type and thus the water use profile of unmetered customers, changes in housing stocks and demographics made the rateable value as an indicator of water consumption out-dated and irrelevant (Parker, 2014).

Out of all weather variables, the sunshine hours, as well as air and soil temperature show a direct relationship to consumption, whereas rainfall and humidity are inversely related to it, i.e. an increase in either of them will likely cause a decrease in consumption. The inverse relationship between humidity and consumption is in agreement with previous studies (Al-Qunaibet and Johnston, 1985), likely due to increased evapotranspiration in both humans and plants, associated with low humidity. As pointed out by Al-Qunaibet and Hohnston (1985), this effect probably outweighs the argument that high humidity intensifies the feeling of heat, leading to increased water use. Similarly, the occurrence of rainfall eliminates in most cases the need for irrigation and can therefore cause a reduction in outdoor use as well as overall consumption.

Sunshine duration correlates well with consumption for more segmentations than any other weather variable, whereas humidity and air temperature also influence a large amount of segmentations. A smaller influence is identified for soil temperature, whereas the amount of rainfall has a minimal effect. Previous studies also identified a high correlation between sunshine hours and consumption (Goodchild, 2003), as well as air temperature and consumption (Downing et al., 2003; Goodchild, 2003; Adamowski, 2008; Beal and Stewart, 2013; Cole and Stewart, 2013; Willis et al., 2013;), whereas a much weaker to no effect was found for rainfall (Downing et al., 2003; Goodchild, 2003; Beal and Stewart, 2013; Cole and Stewart, 2013). However, interactions between rainfall and other weather variables demonstrate that the same rainfall amount could trigger different reactions for different temperatures or sunshine durations.

Finally, a non-linear relationship exists between air and soil temperature and consumption. The effect of temperature only becomes visible when temperature values exceeds a certain threshold, which varies (~10°C-15°C) for different customer types, days and seasons. This effect was previously observed by Parker (2014), who found that outdoor consumption considerably increased after ~15°C and raised the question if this threshold value could change in the future. This study found that this threshold can vary due to multiple factors and should be identified separately for each individual case study and customer group.

2.7. Summary and conclusions

Ensuring the water supply-demand balance is a topic of increasing concern, especially under the threat of climate, population and other uncertain future changes. Understanding the link between weather and water consumption, with demographics, property and socio-economic factors brought into the equation is essential for satisfying this balance.

The current study analyses the correlation between five weather variables (sunshine hours, humidity, rainfall, air and soil temperature) and water consumption, taking into account household, resident and temporal characteristics. This analysis is based on real smart demand metering data, collected every 15-30 minutes for 1,793 properties in the UK, over a period of two years and eleven months. This data is accompanied by data on weather and customers living in the analysed households.

Unlike previous studies, this work accounts for the varying effect that weather changes have across time and space, by aggregating consumption into homogenous groups. Each group contains consumption with the same temporal, resident and property characteristics, averaged over all properties in the group, for each day in the data. The purpose of this is to smooth the erratic consumption signal of individual households, without losing information relating to the drivers of weather induced demand. The approach adopted here can be used in any area where data relating to consumption, weather, as well as customer characteristics are available.

Results lead to the following observations:

- In moderate UK climate, water consumption is only partially influenced by weather changes;
- Sunshine duration has the most widespread (across properties and days in the data) influence on water consumption in the UK, followed by humidity and air temperature. Rainfall has the smallest effect;
- An increase in sunshine duration, as well as air and soil temperature, is likely to cause an increase in water consumption, whereas an increase in humidity and rainfall will likely have the opposite effect;
- The influence of air and soil temperature on water demand becomes noticeable only after temperature exceeds a certain threshold value. This threshold varies for different customer types;
- Although rainfall amount does not correlate well with consumption, high water demand is almost always associated with no rainfall. This is likely due to increased watering requirements, associated with dry weather;

- Water consumption during working days, summers and evenings is affected by weather changes more than during other time periods. This clearly demonstrates the significance of the temporal aspect of water consumption;
- Affluent residents with high variation in their monthly consumption, in medium occupancy households, show higher sensitivity to weather. This could be because they are more likely to use water for showering and watering the gardens during hot and sunny weather;
- Properties with larger gardens and unmetered status are also more prone to be affected by weather changes, whereas the rateable value seems almost irrelevant. Larger gardens justify increased watering requirements, whereas unmetered customers are more likely to use water when the weather is warmer, as they are not billed based on their water usage.

The results in this chapter can assist with managing demand by accounting for the effect of weather on water consumption. Specifically, they can assist with developing improved water demand forecasting models, as well as targeting water conservation campaigns and legislation towards the right customer groups. However, the present work is not without certain limitations.

Acquiring more data could provide additional context to these results. Although all of the available weather factors are investigated in this study, variables such as soil moisture and wind speed, as well as days without rain could further explain water demand fluctuations. In addition, information about indoor and outdoor water use as well as data related to consumption micro-components, could explicitly link certain weather variables to certain types of water uses. Data related to vegetation types and irrigation systems, as well as the calculation of daily potential evapotranspiration could also provide further insights. However, this data for such a range of properties is difficult to collate and maintain.

Furthermore, it is not clear how the distance between the households and the weather stations has influenced results. The properties in this study are scattered over a relatively large area, across several towns in the southwest of England. Although weighted weather averages from nearby weather stations are used, more nearby weather data might lead to stronger relationships, particularly for weather variables that show weaker spatial correlations.

Finally, more work is needed to identify by how much consumption increases on average, for a change in each weather variable. As it is observed here, the increase in water consumption occurs after a weather variable exceeds a certain threshold, which varies for different temporal, property and customer characteristics. Although some examples are provided for certain segmentations of consumption, it is important to identify what is the general response to weather and how this threshold varies for different segmentation categories.

The next chapter aims to address some of these questions by developing water demand forecasting models and assessing how a variety of model predictors (temporal, property, resident and weather characteristics) influence the model's response. As part of this work, the next chapter will identify which weather variable causes the highest spike in consumption, at which threshold and for how many customers and days in the data.

CHAPTER 3

THE INFLUENCE OF HOUSEHOLD, TEMPORAL, AND WEATHER VARIABLES ON WATER DEMAND FORECASTING

This chapter was submitted as a Technical Paper to the Journal of Water Resources, Planning and Management (ISSN: 1943-5452). The chapter has been written by Maria Xenochristou but has benefited from the comments of the co-authors, Zoran Kapelan, Chris Hutton and Jan Hofman.

Citation: Xenochristou, M., Hutton, C., Hofman, J., and Kapelan, Z. (2019). The influence of household, temporal, and weather variables on water demand forecasting in the UK. *J. Water Resources, Planning and Management* (under review).

3.1. Introduction

Ensuring water availability for the future is a matter of increasing concern, especially in the context of a rapidly changing world. Understanding water consumption, as well as the drivers behind it, is the first step towards developing accurate demand forecasts and effective water demand management strategies. However, this is a difficult task as household water use reflects many time and space dependent factors, and research is often limited by data availability (Parker and Wilby, 2013).

In addition, the implementation of smart metering programs is costly, as it requires communications infrastructure and data management applications to support the volume of data and communication between devices, on top of the cost of the metering modules (Hope et al., 2011). Thus, cost-benefit studies in the UK (DEFRA, 2011) and the US (Hope *et al.*, 2011) found that there is no economic case for the roll out of blanket metering programs. These results need to be accounted for by engineers and researchers, who should aim to develop alternative approaches towards water demand sustainability that do not require these data.

Finally, as technology advances, data availability increases rapidly and models become more sophisticated, time-consuming and data-intense, it is important to identify the point where an increase in complexity does not offer any practical value, or even causes model overfitting problems. Donkor et al. (2014) highlights the importance of creating models that are as parsimonious and rudimentary as possible, whilst maintaining high forecasting accuracy.

The overall aim of this chapter is to determine whether credible, short-term forecasting models can be developed for households lacking smart demand metering data but where a variety of other information is available (household characteristics, temporal and weather data). This will be done by identifying the best set of predictors and assessing the level of accuracy that can be achieved, with and without smart metering data. Finally, the influence of each predictor on the model's response, i.e. the water consumption, will be explored using interpretable machine learning techniques. As it becomes apparent from the next section, no previous studies attempted to predict household water demand at the daily level, using household, weather and temporal characteristics.

This chapter is organised as follows. The next section outlines a summary of the literature and highlights the key gaps and limitations. Then, the available dataset is described, in terms of the water consumption data, household characteristics and weather data. The methodology section outlines the model input variables, household grouping, modelling technique, as well as model assessment and implementation. The results section includes the main outcomes of the study, in terms of the model performance and influence of a variety of predictors on water consumption. Finally, the chapter concludes with a critical discussion and summary of key results, the limitations of the study and recommendations for future research.

3.1.1. Water demand studies

According to Cominola et al. (2015), the existing water demand modelling literature can be divided into two approaches; one that focuses on pattern analysis and understanding (descriptive models) and one that provides estimations of water consumption (predictive models). Both approaches have their benefits and shortcomings and find typically different applications.

A few qualitative or descriptive studies (Russac et al., 1991; Edwards and Martin, 1995; Parker and Wilby, 2013) have investigated the large spatial and temporal variations in water demand that occur among households and customers with different characteristics, over different months or days of the week. This was further facilitated by the advance of smart metering that made data available at high temporal and spatial resolution. However, most of these studies used historical data to identify relationships between a set of explanatory variables and water demand, not to make predictions.

Furthermore, a large number of studies have focused on the development of demand predictive techniques. From simple linear regression models (Clarke et al., 1997; Goodchild, 2003; Wong et al., 2010) to sophisticated machine learning algorithms (Herrera et al., 2010; Anele et al., 2017; Chen et al., 2017; Zubaidi et al., 2018). However, few studies (Clarke et al., 1997; Fox et al., 2009; Matos et al., 2014) provided deeper insights into what drives water consumption (Brentan et al., 2017). In addition, models could further improve by treating separately different occupancies, property characteristics (Fox et al., 2009) and temporal factors, such as the month or the day of the week (Parker and Wilby, 2013).

Finally, due to the difficulty of modelling household consumption and the variety of factors that can influence it, this topic has been significantly underrepresented in the water demand forecasting literature. Two studies (Williamson, 2002; Duerr, 2018) attempted to predict single-household water demand using a variety of property characteristics, weather and other data. However, in both cases predictions were made at the monthly scale.

Williamson (2002) used a number of property characteristics (e.g. number of residents, appliance ownership and property type) to predict monthly household consumption using a regression-based function. This method had the potential to distinguish between millions of household types and explained 44% of the variance in water demand, while the rest was attributed to factors that were not included in the model, such as the garden size.

Duerr (2018) also developed a water demand forecasting model using property (e.g. land and building value, green space), temporal (e.g. month and year) and weather (e.g. temperature, precipitation) characteristics. Several methods were compared, including machine learning, linear regression and time series models, for their ability to forecast household monthly consumption. The one that performed best was the time series model, with a minimum Root Mean Square Error (RMSE) of 1,246, for predictions 1 month ahead.

3.1.2. Overview, limitations and scope

Water demand modelling that reconstructs detailed household characteristics would enable planners to predict small area demands, assess the impacts of population changes and test new tariffs (Clarke, 1997). However, most UK water demand studies rely on water-into-supply data (Parker and Wilby, 2013).

Even when explanatory variables (e.g. household and climatic variables) are employed to produce water demand forecasts, this is done using linear regression analysis or geodemographic profiling based on census data (Parker and Wilby, 2013). These techniques (Goodchild, 2003; Wong et al., 2010) have traditionally been used because they are simple and able to capture the relationships between the predictors and water demand in a transparent way. However, their ability to model the complicated relationships between a set of predictors and water consumption may be limited.

Machine learning models are able to provide accurate water demand forecasts (Herrera et al., 2010; Anele et al., 2017; Chen et al., 2017; Zubaidi et al., 2018) but they have been traditionally considered 'black box'. This means that they are not easy to interpret and sometimes even their structure and functionality is not well understood. For this reason, their ability to explain water consumption and provide guidance to water utilities has been limited. Combining both accuracy and interpretability is essential in order to produce accurate forecasts and provide water utilities with the knowledge to improve network operations and secure water for the future.

In addition, surprisingly few studies attempted to estimate and predict water consumption at the household level under potential changes in the climate, which likely reflects the difficulty of understanding and predicting household water use (Parker and Wilby, 2013). At the same time, the non-linear effect of weather on water demand, which could be of particular importance on peak demand days, needs to be further investigated (Parker and Wilby, 2013; Xenochristou et al., 2019a).

This chapter addresses few of the above key gaps in the literature, by developing a novel methodology that combines machine learning models with interpretability techniques.

3.2. Data

The current study utilises a dataset from the southwest of England. This comprises of water demand data and household characteristics that became available by Wessex Water, one of the UK water companies, as well as weather data that were provided by the Met Office. A detailed description of each data type is available in the following.

3.2.1. Past consumption

Water demand data were collected at the household level by the water company using smart meters, recording consumption every 15-30 minutes over a three year period (10/2014 - 09/2017). The above raw data was carefully cleaned and processed before used in any further analysis. A process was implemented, comprising of logical rules that aimed to exclude inconsistent or false data whilst maintaining the natural variability of water demand. This process is outlined in detail in chapter 1. After the pre-processing of the data, 1,793 properties are included in the dataset with recordings corresponding to a duration of 1,019 days.

3.2.2. Household characteristics

The water company also collected household data relating to property and customer characteristics (garden size, rateable value, metering status, council tax band, acorn groups and types, and occupancy rates).

In order to limit the processing time as well as reduce complexity, the properties in the dataset are grouped in two to three segmentation categories for each household characteristic. Garden sizes were divided into small (<60m^2), medium (61-165m^2) and large (>165m^2) by the water company. Properties that are classed as unmetered are a representative sample of all unmetered customers and are not charged based on their meter readings. The water bill of unmetered properties in the UK is adjusted according to the property's rateable value, which is indicative of its rental value and was last updated in the 1970s (UKWIR, 2015). The cutting points for the categories of the rateable value are chosen in order to acquire relatively equal groups that are at the same time distinct enough to identify any differences in their water consumption. The top and bottom 30% of the rateable values are classified as high and low, respectively, whereas the rest are classified as medium. Acorn is a geodemographic segmentation of the UK's population based on social factors and population behaviour (CACI Limited, 2014). According to the acorn guide, consumer groups A, B and C are classified as 'Affluent Achievers' and groups D and E as 'Rising Prosperity' (CACI, 2014). All groups A to E are classified as 'Affluent' in the following. Groups F to J are classified as 'Comfortable Communities', whereas groups K to Q are 'Financially Stretched' (similar to the same guide). Occupancy rate groups are divided into 1, 2 and 3+, based on the corresponding number of occupants living in each household. The council tax bands are divided into three classes containing bands A-C, D-E and F-H, with class A being the lowest and class H the highest paying council tax band.

The cutting points of the new categories for the acorn status, occupancy rate and council tax band are selected based on a z-statistic, according to the following process. Each type of household (e.g. households in tax band C) is associated with a certain water consumption distribution among all days in the data. A z-statistic is used in order to assess the similarity between the consumption distributions for different types of households. Similar consumption distributions that are also in close proximity in terms of the physical meaning of their characteristic (e.g. similarly paying council tax bands) are grouped together into a larger category (e.g. council tax bands A-C).

Many of the household variables described above are indicative of the socio-economic status of the household's residents, thus the correlations between them are evaluated using a chi-square (x^2) test of independence (Table 3.1). The x^2 varies between 1 and -1, indicating a perfect positive or negative correlation, respectively. According to Table 3.1, the council tax band is the most highly interrelated variable. Properties that are under higher paying council tax bands have higher rateable values, larger gardens and residents with higher socio-economic status. The second most correlated variable is the garden size. Properties with larger gardens have a higher rateable value and are occupied by residents in higher acorn groups. Finally, the rateable value and the acorn group, as well as the metering status and the number of occupants show a weaker relationship (Table 3.1). Overall, other than the high correlations identified with the council tax band, all other variables show a much lower degree of association.

Table 3.1. Chi-square correlation statistic between each one of the six household variables.

Chi-square Correlation Table	Garden Size	Rateable Value	Metering Status	Acorn Groups	Occupants	Council Tax Band
Garden Size	1	-0.41	0.16	0.33	-0.12	-0.48
Rateable Value	-0.41	1	0.09	-0.30	-0.07	0.57
Metering Status	0.16	-0.20	1	0.17	0.29	-0.15
Acorn Groups	0.33	-0.30	0.17	1	-0.04	-0.58
Occupants	-0.12	0.10	0.29	-0.04	1	0.13
Council Tax Band	-0.48	0.57	-0.15	-0.58	0.13	1

3.2.3. Weather data

The weather dataset includes Met Office data on air and soil temperature, humidity, sunshine duration and rainfall. This data is recorded at the hourly or daily scale over the same period (10/2014 – 09/2017), from hundreds of weather stations across the study area, as part of the Met Office Integrated Data Archive System (MIDAS) Land and Marine Surface Stations Data (Met Office, 2006a; Met Office, 2006b; Met Office, 2006c; Met Office, 2006d; Met Office, 2006e). The number of consecutive days without rain is also calculated based on the rainfall data.

Figure 3.1 gives a brief overview of the weather over the study period. Weather in England is characterised by mild temperatures and consistent rainfall. Maximum air temperatures vary between 5°C and 25°C, with very few exceptions, mostly over the winter and summer months (Figure 3.1). Springs and summers have generally higher temperatures, increased sunshine hours and lower humidity, although seasonality is not as prominent as in continental climates. Rainfall is reduced over the spring and summer months (Figure 3.1, Rainfall), but the presence of rainfall, which is often more important for water demand, is consistent over all seasons (Figure 3.1, Days Without Rain).

Out of the hundreds of weather stations in the study area, only 56 are included in the analysis, based on their proximity to the properties in the dataset. Since the properties are scattered over a relatively large area, daily and hourly information from multiple weather stations is used to calculate one daily value for each weather variable. In order to do this, a weight is assigned to each weather station, based on the number of properties that are the closest to it geographically (each property is closest to one of the weather stations).

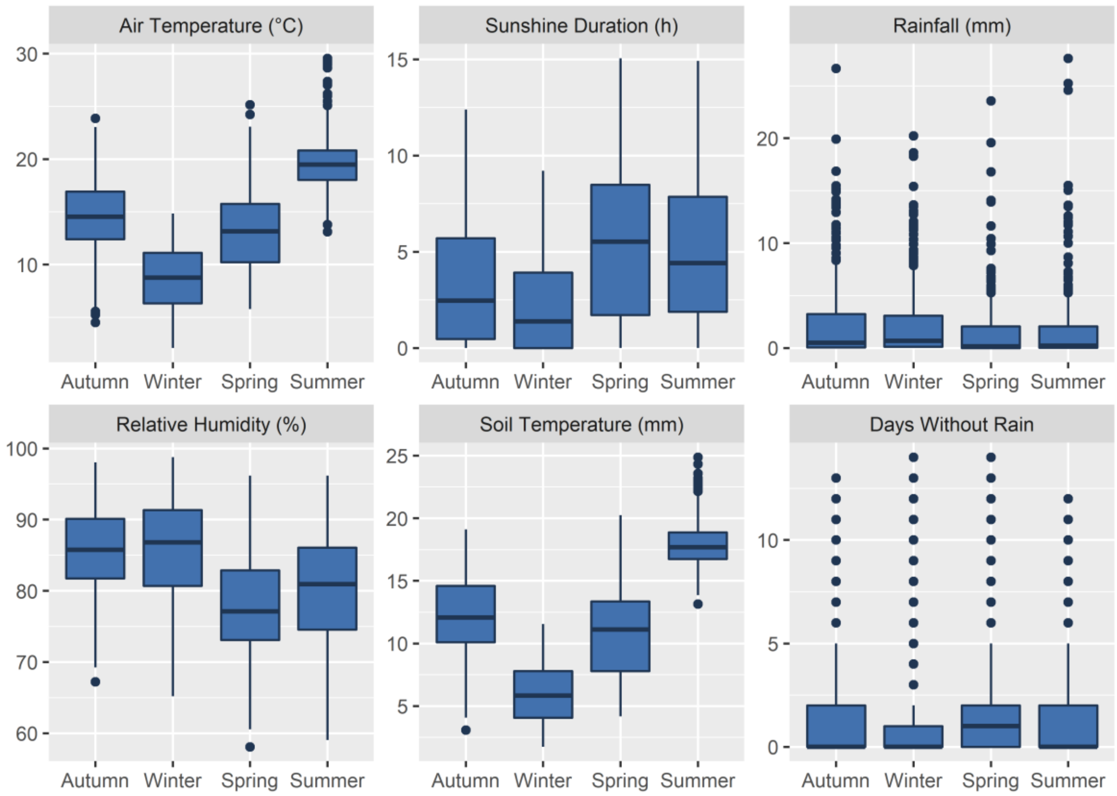

Figure 3.1. Variation of six weather variables within each season over the study period.

3.3. Methodology

3.3.1. Input variables

The first step towards model building is to define the pool of variables that will be included in the analysis. In this study, all available variables are investigated for their influence on water consumption, grouped into the following four types:

- Past consumption: a 7-day window of past consumption is used to capture the repetitive nature of water use over a calendar week. Past consumption consists of seven values, reflecting mean daily consumption for each one of the seven days prior to the prediction day;
- Temporal variables: these refer to the season, month, day of the week and type of day (working day or weekend/holiday) that consumption relates to. They are used as a proxy for time-varying behavioural and weather patterns;
- Household characteristics: the six variables collected by the water company, the garden size, rateable value, metering status, occupancy rate, council tax band and acorn group are also used as predictors, as these variables have been regularly suspected to influence demand;
- Weather variables: six variables relating to daily air and soil temperature, relative humidity, total sunshine hours and rainfall amount as well as the total number of days without rain are used in order to specifically account for the weather induced variance in water consumption.

Each group of variables has unique significance for water utilities. Temporal data are easy to access since they relate to a specific day and are always known to network operators. Information regarding household characteristics on the other hand is sometimes easily accessible (council tax band, metering status, rateable value and acorn) whereas in other cases (garden size and occupancy rate) it needs to be collected through questionnaires or inspections. Finally, weather data that are based on forecasts can be inaccurate as well as expensive to acquire, just as information about past consumption, which requires extensive metering programs as well as processing and storing.

3.3.2. Household grouping

Since one of the main aims of this study is to maintain the heterogeneity of the original dataset, all six household characteristics are used in order to create homogenous groups of properties, for each day in the data. For example, one group could comprise of properties with large gardens, high rateable value, measured consumption, affluent residents, tax bands A-C and occupancy rate 3+. Since each household characteristic has three to four categories, this results in 3,072 groups with homogenous characteristics, as below

$$\mathrm{HG}\ (3{,}072) = \mathrm{GS}\ (4) * \mathrm{RV}\ (4) * \mathrm{MS}\ (3) * \mathrm{Acorn}\ (4) * \mathrm{CT}\ (4) * \mathrm{OR}\ (4),$$

where HG = Household Groups, GS = Garden Size, RV = Rateable Value, MS = Metering Status, CT = Council Tax Band, OR = Occupancy Rate.

However, some groups (3,072 in total) do not include any houses for all or part of the days in the dataset (1,019 in total). In addition, the minimum amount of households in each group is set to two, resulting in a total of 56,020 groups, with 2-24 households each, or ~3.8 households on average, for all days in the dataset.

This grouping is adopted in order to reduce the number of data points and smooth the consumption signal. Instead of having multiple individual households with identical characteristics and high variance in consumption, these are replaced by one representative household, with consumption equal to the mean among all properties in the group. Due to the small size of the final groups and the high variation in their characteristics, daily water consumption varies significantly among days and groups, from ~45 litres/capita/day to ~390 litres/capita/day, with a mean consumption of 127.4 litres/capita/day.

3.3.3. Random Forests

A Random Forest (RF) model is an ensemble of decision trees that can be used for regression or classification purposes (Breiman, 2001). The RF regression used here works by taking a set of input variables, which are then passed onto each of the decision trees in the forest. The uniqueness of a RF model lies in the fact that it implements randomness in the modelling process, as at each node the variable for splitting is chosen among a randomly selected sample of the independent variables (Herrera *et al.*, 2010). Each tree gives a prediction and the mean of these values is the prediction of the RF.

Hyperparameters in machine learning models are parameters whose values are fixed before the learning process begins. RFs' performance depends on three key hyperparameters, the number of eligible features for splitting (mtry), the number of trees that comprise the forest (ntrees), as well as the tree depth, which can also be specified by the number of end points at each node (nodesize). The maximum number of mtry is equal to the total number of input variables. Small values of mtry increase the randomness of the trees and reduce processing time, while small values of nodesize cause the trees to grow deeper, with the danger of overfitting. Although it is commonly believed that default values of these hyperparameters (e.g. mtry = number of variables/3 for regression) can produce good results, there is no theoretical framework that supports this assumption (Scornet, 2017). Therefore, the models are fine-tuned for the optimum set of hyperparameters (mtry, nodesize, ntrees), as the ones that minimize errors whilst not allowing the model to overfit.

RFs are chosen as they have been consistently found to outperform most other models in the literature (Chen et al., 2017), while at the same time they are underrepresented in water demand forecasting (Herrera et al., 2010; Chen et al., 2017; Duerr et al., 2019). In addition, these models are quick to train as the trees are built in parallel and they have limited number of parameters that require tuning.

3.3.4. Model performance assessment

The forecasting accuracy of the models is assessed using the following three performance metrics: the mean square error (MSE), the mean absolute percentage error (MAPE) and the R^2 coefficient of determination. These metrics provide a range of information; the MSE is more sensitive to outliers, the MAPE is biased towards smaller values, whereas the R^2 demonstrates the amount of variance explained by the model.

The variable importance is calculated by assessing by how much accuracy drops when a variable is permutated (i.e. rearranged). Permutating a variable means shuffling its values and thus destroying the link between the predictor and the outcome. For example, shuffling the temperature variable would rearrange the temperature values by randomly assigning each one of them to a day in the dataset. The MSE of the model is calculated before and after the permutation occurs; the higher the increase in MSE, the higher the importance of the variable that was permutated. The shuffling is repeated several times in order to achieve more accurate results. However, this process is affected by variable interactions for two reasons. First, correlated predictors masque each other's effect, since they provide overlapping information to the model. At the same time, shuffling a variable which is strongly correlated with another one could create unrealistic data points (Molnar, 2019a). For example, assuming two correlated predictors, air and soil temperature, shuffling the air temperature values could create a day with soil temperature of 4°C and air temperature of 28°C.

The model predictors are evaluated for their impact on the dependent variable, i.e. the water demand, based on two types of interpretable machine learning methods, the Accumulated Local Effects (ALEs) plots (Apley and Zhu, 2016) and the Individual Conditional Expectation (ICE) curves (Goldstein et al., 2015). In order to explain these methods, it is easier to explain the simpler concept of Partial Dependence Plots (PDPs) first. PDPs work simply by forcing a predictor to take the whole range of its values for each point in the data (each data instance) and calculating the mean response of the model for each value of the predictor. The same happens for categorical predictors, except in this case the variable is forced to take each one of its potential categories, instead of a range of values. PDPs assume non-correlated variables, as in a different scenario this process could create unrealistic data instances, as explained above.

ALE plots also describe how a variable affects the prediction on average by calculating the variation in the model's results within a small window of the predictor. ALE plots are centred at zero, so the value at each point is the difference to the mean prediction. Apley and Zhu (2016) first introduced ALE plots as a faster and non-biased alternative to partial dependence plots (PDP). ALE plots are used here to model the influence of the household and temporal characteristics.

ICE plots are the same as PDPs but instead of averaging, ICEs show one curve for each data instance (each day and household group). Therefore, they are able to capture the variability in the response, for the same change in the predictor. Since there are 56,020 different groups for all days in the data, the same amount of curves are represented in one plot, which makes it very difficult to distinguish between them. Therefore, these curves are aggregated for each plot into three groups, using k-means clustering (Steinley, 2006). Since the weather has a different influence on different types of households and days in the data (see chapter 1), the ICE plots are used to capture this varying effect of the weather variables.

More details and explanations regarding these three methods can be found in Molnar (2019a). All of the above analysis is performed using the R statistical software, particularly the RandomForest (Liaw, 2018) and iml (Molnar, 2019b) packages.

3.3.5. Model implementation

Two groups of RF models are developed and tuned for the optimum set of hyperparameters (mtry, nodesize and ntrees), for daily predictions one day into the future (Table 3.2). Models 1, 2 and 6 incorporate past consumption data whereas models 3, 4, 5 and 7 use a combination of temporal, household and weather characteristics. Consumption data are of high interest for two reasons; firstly, water utilities do not always have access to this data and therefore it is important to account for this scenario and develop an alternative strategy. Secondly, past consumption incorporates many qualities that are characteristic of the household or the day the consumption corresponds to and therefore can masque the effect of other predictors.

As the methods described earlier (variable permutation and ICE curves) are affected by variable interactions, the correlations between the predictors need to be assessed. An investigation into variable interactions (not presented here) showed that sunshine hours and humidity, rainfall and days without rain, as well as air and soil temperature are correlated. On the other hand, temporal variables such as the type of day (working day vs weekend/holiday) and the weekday, as well as the season and the month are by definition also heavily correlated. Past consumption data is also auto-correlated from one day to the next one.

These interactions are taken into account when choosing the model predictors (Table 3.2), thus the input variable configuration for models 1-7 is chosen according to the following. Model 1 (with past consumption) and model 3 (without past consumption) include all temporal, weather and household variables. To reveal the influence of each variable without being concealed by overlapping information, models 2, 4 and 5 exclude strongly correlated inputs (Table 3.2). Finally, results regarding the most important predictors from models 1-5 are used to build models 6 and 7, based on the simplest model configuration that would not compromise the modelling accuracy (Table 3.2).

Table 3.2. Input variables for Models 1-7.

