

## Water Demand Forecasting Accuracy and Influencing Factors at Different Spatial Scales Using a Gradient Boosting Machine

Xenochristou, Maria; Hutton, C.; Hofman, J.; Kapelan, Zoran

**DOI**

[10.1029/2019WR026304](https://doi.org/10.1029/2019WR026304)

**Publication date**

2020

**Document Version**

Final published version

**Published in**

Water Resources Research

**Citation (APA)**

Xenochristou, M., Hutton, C., Hofman, J., & Kapelan, Z. (2020). Water Demand Forecasting Accuracy and Influencing Factors at Different Spatial Scales Using a Gradient Boosting Machine. *Water Resources Research*, 56(8), 1-15. Article e2019WR026304. <https://doi.org/10.1029/2019WR026304>

**Important note**

To cite this publication, please use the final published version (if applicable). Please check the document version above.

**Copyright**

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

**Takedown policy**

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

# Water Resources Research

## RESEARCH ARTICLE

10.1029/2019WR026304

### Key Points:

- The Mean Absolute Percentage Error increases exponentially from 3.2% to 17% for a reduction in group size from 600 to 5 households
- Past consumption data and household characteristics are important predictors of consumption for smaller aggregations of properties
- The weather influence on consumption only becomes visible for larger aggregations of properties

### Supporting Information:

- Supporting Information S1

### Correspondence to:

M. Xenochristou,  
mx220@exeter.ac.uk

### Citation:

Xenochristou, M., Hutton, C., Hofman, J., & Kapelan, Z. (2020). Water demand forecasting accuracy and influencing factors at different spatial scales using a Gradient Boosting Machine. *Water Resources Research*, 56, e2019WR026304. <https://doi.org/10.1029/2019WR026304>

Received 14 SEP 2019

Accepted 4 JUN 2020

Accepted article online 29 JUN 2020

## Water Demand Forecasting Accuracy and Influencing Factors at Different Spatial Scales Using a Gradient Boosting Machine

M. Xenochristou<sup>1</sup> , C. Hutton<sup>2</sup>, J. Hofman<sup>3</sup> , and Z. Kapelan<sup>1,4</sup> 

<sup>1</sup>Centre for Water Systems, University of Exeter, Exeter, UK, <sup>2</sup>Wessex Water, Bath, UK, <sup>3</sup>Water Innovation and Research Centre, University of Bath, Bath, UK, <sup>4</sup>Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, Netherlands

**Abstract** 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 study 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 the relationships between spatial scale, influencing factors, and forecasting accuracy. For this purpose, a Gradient Boosting Machine (GBM) 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 reduces the MAPE up to 20% for the peak 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.

## 1. Introduction

The effectiveness of future efforts, technologies, and conservation strategies is heavily dependent on accurately predicting water demand at the appropriate scale. From emerging technologies (e.g., gray water recycling at the household level) to conservation campaigns (e.g., changing customer's attitudes) or even future investments (e.g. building 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 their success.

As part of the commitment to sustainably manage their water resources, water companies are required to reduce per capita consumption (PCC) and leakage, in order to reduce the impact they have on the environment (Ofwat, 2017). According to the Office for National Statistics, PCC in the United Kingdom is the fifth highest in the EU (Bailey, 2019), amounting to a total of 114 L per capita per day. Gaining a better understanding of the factors that influence water use at different spatial scales can assist with developing improved water demand management strategies and curbing demand. 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 U.K. water industry's regulators, has challenged water companies to reduce this figure by 15% by 2025