Variable Group	**Model Input Variables**	**Model number**						
		1	**2**	**3**	**4**	**5**	**6**	**7**
Past Consumption	Consumption 1-7 days ago	X					X	
	Consumption 1 day ago		X					
Temporal	Type of Day	X	X	X	X		X	X
	Weekday	X		X		X		
	Month	X		X		X		
	Season	X	X	X	X			
Household	Acorn	X	X	X	X	X		X
	Garden Size	X	X	X	X	X		X
	Metering Status	X	X	X	X	X		X
	Rateable Value	X	X	X	X	X		X
	Council Tax Band	X	X	X	x	X		X
	Occupancy Rate	X	X	X	X	X		X
Weather	Sunshine hours	X	X	X	X			
	Soil Temperature	X		X		X		
	Air Temperature	X	X	X	X			
	Humidity	X		X		X		
	Days without rain	X		X		X		
	Rainfall	X	X	X	X			
	Total input variables	**23**	**12**	**16**	**11**	**11**	**8**	**7**

3.4. Results

3.4.1. Preliminary Analysis

The preliminary data analysis is conducted with the aim to investigate how consumption varies across different household and temporal categories. Modelling results can be strongly influenced by interactions between variables as well as the model structure itself. Therefore, it is important to have an initial view of which are the variables with the highest effect on water consumption and see if these conclusions align with the modelling results.

Figure 3.2 shows the distribution of consumption for each household variable category and each day in the dataset. The most distinct difference in consumption is observed when households are grouped based on their occupancy rate, with low occupancy households (1 resident) consuming significantly more compared to high occupancy ones (3+ residents) (Figure 3.2(a)). Differences also appear between households in different council tax bands (Figure 3.2(b)), with houses in bands A-C (lower council tax bands) consuming less water than houses in bands F-H (higher council tax bands).

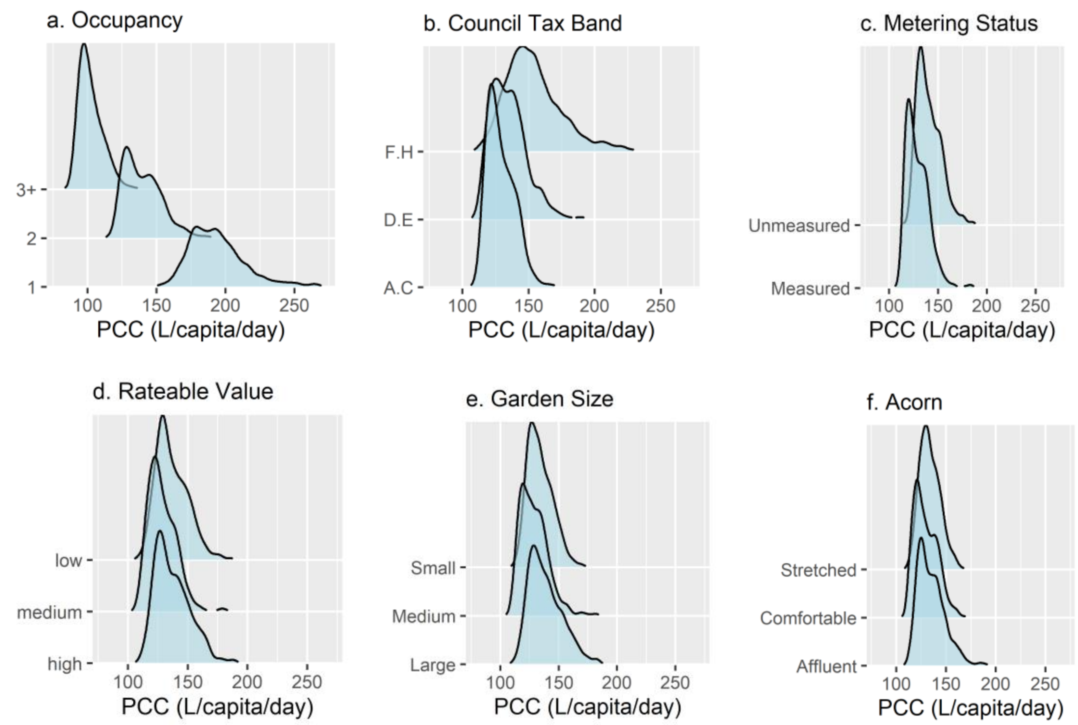

Figure 3.2. Distribution of consumption for different categories of six household characteristics. Each distribution comprises of mean daily consumption, aggregated among all properties with the corresponding characteristic, for each day in the data.

Distributions of household categories that relate to higher consumption are generally more spread out whereas the low consumption curves tend to have a higher peak and a much smaller variance (Figure 3.2). This is likely because lower consumption constitutes base consumption, i.e. water used in order to perform essential day to day activities such as toilet flushing, showering and cooking. Higher demand on the other hand is due to consumption activities that are conditional to a series of other factors. For example, higher council tax bands consume generally more water but they also have a higher spread in their daily consumption. This means that the additional water use could be associated with activities like gardening that occur on some days but not others.

The high variance in the case of the occupancy rate is due to the consumption in single-occupancy properties being more erratic, as it only depends on one person. In the case of two, three or more residents, the PCC is calculated as the mean between the occupants of the property, thus averaging out any differences in consumption behaviour from one day to the next one.

Figure 3.3 shows the distribution of daily PCC for different categories of four temporal characteristics (month, day of the week, type of day and season). Demand is time-dependent as it increases during certain times of the week or the year. Consumption is higher over weekends and holidays as opposed to weekdays, with Sundays claiming the highest weekly consumption (Figure 3.3, (a) and (d)). A milder influence is observed throughout the year, as water demand over the summer months and December is slightly higher than any other time of the year (Figure 3.3, (b) and (c)).

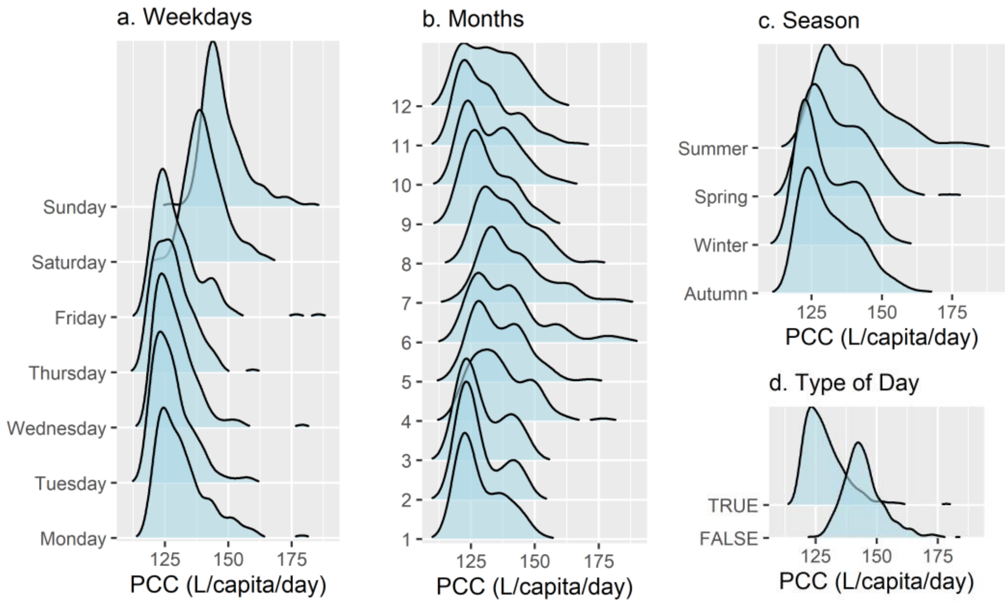

Figure 3.3. Distribution of consumption for different categories of temporal characteristics. Each distribution comprises of mean daily consumption, aggregated among all properties for each day in the data, for different (a) weekdays, (b) months, (c) seasons and (d) day types.

3.4.2. Model tuning

In order to start the modelling process, the dataset is shuffled and divided randomly into a training set (70% of the data) used to train and tune the models and a test set (30% of the data) used to assess their performance on unseen data, i.e. data that is not used during the model-building phase.

Models 1 and 3 are tuned for the optimum set of hyperparameters over a two dimensional grid search space that includes multiple values of mtry and nodesize. To keep the processing time within reasonable limits, the grid search space is built using seven values of mtry for model 1 (5, 8, 11, 14, 17, 20, 23) and model 3 (4, 6, 8, 10, 12, 14, 16), and five values of nodesize (50, 100, 150, 200, 250). The ranges for mtry are selected around the default mtry values (number of predictors/3), which are equal to ~8 for model 1 and ~5 for model 2, whereas the node size range is selected based on expert judgment.

Figure 3.4 shows the model error (MSE) for the test dataset, for various combinations of these parameters. Plot (a) corresponds to model 1, which includes all input variables as explanatory factors (23 variables in total), whereas plot (b) corresponds to model 3, which excludes seven days of past consumption (16 variables in total). The same combinations of mtry and nodesize are tested for multiple numbers of trees but accuracy improvement plateaus after ~300 trees. The optimal MSE values correspond to an mtry of 5 and nodesize of 50 for model 1 (Figure 3.4(a)), as well as an mtry of 8 and nodesize of 200 for model 3 (Figure 3.4(b)). However, the above values for model 1 result in a relatively large difference (not shown here) between the accuracy in the calibration and validation datasets, leading to the conclusion that the model is slightly overfitted, therefore a nodesize of 200 is chosen instead for both models.

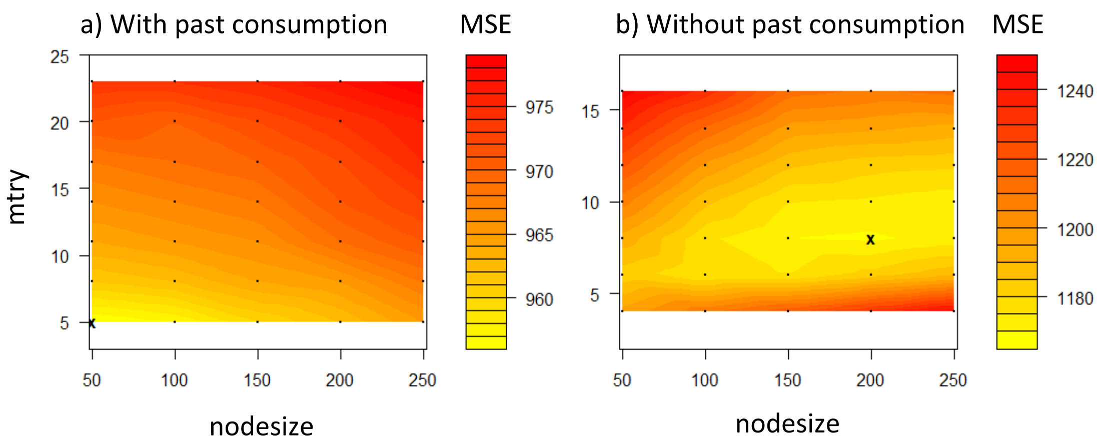

Figure 3.4 Contour plot for the MSE of the validation dataset when (a) all variables including past consumption are included in the model and (b) when past consumption data is not available. The crosses correspond to the point in the grid with the lowest MSE.

The parameter nodesize for the rest of the models is kept at 200 and the number of trees at 300, although all models are tuned for the optimum value of the mtry parameter. This is deemed an acceptable solution based on the above results, since the MSE has a very small range over the search space (Figure 3.4). This confirms the belief that RFs are fairly robust to changes in their hyperparameters, at least when these are varied within reasonable limits.

3.4.3. Variable permutation

Permutating a variable breaks the connection between the predictor and the model's response, therefore it destroys its predictive capability. Here, one variable is permutated at a time for each model and results appear in Figure 3.5 (models with past consumption) and Figure 3.6 (models without past consumption). The x axis demonstrates the importance factor, i.e. the factor by which the MSE increases (denoting decline in model performance), when an input variable is permutated. The variables are ranked on the y axis based on this importance factor. Since the shuffling is repeated multiple times in order to increase the robustness of the outcome, several importance factors are calculated for each variable. The error bar corresponds to the importance at 5% and 95% of the repetitions, whereas the dot corresponds to the median. A factor of one means that excluding the variable from the model does not influence accuracy.

According to Figure 3.5, when seven days of past consumption are included as model input, they are by far the most important predictors (Figure 3.5, Model 1). Demand one day in the past (d.1) has the highest explanatory value, followed by demand on the same day of the week but seven days prior (d.7). The former is because of demand autocorrelation while the latter is because of demand similarity (same day of the week). The day of the week is the only other important variable, whereas the rest has a mild to no influence. However, even when the variable with the highest importance (d.1) loses its predictive capacity, the MSE increases only by a factor of 1.15. Since model 1 already includes seven days of past consumption that carry overlapping information, excluding any one of them does not have a major effect on the output.

However, things are different for model 2 (Figure 3.5), which excludes highly correlated predictors. In this case, both consumption 1 day ago (d.1), as well as the occupancy rate are highly important and excluding either from the model increases the MSE by a factor of 1.50 - 1.53, a much higher rise compared to model 1. In addition, the significance of the rest of the household characteristics as well as the type of day also increases (Figure 3.5, Model 2).

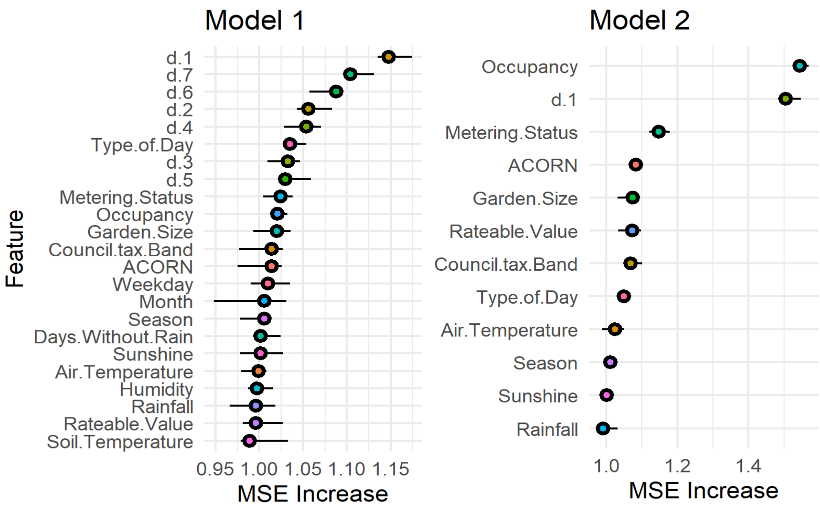

Figure 3.5. Factor by which the MSE increases when each feature is permutated for models 1 and 2.

Figure 3.6 demonstrates the same results, when past consumption data are not used as input (Models 3 - 5). In this case, household characteristics, particularly the occupancy rate, are the most important predictors, followed by temporal information (type of day or weekday) (Figure 3.6). All other variables, including the weather and the rest of the temporal characteristics, are very close to a factor of one. This means that excluding them from the model does not influence the accuracy. Although there are slight differences among models 3-5 (Figure 3.6), the importance factors relating to each predictor are very similar. It is worth noting that the influence of the type of day and weekday slightly increases when these two variables are accounted for separately (Figure 3.5, Models 4 and 5), essentially diminishing the overlap of information that goes in the model.

Notably, there is a large difference in the scale of feature importance between Figure 3.5 (with past consumption) and 3.6 (without past consumption). When the explanatory factors contain overlapping information, excluding one of them only marginally reduces accuracy, resulting in low feature importance factors (Figure 3.5). When information about past consumption data is not available, the occupancy rate is the only variable carrying this information, resulting in an importance factor of up to 2.3 (Figure 3.6, Model 3). This means that excluding information about the occupancy rate of a household, when past consumption is not available, will increase the MSE ~2.3 times.

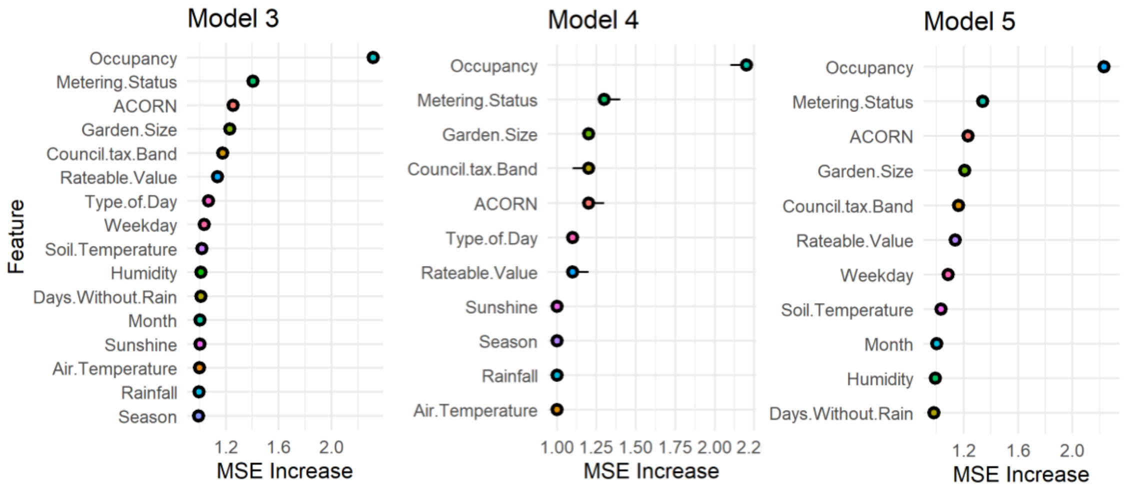

Figure 3.6. Factor by which the MSE increases when a feature is permutated for models 3 - 5.

The above provides a good overview of variable importance and interactions, and can be used as a guide on what variables to include in the model under different conditions, i.e. based on what other relevant information is available in each case.

3.4.4. Prediction accuracy

A summary of the modelling results for the training and test datasets are shown in Table 3.3. Model 6 has the best performance (MAPE = 17.9%, R^2 = 54.9%), although all models have a reasonable accuracy, considering the level of temporal and spatial aggregation (daily consumption, ~3.8 households/group). Model 7, which does not include data on past consumption, can still explain 49% of the variance in the model (MAPE = 19.7%, R^2 = 49.0%).

According to Table 3.3, reducing the number of explanatory variables does not (in most cases) influence the results, whereas in some cases it even improves the model's accuracy. Removing correlated weather and temporal variables has hardly any effect on the result (Table 3.3, Models 3-5), whereas excluding six days of past consumption from model 1 leads to increased errors (Table 3.3, Model 2). Model 7, which includes only six household characteristics and the type of day as input, performs better than model 3, which has additional temporal and weather characteristics. Removing all variables other than past consumption and the type of day from model 1 also slightly increases the prediction accuracy (Table 3.3, Model 6). In both cases, this is likely due to overfitting problems, i.e. the models learning patterns from the variables that do not influence consumption.

Table 3.3: Model configuration and prediction accuracy for models 1-7.

	Model Parameters				Training			Testing		
Models	Cons Data	mtry	nodesize	ntrees	MAPE (%)	MSE (l/hour)	R^2 (%)	MAPE (%)	MSE (l/hour)	R^2 (%)
1	Yes	5	200	300	16.1	742	64.3	17.9	952	54.7
2	Yes	4	200	300	18.1	936	54.7	19.0	1055	50.0
3	No	8	200	300	18.7	983	53.1	19.7	1115	47.6
4	No	6	200	300	19.3	1027	51.3	20.0	1132	47.3
5	No	5	200	300	19.1	1014	52.0	19.8	1126	47.5
6	**Yes**	**3**	**200**	**300**	**16.7**	**809**	**61.0**	**17.9**	**934**	**54.9**
7	**No**	**3**	**200**	**300**	**19.6**	**1069**	**48.5**	**19.7**	**1067**	**49.0**

3.4.5. Influence of household variables

Next, the effect that different household characteristics have in the RF model is uncovered using ALE plots (Figure 3.7). The y axis shows the different categories of each explanatory variable, while the x axis demonstrates the deviation from the mean predicted consumption for each household category (Figure 3.7). When the ALE value of the x axis is positive, the corresponding category is predicted to have a consumption higher than average, whereas the opposite is true when the ALE value is negative.

Results are in agreement with previous analysis that explored the distribution of consumption for each household category (Figure 3.2). Occupancy has by far the highest influence on predicted consumption, as properties with low occupancy rate (1 resident) are predicted to consume ~75 litres/capita/day of water more than properties with high occupancy (3 or more residents) (Figure 3.7(a)). The next most influential variable is the council tax band (Figure 3.7(b)). Higher paying bands (F-H) have a predicted consumption of ~26.5 litres/capita/day more than lower bands (A-C), while unmeasured customers are also on the higher end, with ~19.5 litres/capita/day more than measured customers (Figure 3.7(c)). A smaller influence is identified for the acorn group, garden size and rateable value. Financially stretched customers have the highest predicted consumption, which is ~9 litres/capita/day more than customers in the comfortable acorn group (Figure 3.7(f)). Properties with large gardens are predicted to consume ~5 litres/capita/day more than the ones with small gardens (Figure 3.7(e)), whereas properties with high rateable values are predicted to consume ~3.5 litres/capita/day more than the low ones (Figure 3.7(d)).

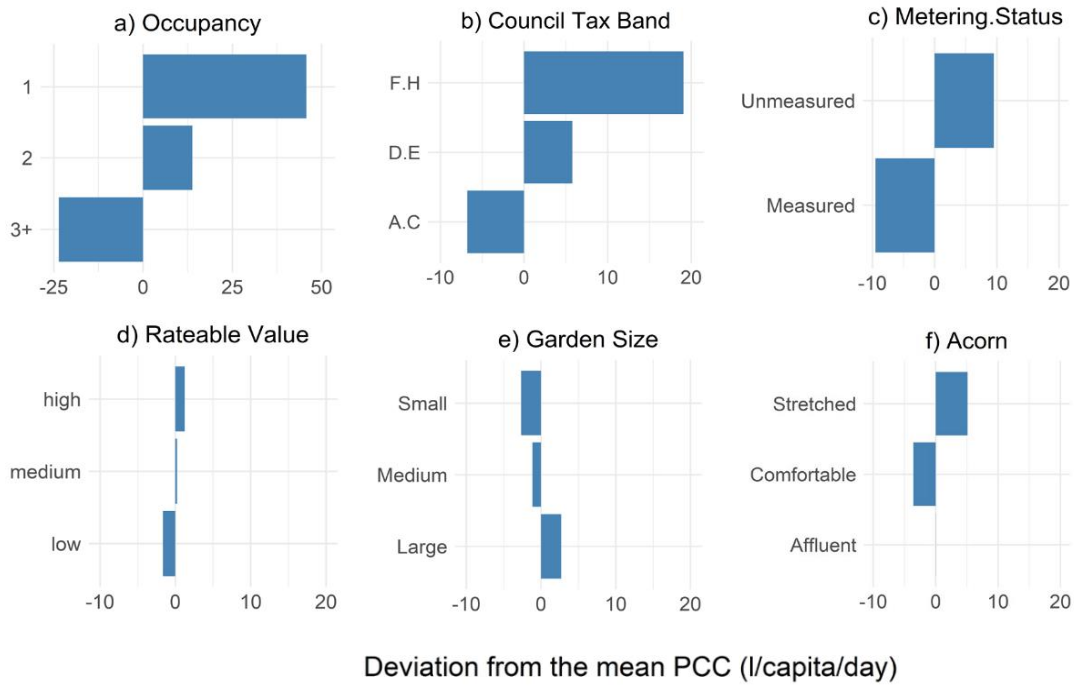

Figure 3.7. Influence of six household characteristics on predicted water consumption – ALE plots.

3.4.6. Influence of temporal variables

The effect of four temporal characteristics on the model's result is also investigated using ALE plots (Figure 3.8). According to Figure 3.8, the type of day and the day of the week have the highest impact on the predicted water demand, whereas the month and the season have almost no influence.

Overall, water consumption on weekends and holidays is predicted to be ~11 litres/capita/day higher than on working days (Figure 3.8(c)). Water demand gradually declines from Monday to Friday, to then increase again on Saturday and Sunday. Sundays claim almost 8 litres/capita/day more on average compared to Fridays, the day with the lowest predicted consumption (Figure 3.8(a)).

Although the month and season have almost no influence on the model's result, summers cause a slight increase (<1 litres/capita/day). An even smaller influence is observed for December (<0.5 litres/capita/day), the month associated with the highest increase in predicted consumption. This is likely due to the holiday season, as people tend to spend more time at home.

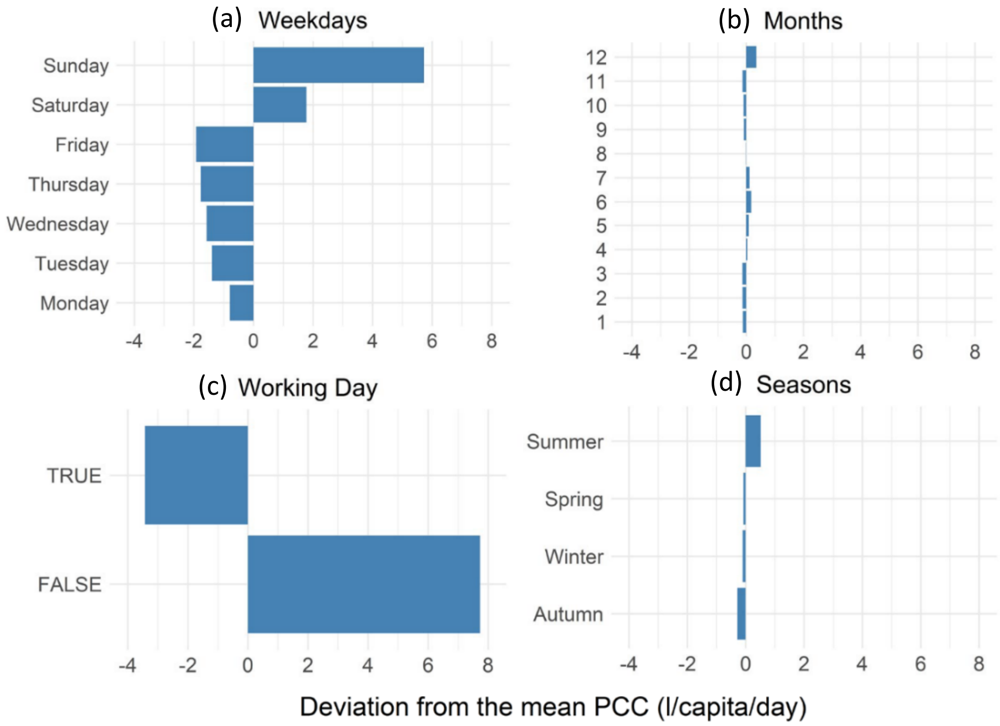

Figure 3.8. Influence of four temporal characteristics on predicted water consumption – ALE plots.

3.4.7. Influence of weather variables

The influence of four weather variables on the model's response variable, i.e. the daily water consumption, is assessed using ICE plots (Figure 3.9). Air and soil temperature are strongly correlated, as is the amount of rainfall and days without rain. In addition, chapter 1 concluded that the rainfall amount and soil temperature have a limited effect on water demand, thus only the ICE curves corresponding to air temperature, humidity, sunshine duration and days without rain are presented in the following. To avoid even small interactions from correlating weather predictors, only one weather variable at a time is considered as model input when creating the ICE plots, along with past consumption data and the type of day. For each plot in Figure 3.9, the y axis represents the change in PCC compared to the mean, when the variable of interest (in this case one of the four weather variables), varies within its whole range of values (x axis). In other words, each plot in Figure 3.9 shows the response of the dependent variable (the daily water consumption), for a change in the independent variable (the weather), for each data instance (one data instance is one day and household type). The percentage associated with each curve represents the percentage of data points that belong to each cluster.

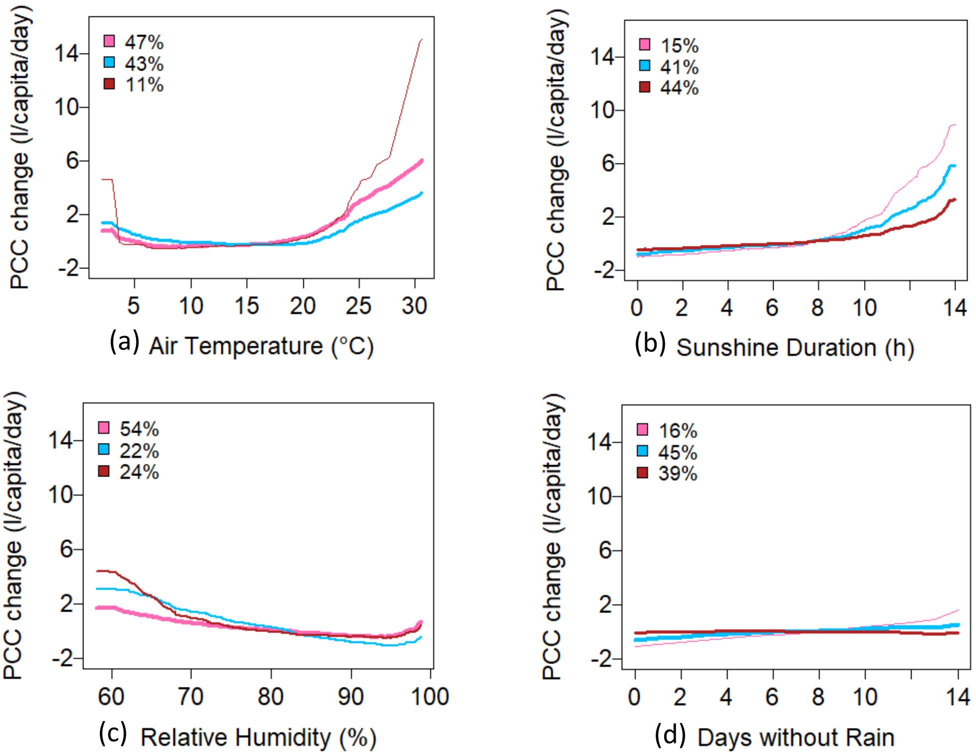

Figure 3.9. Influence of four weather variables on predicted water consumption – ICE plots.

According to Figure 3.9, the weather variable that causes the biggest spike in water consumption is air temperature (Figure 3.9(a)). This effect becomes significant when temperature exceeds ~18°C and to a lesser extent for near-freezing temperatures. Although water consumption starts increasing for temperatures over this threshold, the rate of increase varies significantly (Figure 3.9(a)). As it was pointed out in chapter 1, different days and households have different sensitivity to weather changes. Here, only for 11% of data instances (one data instance is one day and household type), the model predicts an increase in water use of up to 15 litres/capita/day, for an increase in air temperature from 18°C to 30°C. For the rest 89% of the days and household types, the predicted increase in consumption is between 2.5 - 6.0 litres/capita/day (Figure 3.9(a)).

For the rest of the weather variables, the predicted increase in consumption is lower than for air temperature, although the effect is more widespread over household types and days in the data. The maximum increase in water consumption caused by sunshine duration is 9 litres/capita/day, 6 litres/capita/day lower than for air temperature, but this increase relates to 15% of data instances. The relative humidity has an even smaller effect, with a maximum change of 4 litres/capita/day. However, this change applies to ~46% of all days and household types, whereas for 22% of them there is a near-steady decline over the whole range of humidity values (Figure 3.9(c)). For the rest 24% of data points, water consumption drops by 4 litres/capita/day, for an increase in humidity from 60% to 70%, whereas it does not decrease further after this point. The number of consecutive days without rain has the smallest effect on the prediction. Consumption starts increasing after 12 days without rain, reaching a maximum increase of 3 litres/capita/day, for 16% of data points. For the rest of the days and households, the number of days without rain has no effect on consumption.

3.5. Discussion

This chapter attempts to deepen the understanding of water consumption and produce accurate forecasts of demand, with and without past consumption data. However, even for the best model and an abundance of data, the minimum MAPE achieved is 17.9%, while the maximum R^2 is 54.9%. Although these results might seem unimpressive, they need to be put in the right context. In order to maintain the heterogeneity between households with different characteristics, this study resulted in very small aggregations of properties, with ~3.8 households/group. In addition, in order to account for the temporal variability of water consumption (type of day, day of the week), forecasts are made at the daily scale. Thus, taking into account the small temporal and spatial scale for which predictions are made, the models can predict a significant portion of the variance in household consumption, despite the amount of noise and randomness associated with the level of aggregation. As a reference, when predicting household consumption at the monthly scale, previous studies achieved a maximum R^2 of 44% (Williamson, 2002) and a minimum RMSE of 1,246 (Duerr, 2018).

When predicting household demand, past consumption data inherently captures the ‘predictive information’ contained in variables relating to household characteristics. Past consumption has a memory and therefore adding additional information that is already embedded in it does not offer much further benefit. However, in the absence of past consumption data, information about household characteristics can explain a significant amount of variance in the model and produce predictions that are nearly as good as those with past consumption. The implication of this finding is that for the purposes of demand prediction, water utilities do not need to rely heavily on extensive smart metering programs over the whole network. Smaller scale programs may be sufficient to develop useful predictive models that could then be up-scaled with data on customer and property characteristics. This finding is particularly valuable for water utilities in the UK, where almost half of the properties are billed based on the property’s rateable value. It is important to bear in mind that there are other potential benefits of smart metering data beyond demand forecasting, including leakage detection and deriving a greater understanding of household water consumption at the micro-component level.

In this chapter, different approaches are applied to identify the best model predictors. According to Zubaidi et al. (2018), choosing the best set of input variables based on the model’s performance is flawed, due to its dependence on the model’s structure and calibration approach. However, if the objective is solely to maximise the model’s performance, for its current configuration, the model-based approach is the only one that can truly optimise the model’s output. Based on the above, it becomes clear that there are two very distinct aspects when determining the optimum predictors for water demand forecasting. One would be to solely determine the variables that have the highest influence on water consumption, whereas the other would be to determine the ones that can improve forecasting accuracy. Both answers, although distinct, are equally important and could find use in different applications.

Another interesting result is the influence of a variety of predictors on water demand. Household characteristics and particularly the occupancy rate have the strongest effect on predicted PCC, with single-occupancy properties to account for a significantly higher cut of the water supply, followed by customers in high tax bands and unmetered properties. In addition, the temporal variations of water demand over a calendar week as well as a whole year are explored and results show that consumption is predicted to be higher during weekends and holidays. However, no strong seasonal or monthly pattern is identified.

Finally, this study concludes that the weather input cannot increase the accuracy of the modelling results. Out of four weather variables, the air temperature causes the highest spike in water consumption, although sunshine duration and humidity impact more customers and days in the data. In addition, the effect of air temperature and sunshine duration only becomes visible after a certain threshold (~18°C and 8h, respectively). It is worth noting that slightly increased consumption is also associated with temperatures near zero degrees, which is likely because water is used to prevent pipes from freezing (Billings and Jones, 2008). For the case of humidity, the effect is more linear over the whole range of its values, whereas the smallest influence on the predicted consumption is identified for the number of consecutive days without rain. However, consumption starts rising after 12 days without rain, meaning that this could potentially cause problems in the future, if the length of draughts increases.

A reason for the low impact of weather on prediction accuracy could relate to the mild UK climate, which lacks seasonal extremes, as well as the relatively few number of households that are influenced by weather changes. In this region, demand uplifts associated with the weather are typically in the order of 5% during hot summer periods, thus weather induced demand is overall limited. Another reason could be the small size of household groups (~3.8 properties/group). At this level, the random effect of consumption might be too strong to allow for the subtle changes due to weather to show. Overall, this chapter confirms what was observed in chapter 1, that the effect of weather becomes noticeable only for certain households, days and times. Therefore, when looking at the overall influence of the weather over all customer types and days, it is averaged and thus diminished.

3.6. Summary and conclusions

This chapter evaluates the ability of a variety of predictors (household, weather and temporal characteristics) to produce accurate forecasts of short-term demand without information on past consumption. To do this, a number of Random Forest (RF) models are developed using different combinations of input variables, for two general scenarios, with and without past consumption as input. The RF models predict demands one day ahead, for homogenous groups of ~3.8 households on average. In addition, a variety of interpretable machine learning techniques are incorporated in the methodology, in order to assess the contribution of the predictors on the forecasting accuracy and predicted water consumption.

The results obtained show that:

- When past consumption data are not available, household and temporal characteristics can be used to achieve a similar demand forecasting accuracy (MAPE = 19.7%, R^2 = 49.0%) as in the case with known past consumption (MAPE = 17.9%, R^2 = 54.9%). This is of significance to water utilities, as it enables them to make reasonably accurate demand forecasts even for the households where water consumption is not observed. The best performing forecasting model in this case is the model that includes all six household variables as well as the type of day as inputs.
- When past consumption data are included in the demand forecasting model, no other additional variable can significantly improve the prediction results. The reason for this is that the additional information is already embedded in past water use. The best performing demand forecasting model in this case is the one that uses seven days of past consumption and the type of day as input.
- The property's occupancy rate is the most influential input variable, followed by the council tax band and metering status. The acorn group, garden size and rateable value have the smallest effect (Figure 3.7). The weekly pattern of consumption also becomes evident as weekends and holidays have a higher predicted consumption compared to working days (Figure 3.8), although the monthly and seasonal patterns are very weak.

- Although weather input does not improve the model’s accuracy, relationships are identified between water consumption and air temperature, sunshine duration, humidity and to a lesser extent for days without rain. This influence however is limited to only certain household groups and days in the data, and in most cases it is triggered when the weather variable exceeds a certain threshold.

The above results can assist with the effective targeting of water conservation strategies and the development of improved water demand forecasting models. However, they are not without certain limitations.

This study was performed using a certain level of temporal (daily) and spatial (~3.8 households/group) aggregation, which might have influenced the results. Increasing the level of spatial aggregation decreases the range of demand values, as it decreases the randomness of individual household use and thus it should reduce forecasting errors. In addition, it is possible that the variable importance also changes at different aggregation levels. This is the focus of chapter 5, which will explore how the forecasting accuracy and variable importance varies over different scales.

Finally, due to its accuracy, transparency and ease of implementation, a RF model was selected for this analysis. However, results may improve if a different model is used instead. Chapter 4 focuses on comparing and assessing the accuracy of a variety of models, for different forecasting goals. This will help identify the best performing model, with respect to the forecasting aim.

CHAPTER 4

A NEW METHOD FOR WATER DEMAND FORECASTING

This chapter was submitted as a Research Article to Urban Water Journal (ISSN: 1744-9006). This publication has been slightly modified in order to improve consistency throughout the thesis. The chapter was written by Maria Xenochristou but has benefited from the comments of the co-authors, Zoran Kapelan, Chris Hutton and Jan Hofman.

Citation: Xenochristou, M., Hutton, C., Hofman, J., and Kapelan, Z. (2019). An ensemble stacked model with bias correction for improved water demand forecasting. *Urban Water Journal* (under review).

4.1. Introduction

Satisfying the water supply-demand balance is a major challenge in many countries and a topic of increasing concern in the UK. Efforts related to control and management of water networks using modelling techniques are not new and have been the topic of extensive research (Brdys and Ulanicki, 1994). According to the government's water strategy for England report (Defra, 2008), an essential aspect of managing water demand is by ensuring a good forecasting of future patterns. However, forecasting demand is a challenging task, due to the nature and quality of the available data, the numerous factors that influence consumption and the various forecast horizons and spatial scales (Mamade et al., 2014).

With the advancement in technology and computing power, as well as the increasing data availability, machine learning has become a popular approach for water demand forecasting (Froukh, 2001; Cutore et al., 2008; Firat et al., 2009; Bai et al., 2014; Bakker et al., 2014; Romano and Kapelan, 2014; Shabani et al., 2016). There is currently an abundance of methods and models available, from the more researched Artificial Neural Networks (ANNs) to the relatively newer concept of ensemble machine learning.

ANNs have been proven effective to predict short-term, medium-term and long-term demand (Bougadis et al., 2005; Adamowski, 2008; Firat et al., 2009; Herrera et al., 2010; Dos Santos and Pereira, 2014; Mouatadid and Adamowski, 2017; Ghiassi et al., 2017; Altunkaynak and Nigussie, 2018). Adamowski (2008) used an ANN to predict peak daily water demand for ~77,500 consumers in the city of Ottawa and found it performed better (R^2 = 69%) than multiple linear regression and time series analysis. Dos Santos and Pereira (2014) tested eight model configurations of an ANN (3-layer, feed forward, back propagation) for short-term water demand forecasting using weather and temporal characteristics. The ANN was compared with multiple linear regression for hourly predictions at a large metropolitan area in Sao Paulo, Brazil. The best performance was obtained for the ANN that implemented 12-hour averages of the input variables and past consumption data as explanatory factors (R^2 = 67.9%). However, the authors argued that the model could benefit from additional input variables. Ghalehkhondabi et al. (2017) reviewed the water demand forecasting literature between 2005 and 2015 and concluded that although soft computing techniques have been extensively used, deep neural networks (DNNs) have yet to be tested.

In recent years, some of the most successful models in machine learning competitions have been ensemble methods, which create a strong learner by combining multiple, individual, weak learners. There are three ensemble techniques, bagging, boosting and stacking. Bagging is a resampling technique that randomly chooses a sub-sample of the dataset with replacement for training each learner (Mao, 1998). An example of a commonly used bagging algorithm is Random Forests (RFs) (Breiman, 2001), which are based on training multiple decision trees on different samples of the original training set. Boosting is also a resampling technique, but in this case the instances of the training data that got misclassified from previous learners gain additional weight, while the ones that were classified correctly lose weight. This way, the model gradually becomes better, as it focuses on harder areas of the problem. Gradient Boosting Machines

(GBMs) are an example of a commonly used machine learning algorithm that uses this method. Finally, stacking is the process of feeding the outputs of different machine learning models (base models) into one meta-learner (Ngo, 2018). Stacked models have been found to outperform individual models, since they combine the strengths and reduce the negative capabilities of their individual counterparts.

Although proven to perform better than their base models, ensemble techniques have been very rarely tested in water demand forecasting studies (Ghalehkhondabi et al., 2017). Herrera et al. (2010) used RFs for forecasting hourly water demand for a region of ~5,000 consumers and found them to perform worse than Support Vector Regression (SVR), Multivariate Adaptive Regression Splines (MARS) and Projection Pursuit Regression (PPR). However, since not all parameters of the RFs were properly tuned, results could potentially improve. Tiwari et al. (2016) assessed the capacity of extreme learning machines (ELMs) alone, or combined with Wavelet analysis or bootstrap method and compared it with traditional ANN models. The aim was to forecast urban water demand for one day lead for the city of Calgary (~1.1 million consumers). The combined ELM-Wavelet (ELMw) model performed best for short-term forecasting and peak demands, with smaller errors and less computational time. However, in this study there was a clear tendency in all models to over-predict the lower consumption days and under-predict the days with high consumption. Chen et al. (2017) also used RFs as well as a combined Wavelet transform to predict daily water consumption for a supply area of 170,000 households and found that although the combined model performed better (R = 80%), it was still not capable of predicting the daily variations in water demand. Finally, Duerr et al. (2018) compared several time series and machine learning models, including RFs and GBMs, for monthly predictions at the household level and found that machine learning models generally underperformed when predicting monthly averages. However, the authors pointed out that improved data collection, high-resolution covariates, demographic information, as well as capturing the spatial dependence between neighbouring households could improve results.

As it becomes apparent from the above, although machine learning methods have been commonly used for water demand forecasting, the classical methods cannot produce the most accurate results (Ghalehkhondabi *et al.*, 2017). Even when consumption is aggregated at high temporal (e.g. monthly or quarterly) or spatial (e.g. city level) scale, the models commonly used in the literature struggle with accuracy, bias and peak day predictions. Models based on deep learning and ensemble techniques, particularly model stacking, have been consistently found to produce excellent results in other fields. However, they have attracted very little to no attention in the water demand forecasting literature. Even when explored, essential aspects of the modelling and evaluation process like the tuning of the model's parameters or the assessment of its ability to predict outliers are often overlooked.

This chapter aims to address this gap by developing a new methodology based on model stacking and bias correction. This methodology is compared with a selection of ensemble and deep learning models using real data from the UK. A detailed description of the data used in this study is provided in the next section. Then, the overall structure and characteristics of each model are outlined, followed by the bias correction methods. The same section also includes details about the technical implementation of the models, such as the software, programming language and open-source tools. This is followed by the results of the study, in terms of modelling accuracy for all days as well as peak days. Finally, the chapter concludes with a discussion of key findings, followed by a summary of results, conclusions and recommendations for further research.

4.2. Data

An essential aspect of developing machine learning models is getting access to sufficient, high quality data. This study uses real data from the southwest of England (Figure 4.1) that are available at very high temporal and spatial resolutions. Specifically, the dataset comprises of past consumption data and partial postcodes that became available by Wessex Water, one of the UK water companies. In addition, weather data were provided by the Meteorological Office of the United Kingdom (Met Office).

Water consumption data were collected at the household level using smart meters. The smart metering modules recorded consumption every 15-30 minutes over a period of three years (10/2014 – 9/2017), from 1,793 properties scattered around the study area. These data were cleaned and pre-processed in order to remove inconsistencies, errors, empty properties and water-supply leakage. A detailed description of this process is available in chapter 2. For each household in the dataset, a partial postcode indicates its approximate location. The study area includes six postcode areas, with up to 212 properties/day, depending on data availability on the corresponding day and postcode. In order to smooth out the consumption signal, water consumption is aggregated at the daily scale (1,019 days in total) among houses with the same postcode. A spatial analysis of the dataset concluded that smaller groups of properties are associated with increased forecasting errors, thus days and postcodes with less than 60 properties were excluded from the data. This resulted in 5,063 groups with 120 properties/day on average.

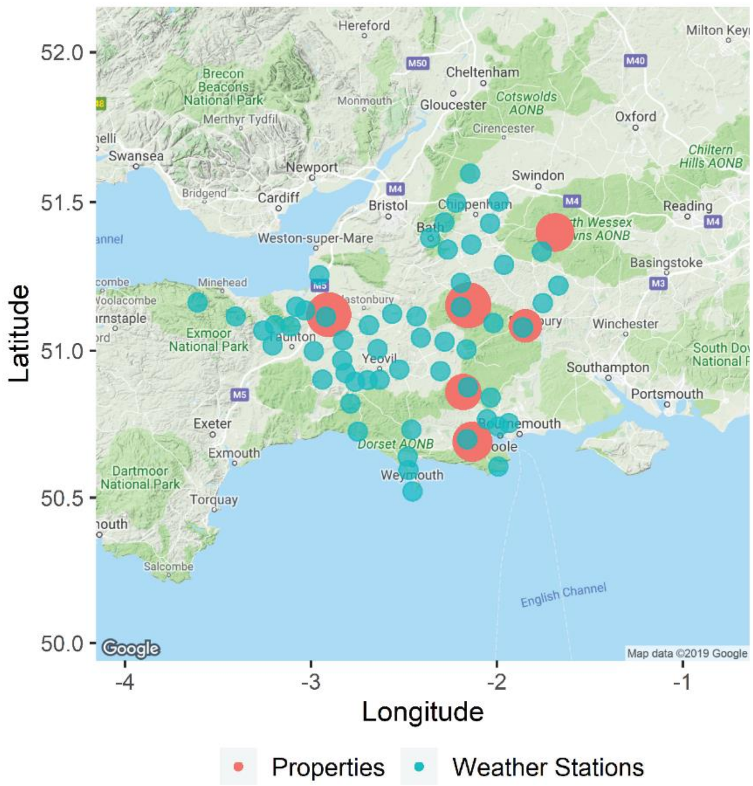

Figure 4.1. Location of property areas (red) and weather stations (blue).

The weather dataset includes four weather variables, maximum air temperature, mean soil temperature at 10cm depth, mean relative humidity and total rainfall. This data was recorded at the hourly or daily scale from hundreds of weather stations across the study area as part of the MIDAS (Met Office Integrated Data Archive System) dataset (Met Office, 2006a; Met Office, 2006b; Met Office, 2006c; Met Office, 2006d; Met Office, 2006e). In addition, the number of consecutive days without rain is calculated based on the daily rainfall. The values recorded at multiple weather stations are combined using weights, based on the station’s proximity to the properties in the study area. Weather stations that are located closer to the properties are assigned a higher weight whereas weather stations with no households in close proximity (closer than any other weather station) are removed from the analysis. Weather records that were not quality checked by the Met Office are also excluded.

4.3. Methodology

4.3.1. Model inputs

All demand forecasting models have a single output (or response) variable and a variety of inputs (or predictors). The predictor variables are a selection of explanatory factors that can influence water use and thus explain part of the variance in the model. In this case, the response variable is the water consumption one day into the future, at a given postcode area. The model inputs are past consumption data, area postcodes, temporal and weather characteristics (Table 4.1).

Two model input configurations are tested in the following, one that includes all predictors (Group 1, Table 4.1) and one that excludes past consumption data (Group 2, Table 4.1). In terms of the practical value of this work, it is important to note that many water utilities do not have access to high resolution consumption records, at least not for the whole extent of their network. Therefore, it is essential when evaluating the best model to also account for its ability to deal with the absence of past consumption data.

Table 4.1. Input variables used to train each group of models.

Variable Group	**Model Input Variables**	**Group 1**	**Group 2**
Past Consumption	1-7 days prior	X	
Temporal	Type of Day	X	X
	Season	X	X
Postcode	Area Postcode	X	X
Weather	Sunshine hours	X	X
	Air Temperature	X	X
	Humidity	X	X
	Days without rain	X	X
Total Variables		**14**	**7**

Each input variable describes a different aspect of water demand variability. Water consumption is highly autocorrelated from one day to the next one, therefore a sliding window of 7 days (one input variable for each day) is chosen to capture the weekly repetition of water use. The postcode is also considered a valuable predictor, since the location of a property is associated with certain socio-economic status and property characteristics that can also influence water behaviour (see chapter 3). Finally, previous work (see chapter 3) concluded that both time-varying factors and weather changes can influence demand. Therefore, the type of day (working day vs weekend/holiday), the season, as well as four weather variables (sunshine hours, air temperature, humidity and days without rain) are used as explanatory factors in the models. Since rainfall and days without rain are highly correlated and previous research (see chapter 2) concluded that rainfall has little influence on water consumption, only the number of days without rain is used as model input.

4.3.2. Model tuning and assessment

Initially, the dataset is shuffled and randomly divided into a training (70%) and a test (30%) dataset. The training set is used to fit and tune the model whilst the test dataset is used to assess the model's ability to perform predictions on unseen data, i.e. data that is not used during the model-building phase.

4.3.2.1. Model tuning

The hyperparameter tuning step is a vital part of building an efficient machine learning model. It assists with defining a set of input parameters that influence the model structure and thus the results. The available parameters for tuning depend on the type of model and can determine how closely the model will fit on the training data. Fitting too closely could mean that the model learns from the noise in the training dataset (overfitting), which will result in a poor prediction on the test dataset. On the other hand, fitting too loosely (underfitting) means that the model has not learnt to represent the patterns in the data.

The models here are tuned for the optimum combination of hyperparameters using a 5-fold cross validation process (Zhang, 1993). This means that in every run, the training data is shuffled and divided into five parts, out of which four are used for training and one for testing. This ensures the model's performance on different sets of data and enhances the robustness of the hyperparameter selection.

Although there are different approaches to select the hyperparameter values (e.g. grid search, random search and evolutionary optimisation), a random search as well as a simple grid search are used here, depending on the number of hyperparameters that need tuning at a time and the tools available. In a grid search, a number of values are defined for each parameter, creating a multi-dimensional grid search space, where each dimension is one variable. In a random search, the hyperparameters are sampled from a pre-defined range of values. Each candidate model is built on a unique set of hyperparameters and the best model is chosen as the one that achieves the lowest error on the test dataset.

The 'autoML' module of the 'h2o' platform can train a number of machine learning models (RF, XRT, GBM, DNN and GLM), as well as tune some of them (GLM, GBM and DNN) for the optimum set of hyperparameters. The model training stops according to a variety of stopping criteria. In this case, these were the stopping tolerance (0), stopping metric (MSE) and stopping rounds (1). This means that 'h2o' stops running when the MSE does not improve more than zero, over two consecutive iterations (for the same or different models). In addition, the maximum runtime is set to two hours, which means that the program stops running and saves the models developed up to this point, if none of the above criteria have been fulfilled.

During this time, 'h2o' trained 335 models without past consumption data and 147 models including past consumption data. Since additional variables add complexity to the model, they consequently increase training time, leading to less than half of models being trained within the same time frame.

Out of the six model types that are presented here, three of them (Random Forests, Extreme Gradient Boosting, Artificial Neural Networks) are tuned using a pre-defined grid search space, whereas the Generalised Linear Model, Gradient Boosting Machine and Deep Neural Network are tuned automatically by 'h2o' using a random search.

4.3.2.2. Model assessment

Three performance criteria are used to assess the model's performance: the mean absolute percentage error (MAPE), mean square error (MSE) and R^2 coefficient of determination. Each one of these provides slightly different, i.e. complementary information about the model's performance. The MAPE is one of the most common metrics, as it is easy to interpret and it scales the error in relation to the actual value. The MSE is sensitive to outliers, while the R^2 shows the variance in the dependent variable (model output) that can be explained by changes in the independent variables (model inputs) (Xenochristou et al., 2019a).

4.3.3. Modelling techniques

A number of modelling techniques such as neural networks and linear models, as well as representatives from every family of ensemble algorithms (bagging, boosting and stacking) are considered in this study. The following is an extensive list of all models that are used, either as a prediction tool or as a component of the stacked model.

4.3.3.1. Random Forests

Random Forests (RFs) were first introduced by Breiman (2001) as an ensemble of (hundreds or thousands) of decision trees. The unique value of RFs is partly due to the implementation of randomness in the modelling process (Herrera *et al.*, 2010). A RF model trains each tree on a slightly different set of data, whilst at each split of the tree it chooses among a different subset of input variables. The final result of the forest is calculated as the mean prediction among all the trees. RFs have been consistently found to perform better than other machine learning techniques while being a method that has not been fully explored in the water demand forecasting literature (Herrera et al., 2010; Chen et al., 2017).

There are three main parameters that need tuning in RFs, the mtry, ntrees and tree depth (Scornet, 2017). The mtry is the number of variables randomly selected at each node and considered for splitting. Reducing the mtry increases the randomness of the tree-building process and therefore creates trees that are less similar to each other. The ntrees parameter is the number of trees used to build the forest. Model accuracy typically plateaus after a number of trees that are required to build a credible model. The tree depth is the point at which the tree should stop growing, sometimes also denoted by the size of the final tree node (nodesize). The higher the tree depth, the closer the model fits on the training data, thus increasing the risk of overfitting.

The optimum (and default) value in regression for the number of random variables used for splitting (mtry) at each node is often considered to be the total number of input variables divided by three. According to Table 4.1, the total number of variables is 14 for the models in Group 1 (with past consumption) and 7 for the models in Group 2 (without past consumption). Therefore, the mtry range tested for Group 1 is 3-7, while for Group 2 is 2-4. The number of trees is varied from 120 to 240, whereas the node size is varied from 20 to 120.

Extremely Randomized Trees (XRT) are a variation of RFs that introduce added randomness in the above process. Similarly to RFs, a random subset of variables is selected for splitting at each node, but in this case a number of cutting-points (thresholds) are also selected at random. The best of these randomly selected thresholds is chosen for splitting at the node. The level of randomness implemented in the process can be tuned and is controlled by the model parameters. In the extreme case, the trees are built completely at random, independent of the training sample (Geurts et al., 2006).

4.3.3.2. Gradient Boosting

Gradient Boosting Machines (GBMs) were first introduced by Friedman (2001) as an implementation of gradient boosting that explicitly deals with regression problems. In the GBM implemented here, the base learner is also a decision tree. The boosting algorithm starts with one tree and at each iteration step, a new decision tree is fitted on the residuals of the previous tree and subsequently added to the model (Touzani et al., 2018). This is an iterative process that is built as a simple optimisation problem, where the objective is to minimise the loss function, i.e. the model error. Since the new trees are trained on the residuals of the old trees, the model focuses on areas of the problem that did not perform well (Touzani et al., 2018). A shrinkage rate can also be applied on the algorithm, meaning that the new trees that are added to the model are gradually assigned lower weights. This increases the steps required for the algorithm to converge to a solution and reduces the risk of overfitting. The final result of the GBM is the weighted sum of the individual trees that were trained on weighted parts of the dataset (based on the accuracy achieved at the previous step).

There is a variety of hyperparameters available for tuning GBMs that aim to assist the algorithm with arriving at the best solution, by implementing randomness in the modelling process or avoiding overfitting. In addition to the number of trees (ntrees), maximum tree depth (max_depth), and number of variables sampled for splitting (col_sample_rate), the number of variables sampled for each tree (col_sample_rate_per_tree) is also a hyperparameter. The number of variables sampled at each node is then calculated as the product of the variables sampled for the tree, multiplied by the variables sampled for splitting. The learning rate of the algorithm (learn_rate) is the factor by which the contribution of each consecutive tree is reduced compared to the previous tree. Another parameter (histogram_type) defines the type of histogram used to sample values that are tested for splitting at each node, thus speeding up the selection of the best splitting point. The subsample size (sample_rate) determines the size of the random sample used to train a new tree at each iteration. Smaller samples result in lower testing errors whereas higher samples improve the training accuracy. Finally, two hyperparameters determine if a further split in a tree will occur, based on the minimum required relative improvement in squared error (min_split_improvement) and the minimum number of observations in a leaf node to allow further splitting (min_rows). More details regarding the implementation of the GBM algorithm can be found in Malohlava and Candel (2017).

A total of nine hyperparameters are tuned for the GBM model, using the ‘h2o autoML’ platform. The selected hyperparameter values for the models with and without past consumption data appear in Table B2. The ‘auto’ histogram type means that the cutting points tested for splitting are chosen by dividing the range of values of each variable in equal steps. Here, the values tested for splitting are selected by dividing the variable range into twenty equal steps.

Extreme Gradient Boosting (XGBoost) is another implementation of a boosting algorithm. It was introduced by Chen and Guestrin (2016) as ‘an efficient and scalable implementation of the Gradient Boosting framework by

Friedman (2001)' (Chen and He, 2015). XGboost aims to prevent overfitting and maximise the efficiency of computer resources (Fan *et al.*, 2018). According to Chen and Guestrin (2016), 17 out of the 29 winning solutions published by Kaggle, an online coding competition platform, used XGBoost, either as a single model or as part of a stacked model.

The number of iterations (nround), the subsample size (subsample), maximum tree depth (max_depth) and fraction of explanatory variables sampled at each tree (colsample_bytree) are also hyperparameters of the XGBoost algorithm. In addition, the shrinkage rate (eta) defines the learning rate of the algorithm in the training step, i.e. the amount by which the contribution of each consecutive tree is reduced compared to the previous tree. Additional parameters that need tuning for this algorithm are the gamma and min_child_weight that determine how conservative the algorithm is in terms of further partitioning at a leaf node. The larger these parameters, the more conservative the algorithm. More details about the implementation of the XGBoost package can be found in Chen and Guestrin (2016).

The XGBoost model is tuned using a user-defined grid search space. However, the number of hyperparameters required for the XGB algorithm makes it difficult to define an extended search range for each parameter, due to the high dimensionality of the problem. Here, the XGBoost algorithm is tuned for six input parameters, whereas the search range for each parameter and the selection of the subsample size are based on trial and error.

4.3.3.3. Artificial Neural Networks

Artificial Neural Networks (ANNs) are a family of machine learning algorithms inspired by nature, specifically biological neural networks, and are comprised of nodes, organised into layers. Each node receives information with a certain weight from another node or external stimuli, transforms it and then passes it to the next node, or transfers it as external output (Zhang et al., 1997). Nodes that belong in the same layer, work collectively within the same depth of the network. The higher the number of layers, the deeper the ANN.

The ANN implemented here is a feed-forward, single hidden layer network. This means that information travels through the network one way, from the input, through the hidden layers and to the output layer, calculating the model weights through this learning process. It is tuned for the number of units (size) in the hidden layer, as well as a gradient decay (decay), i.e. a factor less than one by which the weights are multiplied at each iteration of the algorithm.

The ANN implemented here has a single layer. The hyperparameters used for tuning the model are the size of the hidden layer, which is varied between 5 and 20, with a step of 1, and the decay, which is varied between 0.01 and 0.1 with a step of 0.001. The above ranges for the grid space are chosen based on trial and error.

Deep Neural Networks (DNNs) are ANNs composed of multiple layers, which allow them to transform information and learn from data with multiple abstraction levels (LeCun et al., 2015). The DNN implemented here is a multi-layer, feedforward ANN trained using stochastic gradient descent and back-propagation (Candel et al., 2014). In back-propagation, the model's error is fed back into the model in order to update the weights and further improve results. This process evolves as an optimisation problem, where the objective is to minimise the model's error using stochastic gradient descent (Bottou, 2010).

Although there are many hyperparameters in a DNN, the following eight are tuned using a random search in this study. The number of epochs indicates how many times the whole dataset, divided into smaller batches, will go back and forth through the neural network during the training process. The higher the number of epochs, the higher the risk of overfitting while too few could lead to underfitting. The activation functions (activation) transform the input in a node to a certain output, while the size of the hidden layers (hidden) determines the number of nodes in each one. The dropout ratio of the input (input_dropout_ratio), as well as the dropout ratio of the hidden layers (hidden_dropout_ratio) aim to prevent model overfitting. At each training example, they suppress the activation of the nodes in the input or hidden layers by a certain probability (dropout ratio). As a result, each training example creates a different model. The combination of these learners resembles an ensemble model (Candel et al., 2014). There is also the option to activate an adaptive learning rate (adaptive_rate) method for gradient descent that determines how quickly the algorithm converges to an optimum solution. The momentum of the learning rate is determined by two more hyperparameters, the rho and epsilon (Candel et al., 2014). The DNN model is tuned for eight hyperparameters, using a random search implemented by the ‘h2o autoML’ platform.

More information regarding the algorithm implementation and tuning parameters of the DNN can be found in Candel et al. (2014).

4.3.3.4. Generalised Linear Models

Generalised Linear Models (GLMs) are an extension of simple linear models, for errors that do not follow the normal distribution or predictors whose influence is not linear (Aiello et al., 2016). GLMs typically create regression models that follow an exponential distribution (Aiello et al., 2016).

There are two parameters tuned for the GLM, one that determines how the model deals with missing values and the alpha regularization parameter. The value of alpha determines the penalisation function used in order to avoid model overfitting, reduce the variance in the error and deal with correlated predictors (h2o.ai, 2019a). More information regarding the meaning of these two parameters can be found in Nykodym et al. (2019).

The GLM model is tuned for two hyperparameters using a grid search. The alpha hyperparameter is varied between 0-1, with a step size of 0.2. An alpha value of zero indicates that a ridge regression (regularised linear regression) model is used to introduce penalties to the model building process, while ‘MeanImputation’ means that the model replaces missing values with the mean.

4.3.3.5. Model stacking

Stacking is the process of combining the results of individual learners into one super-learner. The way of combining them could be using a simple weighted average or a machine learning model such as a RF or ANN to learn the best combination based on the residual errors.

4.3.4. Bias correction methods

The concept of model bias is well-documented in the machine learning literature (Zhang and Lu, 2012; Nguyen et al, 2015; Song, 2015; Ghosal and Hooker, 2018; Hooker and Mentch, 2018). Especially in methods such as RF, where the final prediction is estimated as the mean among the predictions of the individual trees, the range of the prediction values becomes smaller due to averaging, compared to the actual range. This leads to overestimating the smaller values and underestimating the larger values in the dataset, referred to as bias towards the mean in the following. As opposed to the above, which is a fundamental statistical concept, the systematic bias in the model's results refers to a consistent overprediction or underprediction of the response variable. A well-performing model should ideally exhibit a zero or near-zero systematic bias.

In this chapter, four methods for bias correction (BC) described in Song (2015) are tested for their ability to reduce the bias towards the mean. In the first BC method (BC1), a RF model is used to predict the residual errors based on a set of predictors in the training dataset that include the predicted values of the response variable. The final prediction of the model is then adjusted by adding the predicted residuals to the predicted outcome. In the second BC method (BC2), a simple linear model is fitted on the residuals of the training set but this time only the predicted values are used as input. The same linear model is then used to predict the residuals in the test dataset. As with the first method, the final prediction is calculated by adding the residuals to the model's output to adjust it. BC methods 3 and 4 (BC3 and BC4) use a residual rotation approach. They first calculate the prediction and the residuals based on BC1. Then a simple linear model is fitted on the residuals against the predicted values. In BC3, the residuals are rotated so that y=0, while in BC4 the best rotation angle is determined sequentially as the one that achieves the minimum MSE.

An extensive description of the four methods can be found in Song (2015). The code used for the implementation of the four BC methods is adapted by Song (2015).

4.3.5. Technical implementation

All models, analysis and results produced in this work are created using R (R core team, 2013). The RF, XGBoost and ANN models are trained using the algorithms implemented in the 'randomForest' (Liaw and Wiener, 2018), 'xgboost' (Chen et al., 2019) and 'nnet' (Ripley and Venables, 2016) packages, respectively. All three models are tuned using 'caret' (Kuhn, 2019), which allows to perform a grid search for the optimum hyperparameter values. The GBM, DNN, GLM and stacked models are built using an open source machine learning platform, ‘h2o’, and specifically its automated machine learning capability (autoML). This is accessed through an R interface using package ‘h2o‘ (LeDell et al., 2019). The ‘autoML’ function of ‘h2o’ can automatically train a selection of models and perform hyperparameter tuning within a user-defined limit. This method is implemented due to its high performance, speed, automation and efficiency.

The ‘h2o’ platform currently provides support for automated implementation of five machine learning methods, RF, XRT, GBM, DNN, GLM and in some cases also for the XGBoost algorithm, which is not available here. However, it only tunes the GBM, DNN and GLM models over a random grid, whereas it uses default versions of the XRT and RF models (h2o.ai, 2019a). In addition to this, ‘h2o autoML’ trains two stacked ensemble models. The first stacked model includes the best combination among a selection of model types, including multiple models from the same family (e.g. RF) that are trained as part of the hyperparameter tuning process. The second stacked model is based only on the best model from each family (h2o.ai, 2019a). The metalearner algorithm that is used to combine the models for the automated machine learning capability of ‘h2o’ is a GLM model with non-negative weights. Only the three properly tuned ‘h2o’ models (GBM, DNN and GLM) are presented in the results section, although both the default XRT and RF are used as components to build the stacked ensemble models.

4.4. Results

4.4.1. Model parameters

The following section outlines the hyperparameter values selected for each model, as a result of the tuning process. The default DRF and XRT implementations (h2o.ai, 2019b) are used to build the stacked model, therefore these are not described in the following.

More details regarding the hyperparameters available for tuning, their meaning, as well as the default hyperparameters of the models that are not mentioned here can be found in the online ‘h2o’ documentation (h2o.ai, 2019c).

The best parameter values for each model type appear in Tables 4.2-4.7.

Table 4.2. Hyperparameter values selected for the RF model, for Groups 1 and 2.

Hyperparameters	Group 1	Group 2
mtry	6	7
nodesize	100	40
ntrees	160	200

Table 4.3. Hyperparameter values selected for the GBM model, for Groups 1 and 2.

Hyperparameters	Group 1	Group 2
ntrees	104	109
max_depth	13	8
learn_rate	0.05	0.05
sample_rate	0.9	0.8
col_sample_rate	0.4	0.4
col_saple_rate_per_tree	0.4	1
histogram_type	Auto	Auto
min_split_imrpovement	1e-04	1e-05
min_rows	10	15

Table 4.4. Hyperparameters values selected for the XGBoost model, for Groups 1 and 2.

Hyperparameters	Group 1	Group 2
nrounds	140	120
max_depth	6	5
colsample_bytree	0.4	0.7
eta	0.05	1
gamma	1	1
min_child_weight	1.3	1.3
subsample	0.6	0.6

Table 4.5. Hyperparameter values selected for the ANN model, for Groups 1 and 2.

Hyperparameters	Group 1	Group 2
size	11	16
decay	0.002	0.006

Table 4.6. Hyperparameter values selected for the GLM model, for Groups 1 and 2.

Hyperparameters	Group 1	Group 2
alpha	0	0
missing values	MeanImputation	MeanImputation

Table 4.7. Hyperparameter values selected for the DNN model, for Groups 1 and 2.

Hyperparameters	Group 1	Group 2
epochs	270.4	131.2
adaptive_rate	TRUE	TRUE
activation	RectifierWithDropout	RectifierWithDropout
rho	0.9	0.95
epsilon	1e-08	1e-08
input_dropout_ratio	0.2	0.1
hidden	500	200 200 200
hidden_dropout_ratios	0.4	0.2 0.2 0.2

All of the above hyperparameters are provided for reference only and for comparison purposes and do not replace the need to properly tune the above models based on the respective dataset.

4.4.2. Model performance

In this section, the forecasting performance of seven models (RF, XGB, GBM, GLM, ANN, DNN and stacked) is compared based on four evaluation metrics, the MAPE for all days as well as peak days, the R^2 and the MSE. For comparison, the error of the 'naïve' model (the model that assumes forecasted consumption for each day is equal to the mean consumption among all days in the dataset) is 10.1% for all days and 19.8% for peak days, i.e. the 10% of the days with the highest consumption. All models are implemented for two different configurations, with (Group1) and without (Group 2) past consumption data as input. In addition to this, four BC methods are applied on top of the best performing model (BC1-BC4). Only the best models acquired from each family after tuning are presented in the following. Table 4.2(a) summarises the results of the models that include past consumption (Group 1), whereas Table 4.2(b) demonstrates the results of the models that include only postcode location, temporal and weather characteristics as input (Group 2).

According to Table 4.2(a), when past consumption is included as input, the model with the best performance (R^2 = 74.1%, MAPE = 4%) is the stacked model created by 'h2o' as an ensemble of five individual learners (the best from each family). Specifically, the stacked model comprises of a GBM, XRT, GLM, DRF and DNN model, with a corresponding contribution to the output of 31%, 24%, 19%, 14% and 12%, respectively. Out of the rest, the GBM (R^2 = 74.1%, MAPE = 4.1%) and RF (R^2 = 72.8%, MAPE = 4.1%) models have the highest forecasting accuracy for all days in the data. The neural network based models have the lowest peak day errors, with a MAPE of 4.8% for the ANN and 5.2% for the DNN. However, the ANN model does not perform equally well for the other two performance metrics (R^2 = 70.8%, MSE = 55). This implies that the reason that the model performes better for peak days might be that it systematically overpredicts consumption, especially due to the high MSE value, which is an indicator of bias in the model. Finally, the GLM is the worst performing model across most metrics (MAPE = 4.2%, R^2 = 70.6%, MSE = 55).

Out of the four BC methods tested here, the second method (BC2, Table 4.2), which predicts residual errors based on the predicted value of the response variable performs best. Although applying the BC2 method on top of the stacked model's results does not improve the overall model performance (Models 7 & 9, Table 4.2(a)), it reduces the MAPE on peak days from 5.1% to 4.6% (Models 7 & 9, Table 4.2(a)).

Table 4.8. Model comparison (a) with and (b) without past consumption as input, for the test dataset, for seven model types and four bias correction methods.

Model Groups	ID	Model Type	Bias Correction Method	MAPE (%) All days		MAPE (%) Peak days		R^2 (%)		MSE (l/postcode/day)	
				Train	Test	Train	Test	Train	Test	Train	Test
(a) Group 1	1	RF	-	1.8	4.1	2.8	5.6	95.5	72.8	10	51
	2	XGBoost	-	3.0	4.2	4.5	6.0	86.3	72.5	27	53
	3	ANN	-	3.9	4.2	4.8	4.8	74.9	70.8	45	55
	4	GLM		4.1	4.2	5.8	5.8	71.3	70.6	51	55
	5	GBM	-	2.0	4.1	2.9	5.4	93.7	**74.1**	12	49
	6	DNN	-	3.5	4.2	4.7	5.2	79.7	72.5	36	51
	7	Stacked	-	2.2	**4.0**	3.2	5.1	91.8	**74.1**	15	**48**
	8	Stacked	BC1	2.2	**4.0**	2.8	4.8	91.4	**74.1**	16	**48**
	9	Stacked	BC2	2.6	**4.0**	3.3	**4.6**	88.7	**74.1**	20	**48**
	10	Stacked	BC3	2.2	**4.0**	3.0	5.1	91.6	**74.1**	15	**48**
	11	Stacked	BC4	2.2	**4.0**	2.9	4.8	91.5	**74.1**	15	**48**
(b) Group 2	1	RF	-	2.3	4.6	3.5	6.0	92.2	68.0	16	60
	2	XGBoost	-	3.3	4.4	4.9	6.1	82.7	70.7	33	55
	3	ANN	-	4.3	4.7	5.9	6.0	68.5	65.1	56	65
	4	GLM		4.6	4.7	6.8	6.8	64.7	63.8	63	67
	5	GBM	-	3.1	**4.3**	4.2	5.6	84.0	70.9	29	54
	6	DNN	-	3.7	4.5	5.4	6.2	76.6	68.5	43	59
	7	Stacked	-	3.0	**4.3**	4.0	5.5	85.5	**71.1**	26	54
	8	Stacked	BC1	2.7	4.4	3.3	**5.1**	87.9	70.2	22	**51**
	9	Stacked	BC2	2.9	**4.3**	3.6	**5.1**	85.5	**71.1**	26	54
	10	Stacked	BC3	2.7	4.4	3.7	5.5	88.1	70.0	22	56
	11	Stacked	BC4	2.7	4.4	3.6	5.4	88.1	70.0	22	56

When past consumption is not included as input (Table 4.2(b)), the best performing model is again the stacked model (MAPE = 4.3% for all days and 5.5% for peak days, R^2 = 71.2%, MSE = 54). This time, it comprises of a GBM, DNN, DRF, GLM and XRT model with a percentage contribution to the output of 53%, 15%, 11%, 11% and 10%, respectively. Adding BC2 further reduces the MAPE to 5.1% for peak days. The second best performing model in this case is again the GBM (R^2 =70.9%, MAPE = 4.3%), which has the same MAPE for all days and slightly higher (MAPE = 5.6%) for peak days. It is worth noting that the

ANN model, which performed relatively well with past consumption input and is the model most commonly used in the literature, underperformed in this case (MAPE = 4.7% for all days and 6% for peak days, R^2 = 65.1%, MSE = 65). Similar results apply for the GLM model, which performed reasonably well with past consumption data (MAPE = 4.2% for all days and 5.8% for peak days, R^2 = 70.6%, MSE = 55), but whose error increases significantly without (MAPE = 4.7% for all days and 6.8% for peak days, R^2 = 63.8%, MSE = 67).

Figure 4.2 demonstrates an example of the actual against the predicted values for two model types, the GLM and stacked-BC2 (stacked with Bias Correction method 2), without past consumption data. According to Figure 4.2, the days with the lowest consumption are most of the times overpredicted, while the days with unusually high consumption are underpredicted. Although this effect is particularly prominent for the GLM (Figure 4.2(a)), it improves in the case of the stacked-BC2 model (Figure 4.2(b)).

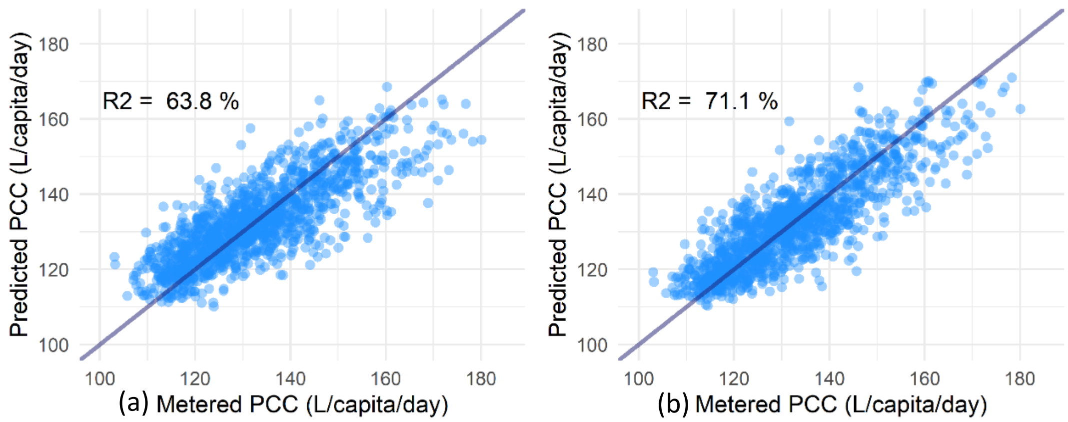

Figure 4.2. Metered against predicted values for (a) the GLM and (b) the stacked-BC2 model, without past consumption as input.

Overall, predicting demand becomes slightly more challenging when past consumption data is not available, as well as for peak days (Table 4.2). However, certain models are able to deal significantly better with the lack of additional information (e.g. XGB, GBM) compared to others (ANN, GLM). The method that seems to be affected the most with forecasting demands without past consumption is the method that is frequently suggested as best in the literature - the ANN model. The MAPE for this method increases for the peak days from 4.8% to 6%, when comparing the models with and without past consumption as input. Finally, although slight differences exist, most models have very similar results for all days in the data, with a range in MAPE between 0.2% (with past consumption) and 0.4% (without past consumption) across the test dataset (MAPE – All days, Table 4.2). However, the range of errors increases significantly for peak days, i.e. the 10% of the days with the highest consumption, with a range in MAPE of 1.4% (with past consumption) to 1.7% (without past consumption) (MAPE – Peak days, Table 4.2).

4.5. Discussion

One of the main observations of this study is the power of stacked models to improve the prediction accuracy of their counterparts by adding up their individual strengths and overcoming their weaknesses. However, there is a time and cost sacrifice to be made in exchange for improving the results' accuracy. No machine learning technique is universally best for all types of data, purposes and datasets. Therefore, it is important to account for the computational power, effort and expertise that is required to identify and tailor the machine learning technique that will produce the best outcome.

Another important point is the level of transparency and interpretability associated with each model. Generally, the fewer the number of model parameters, the simpler the model, therefore the easier it is to understand, explain and interpret. According to Molnar (2019a), transparency refers to understanding how the algorithm learns from the data and is independent of the trained model, whereas interpretability is the knowledge of how the model makes decisions, based on its features, weights and parameters. A linear regression model is transparent as the way the algorithm is built is thoroughly explored and understood and at the same time it is interpretable, as the weight of each predictor indicates its influence on the response variable. DNNs on the other hand are neither transparent nor interpretable due to the complexity and number of hyperparameters and hidden layers (Molnar, 2019a). Tree-based models are relatively easy to interpret and explain as they are essentially an ensemble of decision trees. Stacked models can achieve high accuracy as they combine the strengths of different models but at the same time they have limited interpretability, as they lack a model structure. In some cases, sacrificing some accuracy in order to increase the level of model interpretability is the preferred solution.

Another interesting concept that has not been highlighted in previous water demand forecasting attempts is the concept of bias towards the mean. This is a combination of the elementary statistical concept of regression towards the mean, which is often exaggerated by certain model structures (e.g. RF), prone to create biased results (Zhang and Lu, 2012). Regression towards the mean is the term for a statistical phenomenon that can be illustrated by a simple example as follows. For an extreme measurement of a variable, e.g. an unusually high daily temperature, it is unlikely that a second measurement will result in a similar or higher value. The most likely scenario is that the second measurement is going to be closer to the mean annual temperature. Another example described by Stigler (1997) is a student that scored really high at a test. In order for this high score to occur, it is likely that not only skill, but also luck was involved, a factor that might diminish if another test was taking place, resulting in a lower score. A similar concept can be applied to water demand. In order for a very high consumption to occur on a certain day for a population of 120 households, a number of factors need to contribute. For example, chapter 1 concluded that an affluent area on a Saturday with high air temperature, is likely to result in high consumption. However, there are a number of additional factors that will determine how high exactly. This means that although days with the same weather characteristics, the same past consumption, in the same location, are likely to have a higher than normal demand, for only one of these days consumption will be high enough to be an outlier in the data. As the model learns from all days that had the same characteristics, but not as an extreme consumption, the predictions are likely to gravitate towards mean values. This will naturally result in underpredicting and overpredicting the highest and the lowest values in the dataset, respectively. This effect is exaggerated by certain models such as RFs due to their structure, which is based around averaging among hundreds or thousands of individual predictions. Stacked models on the other hand, are able to deal with outliers much better. A simple bias correction technique could achieve an additional reduction in errors for the days with the highest consumption. Therefore, being aware of the problem and choosing wisely the model structure and the tools available could significantly improve predictions on critical days.

This research also demonstrates how a simple tool, ‘h2o.ai’, can assist with the water demand forecasting model development process. As machine learning becomes the mainstream approach in many sectors, there is an increasing need for people that are not trained in the field of computer science to use these tools efficiently. The ‘h2o’ platform can be useful not only in order to choose the best algorithm, but also in order to efficiently tune the model’s hyperparameters. One of the problems in previous studies was the lack of proper tuning of the machine learning algorithms that were used for forecasting. In addition, creating a grid space for hyperparameter tuning is a brute-force approach that is time-consuming and not computationally efficient for high-dimensional problems, even when it is parallelised, while it requires a thorough understanding of the model parameters and how exactly they influence the results. Using the ‘autoML’ function of ‘h2o’, even when the preferred algorithm is known, could significantly reduce complexity, computational time, as well as improve the model’s results.

4.6. Summary and conclusions

This study explores the potential of a stacked ensemble model with added bias correction (BC) to produce improved water demand forecasts. The proposed model is compared with several traditional (e.g. GLM, ANN) as well as emerging (e.g. DNN, GBM, XGB) methods in the water demand forecasting literature. Finally, the potential of automating this process using the machine learning platform ‘h2o’ is explored and compared to model development using methods that require extensive user engagement and expertise.

Results show that the new methodology performs best, especially for peak days and lack of past consumption data. The MAPE of the stacked-BC2 model (stacked model with bias correction method 2) is 4% for all days and 4.6% for peak days, when past consumption data is included as input, as opposed to 4.3% and 5.1%, respectively, when past consumption data is not available.

The GBM model has a similar prediction accuracy (MAPE = 4.1% for all days and 5.4% for peak days), especially when past consumption data is not available (MAPE = 4.3% for all days and 5.6% for peak days). At the same time, the GBM model turned out to be quicker and easier to build since it requires tuning only one set of parameters. The stacked model on the other hand requires the development and tuning of multiple individual learners that are combined to create a super-learner. The GBM model also has a higher level of transparency and interpretability. This means that in situations where demand forecasting accuracy is not of the utmost importance, the GBM model is a viable alternative to the stacked model.

Depending on the scenario, in terms of the data availability and forecasting goal, the choice of model could significantly alter results. For easier tasks (e.g. when past consumption data is available and when the focus is not on predicting outliers) most models perform well. However, in situations where data availability is limited and the goal is to predict days with abnormal consumption, different models produce a wide range of accuracy. Specifically, when predicting demand using past consumption data over all days in the dataset, all models perform very similar with a range in MAPE from 4.0% to 4.2%. However, when focusing on harder aspects of the same problem, e.g. when past consumption data is not available and for peak consumption days, the MAPE among different models varies from 5.1% (stacked-BC2) to 6.8% (GLM), and increase of 33% of the MAPE.

Finally, this study concludes that applying simple techniques such as bias correction on top of the model's results can improve predictions for the peak days. Although most demand forecasting models reached a good accuracy (MAPE lower than 5%), they struggled to predict outliers. This fact could be particularly problematic in the context of water demand forecasting, as days with unusually high consumption are usually the critical ones for water utilities. This technique, although it does not alter the overall accuracy of the model, it improves predictions for the 10% of the days with the highest consumption (Table 4.2).

Although the above models were tested under two scenarios, with and without past consumption data, as well as for peak consumption days, it is not clear how the models would perform with a less rich or more noisy dataset. An uncertainty analysis around the amount and quality of data necessary for each model type to perform well is needed to assess the model's robustness and suitability to produce accurate forecasts under different data availability scenarios.

This chapter focused on identifying models and techniques that can be used to improve predictions in water demand forecasting. However, this analysis was performed at a certain spatial and temporal scale. Chapter 5 uses the above results to explore what is the best accuracy that can be achieved at different spatial scales, as well as assess the contribution of several types of predictors (weather, temporal and household characteristics) at different spatial scales.

CHAPTER 5

WATER DEMAND FORECASTING AT DIFFERENT SPATIAL SCALES

This chapter was submitted as a research paper at the Journal of Water Resources Research (ISSN: 1944-7973). This publication has been slightly modified in order to improve consistency throughout the thesis. The chapter was written by Maria Xenochristou but has benefited from the comments of the co-authors, Zoran Kapelan, Chris Hutton and Jan Hofman.

Citation: Xenochristou, M., Hutton, C., Hofman, J., and Kapelan, Z. (2019). Water demand forecasting accuracy and influencing factors at different spatial scales using a Gradient Boosting Machine. *J. Water Resources Research* (under review).

5.1. Introduction

The effectiveness of future efforts, technologies and conservation strategies in water management depends heavily on accurate predictions of water demand, at the appropriate scale. From emerging technologies (e.g. grey water recycling at the household level) to conservation campaigns (e.g. changing customer's attitudes) or even future investments (e.g. building of new reservoirs), solutions are typically targeted at a certain level of spatial aggregation. Thus, accurately predicting demand at the appropriate scale is of the utmost importance for the success of these solutions.

As part of the commitment to sustainably manage their water resources and to reduce their environmental impact, water companies are required to reduce per capita consumption (PCC) and leakage (Ofwat, 2017). According to the Office for National Statistics, PCC in the UK is the 5th highest in the EU, amounting to a total of 114 litres/capita/day (Bailey, 2019). Leakage also remains at relatively high rates, as approximately 23% of the total inflow into the network is lost through leaks (Ulanicki et al., 2009). Ofwat, one of the UK water industry's regulators, has challenged water companies to reduce this figure by 15% by 2025 (Ofwat, 2019).

Over meetings and discussions with water companies in the UK and the Netherlands that took place during this study, leakage was often brought up as one of the most prominent problems in the water industry. Operators can choose to estimate leakage at different reporting levels, such as district meter areas (DMAs), water resource zone levels or even an intermediate zone level within the distribution network (Ofwat, 2018). In order to do this, they need to be able to accurately forecast water demand at different levels within the network. Therefore, the forecasting accuracy that can be achieved at each level, as well as the factors that determine it need to be assessed. This will allow water companies to make informed decisions and their regulator to accurately assess their performance.

However, predicting water demand is not an easy task as there are many uncertainties involved in the process. The main challenges arise due to the tight relationship between the human and natural systems in urban environments, where more than half of the population currently resides (House-Peters and Chang, 2011). Furthermore, the maximum prediction accuracy that can be achieved, as well as the most influential explanatory factors, can vary greatly depending on the spatial scale. When aggregating large areas, the demand signal is fairly smooth, since it averages out over a large number of water users. On the other hand, small levels of spatial aggregation are likely to be associated with increased variability, due to small-scale water use, leading to a higher uncertainty and thus increased errors.

This study aims to answer two main questions:

- What is the maximum demand forecasting accuracy that can be achieved at different spatial scales?
- What are the most important influencing factors at each spatial scale?

In order to do this, several GBMs are trained here using different sets of explanatory factors as input, with the aim to predict consumption 1-7 days into the future, for different household group sizes. Chapter 4 compared multiple machine learning models and concluded that Gradient Boosting Machines (GBMs) combine high prediction accuracy with ease of implementation, thus they are the models selected for this analysis.

The rest of this chapter is organised as follows. The next section discusses the results and shortfalls of previous studies that implemented some sort of spatial variability in their water demand forecasting models. This is followed by a brief description of the dataset and an overview of the model building process. The results of the study present the modelling accuracy that is achieved at each spatial scale, along with the corresponding variables of interest. Finally, the chapter concludes with a discussion of the key messages and a brief summary of results, conclusions and recommendations for further research.

5.2. Background

Several studies attempted to predict water demand, using a great variety of data, models, methods and explanatory variables (Prescott and Ulanicki, 2008; Herrera et al., 2010; Adamowski et al., 2012; Tiwari and Adamowski, 2013; Matos et al., 2014; Romano and Kapelan, 2014; Hutton and Kapelan, 2015; Anele et al., 2017; Brentan et al., 2017; Zubaidi et al., 2018; Xenochristou et al., 2019b). Some studies in the literature even accounted for the spatial variability of water demand (Balling at al., 2008; Lee et al., 2009; House-Peters et al., 2010; Polebitski and Palmer, 2010; House-Peters and Chang, 2011; Maheepala et al., 2011; Rathnayaka et al., 2017a; Chen and Boccelli, 2018).

Lee et al. (2010) used space-time variation and projections on population density to forecast water demand for the city of Phoenix over a time-space dependent grid. Although integrating future density estimates in the forecasting methodology improved accuracy, Lee et al. (2010) argued that additional input factors (other than population density) could further improve results.

Rathnayaka et al. (2017a) introduced a model that predicts water end-uses for different types of households at multiple temporal and spatial scales. Although this approach made use of a variety of household, temporal and weather characteristics, it did not deal with consumption at each scale as a separate problem. Instead, the total consumption was constructed by adding the individual end-uses and households at each aggregation level.

A study by Balling et al. (2008) investigated water consumption among census tracts and how this is influenced by several weather variables. Using a variety of explanatory factors, it concluded that census tracts' sensitivity to drought depends heavily on their socio-economic and land-use characteristics, particularly the presence of pools. However, results were only tested at the census tract scale.

House-Peters et al. (2010) investigated the drivers of water demand in Hillsboro, Oregon and concluded that drought condition was not a good predictor of water use at the study area level. However, it was a good predictor for certain census blocks containing large, new, affluent and well-educated households.

As it becomes apparent from the above, although some studies implemented spatial variability in their forecasting models, there are certain limitations. One of the limits for comprehensive spatial analysis of water demand has been data availability at high spatial resolutions. On the other hand, the level of spatial aggregation of water consumption data often does not match the scale of the explanatory variables. In order to overcome this problem, researchers often have to rely on interpolating or extrapolating data (Lee at al., 2010; House-Peters and Chang, 2011), i.e. estimating values for locations within the study area or outside the study area, respectively, which can be a challenging process (Lee at al., 2010). Even when data are available at the household level, it often lacks spatial coordinates (House-Peters and Chang, 2011), sometimes due to privacy concerns. Another main problem is the lack of a systematic comparison of predictions and influencing factors at various spatial scales. Since the variables that influence water consumption and the range of temporal and spatial scales can vary greatly at different settings and case studies, this comparison cannot be derived by merely comparing the results of different studies in the literature.

To summarise, data availability, computational power and new technologies have substantially increased in recent years. This has contributed in developing spatially explicit demand forecasting models and identifying and quantifying relationships among a variety of weather, social and water consumption data (House-Peters and Chang, 2011; Rathnayaka et al., 2017b). However, there is still the need to develop methodologies that incorporate this information at multiple spatial scales (House-Peters and Chang, 2011).

This study aims to address this gap by making use of a very rich dataset comprising of a variety of household characteristics, weather data, temporal characteristics and past consumption. The aim is to identify and quantify the influence of the drivers of water demand at multiple spatial scales and determine how they contribute to the accuracy of demand forecasting models.

5.3. Data

This section provides a brief overview of the data that are used in this study. Additional details are provided in chapter 3.

The data comes from a region in the southwest of England and includes 1,793 properties. These were monitored by the water company at 15-30 minute intervals over a period of almost three years (October 2014 to September 2017), using smart meters. The raw dataset was carefully cleaned in order to exclude incorrect and missing data, empty properties and leakage. A detailed description of this process is provided in chapter 1.

The water company also collected data related to household characteristics and postcodes. Information regarding the garden size, occupancy rate, metering status, rateable value of the property, acorn group (customer socio-economic classification) and council tax band became available at the household level. In addition, partial postcodes were used to identify the properties' location in the study area. Postcodes in the UK are comprised of four parts, indicating the area, district, sector and unit the house belongs to (Royal Mail, 2012). In this study, only the first two parts of the postcode, corresponding to the area and district, were available to group the properties.

Each one of the above six household characteristics (garden size, rateable value, occupancy rate, council tax band, rateable value and acorn group) divides the dataset into different categories (Table 5.1). For example, depending on the characteristic 'garden size', the households are divided into three categories, ‘large’, ‘medium’ and ‘small’, reflecting the size of the garden of the corresponding household. The categories created for each household characteristic are available in Table 5.1.

Table 5.1. Categories formed for each household characteristic.

Garden Size	**Rateable Value**	**Metering Status**
Large (> 165m^2)	High (top 30%)	Metered (billed on meter reading)
Medium (61-165m^2)	Medium (mid 40%)	
Small (< 60m2)	Low (bottom 30%)	Unmetered (billed on an estimation)
Acorn Group	**Occupancy Rate**	**Council Tax Band**
Affluent (A - E)	High (3+ occupants)	High (tax groups A - C)
Comfortable (F - J)	Medium (2-3 occupants)	Medium (tax groups D - E)
Financially Stretched (K - Q)	Low (1 occupants)	Low (tax groups F - H)

Finally, weather data were provided by the Met Office. These include information about air and soil temperature, sunshine hours, relative humidity and rainfall (Met Office, 2006a; Met Office, 2006b; Met Office, 2006c; Met Office, 2006d; Met Office, 2006e), collected at hourly to daily intervals for the same period (October 2014 to September 2017). These data were recorded at hundreds of weather stations within the study area. One additional variable representing the number of consecutive days without rain was also calculated based on the rainfall data.

5.4. Methodology

This section describes the main steps of the model development process, which include the selection of the aggregation levels and candidate input variables as well as the modelling technique.

5.4.1. Spatial aggregation

The households are grouped based on their postcodes into the following three levels of spatial aggregation:

- Network grouping: No grouping criteria are used. Consumption is aggregated among all properties for each day in the data (Network, Figure 5.1(a)). Due to errors and inconsistencies, consumption is not available for every property over each day. Therefore, this group can vary in composition, i.e. include a slightly different collection of properties on each day. The network group consists of 1,056 data points (each point represents one day), with 64-804 properties in each one, depending on data availability on the corresponding day.

- Area-based grouping: The first part of the postcode (e.g. BA) is used to group the properties into one of six areas. This group consists of 6,336 data points (Areas, Figure 5.1(a)), with 1-212 properties in each one (depending on data availability on the corresponding postcode and day). Each data point represents the consumption of an area for one day.
- District-based grouping: The first and second part of the postcode (e.g. BA1) is used to group the properties into 63 districts. This group consists of 76,032 data points (Districts, Figure 5.1(a)), with 1-56 properties in each one (depending on data availability for the corresponding postcode and day). Each data point represents the consumption of a district for one day.

The three aggregation levels have a different range in household composition (i.e. the types of households they consist of), among the groups. The smaller (district) groups are a lot more diverse in terms of the types of households they contain, compared to the relatively homogenous network grouping. If there were no gaps in the data and information for all households was available for each day in the dataset, all days would contain information about the same properties. Therefore, no variation would exist when aggregating the whole network. More details regarding the household composition of each aggregation of properties are available in Appendix B.

In order to create additional spatial scales, the household group size is set to a fixed number (from 5 to 600), for each postcode and level of spatial aggregation (Figure 5.1(b)). Each aggregation level has a set number of household groups for each day (this might slightly vary due to missing data), which is 63 for the district level, six for the area level and one for the network level. The number of households in each group depends on data availability for the corresponding postcode and day in the dataset and can vary significantly. When the household group size is set to a fixed number, the groups that are smaller than the threshold are excluded from the dataset, whereas the groups that are larger are reduced to the fixed number of properties.

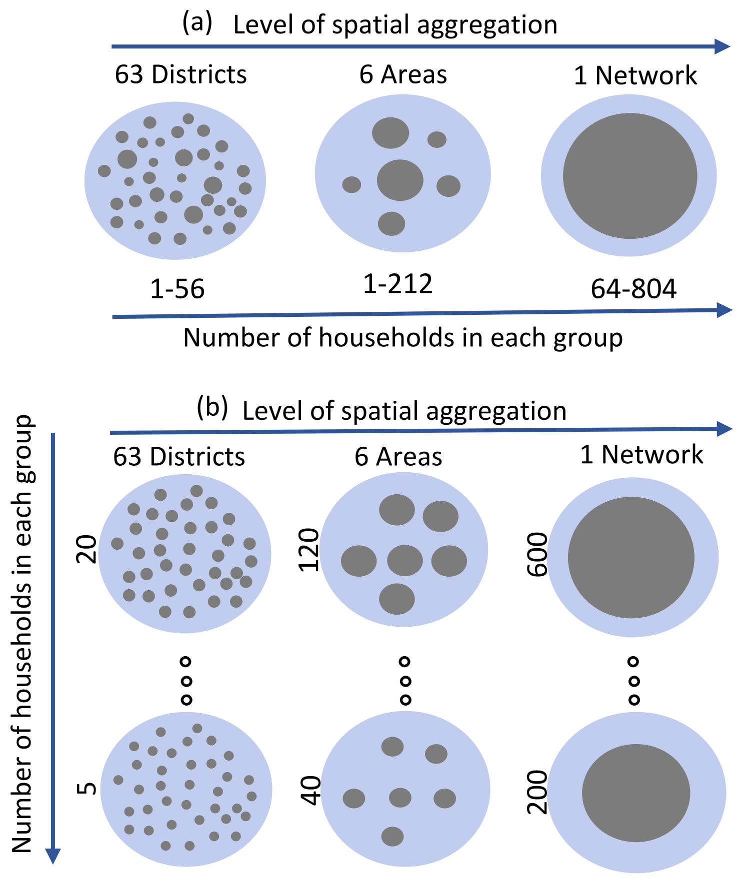

Figure 5.1. (a) Range of household group sizes for each level of spatial aggregation among different days and groups. (b) Spatial scales created using the level of spatial aggregation and a fixed group size, varying from 5 households for the district level to 600 for the network level. Each disc illustrates the size and number of groups for one day in the data.

The result is nine different spatial scales, comprising of different household group sizes (Figure 5.1(b)). The group sizes are set to 5, 10 and 20 for the district groups, to 40, 80 and 120 for the area groupings and to 200, 400 and 600 for the whole network. This way, it is easy to ensure that the properties grouped together are actually in close geographical proximity. The disks in Figure 5.1 illustrate the number and size of household groups that correspond to each spatial aggregation, for one day in the data.

5.4.2. Variable selection

Based on their nature, the variables described in the data section are divided into four types:

- Past consumption data: Consumption data are aggregated temporally at the daily level and spatially at multiple scales. A sliding, 7-day window of past consumption is used as input in order to capture the weekly repetition of demand patterns.
- Household characteristics: These refer to the occupancy rate, acorn group, garden size, rateable value, council tax band and metering status. Since each household group is composed of a variety of households with different characteristics, the percentage of households in each category is used as an explanatory variable, rather than the characteristic itself. For example, for the characteristic 'garden size', there are three possible categories, 'large', 'medium' and 'small'. Each category is used as a continuous explanatory variable in the model, with values varying from zero (0% of households) to one (100% of households). In the case of the garden size, a possible composition for a household group is 30% large gardens, 60% medium gardens and 10% small gardens. Thus, the garden size is represented by three values (0.30, 0.60 and 0.10), one for each category. The same applies to the rest of the household variables.
- Temporal characteristics: These relate to the season and type of day (working day or weekend/holiday). People tend to have different habits over different times of the year as well as the week, thus temporal variables can be helpful in capturing the time variability of demand.
- Weather: Weather information includes four weather variables, air temperature, sunshine hours, relative humidity and number of consecutive days without rain. These can capture the weather-dependent variability of demand.

The above four variable types are treated as separate entities in the demand forecasting models, as they have very distinct characteristics that relate to their availability, accessibility, reliability and thus importance for network operators. Some of the variables are always easily accessible, reliable and ready to use (temporal characteristics). Others can be expensive to acquire, store and process, or even inaccurate, especially when they are based on forecasts and estimations (weather and past consumption data). Information about household characteristics can be anywhere in between; some are relatively easily accessible (council tax band, metering status, rateable value, acorn), whereas others need to be collected through questionnaires and inspections (Xenochristou et al., 2019a).

Eight models with different configurations of the above input variables are tested at each level of spatial aggregation (Table 5.2). Models 1 to 4 include a combination of past consumption data and other characteristics as input whereas models 5 to 8 are built using only temporal, weather and household characteristics. Each model is trained and tuned separately for the optimum set of input parameters, at each aggregation level, but on the same training dataset.

Table 5.2. Model configurations tested at each level of spatial aggregation.

Variable group	Model input variables	Model number							
		1	2	3	4	5	6	7	8
Past Consumption	1-7 days prior	X	X	X	X				
Temporal	Type of Day	X	X	X		X	X	X	X
	Month	X	X	X		X	X	X	X
Household	Acorn	X				X	X		
	Garden Size	X				X	X		
	Metering Status	X				X	X		
	Rateable Value	X				X	X		
	Council Tax Band	X				X	X		
	Occupancy Rate	X				X	X		
Weather	Sunshine hours	X	X			X		X	
	Air Temperature	X	X			X		X	
	Humidity	X	X			X		X	
	Days without rain	X	X			X		X	

Chapters 2 and 3 concluded that all of the above variables have an influence on water consumption. Although weather input did not improve the forecasting accuracy at the small aggregation level (~3.8 households/group) tested in chapter 3, this chapter will explore if weather can improve predictions for larger household groups. For this reason, the four weather variables (sunshine hours, air temperature, humidity, days without rain) that were found to have some sort of influence on water consumption (see chapters 2 and 3) are used here to capture the effect of weather. Soil temperature and rainfall are strongly correlated with air temperature and days without rain, respectively, and thus were excluded from any further analysis.

5.4.3. Demand forecasting model

Chapter 4 compared a selection of machine learning models for water demand forecasting and concluded that the Gradient Boosting Machine (GBM) method combines high prediction accuracy with ease of implementation, hence was chosen for this work. A brief description of the characteristics and implementation of the GBM is provided in the following. More details regarding the type of GBM algorithm implemented here, including the hyperparameters and modelling process can be found in Click et al. (2017).

5.4.3.1. Gradient Boosting Machines

The idea behind GBMs is to combine a set of weak (base) learners in order to create one strong learner. In this study, the base learner is decision trees. The way decision trees work is by dividing the dataset at each branch in a way that maximises entropy, i.e. the homogeneity within each of the split groups. At each branch (node) of the tree, a variable and a threshold value are chosen for splitting the dataset. The tree keeps dividing until it reaches a limit, typically defined by the user, such as the maximum tree depth or minimum final node size.

The GBM algorithm uses bagging, as well as boosting in order to achieve the best results. Each tree is trained on a subset of the original data and at each node of the tree, the best variable for splitting is chosen among a random sample of the input variables (bagging). In addition, at each step of the algorithm one regression tree is built on the residual errors of the previous tree, with the aim to improve the final result. In this way, the model gradually learns harder parts of the problem, as higher weights are assigned to the areas of the training set where the highest errors occurred (boosting). The result is altered at each step of the process by adjusting the overall prediction based on the new tree that is added to the model. The overall process in regression is set up as a simple optimisation problem, where the objective is to minimise the error in the objective function (gradient descent).

The nine hyperparameters that require tuning for the GBM algorithm are: the total number of trees that construct the final model (ntrees); the size of the subsample of the training dataset used to train each tree (sample_rate); the maximum tree depth (max_depth); the number of variables that are sampled and tested for splitting at each node, for the overall model as well as for each tree (col_sample_rate, col_sample_rate_per_tree); the learning rate (learn_rate) of the algorithm, which is used to reduce the contribution of subsequent trees to the final result; the histogram type used to assist with the splitting selection process (histogram_type); the minimum requirements for splitting at each node (min_split_improvement and min_rows). More information regarding the model hyperparameters are provided in chapter 4.

5.4.3.2. Model implementation and assessment

In order to build the model, the dataset is randomly shuffled and divided into a training (70%) and a test (30%) dataset. The training data is used to train and tune the model for the optimum set of hyperparameters, through a 5-fold cross validation process (Zhang, 1993)**.** The test dataset does not participate in the model-building phase and is used to carry an unbiased evaluation of the model's prediction accuracy based on unseen data, i.e. data that is not used during the model-building phase.

The 'h2o' machine learning platform (Aiello et al., 2019) is used here to train and tune a range of GBM models for the optimum set of hyperparameters, through a random search (Bergstra and Bengio, 2012). The high number of hyperparameters that require tuning (nine in total) increases significantly the dimensionality of the search space. Thus, any exhaustive grid search, manually implemented by the user, would be counter-productive, especially since the aim is to train, tune and compare a large number of models. Thus, 'h2o' is used instead to perform a random search for the best hyperparameter values.

After the model is properly trained and tuned, it is used on the test dataset to make predictions for daily consumption 1-7 days ahead. The model performance is assessed by comparing the model predictions with real data, based on three criteria, the mean absolute percentage error (MAPE), mean square error (MSE), and R^2 coefficient of determination. The MAPE is intuitive and independent of the scale of the dependent variable, thus it can be used to compare results from different studies and variables of interest (e.g. PCC and PHC). The MSE is sensitive to outliers whereas the R^2 indicates the variance in the dependent variable that can be explained by changes in the independent variables.

5.5. Results

5.5.1. Demand forecasting accuracy at different spatial scales

Increasing the level of spatial aggregation decreases the randomness and variability of the water demand signal, making it easier to predict. However, it is unclear by how much. In the following, the relationship between household group size and prediction accuracy is investigated in detail.

First, nine models are trained, tuned and assessed for their ability to predict demand for different household group sizes, one day into the future. For comparison purposes, each model is trained using the same input, seven days of past consumption. Table 5.3 shows the aggregation level, group size and number of data points that are used to train each model, as well as the results acquired based on three assessment criteria, the MAPE, MSE and R^2, for the training and test dataset.

Table 5.3. Prediction accuracy for nine models, trained on different household groups.

Aggregation level	Data points	Group size	MAPE (%)		MSE (l/capita/day)^2		R^2 (%)	
			Train	*Test*	*Train*	*Test*	*Train*	*Test*
District	43,875	5	16.2	17.0	1047	1133	59.3	**55.0**
District	26,153	10	12.6	12.9	536	612	59.2	**55.2**
District	8,537	20	9.1	10.0	247	308	61.4	**56.4**
Area	5,729	40	6.9	7.7	148	186	59.3	**51.8**
Area	4,349	80	5.4	5.9	92	105	60.7	**55.5**
Area	1,915	120	3.2	5.1	32	83	85.7	**61.7**
Network	978	200	2.9	4.5	28	57	80.4	**60.6**
Network	922	400	3.1	3.8	34	49	70.0	**64.8**
Network	806	600	3.0	3.2	34	39	73.2	**65.3**

According to Table 5.3, the prediction error (MAPE and MSE) reduces as the group size increases. The minimum MAPE corresponds to the largest aggregation, at the network level, with a group size of 600 households, which has an error of 3.2% for the test dataset (Group size = 600, Table 5.3). The largest MAPE on the other hand relates to the smallest aggregation scale, at the district level, with a group size of 5 households (Group size = 5, Table 5.3). The R^2 value also increases with the group size, but only within the same aggregation level.

However, it is still not clear which point represents a good balance between prediction accuracy and household group size. In other words, at which spatial scale, a further increase in group size does not offer a significant reduction in prediction errors. This is depicted in Figure 5.2, which represents the balance between the MAPE and the spatial scale, for the test dataset.

According to Figure 5.2, the model error increases exponentially as the household group size decreases. When everything else remains the same (model structure, input variables), increasing the prediction group size from 40 to 120 households reduces the MAPE by 2.6% (Figure 5.2). However, for group sizes below ~20 households, the MAPE increases significantly, for a rather small decrease in group size. For example, the MAPE increases an additional 7%, from 10% to 17%, for a decrease of 15 households per group (from 20 to 5). On the other hand, for group sizes above ~200 households, the MAPE reduces marginally for a high increase in group size (Figure 5.2).

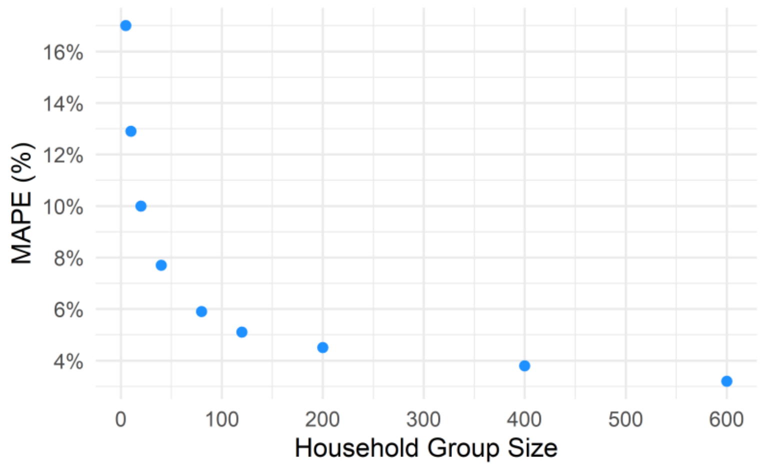

Figure 5.2. Model accuracy (MAPE) for each household group size, for the test dataset.

5.5.2. Variable importance at different spatial scales

The three aggregation levels have different household group sizes, different ranges in their daily consumption and different amounts of data points (Table 5.4). In order to avoid increased prediction errors associated with very small groups (<20 households), the minimum group size is set to 20, 60 and 100, for the districts, areas and network, respectively. The smaller the aggregation level, the smaller the mean group size and the larger the number of data points. In addition, as consumption becomes more erratic for smaller household groups, the range in daily consumption also increases (Table 5.4).

Table 5.4. Household group sizes, number of data points and daily water consumption range, for each spatial aggregation level.

Spatial aggregation	Min group size	Mean group size	Number of data points	Daily consumption range (l/capita/day)
Network	100	657	992	117-175
Areas	60	114	5,592	100-195
Districts	20	29	8,537	80-250

Results are summarised in Figure 5.3 and Table 5.5. Figure 5.3 shows the model accuracy, in terms of MAPE, for predictions 1-7 days ahead, over all days in the data (plots a-c, Figure 5.3), as well as peak days, i.e. the 10% of days with the highest consumption (plots d-f, Figure 5.3). Each plot represents the MAPE for eight models and one aggregation level (network, areas and districts). Table 5.5 shows the MAPE for each model and aggregation level, for one as well as seven days into the future, for all days and peak days. The final hyperparameter values selected for each model are provided in Appendix B.

The best performing model for the network level is the one that uses all explanatory variables to make predictions (model 1). When past consumption data is included in the model (models 1-4), temporal characteristics reduce the MAPE by 0.5%, for predictions 1 day ahead (model 3), while weather input further reduces errors by 0.4% (model 2) and household characteristics by 0.1% (model 1). For models 5-8 (no past consumption data), weather input reduces the MAPE by 0.4% (Model 7), while household characteristics reduce it by 0.1% (Model 6). Adding both household and temporal characteristics (Model 5) reduces model errors by 0.9% (Table 5.5).

Although the MAPE value and variance increase for peak days, results are very similar. The best performing model (MAPE = 4.6%), for one day lead time, is the one that uses all predictors (model 1). However, for predictions seven days into the future, the model with temporal, household and weather characteristics (model 5) performs better (MAPE = 6.1%) than the model (model 1) that also incorporates past consumption data (MAPE = 6.4%) (Table 5.5). Temporal characteristics, on top of past consumption, improve the MAPE by 2.5% (model 3), for one day lead time. Weather input further reduces errors by 0.2% (model 2) and household characteristics by 0.6% (model 1).

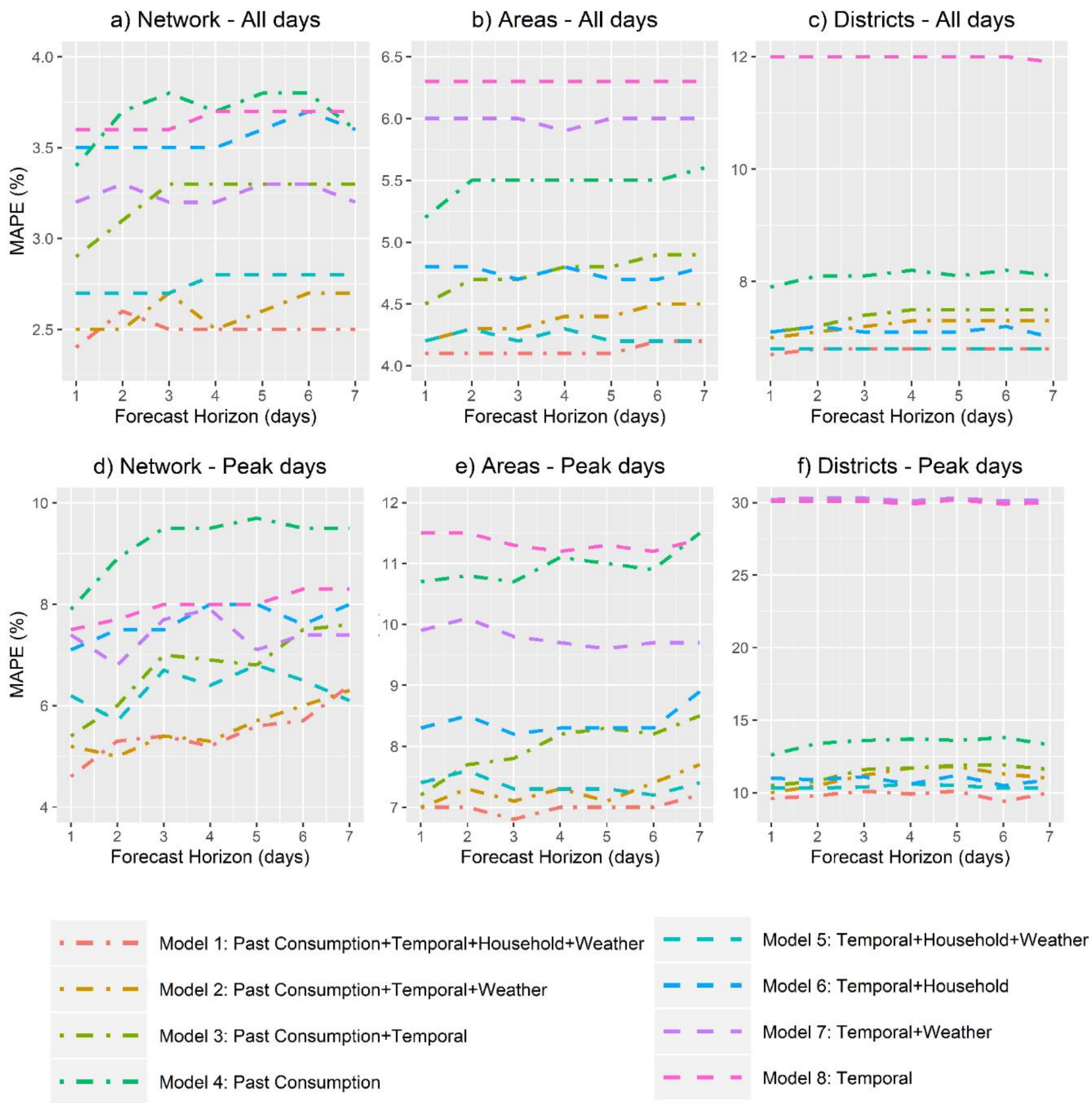

Figure 5.3. Mean Absolute Percentage Error (MAPE) for different model configurations (Models 1-8) and different spatial aggregations (network, areas, districts), for all days in the data (plots a-c), as well as peak days (plots d-f).

For models 5-8 however (the ones excluding past consumption data), weather and household input reduce errors by 0.1% (model 7) and 0.4% (model 6), respectively, for predictions 1 day ahead. Both of the above reduce the MAPE by 1.3%, a reduction much higher than the simple addition of their individual contributions (model 5). In both cases (all days and peak days), the model that includes only temporal and weather variables (model 7) performs better than the model that includes only past consumption data (model 4) (Table 5.5).

As the level of spatial aggregation decreases, the range in errors among the models drastically increases. The best performing model for the areas is still the one that includes all variables (model 1), for all days and peak days (Figure 5.3, (b) and (e)). In this case, temporal, weather and household characteristics, on top of past consumption data, reduce errors by 0.7%, 0.3% and 0.1%, respectively, for all days and 3.5%, 0.2% and 0%, respectively, for peak days. Weather input for the models without past consumption reduces the MAPE by 0.3% (model 7), for one day lead time, whereas household characteristics reduce it by 1.5% (model 6), for all days (Table 5.5). The combined effect of both household and weather characteristics outperforms again the mere addition of their individual contributions; the model that includes temporal, household and weather variables (model 5) has a MAPE of 4.2% for predictions 1 day ahead (an improvement of 2.1%), an error almost as low as the best performing model (model 1) (Table 5.5). The same is true for peak days; weather (model 6) and household (model 7) input reduce errors by 1.6% each, whereas the combination of the two contributes to an error reduction of 4.1% (Table 5.5). Finally, for peak days, the model with temporal and weather input (model 7, MAPE = 9.9%) performs better than the model with past consumption data (model 4, MAPE = 10.7%), for 1 day lead time.

Table 5.5. MAPE for eight model configurations, for predictions one and seven days into the future, for three spatial aggregations of properties (network, areas, districts).

	NETWORK – MAPE (%)				**AREAS – MAPE (%)**				**DISTRICTS – MAPE (%)**			
Model	**All days**		**Peak days**		**All days**		**Peak days**		**All days**		**Peak days**	
	1 day	*7 days*	*1 day*	*7 days*	*1 day*	*7 days*	*1 day*	*7 days*	*1 day*	*7 days*	*1 day*	*7 days*
1	2.4	2.5	4.6	6.4	4.1	4.2	7.0	7.2	6.7	6.8	9.6	**10.0**
2	2.5	2.7	5.2	6.3	4.2	4.5	7.0	7.7	7.0	7.3	10.0	**11.0**
3	2.9	3.3	5.4	7.6	4.5	4.9	7.2	8.5	7.1	7.5	10.5	**11.6**
4	3.4	3.6	7.9	9.5	5.2	5.6	10.7	11.5	7.9	8.1	12.6	**13.3**
5	2.7	2.8	6.2	6.1	4.2	4.2	7.4	7.4	6.8	6.8	10.3	**10.3**
6	3.5	3.6	7.1	8.0	4.8	4.8	8.3	8.9	7.1	7.0	11.0	**10.9**
7	3.2	3.2	7.4	7.4	6.0	6.0	9.9	9.7	12.0	11.9	30.2	**30.2**
8	3.6	3.7	7.5	8.3	6.3	6.3	11.5	11.4	12.0	11.9	30.1	**30.0**

For the district groups, the MAPE range increases further, varying from 6.7% to 12%, for predictions 1 day ahead, for all days. In this case, past consumption data and household characteristics offer significant improvements, whereas weather is rather irrelevant (Figure 5.3(c)). The model that includes all variables as input (model 1) has once again the best performance (MAPE = 6.7%, for 1 day lead), although temporal, household and weather input (model 5) can achieve a similar accuracy (MAPE = 6.8%), for all days in the data. For seven days ahead, models 1 and 5 perform equally well for all days in the data (MAPE = 6.8%), whereas model 5 performs slightly worse (MAPE = 10.3%) compared to model 1 (MAPE = 10.0%) for peak days. Past consumption data (model 3) and household characteristics (model 6), on top of temporal characteristics, reduce errors by 4.9%, from 12.0% to 7.1%, for 1 day lead time (Table 5.5). Weather input (models 2 and 7) offers hardly any benefit to the model for predictions across all days. However, it does improve the MAPE by a maximum of 0.6% on peak days (model 2), for predictions seven days ahead. Finally, the model that uses only weather and temporal characteristics (model 7) has almost double the MAPE for all days (MAPE = 12.0%) and triple for peak days (MAPE = 30.2%), compared to the best performing model (model 1).

5.6. Discussion

In this work, water demand forecasting errors improve for larger aggregations of properties, since it means that water demand becomes less erratic and therefore easier to predict. This is illustrated by the level of water demand variability (Table 5.4), which is clearly associated with the level of spatial aggregation; smaller groups have a much wider daily water consumption range compared to larger ones. Here, a constant prediction accuracy is reached for groups larger than ~200 houses, whereas errors start to increase exponentially for groups smaller than 20-40 properties.

As errors reduce for larger group sizes, the R^2 value increases, but only within the same aggregation level (e.g. areas). As the group size increases, the variance in the response variable (i.e. water consumption) decreases, making consumption easier to predict. However, increasing the aggregation level (e.g. from districts to areas) also means that houses that are further away from each other are grouped together, creating less homogenous groups and thus reducing the explanatory value of past consumption data. This is likely the reason that the R^2 decreases when moving to a larger aggregation of properties, even though the household group size decreases (Table 5.3).

Even though errors increase for smaller household groups, as water demand becomes more erratic, the large variance in water use is largely explainable by identifying the right predictors. The larger the level of spatial aggregation, the closer the models are in terms of their performance, thus the less important the input variables. When all household groups have similar characteristics (e.g. at the network level), these characteristics cannot explain the variance in consumption (Figure 5.3, Network). In order for a variable to be a useful input to a forecasting model, it needs to have an influence on the model's response but also a wide range of values among the groups in the dataset (Figure 5.3, Districts). When groups are rather homogenous, the potential for error reduction is significantly smaller. For this reason, household characteristics and past consumption become more important for smaller household groups. Smaller groups are associated with higher variations in water demand (Table 5.4) but also higher variations in their household composition (Figure C1). The higher the variation in household composition and past consumption between the groups, the higher the importance of these variables as predictors.

On the other hand, household characteristics are embedded into past consumption, along with other factors that define the consumption behaviour of a property or group of properties. Therefore, using past consumption data can be particularly valuable for smaller groups, as a proxy of the consumption behaviour that relates to their individual characteristics. This is demonstrated by examining the influence of the predictors of the district areas (Figure 5.3, Districts). When past consumption data are available, household characteristics do not further improve predictions. However, when past consumption is not used as input, a combination of household, weather and temporal characteristics can adequately characterise and thus predict water demand with the same accuracy. For example, weather and household input, on top of past consumption, reduces the MAPE by a maximum of 1.6% for peak days and district areas. When the same variables are used on top of temporal characteristics, they reduce the MAPE by 19.7%, from 30% to 10.3%.

The combined contribution of household and weather characteristics in the model is in most cases much higher than their individual contributions. This result confirms further what was already concluded in chapters 2 and 3, that the influence of weather on water consumption is variable and strongly depends on the type of property and residents. Therefore, providing additional context in terms of household characteristics, on top of weather information, can improve results.

Although weather does not improve results for smaller household groups (Figure 5.3, Districts), it does improve accuracy for larger groups of properties (Figure 5.3, Network and Areas). Chapter 2 showed that the effect of weather on water consumption varies between households, days and times in the year. Out of all households in the dataset, few of them alter their consumption due to weather changes, for few days in the data. Therefore, the model learns based on the majority of the data points (household groups and days in the data), for which weather does not actually influence consumption. When aggregating all properties, the effect of weather, although mild, is visible for many more data points (days) used to train the model, therefore weather is found to have a (slight) impact on consumption.

Finally, it is worth noting the upward trend of all models that include past consumption data (models 1-4), for predictions further into the future (Figure 5.3). Since water consumption is highly auto-correlated from one day to the next one, predictions one day ahead are more accurate than seven days ahead. However, adding weather and household input reduces errors for predictions further into the future. On the other hand, for models 5-8 (no past consumption input), the forecast horizon does not have an effect on the model's accuracy (Figure 5.3). As a result, the best model sometimes shifts depending on the forecast horizon. The models that include past consumption often perform best for one day lead time, but worse than the ones that use temporal, household and weather input for increased lead times (e.g. 7 days).

5.7. Summary and conclusions

This study explores the effect of spatial aggregation on water demand forecasting, both in terms of prediction accuracy and influencing factors. In order to achieve this, multiple models with different input configurations are trained on real-life UK daily consumption records, for different aggregations of consumption.

Initially, the effect of spatial aggregation on forecasting accuracy is determined for nine different group sizes, varrying from 5 to 600 households. A GBM model with only past consumption data as input is used to compare the modeling accuracy for daily forecasts, one day ahead. Then, the predictive capability of several variable types (temporal, household, weather and past consumption) is determined at three spatial scales, at the network level (up to 804 properties/group), area level (up to 262 households/group) and district level (up to 56 households/group), for each day in the data.

Results show that:

- The level of spatial aggregation has a direct influence on the demand forecasting accuracy; the larger the spatial scale, the more accurate the demand forecast. For groups smaller than 20-40 households, the MAPE increases exponentially for a further decrease in household group size. For group sizes above ~200 households, an increase in group size only marginally reduces the MAPE.
- Using the right predictors can significantly reduce forecasting errors, especially for smaller household groups. In this study, the most influential input variables vary for different levels of spatial aggregation. Past consumption data and household characteristics become more important for smaller aggregations, while weather data contribute to the model's accuracy only for larger household groups.

This work is particularly important in the UK, where water networks are decomposed into district metered areas (DMAs). Results show that at the DMA level, i.e. for larger aggregations of properties, using past consumption, along with temporal and weather variables, results in very low MAPE for predictions 1-7 days into the future. This can be particularly useful in optimising network operations as well as estimating leakage.

Although the effect of different levels of spatial aggregation is investigated here in detail, this is done within a fixed set of environmental conditions. All of the above analysis reflects the consumption of houses in the southwest of England. In a different setting, with different prominent household and customer characteristics and different climate, these results may be different. Although the above methodology can be replicated anywhere where the related data is available, results may vary.

Another interesting aspect that is not explored in this chapter is the effect that the temporal aggregation of consumption has on forecasting, particularly in conjunction with the spatial aggregation. Further work is needed to develop a grid of spatial and temporal aggregations of consumption that will demonstrate the limitations and opportunities that arise at each scale.

CHAPTER 6

SUMMARY, CONCLUSIONS AND FUTURE WORK RECOMMENDATIONS

This thesis investigates the topic of water demand forecasting in terms of models and influencing factors, over a range of scenarios. The ultimate aim is to provide solutions to real-world problems and improve the current engineering practice. For this reason, the influence of weather on water consumption, which is an uncertain factor and cause of concern for the future of water resources, is given particular importance. In addition, aspects such as improving predictions on peak demand days, which are the critical ones for water utilities, or dealing with the lack of past consumption data and producing forecasts for different levels of spatial aggregations are addressed here. Ultimately, this study developes an improved methodology that can inform decisions regarding the models and data needed to produce accurate demand forecasts.

6.1. Thesis summary

The following is a summary of each methodological chapter, including the aim, approach and key results.

Chapter 2. Predicting water demand is necessary to ensure a secure water supply to homes and businesses. With great uncertainty around future changes in the climate and the UK households, it is essential to accurately determine the effect of weather on water consumption. A systematic approach based on smart demand metering data and customer characteristics (e.g. metering status, garden ownership) is used to investigate the sensitivity of household water *+ consumption to weather, for different consumer types and time-varying parameters. The following weather variables are analysed: air and soil temperature, humidity, rainfall and sunshine hours. Results indicate that the effect of weather on water consumption is moderate in the UK. This effect becomes more significant for affluent customers with high variation in their mean monthly consumption and medium occupancy households, as well as working days, summers and evenings. Sunshine hours, humidity and air temperature are the weather variables with the most widespread influence in the UK. Soil temperature has a milder effect, whereas daily rainfall shows minimal impact.

Chapter 3. Smart demand metering data at the household level is becoming increasingly available but not all households are currently monitored. This chapter compares two modelling approaches, one with and one without past consumption data as input, with the aim to predict daily demands for different household types, one day ahead. The methodology developed combines Random Forests with a variety of interpretable machine learning techniques (Variable permutation, Accumulated Local Effects plots and Individual Conditional Expectation curves). These techniques are used to quantify the influence of several model predictors (household, weather and temporal characteristics) on water consumption and forecasting accuracy. Results show that when past consumption data are available, it is by far the most important explanatory factor. However, when it is not, a combination of household and temporal characteristics can be used to produce a credible model, with forecasting accuracy similar to the model that includes past consumption data. In this case, the household characteristics are the best predictors of consumption, whereas the weather has little to no influence on the model's output, under the current UK climate. This methodology is of high value to the engineering practice as it combines accuracy with interpretability.

Chapter 4. Water demand forecasting is an essential task for water utilities, with increasing importance due to future societal and environmental changes. This chapter suggests a new methodology for water demand forecasting, based on model stacking and bias correction that predicts daily demands for groups of ~120 properties. This methodology is compared to a number of models (Artificial Neural Network, Generalised Linear Model, Random Forest, Gradient Boosting Machine, Extreme Gradient Boosting and Deep Neural Network) using real consumption data from the UK, collected at 15-30 minute intervals from 1,793 properties. Results show that the newly proposed method consistently outperforms other water demand forecasting techniques, for one day lead time (peak R^2 = 74.1%), especially for peak consumption days and limited input data.

Chapter 5. Understanding, comparing and accurately predicting water demand at different spatial scales is an important goal that will allow effective targeting of the appropriate operational and conservation efforts under an uncertain future. This chapter uses data relating to water consumption, available at the household level, as well as postcode locations, household characteristics and weather data in order to identify relationships between spatial scale, influencing factors and forecasting accuracy. For this purpose, a Gradient Boosting Machine is used to predict water demand 1-7 days into the future. Results show an exponential decay in prediction accuracy from a Mean Absolute Percentage Error (MAPE) of 3.2% to 17%, for a reduction in group size from 600 to 5 households. Adding explanatory variables to the forecasting model achieves a reduction in MAPE of up to 20% for the peak consumption days and smaller household groups (20-56 households), whereas for larger aggregations of properties (100-804 households) the range of improvement is much smaller (up to 1.2%). Results also show that certain types of input variables (past consumption and household characteristics) become more important for smaller aggregations of properties whereas others (weather data) become less important.

6.2. Thesis contributions

This thesis has the following key contributions:

1. **The first contribution is a new, improved demand forecasting methodology, which is tested and demonstrated on real data.** This model is based on stacking and is built as a combination of five base models. A bias correction method, applied on the model’s output, assists

with predicting outliers. This method can be used to improve the accuracy of water demand forecasting (see chapter 4).

2. **The second contribution is an improved understanding of the link between weather and water consumption (in a UK context).** This is done using both a big-data statistical analysis (see chapter 2) that specifically addresses the influence of the weather over space and time, as well as a machine learning approach (see chapter 3). These results can assist with addressing regulatory requirements that relate to climate change planning and mitigation.

3. **The third contribution is an improved understanding of key water demand explanatory factors and how these could be used to make more accurate demand forecasts, especially when data are limited.** A Random Forest model and three interpretable machine learning techniques are used to produce demand forecasts using a variety of property, temporal and weather predictors. These results can assist with predicting demand for the unmetered customers, long-term water demand projections, leakage estimations and new water billing incentives (see chapter 3).

4. **The fourth contribution is an improved understanding of the limitations in demand forecasting accuracy at different spatial scales (i.e. household groupings), together with the best predictors, which tend to change at each scale.** A Gradient Boosting Machine model is used with different input configurations to make predictions at different spatial scales. Results can assist with benchmarking the accuracy of forecasting models and as a guidance for water utilities in order to select the appropriate predictors at the right scale (see chapter 5).

5. **New technical guidance for water utilities.** The results produced in this thesis can be used as guidance for water utilities to help them identify the best model and input variables, with respect to the characteristics of the problem and the forecasting target.

6.3. Thesis conclusions

This work attempts to provide an improved understanding of water demand, in terms of the variables and factors that influence it, as well as the necessary data, models, and techniques that can improve demand forecasts. The first aspect of this thesis investigates the variables that influence water consumption, taking into account their interactions with a variety of other household, weather, and temporal characteristics (see chapters 2 and 3). The second part focuses on comparing different water demand forecasting models, under different scenarios, and developing an improved methodology that combines the benefits of individual models. Finally, the best models and types of inputs are used in order to improve water demand forecasting in practice, for different sizes of network sections.

One of the main benefits of the methodology adopted here is that it assesses the drivers of water demand in a multidimensional context. When examining the influence of a variable on water demand, the interactions between this variable and a variety of other temporal, household and weather characteristics are taken into account. In chapter 2, a disaggregated approach is adopted to separately assess the influence of weather on water demand for different household (e.g. garden size, occupancy rate) and time-varying (e.g. season, month) characteristics. In chapter 3, a machine learning model with several inputs as explanatory variables is used to evaluate the influence of these variables on water demand. Results show that the variables that have the highest influence on consumption are the household characteristics, particularly the occupancy rate and council tax band, followed by the type of day (working day or weekend/holiday). The weather has a non-linear effect on consumption that can vary significantly for different household and time-varying characteristics. Using the above temporal and household characteristics as model predictors can achieve a similar accuracy to using past consumption data. Unlike this work, most studies in the literature assessed the influence of several household, weather, and temporal characteristics on water consumption using simple statistical techniques that did not account for the interactions between them.

In addition, the modelling accuracy and best types of model variables are assessed under multiple forecasting scenarios. Most studies in the literature made conclusions within a specific context. This work takes into account several factors, such as the type of available data, the level of spatial aggregation, the forecasting aim, as well as the forecast horizon in order to make conclusions about the best type of model and variables, with respect to the individual problem. Results highlight that the modelling accuracy, as well as the best model and types of variables depend on all of the above factors (see chapters 4 and 5). The prediction accuracy decreases exponentially together with the level of spatial aggregation. In addition, certain models perform better for peak consumption days as well as limited input data (stacked and GBM models). On the other hand, certain types of variables (household and past consumption data) become significant for smaller aggregations of properties whereas others (weather variables) improve modelling accuracy only for larger aggregations.

Another benefit of this work is the usage of a variety of new, emerging machine learning methods and other techniques that can facilitate and improve the accuracy of water demand forecasting. Several ensemble (e.g. model stacking and gradient boosting) and other (e.g. deep learning) machine learning models are tested for their ability to produce accurate demand forecasts (see chapter 4). Results show that the new method developed in this thesis using model stacking and bias correction performs best, especially for peak days and when past consumption data are not available. In addition to this, new interpretable machine learning techniques are used to uncover the drivers of water demand and enhance the value of 'black box' models (see chapter 3). Finally, the ability of machine learning platforms that facilitate the modelling process, including the model building and selection, is demonstrated in chapter 4.

However, there is no model or input factors that are universally best, under all scenarios. Sophisticated models, new techniques and increasing data availability can improve the accuracy of forecasting models, especially under certain scenarios. Nevertheless, the level of spatial aggregation (household group size), forecasting aim (peak days or all days), forecast horizon (one or seven days lead time) and data availability can determine what is the appropriate model and predictors required to produce credible forecasts (see chapters 2-5). A cost-benefit analysis needs to be performed in order to determine the best model structure based on all of the above factors. The methodology developed here can be used as a guide for water utilities in order to build the best forecasting model, taking into account the characteristics of the individual problem.

A source of uncertainty that is not considered here relates to the quality of data acquired using smart meters. Although every effort is made in order to ensure the quality of these data and remove any inconsistencies, leakage, errors, as well as empty properties, the extent of errors remaining in the dataset is uncertain. These errors could relate to inconsistencies that were not removed during the data cleaning process as well as systematic meter recording errors. The results presented here are based on the assumption of the water company that the smart meters record consumption with an error of 2%. The extent to which potential errors could have influenced the results is unclear.

In addition, results are based on the available dataset, which is derived from residential properties in the UK and specifically a particular region in the southwest of England. In a different climate with high seasonal variations, different culture, different societal structure, or even different household types, results could be very different. Although the methodology developed here can be transferred, results are topical. Furthermore, since the case study includes only residential properties, it is unclear how results would differ for industrial or commercial buildings. Since consumption in these buildings has different time patterns and end-uses, further analysis is needed to assess the best model types and influencing factors in this case.

Finally, the performance metric used to assess the model can also have a strong impact on the results. In this study, three metrics are used in order to account for different aspects of modelling accuracy. In addition, the model's ability to predict demand on peak days, which are the critical ones for water utilities, is also treated separately. However, in a different scenario that would require the models to perform well on certain times or for certain customers, results could vary. Even if the forecasting aim is the same, using a different metric could potentially lead to different conclusions.

Overall, above results can find use in short-term operational optimisation and forecasting as well as long-term planning. Although the influence of the weather is currently limited, it could cause problems under a different climate, with more hot, sunny and humid days. On the other hand, changes in the customer base and societal reforms, such as increases in single-occupancy properties and standard of living could also cause water availability concerns in the long-term. Above results can help water companies develop targeted water demand management strategies, plan infrastructure investments and secure water for the future. In the short-term, the methodologies and results within this thesis can provide practical guidance to network operators in order to develop improved water demand forecasting models, with respect to the characteristics of the problem. Specifically, results demonstrate alternative ways to predict water consumption using new models and techniques and without utilising past consumption data, for different levels of spatial aggregation and peak consumption days. Above results can assist water utilities as well as their operators with optimising network operations, avoiding over abstractions and reducing energy spending and carbon emissions, as well as assessing the amount of water that is lost through leakage.

6.4. Future work recommendations

This thesis addresses a few important topics in water demand forecasting. Through this work, many opportunities for future research projects emerge.

Water utilities' operations, strategies and investments are targeted at different temporal (e.g. week, month or year) and spatial scales. Therefore, it is important to assess the forecasting accuracy and influencing factors that relate to each one. This work addresses the topic of water demand forecasting at various spatial scales but the temporal one is set to daily. More work is required to define a temporal and spatial grid space and assess the limitations and opportunities that arise at each aggregation level.

Another topic of interest is understanding when, where and for whom models perform poorly. This will provide a good basis to further improve predictions. In this work, the model errors are higher for peak consumption days, but also for days with unusually low consumption. Therefore, a bias correction method is applied in order to improve results. Identifying the accuracy that can be achieved for different customer groups and days in the data is the first step towards understanding how to improve water demand forecasting.

Although developing models that perform real-time predictions were out of the scope of this work, the methodologies developed here can be used for real-time forecasts. This would require the model to be calibrated offline using the necessary data, before it is implemented online. Future work could focus on expanding the methodology developed here and applying it to a different context, for real-time demand forecasting.

Finally, no risk or ‘what if’ analysis is conducted as part of this thesis. Results show that water demand is heavily influenced by household characteristics that are rapidly changing due to lifestyle, societal and economic restructures. However, the effect that this could have on the future of water resources remains uncertain. Although the weather has a minor overall influence on consumption, partly due to the mild UK climate, water consumption increases significantly when air temperature or sunshine hours exceed certain threshold values. This means that if weather extremes occur more often in the future, this could have a huge impact on the water supply-demand balance. Future projects should focus on using these results to create projections of demand, based on future societal and climate scenarios.

APPENDIX A

SUPPORTING INFORMATION – CHAPTER 2

A.1 Distribution of correlation coefficients and gradients for each model in the data

For each segment (115,200) and weather variable (5), the relationship of consumption and the weather is evaluated using the Spearman's ρ correlation coefficient, the p-value and the gradient of the linear curve that is fitted on the data. This results in 576,000 (115,200*5) relationships. Figures A1 to A9 demonstrate the range of the correlation coefficient values and gradients that correspond to all relationships that have a p-value less than 1%, i.e. for all the statistically significant relationships. Each point corresponds to one segmentation of consumption and one weather variable.

Figure A1 shows the distribution of the correlation coefficient and gradient values that relate to consumption that occurred each season of the year. Results clearly indicate that over the summer months, the gradients for all weather variables are significantly steeper, whereas almost no strong relationships ($\rho > 0.5$) are identified between the weather and consumption over the winter months. This means that consumption over the summer is much more sensitive to weather changes compared to all other seasons, whereas the effect of weather becomes rather irrelevant in the winter. Results also clearly demonstrate that humidity and rainfall are inversely related to consumption, as the vast majority of gradients and correlation coefficients that relate to these two variables are negative. Rainfall is the variable with the smallest amount of statistically significant relationships.

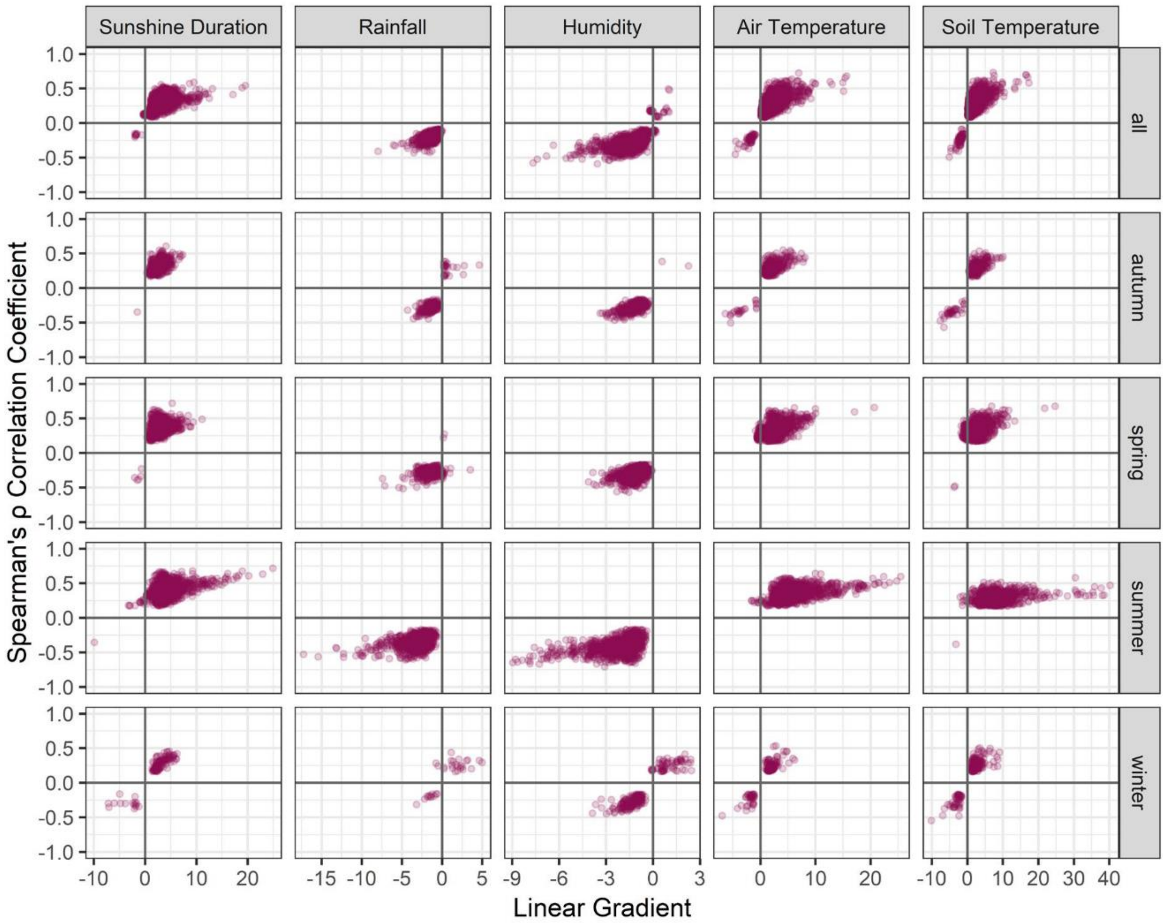

Figure A1. Distribution of correlation coefficients and gradients for segmentations that correspond to autumn, spring, summer and winter consumption, as well as all of the above. Each point in the plot corresponds to the relationship between consumption and one weather variable, for one segmentation of consumption.

Figure A2 shows that more statistically significant relationships, as well as higher gradients and ρ values relate to working days, compared to holidays and weekends.

Regarding the time of the day (Figure A3), by far the steepest gradients among all weather variables are identified for the evening hours.

Results indicate that both properties with large and medium gardens fluctuate their consumption due to weather changes, although larger gardens are linked to higher gradients, i.e. higher increase in water consumption (Figure A4).

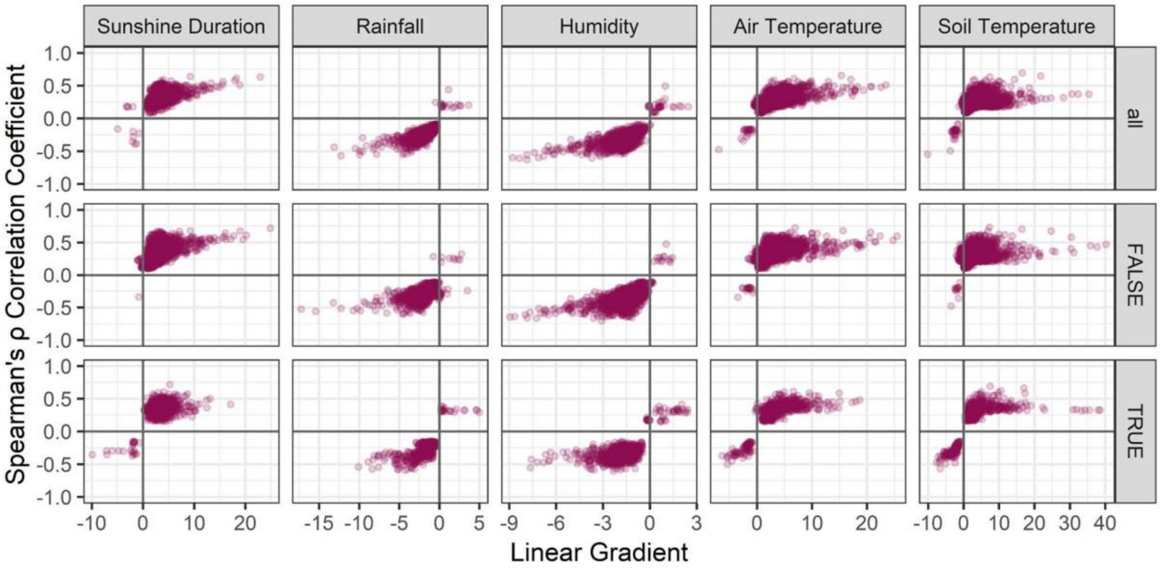

Figure A2. Distribution of correlation coefficients and gradients for segmentations that correspond to working days and weekends/holidays, as well as all of the above. Each point in the plot corresponds to the relationship between consumption and one weather variable, for one segmentation of consumption.

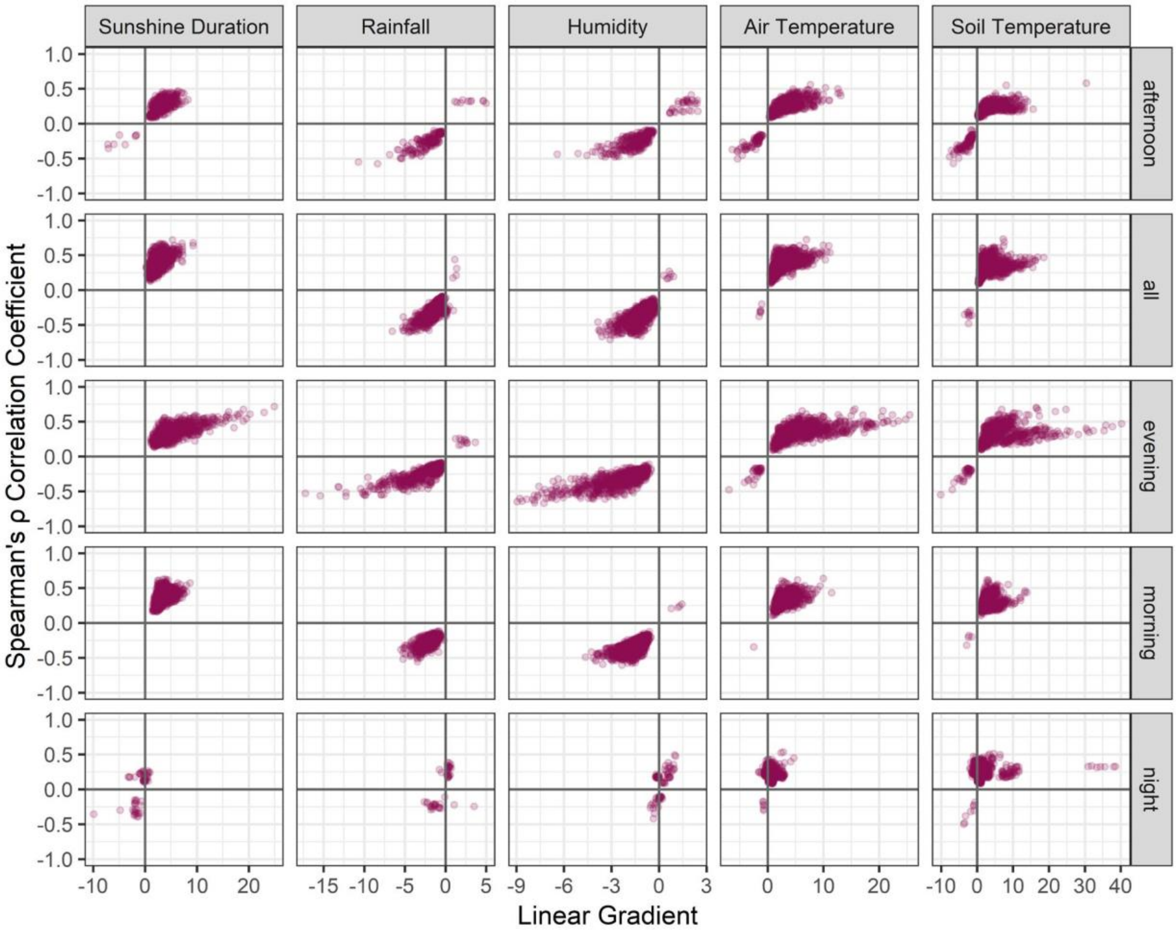

Figure A3. Distribution of correlation coefficients and gradients for segmentations that correspond to afternoons, evenings, mornings and nights, as well as all of the above. Each point in the plot corresponds to the relationship between consumption and one weather variable, for one segmentation of consumption.

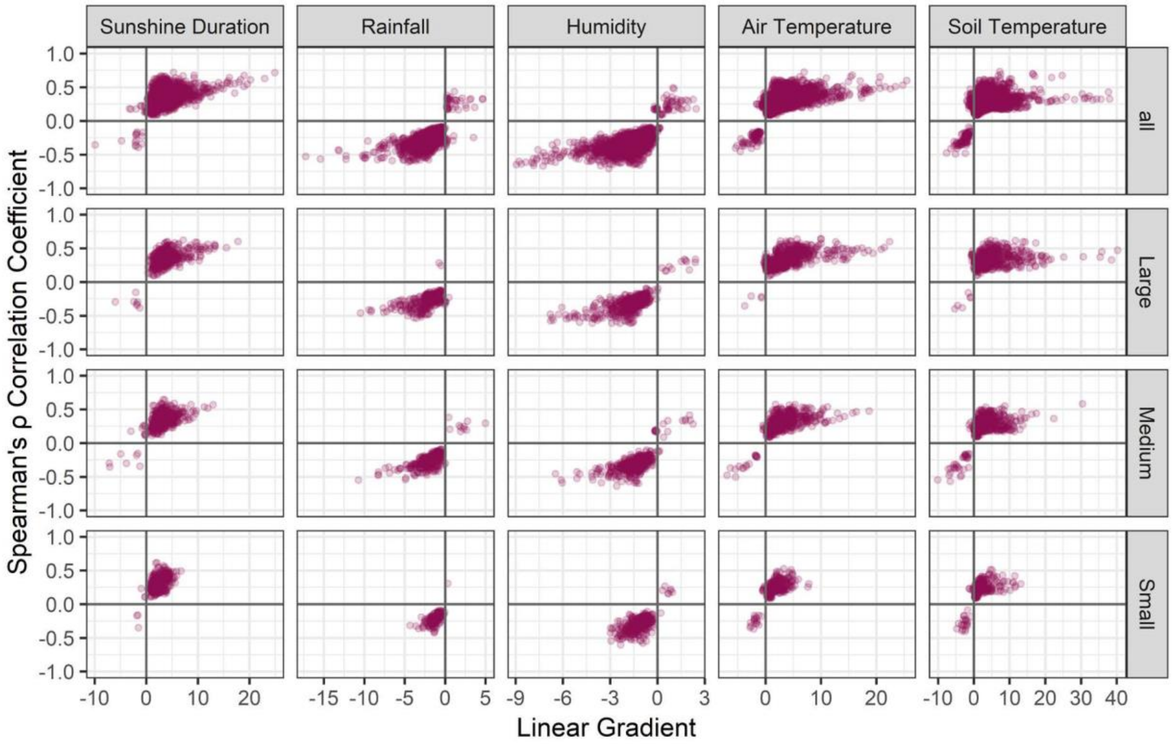

Figure A4. Distribution of correlation coefficients and gradients for segmentations that correspond to large, medium and small gardens, as well as all of the above. Each point in the plot corresponds to the relationship between consumption and one weather variable, for one segmentation of consumption.

The same applies to unmetered customers that also appear more sensitive to weather changes (Figure A5).

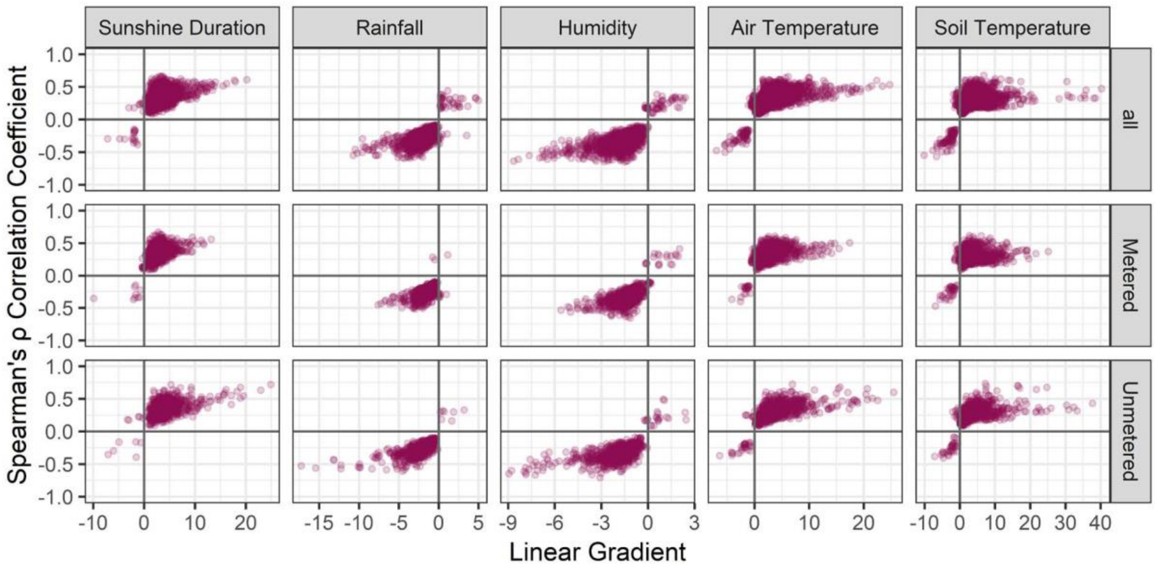

Figure A5. Distribution of correlation coefficients and gradients achieved for segmentations that corresponded to metered and unmetered customers, as well as all of the above. Each point in the plot corresponds to the relationship between consumption and one weather variable for one segmentation of consumption.

However, little change appears among households with varying rateable values (Figure A6).

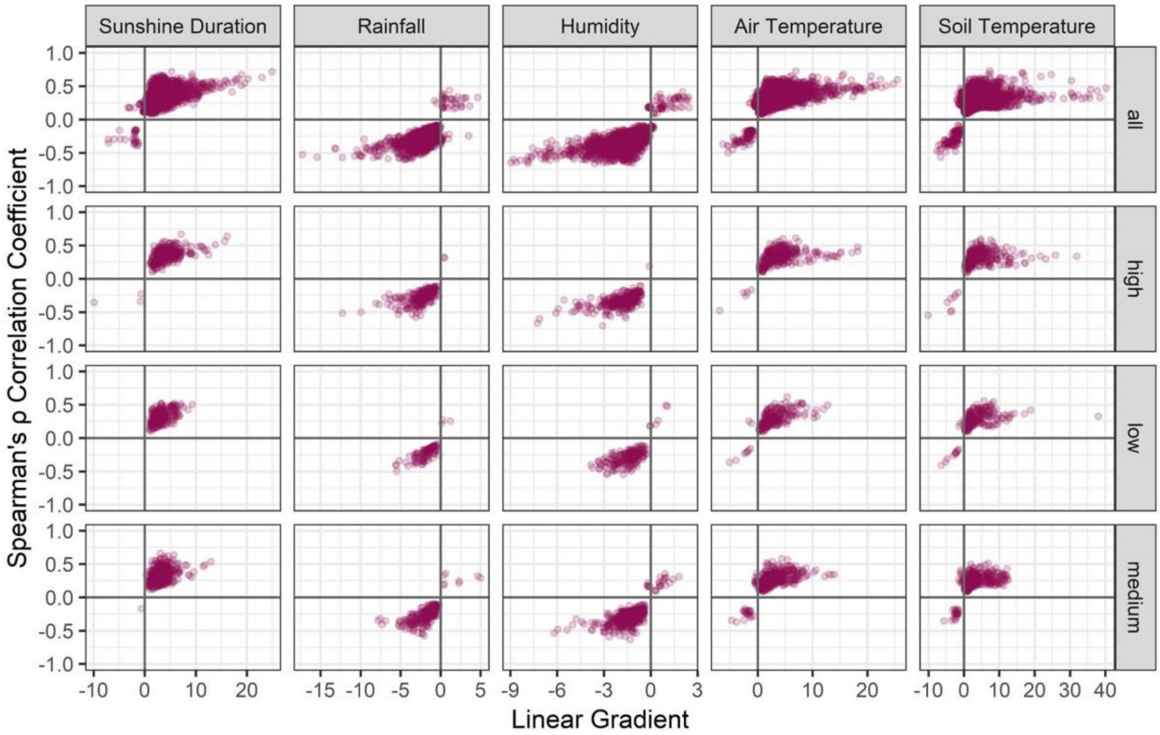

Figure A6. Distribution of correlation coefficients and gradients achieved for segmentations that corresponded to high, medium, and low rateable values, as well as all of the above. Each point in the plot corresponds to the relationship between consumption and one weather variable for one segmentation of consumption.

Finally, Figures A7 to A9 show that the clouds of points that correspond to affluent residents (Figure A7), customers with high variation in their monthly consumption (Figure A8), as well as medium occupancy (Figure A9) are also shifted towards the higher gradients and ρ values.

It is worth noting that for most of the plots in Figures A1-A9, the points cover only two of the four quarters of the Euclidean space. Positive ρ values correspond to positive gradients whereas the same applies for negative ones, reflecting the direct or inverse, respectively, relationship between consumption and the corresponding weather variable.

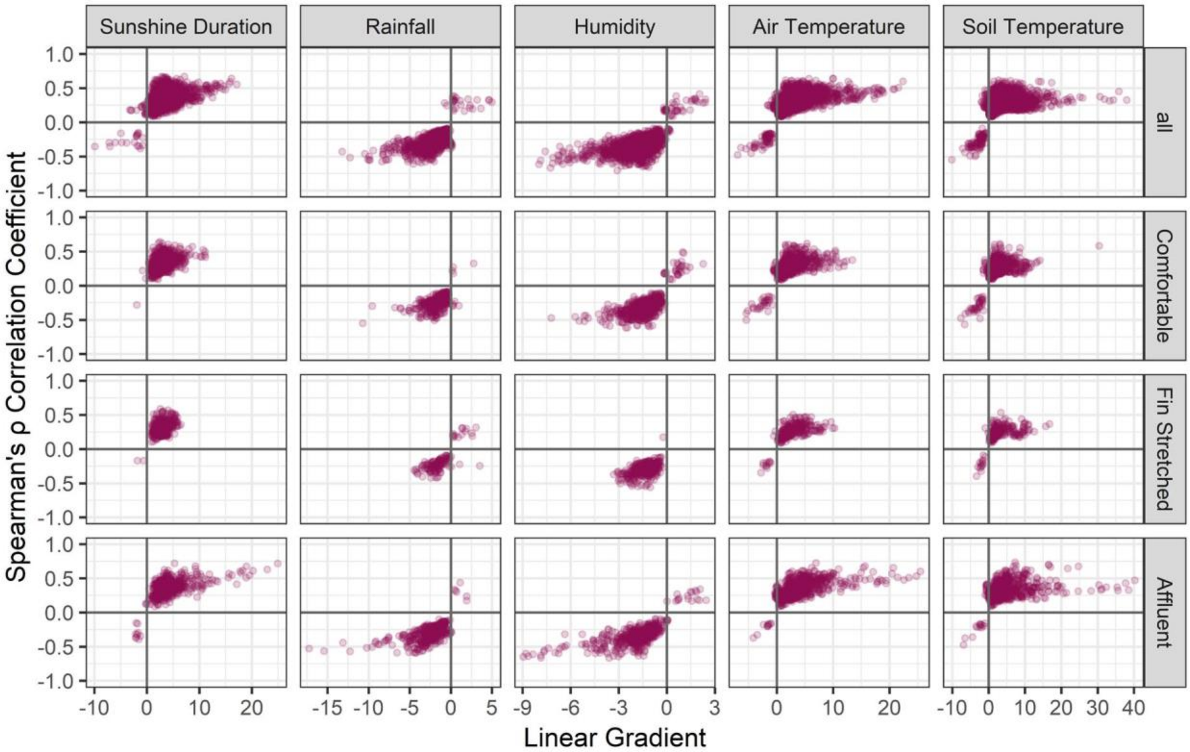

Figure A7. Distribution of correlation coefficients and gradients for segmentations that correspond to comfortable, financially stretched and affluent customers, as well as all of the above. Each point in the plot corresponds to the relationship between consumption and one weather variable, for one segmentation of consumption.

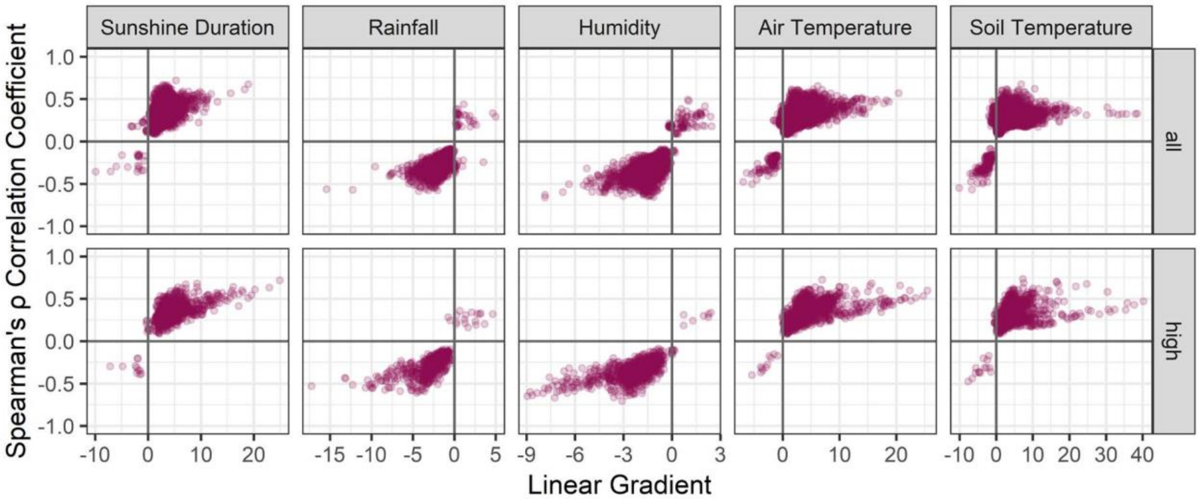

Figure A8. Distribution of correlation coefficients and gradients for segmentations that correspond to customers with high variation in their monthly consumption, as well as all customers. Each point in the plot corresponds to the relationship between consumption and one weather variable, for one segmentation of consumption.

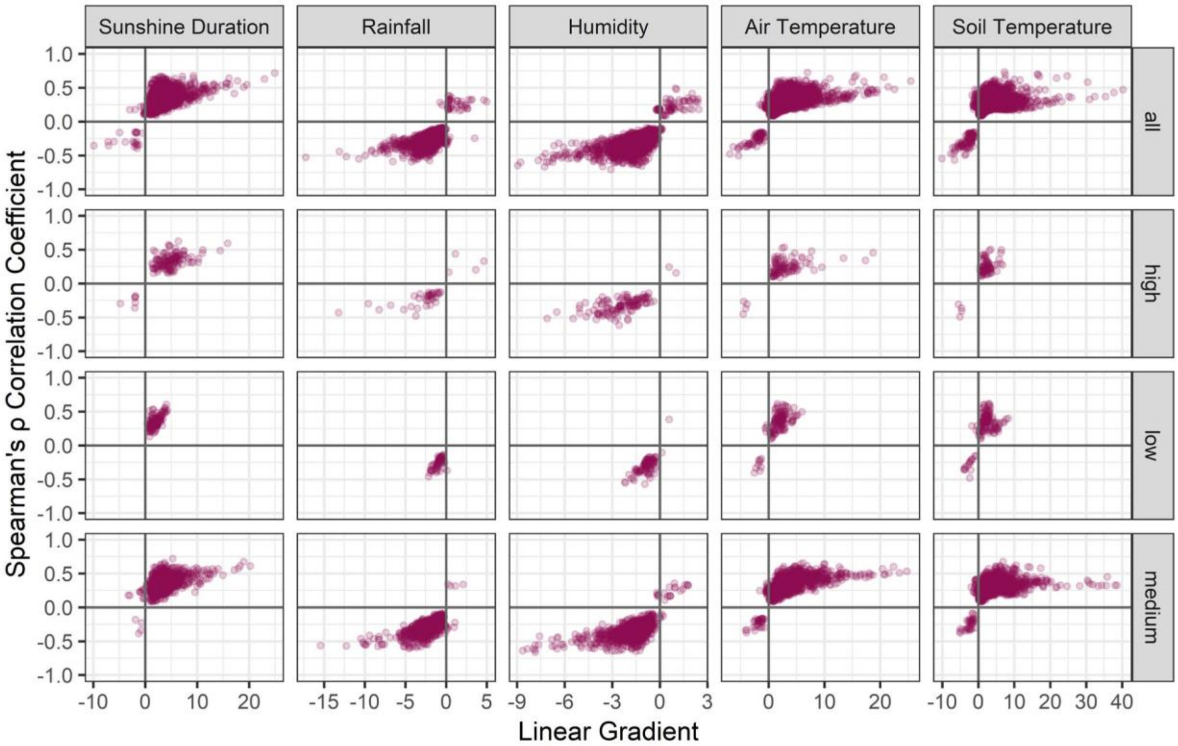

Figure A9. Distribution of correlation coefficients and gradients for segmentations that correspond to high, low and medium occupancy households, as well as all of the above. Each point in the plot corresponds to the relationship between consumption and one weather variable, for one segmentation of consumption.

APPENDIX

B

SUPPORTING INFORMATION – CHAPTER 5

B.1 Model Hyperparameters

'H2o' is used to tune the models for nine hyperparameters using a random grid search. The final hyperparameter values for each model are presented in the following. Table B1 shows the chosen parameters for each one of the models that are trained with only seven days of past consumption as input, for nine aggregations of properties.

Table B1. Hyperparameter values selected for the GBM, for different group sizes.

Hyperparameters	Household group sizes								
	5	**10**	**20**	**40**	**80**	**120**	**200**	**400**	**600**
Ntrees	44	54	329	3545	45	605	30	34	758
Max_depth	6	3	5	5	15	8	10	10	16
Learn_rate	0.1	0.1	0.01	0.001	0.1	0.005	0.1	0.1	0.005
Sample_rate	0.8	0.8	0.7	0.9	0.8	0.5	0.8	0.8	0.9
Col_sample_rate	0.8	1	1	0.4	0.8	0.7	0.8	0.8	1
Col_saple_rate_per_tree	0.8	1	0.7	1	0.8	0.7	0.8	0.8	0.4
Histogram_type	auto	auto	auto	auto	auto	auto	auto	auto	auto
Min_split_imrpovement	1e-05	1e-05	1e-05	1e-04	1e-05	1e-05	1e-05	1e-05	1e-04
Min_rows	1	5	10	100	100	15	10	10	30

Tables B2 to B4 show the hyperparameter values for models 1-8, for each spatial aggregation of properties. Table B2 refers to aggregation at the district level, Table B3 at the area level and Table B4 at the network level. The 'auto' tag under the histogram type means that the cuts that are tested for splitting at each node of the decision trees are chosen by dividing the variable range in equal steps, which here are 20. As it can be seen from the following tables, when the learning rate of the algorithm decreases, the number of trees increases, as the model requires more trees to converge to a solution when the trees have smaller contributions to the final result.

Table B2. Hyperparameter values for models 1-8, for the district level.

Hyperparameters	1	2	3	4	5	6	7	8
Ntrees	81	388	342	352	104	81	28	104
Max_depth	3	5	5	5	3	3	3	3
Learn_rate	0.1	0.01	0.01	0.01	0.08	0.1	0.1	0.08
Sample_rate	0.8	0.7	0.7	0.7	1	0.8	0.8	1
Col_sample_rate	1	1	1	1	0.4	1	1	0.4
Col_saple_rate_per_tree	1	0.7	0.7	0.7	0.4	1	1	0.4
Histogram_type	auto	auto	auto	auto	auto	auto	auto	auto
Min_split_imrpovement	1e-05	1e-05	1e-05	1e-05	1e-05	1e-05	1e-05	1e-05
Min_rows	5	10	10	10	5	5	5	5

Table B3. Hyperparameter values for models 1-8, for the area level.

Hyperparameters	1	2	3	4	5	6	7	8
Ntrees	56	446	46	52	59	47	55	36
Max_depth	10	5	15	15	10	8	3	8
Learn_rate	0.1	0.01	0.1	0.1	0.1	0.1	0.1	0.1
Sample_rate	0.8	0.7	0.8	0.8	0.8	0.8	0.8	0.8
Col_sample_rate	0.8	1	0.8	0.8	0.8	0.8	1	0.8
Col_saple_rate_per_tree	0.8	0.7	0.8	0.8	0.8	0.8	1	0.8
Histogram_type	auto	auto	auto	auto	auto	auto	auto	auto
Min_split_imrpovement	1e-05	1e-05	1e-05	1e-05	1e-05	1e-05	1e-05	1e-05
Min_rows	10	10	100	100	10	10	5	10

Table B4. Hyperparameter values, for models 1-8, for the network level.

Hyperparameters	1	2	3	4	5	6	7	8
Ntrees	118	62	5708	98	42	85	42	35
Max_depth	8	3	12	15	8	15	3	3
Learn_rate	0.05	0.1	0.001	0.1	0.1	0.1	0.1	0.1
Sample_rate	1	0.8	0.9	0.8	0.8	0.8	0.8	0.8
Col_sample_rate	1	1	0.7	0.8	0.8	0.8	1	1
Col_saple_rate_per_tree	1	1	0.7	0.8	0.8	0.8	1	1
Histogram_type	auto	auto	auto	auto	auto	auto	auto	auto
Min_split_imrpovement	1e-04	1e-05	1e-04	1e-05	1e-05	1e-05	1e-05	1e-05
Min_rows	5	5	100	100	10	100	5	5

These values are provided for guidance only and as a good starting point for the hyperparameter values but they do not replace the need for tuning the model based on the corresponding dataset.

B.2 Household composition

In order to provide additional insight into the above results, Figure B1 demonstrates the variation in household composition among the groups in each one of three aggregation levels (network, area, district). In the boxplots presented in Figure B1, the lower and upper hinges correspond to the first and third quantiles (the 25th and 75th percentiles). If IQR is the distance between them, the lower and upper whiskers are calculated as follows:

Lower whisker = max (lower hinge – 1.5*IQR, min value)

Upper whisker = min (upper hinge + 1.5*IQR, max value)

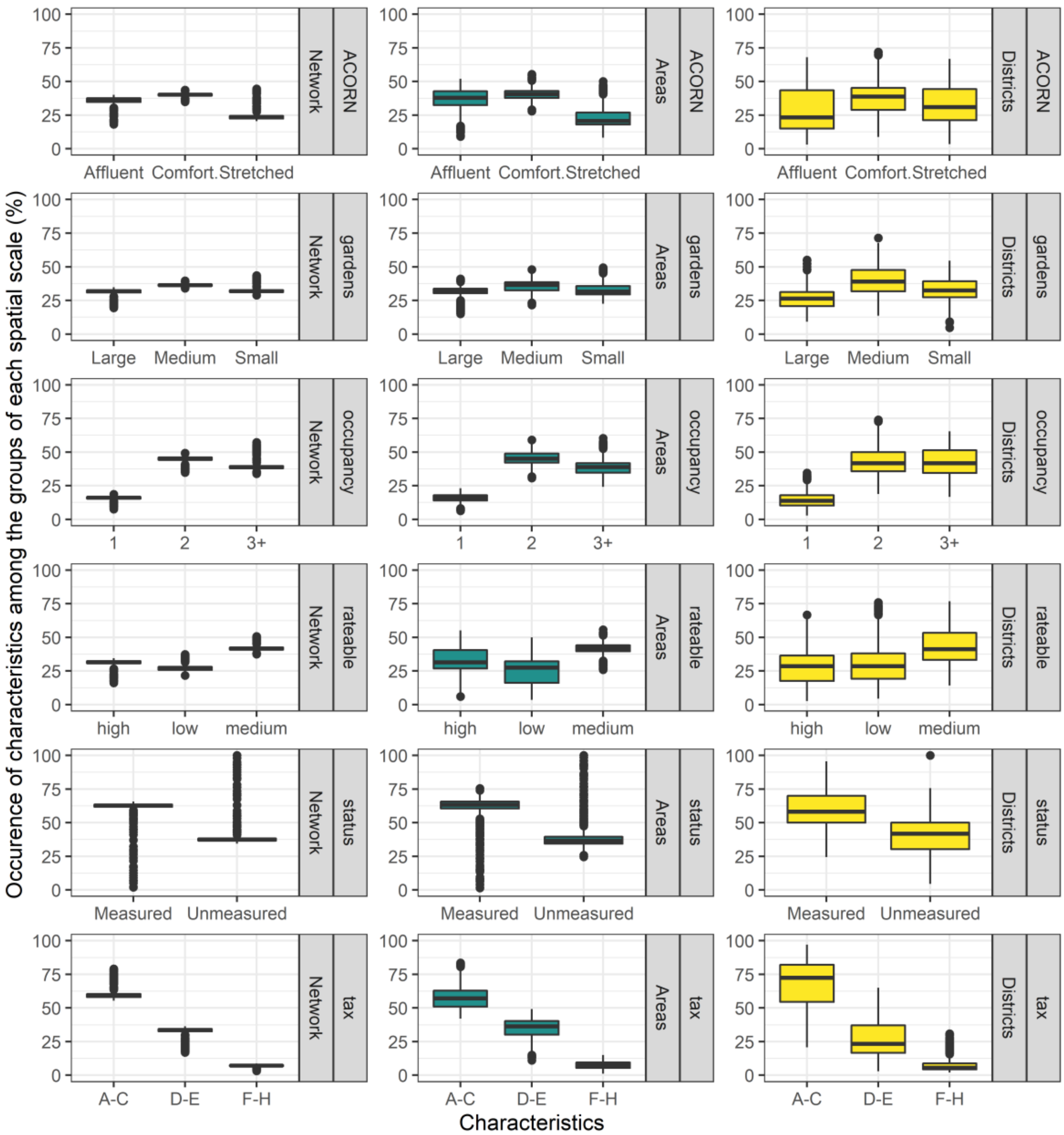

Figure B1. Group composition in terms of household types among the groups, for each level of spatial aggregation.

All values outside the lower and upper whiskers are considered outliers and are plotted individually on each boxplot.

According to Figure B1, when grouping all households together (Figure B1, Network), the household composition among each group does not vary greatly, while for the area and district aggregations, the variation gradually increases (Figure B1, Area and District).

Bibliography

Adamowski, J. F. (2008). Peak Daily Water Demand Forecast Modeling Using Artificial Neural Networks. *Journal of Water Resources Planning and Management*, *134*(2), pp. 119–128. https://doi.org/10.1061/(asce)0733-9496(2008)134:2(119).

Agthe, D.E. and Billings, R.B. (2002). Water Price Influence on Apartment Complex Water Use. *J. Water Resources Planning and Management,* 128(5).

Aiello, S., Eckstrand, E., Fu, A., Landry, M., and Aboyoun, P. (2016). Machine Learning with R and H2O. http://h2o.ai/resources/.

Allon, F. and Sofoulis, Z. (2006). Everyday water: cultures in transition. *Australian Geographer, 37*(1), pp. 45-55.

Altunkaynak, A. and Nigussie, T. A. (2018). Monthly water demand prediction using wavelet transform, first-order differencing and linear detrending techniques based on multilayer perceptron models. *Urban Water Journal*, *15*(2), pp. 177–181. doi: 10.1080/1573062X.2018.1424219.

Al-Qunaibet, M. H., and Johnston, R. S. (1985). Municipal demand for water in Kuwait: Methodological issues and empirical results. *Water Resources Research.*

Anele, A. O., Hamam, Y., Abu-Mahfouz, A. M., and Todini, E. (2017). Overview, comparative assessment and recommendations of forecasting models for short-term water demand prediction. *Water (Switzerland)*, *9*(11). https://doi.org/10.3390/w9110887.

Antunes, A., Andrade-Campos, A., Sardinha-Lourenço, and Oliveira, M. S. (2018). Short-term water demand forecasting using machine learning techniques. *Journal of Hydroinformatics*, *20*(6), pp. 1343-1366. https://doi.org/10.2166/hydro.2018.163**.**

Apley, D. W. and Zhu, J. (2016). Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models. pp. 1–36. Retrieved from: http://arxiv.org/abs/1612.08468 [Accessed 27 September 2019].

Babel, M. S., Gupta, A. D., and Pradhan, P. (2007). A multivariate econometric approach for domestic water demand modeling: An application to Kathmandu, Nepal. *Water Resources Management*, *21*(3), pp. 573–589. https://doi.org/10.1007/s11269-006-9030-6.

Bai, Y., Wang, P., Li, C., Xie, J., and Wang, Y. (2014). A multi-scale relevance vector regression approach for daily urban water demand forecasting. *Journal of Hydrology*, *517*, pp. 236–245. https://doi.org/10.1016/j.jhydrol.2014.05.033.

Bailey, S. (2019). Consumption per head in the UK : 2017. Office for National Statistics, pp. 1–9.

Bakker, M., Van Duist, H., Van Schagen, K., Vreeburg, J., and Rietveld, L. (2014). Improving the performance of water demand forecasting models by using weather input. *Procedia Engineering*, *70*, pp. 93–102. https://doi.org/10.1016/j.proeng.2014.02.012.

Balling, R. C., Gober, P., and Jones, N. (2008). Sensitivity of residential water consumption to variations in climate: An intraurban analysis of Phoenix, Arizona. *Water Resources Research*, *44*(10), pp. 1–11. https://doi.org/10.1029/2007WR006722.

Beal, C. D. and Stewart A. S. (2013). Identifying residential water end uses underpinning peak day and peak hour demand. *J. Water Resources Planning and Management, 140*(7). https://doi.org/10.1061/(ASCE)WR.1943-5452.0000357.

Bellfield, S. (2001). Short-term domestic water demand: estimation, forecasting and management. Retrieved from: http://etheses.whiterose.ac.uk/2576/ [Accessed 27 September 2019].

Bello-Dambatta, A., Kapelan, Z., and Butler, D. (2014). Impact assessment of household demand saving technologies on system water and energy use. *British J. Environment and Climate Change, 4*(2), pp. 243-260.

Beran, M. A. and Arnell, N. W. (1989). Effect of climatic change on quantitative aspects of United Kingdom water resources, Institute of Hydrology, Wallingford. Report to Department of the Environment.

Bergstra, J. and Bengio, Y. (2012). Random Search for Hyper-Parameter Optimization. *J. Machine Learning Research*, 13, pp. 281-305.

Billings, R. B. and Jones, C. V. (2008). Forecasting urban water demand. Denver (Colorado): s.n.

Bottou, L. (2010). Large-scale machine learning with stochastic gradient descent. *International Conference on Computational Statistics*, pp. 177–187.

Bougadis, J., Adamowski, K., and Diduch, R. (2005). Short-term municipal water demand forecasting. *Hydrological Processes*, *19*(1), pp. 137–148. https://doi.org/10.1002/hyp.5763.

Brdys, M. and Ulanicki, B. (1994). Operational Control of Water Systems: Structures, algorithms and applications. Prentice Hall International.

Breiman, L. (2001). Random forests. *Ensemble Machine Learning: Methods and Applications*, pp. 157–175. https://doi.org/10.1007/9781441993267_5.

Brentan, B. M., Meirelles, G., Herrera, M., Luvizotto, E., and Izquierdo, J. (2017). Correlation Analysis of Water Demand and Predictive Variables for Short-Term Forecasting Models. *Mathematical Problems in Engineering*, pp. 1–10. https://doi.org/10.1155/2017/6343625.

Breyer, B. and Chang, H. (2014). Urban water consumption and weather variation in the Portland, Oregon Metropolitan area. *J. Urban Climate, 9*, pp. 1-18.

Browne, A. L., Medd, W., and Anderson, B. (2013). Developing novel approaches to tracking domestic water demand under uncertainty - A reflection on the "Up scaling" of social science approaches in the Unitek Kingdom. *Water Resources Management, 27*, pp. 1013-1035.

Butler, D. (1993). The influence of dwelling occupancy and day of the week on domestic appliance wastewater discharges. *Build Environment, 28*, pp. 73–79.

CACI Limited. (2014). The Acorn User Guide. *The Consumer Classification*, 108. Retrieved from: https://acorn.caci.co.uk/downloads/Acorn-User-guide.pdf [Accessed 27 September 2019].

Candel, A. and Parmar, V. (2014). Deep Learning with H2O, pp. 1–21.

Candelieri, A., Soldi, D., and Archetti, F. (2015). Short-term forecasting of hourly water consumption by using automatic meter readers data. *Procedia Engineering, 119*, pp. 844-853.

Chang, H., Parandvash, G. H., and Shandas, V. (2010). Spatial variations of single-family residential water consumption in Portland, Oregon. *J. Urban Geography, 31*(7), pp. 953-972.

Chen, G., Long, T., Xiong, J., and Bai, Y. (2017). Multiple Random Forests Modelling for Urban Water Consumption Forecasting. *Water Resources Management*, *31*(15), pp. 4715–4729. https://doi.org/10.1007/s11269-017-1774-7.

Chen, J. and Boccelli, D. L. (2018). Forecasting Hourly Water Demands With Seasonal Autoregressive Models for Real-Time Application. *Water Resources Research*, *54*(2), pp. 879–894. https://doi.org/10.1002/2017WR022007.

Chen, T. and Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794. https://doi.org/10.1145/2939672.2939785.

Chen, T. and He, T. (2015). xgboost : eXtreme Gradient Boosting. *R Version 0.4-2*, 1–4.

Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., et al. (2019). Package 'xgboost', *R version 0.90.0.2.*

CIWEM (2015). Water Efficiency in the Home.

Clarke, G. P., Kashti, A., Mcdonald, A., and Williamson, P. (1997). Estimating Small Area Demand for Water: A New Methodology. *J. CIWEM*, pp. 186–192.

Click, C., Malohlava, M., Parmar, V., Roark, H., and Candel, A. (2017). Gradient Boosting Machine with H2O. http://h2o.ai/resources/.

Cole, G., and Stewart, R. A. (2013). Smart meter enabled disaggregation of urban peak water demand: Precursor to effective urban water planning. *Urban Water Journal, 10*(3), pp. 174-194.

Cominola, A., Giuliani, M., Piga, D., Castelletti, A., and Rizzoli, A. E. (2015). Benefits and challenges of using smart meters for advancing residential water demand modeling and management: A review. *Environmental Modelling and Software*, *72*, pp. 198–214. https://doi.org/10.1016/j.envsoft.2015.07.012.

Cutore, P., Campisano, A., Kapelan, Z., Modica, C., and Savic, D. (2008). Probabilistic prediction of urban water consumption using the SCEM-UA algorithm. *Urban Water Journal*, *5*(2), pp. 125–132. https://doi.org/10.1080/15730620701754434.

Defra (2003) Initial guidance from the secretary of state to the Director-General of water services. Department for Environment, Food and Rural Affairs.

Defra (2008). Future Water The Government's water strategy for England. Department for Environment, Food and Rural Affairs. Retrieved from: www.tsoshop.co.uk [Accessed 27 September 2019].

Defra (2011). Water for life. Department for Environment, Food and Rural Affairs. Retrieved from: http://static.ow.ly/docs/WaterisLife_1hi.pdf [Accessed 27 September 2019].

Domene, E. and Sauri, D. (2006). Urbanisation and water consumption: Influencing factors in the metropolitan region of Barcelona. *Urban Studies, 43*(9), pp. 1605-1623.

Donkor, E. A., Mazzuchi, T. A., Soyer, R., and Roberson, J. A. (2014). Urban Water Demand Forecasting: Review of Methods and Models. *Journal of Water Resources Planning and Management*, *140*(2), pp. 146–159. https://doi.org/10.1061/(ASCE)WR.1943-5452.

Dos Santos, C. C. and Pereira, F. A. J. (2014). Water Demand Forecasting Model for the Metropolitan Area of São Paulo, Brazil. *Water Resources Management*, *28*(13), pp. 4401–4414. https://doi.org/10.1007/s11269-014-0743-7.

Downing, T. E., Butterfield, R. E., Edmonds, B., Knox, J. W., Moss, S., Piper, B. S., and Weatherhead, E. K. (2003). Climate change and the demand for water. Oxford, UK: Stockholm Environment Institute.

Duerr, I., Merrill, H. R., Wang, C., Bai, R., Boyer, M., Dukes, M. D., and Bliznyuk, N. (2018). Forecasting urban household water demand with statistical and machine learning methods using large space-time data: A Comparative study. *Environmental Modelling and Software*, *102*, pp. 29–38. https://doi.org/10.1016/j.envsoft.2018.01.002.

Edwards, K. and Martin, L. (1995). A Methodology for Surveying Domestic Water Consumption. *Water and Environment Journal*, *9*(5), pp. 477–488. https://doi.org/10.1111/j.1747-6593.1995.tb01486.x.

Environment Agency (2003) Water resources planning guidelines, version 3.2. Environment Agency, Bristol.

Fan, J., Wang, X., Wu, L., Zhou, H., Zhang, F., Yu, X., Lu, X., and Xiang, Y. (2018). Comparison of Support Vector Machine and Extreme Gradient Boosting for predicting daily global solar radiation using temperature and precipitation in humid subtropical climates: A case study in China. *Energy Conversion and Management*, *164*, pp. 102–111. https://doi.org/10.1016/j.enconman.2018.02.087.

Fidar, A., Memon, F., and Butler, D. (2010). Environmental implications of water efficient micro-components in residential buildings. *Science of the Total Environment*, *408*, pp. 5828–5835.

Firat, M., Turan, M. E., and Yurdusev, M. A. (2009). Comparative analysis of fuzzy inference systems for water consumption time series prediction. *Journal of Hydrology*, *374*(3–4), 235–241. https://doi.org/10.1016/j.jhydrol.2009.06.013.

Fisher, A., Rudin, C., and Domici, F. (2019). All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. Retrieved from: https://arxiv.org/pdf/1801.01489.pdf [Accessed 06/09/2019].

Fontanazza, C. M., Notaro, V., Puleo, V., and Freni, G. (2014). Multivariate statistical analysis for water demand modelling. *Procedia Engineering*, *89*, pp. 901-908.

Fox, C., McIntosh, B. S., and Jeffrey, P. (2009). Classifying households for water demand forecasting using physical property characteristics. *Land Use Policy*, *26*(3), pp. 558–568. https://doi.org/10.1016/j.landusepol.2008.08.004.

Friedman, J. (2001). Greedy Function Approximation : A Gradient Boosting Machine, *The Annals of Statistics*, *29*(5), pp. 1189–1232. https://doi.org/10.1214/009053606000000795.

Froukh, M. L. (2001). Decision-support system for domestic water demand forecasting and management. *Water Resources Management*, *15*(6), pp. 363–382. https://doi.org/10.1023/A:1015527117823.

Garcia, D., González Vidal, D., Quevedo, J., Puig, V., and Saludes, J. (2015). Water Demand Estimation and Outlier Detection from Smart Meter Data Using Classification and Big Data Methods. *2nd New Developments in IT & Water Conference*, pp. 1–8. Retrieved from: http://hdl.handle.net/2117/26473%0Ahttps://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0ahUKEwi67_KVkdPPAhUs6oMKHVo4BPsQFgghMAA&url=http://upcommons.upc.edu/bitstream/handle/2117/26473/IWC 2015 full paper 036 J. Quevedo.pdf [Accessed 27 September 2019].

Gato, S., Jayasuriya, N., and Roberts, P. (2007). Temperature and rainfall thresholds for base use urban water demand modelling. *J. Hydrology, 337*, pp. 364-376.

Geurts, P., Ernst, D., and Wehenkel, L. (2006). Extremely randomized trees. *Machine Learning*, *63*(1), pp. 3–42. https://doi.org/10.1007/s10994-006-6226-1.

Ghalehkhondabi, I., Ardjmand, E., Young, W. A., and Weckman, G. R. (2017). Water demand forecasting: review of soft computing methods. *Environmental Monitoring and Assessment*, *189*(7). https://doi.org/10.1007/s10661-017-6030-3.

Ghiassi, M., Fa'al, F., and Abrishamchi, A. (2017). Large metropolitan water demand forecasting using DAN2, FTDNN, and KNN models: A case study of the city of Tehran, Iran. *Urban Water Journal, 14*(6), pp. 655–659. doi: 10.1080/1573062X.2016.1223858.

Ghosal, I. and Hooker, G. (2018). Boosting Random Forests to Reduce Bias; One-Step Boosted Forest and its Variance Estimate. Retrieved from: http://arxiv.org/abs/1803.08000 [Accessed 27 September 2019].

Goodchild, C. (2003). Modelling the impact of climate change on domestic water demand. *J. Chart. Ins. Water Environ. Manag, 17*(1), pp. 8-12.

Goldstein, A., Kapelner, A., Bleich, J., and Pitkin, E. (2015). Peeking Inside the Black Box: Visualizing Statistical Learning With Plots of Individual Conditional

Expectation. *Journal of Computational and Graphical Statistics, 24*(1), pp. 44-65, https://doi.org/10.1080/10618600.2014.907095.

Griffin, R. C. and Chang, C. (1991). Seasonality in community water demand. *Western J. of Agricultural Economics, 16*(2), pp. 207-217.

Gurung, T. R., Stewart, R. A., Sharma, A. K., and Beal, C. D. (2014). Smart meters for enhanced water supply network modelling and infrastructure planning. *Resources, Conservation and Recycling*, *90*, pp. 34–50. https://doi.org/10.1016/j.resconrec.2014.06.005.

H2O.ai (2019a), Appendix A -Parameters: alpha. Retrieved from: https://h2o-release.s3.amazonaws.com/h2o/master/3908/docs-website/h2o-docs/data-science/algo-params/alpha.html#description [Accessed 27 September 2019].

H2O.ai (2019b), AutoML: Automatic Machine Learning. Retrieved from: http://docs.h2o.ai/h2o/latest-stable/h2o-docs/automl.html [Accessed 27 September 2019].

H2O.ai (2019c), Overview. Retrieved from: http://docs.h2o.ai/h2o/latest-stable/h2o-docs/index.html [Accessed 27 September 2019].

Hanke, S. H. and Flack, J. E. (1968). Effects of metering urban water. *J. American Water Works Association, 60*(12), pp. 1359-1366.

Haque, M., Hagare, D., Rahman, A., and Kibria, G. (2014). Quantification of water savings due to drought restrictions in water demand forecasting models. *Journal of Water Resources Planning and Management*, *140*(11). https://doi.org/10.1061/(ASCE)WR.1943-5452.0000423.

Herrera, M., Torgo, L., Izquierdo, J., and Pérez-García, R. (2010). Predictive models for forecasting hourly urban water demand. *Journal of Hydrology*, *387*(1–2), pp. 141–150. https://doi.org/10.1016/j.jhydrol.2010.04.005.

Herrington, P. R. (1996). Climate change and the demand for water. Department of the Environment.

Hooker, G. and Mentch, L. (2018). Bootstrap bias corrections for ensemble methods, *Statistics and Computing*, *28*(1), pp. 77–86. https://doi.org/10.1007/s11222-016-9717-3.

Hope, R., Foster, T., Money, A., Rouse, M. , Money, N., and Thomas, M. (2011). Smart Water Systems. Project report to UK DFID. Oxford University, Oxford.

House-Peters, L. A. and Chang, H. (2011). Urban water demand modeling: Review of concepts, methods, and organizing principles. *Water Resources Research*, *47*(5). https://doi.org/10.1029/2010WR009624.

House-Peters, L., Pratt, B., and Chang, H. (2010). Effects of Urban spatial structure, sociodemographics, and climate on residential water consumption in hillsboro, Oregon. *Journal of the American Water Resources Association*, *46*(3), pp. 461–472. https://doi.org/10.1111/j.1752-1688.2009.00415.x.

Hussien, W. A., Memon, F. A., and Savic, D. A. (2016). Assessing and modelling the influence of household characteristics on per capita water consumption. *Water Resources Management, 30*, pp. 2931-2955. https://doi.org/10.1007/s11269-016-1314-x.

Hutton, C. J. and Kapelan, Z. (2015). A probabilistic methodology for quantifying, diagnosing and reducing model structural and predictive errors in short term water demand forecasting. *Environmental Modelling and Software*, *66*, pp. 87–97. https://doi.org/10.1016/j.envsoft.2014.12.021.

Jain, A., Varshney, A. K., and Joshi U. C. (2001). Short-term water demand forecast modelling at IIT Kanpur using artificial neural networks. *Water Resources Management.*

Khatri, K. B. and Vairavamoothry, K. (2009). Water demand forecasting for the city of the future against the uncertainties and the global change pressures: Case of Birmingham. *World Environmental and Water Resources Congress*.

Kofinas, D., Mellios, N., Papageorgiou, E., and Laspidou, C. (2014). Urban water demand forecasting for the island of skiathos. *Procedia Engineering*, *89*, pp. 1023–1030. https://doi.org/10.1016/j.proeng.2014.11.220.

Kuhn, M. (2019). Package ‘caret’. R version 6.0-84.

Lecun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), pp. 436–444. https://doi.org/10.1038/nature14539.

LeDell, E., Gill, N., Aiello, S., Fu, A., Candel, A., Click, C., Kraljevic, T., et al. (2019). Package 'h2o', R package version 3.26.0.2.

Lee, S. J., Wentz, E. A., and Gober, P. (2010). Space-time forecasting using soft geostatistics: A case study in forecasting municipal water demand for Phoenix, Arizona. *Stochastic Environmental Research and Risk Assessment*, *24*(2), pp. 283–295. https://doi.org/10.1007/s00477-009-0317-z.

Liaw, A. (2018). Package 'randomForest'. R version 4.6-14.

Linaweaver, F. P. Jr., Geyer, J. C., and Wolff, J. B. (1967). Summary report on the residential water use research project. *American Water Works Association, 59*(3), pp. 267-282.

Loh, M. and Coghlan, P. (2003). Domestic water use study: Perth, Western Australia, Perth, Western Australia: Water Corporations.

Maheepala, S., Blackmore, J., Diaper, C., Moglia, M., Sharma, A., and Kenway, S. (2011). Integrated Urban Water Management Planning Manual.

Malohlava, M., Candel, A., Bartz, A., Roark, H., and Parmar, V. (2017). Gradient Boosting Machine with H2O: Seventh Edition. Retrieved from: http://h2o.ai/resources/ [Accessed 27 September 2019].

Mamade, A., Loureiro, D., Covas, D., Coelho, S. T., and Amado, C. (2014). Spatial and temporal forecasting of water consumption at the DMA level using extensive measurements. *Procedia Engineering*, *70*, pp. 1063–1073. https://doi.org/10.1016/j.proeng.2014.02.118.

Mao, J. (1998). Case study on Bagging, Boosting, and Basic ensembles of neural networks for OCR. *IEEE International Conference on Neural Networks - Conference Proceedings*, *3*, pp. 1828–1833.

Matos, C., Teixeira, C. A., Bento, R., Varajão, J., and Bentes, I. (2014). An exploratory study on the influence of socio-demographic characteristics on water end uses inside buildings. *Science of the Total Environment*, pp. 467–474. https://doi.org/10.1016/j.scitotenv.2013.07.036.

McKenna, S. A., Fusco, F., and Eck, B. J. (2014). Water demand pattern classification from smart meter data. *Procedia Engineering*, *70*, pp. 1121–1130. https://doi.org/10.1016/j.proeng.2014.02.124.

Merrett, S. (2004). The Demand for Water: Four Interpretations. *Water International*, *29*(1), pp. 27–29. https://doi.org/10.1080/02508060408691745.

Met Office (2012). England Averages*.*
Retrieved from:
https://web.archive.org/web/20130410093339/http://www.metoffice.gov.uk/climate/uk/averages/19812010/areal/england.html
[Accessed 15 February 2019].

Met Office (2006a): MIDAS: UK Daily Rainfall Data. NCAS British Atmospheric Data Centre, Retrieved from:
http://catalogue.ceda.ac.uk/uuid/c732716511d3442f05cdeccbe99b8f90
[Accessed 15 February 2019].

Met Office (2006b): MIDAS: UK Daily Temperature Data. NCAS British Atmospheric Data Centre, Retrieved from:
2019]. http://catalogue.ceda.ac.uk/uuid/1bb479d3b1e38c339adb9c82c15579d8
[Accessed 15 February 2019].

Met Office (2006c): MIDAS: UK Daily Weather Observation Data. NCAS British Atmospheric Data Centre, Retrieved from:
http://catalogue.ceda.ac.uk/uuid/954d743d1c07d1dd034c131935db54e0
[Accessed 27 September 2019].

Met Office (2006d): MIDAS: UK Hourly Weather Observation Data. NCAS British Atmospheric Data Centre, Retrieved from:
http://catalogue.ceda.ac.uk/uuid/916ac4bbc46f7685ae9a5e10451bae7c
[Accessed 27 September 2019].

Met Office (2006e): MIDAS: UK Soil Temperature Data. NCAS British Atmospheric Data Centre, Retrieved from:
http://catalogue.ceda.ac.uk/uuid/8dc05f6ecc6065a5d10fc7b8829589ec
[Accessed 27 September 2019].

Miaou, S. P. (1990). A class of time series urban water demand models with nonlinear climatic effects. *Water Resources Research, 26*(2), pp. 169-178.

Mohamed, M. M. and Al-Mualla, A. A. (2010). Water demand forecasting in Umm Al-Quwain using the constant rate model. *Desalination*, *259*(1–3), pp. 161–168. https://doi.org/10.1016/j.desal.2010.04.014.

Molnar, C. (2019a). Interpretable machine learning. A Guide for Making Black Box Models Explainable. Retrieved from: https://christophm.github.io/interpretable-ml-book/ [Accessed 27 September 2019].

Molnar, C. (2019b). Package 'iml'. R version 0.9.0.

Butler, D. and Memon, F. A., (2006). Water consumption trends and demand forecasting techniques. *Water Demand Management*, pp. 1–26.

Mouatadid, S. and Adamowski, J. (2017). Using extreme learning machines for short-term urban water demand forecasting. *Urban Water Journal*, *14*(6), pp. 630–638. https://doi.org/10.1080/1573062X.2016.1236133.

Ngo, K. T. (2018). Stacking Ensemble for auto_ml.

Nguyen, T. T., Huang, J. Z., and Nguyen, T. T. (2015). Two-level quantile regression forests for bias correction in range prediction. *Machine Learning*, *101*(1–3), pp. 325–343. https://doi.org/10.1007/s10994-014-5452-1.

Nykodym, T., Kraljevic, T., Hussami, N., Rao, A., and Wang, A. (2019). Generalized Linear Modeling with h2o. Retrieved from: http://h2o.ai/resources/ [Accessed 27 September 2019].

Ofwat (2013). Water meters – your questions answered.

Ofwat (2017). PR19. Appendix 2: Delivering outcomes for customers.

Ofwat (2018). Reporting Guidance Leakage. pp. 1–44. Retrieved from: https://www.ofwat.gov.uk/publication/reporting-guidance-leakage/ [Accessed 27 September 2019].

Ofwat (2019). Ofwat’s forward programme 2019-20.

Parker J. (2014). Assessing the sensitivity of historic micro-component household water-use to climatic drivers. Loughborough University.

Parker, J. M. and Wilby, R. L. (2013). Quantifying Household Water Demand: A Review of Theory and Practice in the UK. *Water Resources Management*, *27*(4), 981–1011. https://doi.org/10.1007/s11269-012-0190-2.

Pericli, A. and Jenkins, J. O. (2015). Smart Meters and Domestic Water usage. 44. Retrieved from: http://www.fwr.org/econom/SmartMeters.pdf [Accessed 27 September 2019].

Polebitski, A. S. and Palmer, R. N. (2010). Seasonal Residential Water Demand Forecasting for Census Tracts. *Journal of Water Resources Planning and Management*, *136*, pp. 27–36.

Prescott, S. and Ulanicki, B. (2008) Improved control of pressure reducing valves in water distribution networks. *J Hydraulic Engineering*, *134*(1), pp. 56–65.

Pullinger, M., Browne, A., Anderson, B., & Medd, W. (2013). Patterns of water: The water related practices of households in southern England, and their influence on water consumption and demand management. Lancaster University: Lancaster UK. Retrieved from: https://www.escholar.manchester.ac.uk/uk-ac-man-scw:187780 [Accessed 27 September 2019].

Quevedo, J., Saludes, J., Puig, V., and Blanch, J. (2014). Short-term demand forecasting for real-time operational control of the Barcelona water transport network. 22nd Mediterranean conference on control and automation (MED) - Conference Proceedings. https://doi.org/10.1109/MED.2014.6961503.

R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Rathnayaka, K., Malano, H., Arora, M., George, B., Maheepala, S., and Nawarathna, B. (2017a). Prediction of urban residential end-use water demands by integrating known and unknown water demand drivers at multiple scales I: Model development. *Resources, Conservation and Recycling*, *117*, 85–92. https://doi.org/10.1016/j.resconrec.2016.11.014.

Rathnayaka, K., Malano, H., Arora, M., George, B., Maheepala, S., and Nawarathna, B. (2017b). Prediction of urban residential end-use water demands by integrating known and unknown water demand drivers at multiple scales I: Model development. *Resources, Conservation and Recycling*, *117*, pp. 85–92. https://doi.org/10.1016/j.resconrec.2016.11.014.

Rinaudo, J. (2015). Long-Term Water Demand Forecasting. Understanding and Managing Urban Water in Transition, pp. 239-268, https://doi.org/10.1007/978-94-017-9801-3 11>.

Ripley, B. and Venables, W. (2016). Package ‘nnet’. R version 7.3-12.

Romano, M. and Kapelan, Z. (2014). Adaptive water demand forecasting for near real-time management of smart water distribution systems. *Environmental Modelling and Software*, *60*, pp. 265–276. https://doi.org/10.1016/j.envsoft.2014.06.016.

Royal Mail (2012). Postcodes and Addresses explained, Available from: https://web.archive.org/web/20120716211307/http://www.royalmail.com/delivery/mail-advice/postcodes-addresses-explained [Accessed 27 September 2019].

Russac, D. A. V, Rushton, K. R., and Simpson, R. J. (1991). Insights into Domestic Demand from a Metering Trial. *Water and Environment Journal*, *5*(3), 342–351. https://doi.org/10.1111/j.1747-6593.1991.tb00628.x.

Schleich, J. and Hillenbrand, T. (2008). Determinants of residential water demand in Germany. *Ecological Economics,* 68, pp. 1756-1769.

Scornet, E. (2017). Tuning parameters in random forests. *ESAIM: Proceedings and Surveys*, *60*(2001), pp. 144–162. https://doi.org/10.1051/proc/201760144.

Shabani, S., Yousefi, P., Adamowski, J., and Naser, G. (2016). Intelligent Soft Computing Models in Water Demand Forecasting. *Water Stress in Plants*. https://doi.org/10.5772/63675.

Song, J. (2015). Bias corrections for Random Forest in regression using residual rotation. *Journal of the Korean Statistical Society*, *44*(2), pp. 321–326. https://doi.org/10.1016/j.jkss.2015.01.003.

Steinley, D. (2006). K-means Clustering: A Half-century Synthesis. *British Journal of Mathematical and Statistical Psychology, 59*, pp. 1–34.

Stigler, S. M. (1997). Regression towards the mean, historically considered. *Statistical Methods in Medical Research*, *6*(2), pp. 103–114. https://doi.org/10.1177/096228029700600202.

The UK government (2008). Climate Change Act 2008. London: The Stationery Office Limited.

Tiwari, M., Adamowski, J., and Adamowski, K. (2016). Water demand forecasting using extreme learning machines. *Journal of Water and Land Development*, *28*(1), pp. 37–52. https://doi.org/10.1515/jwld-2016-0004.

Touzani, S., Granderson, J., and Fernandes, S. (2018). Gradient boosting machine for modeling the energy consumption of commercial buildings. *Energy and Buildings*, *158*, pp. 1533–1543. https://doi.org/10.1016/j.enbuild.2017.11.039.

UKWIR (2015). UK Water Industry Research Limited Wrmp19 Methods – Household Consumption Forecasting – Guidance Manual. (15). Retrieved from: https://corporate.thameswater.co.uk/-/media/Site-Content/Thames-Water/Corporate/AboutUs/Our-strategies-and-plans/Water-resources/Document-library/External-reports/demandforecastingjan16.pdf.

Ulanicki, B., AbdelMeguid, H., Bounds, P., and Patel, R. (2009). Pressure control in district metering areas with boundary and internal pressure reducing valves. *Geotechnical Special Publication*, (187), pp. 691–703. https://doi.org/10.1061/41023(337)58.

Water U.K. (2016). Water resources long term planning framework (2015-2065).

Waterwise (2009). RENEW Project, Action a2a, Development of concept and offering., s.l.: s.n.

Waterwise (2019). Water Saving Facts.
Available at: https://www.waterwise.org.uk/save-water/
[Accessed 27 September 2019].

Williamson, P., Mitchell, G., McDonald, A., and Wattage, P. (2002). Domestic water demand forecasting: a static microsimulation approach. *J Chart Inst Water Environ Manag*, *16*, pp. 243–248.

Willis, R. M., Stewart, R. A., Panuwatwanich, K., Capati, B., and Giurco, D. (2013). End-use water consumption in households: impact of socio-demographic factors and efficient devices. *J. Clean. Prod., 60*, pp. 107–115.

Wong, J., Zhang, Q., and Chen, Y. D. (2010). Statistical modeling of daily urban water consumption in Hong Kong: Trend, changing patterns, and forecast. *Water Resources Research*, *46*(3). https://doi.org/10.1029/2009WR00814.

Xenochristou, M., Hutton, C., and Kapelan, Z. (2019a). Using smart demand-metering data and customer characteristics to investigate the influence of weather on water consumption in the UK. *Journal of Water Resources Planning and Management*, doi: 10.1061/(ASCE)WR.1943-5452.0001148.

Xenochristou, M., Hutton, C., Hofman, J., and Kapelan, Z. (2019b). A new approach to forecasting household water consumption (under review).

Zhang, G. and Lu, Y. (2012). Bias-corrected random forests in regression. *Journal of Applied Statistics*, *39*(1), pp. 151–160. https://doi.org/10.1080/02664763.2011.578621.

Zhang, P. (1993). Model Selection Via Multifold Cross Validation. *The Annals of Statistics, 21*(1), pp. 299-313.

Zhao, Q. and Hastie, T. (2016). Casual interpretations of black-box models. Retrieved from: https://web.stanford.edu/~hastie/Papers/pdp_zhao_final.pdf [Accessed 27 September 2019].

Zubaidi, S. L., Gharghan, S. K., Dooley, J., Alkhaddar, R. M., and Abdellatif, M. (2018). Short-Term Urban Water Demand Prediction Considering Weather Factors. *Water Resources Management*, *32*(14), pp. 4527–4542. https://doi.org/10.1007/s11269-018-2061-y.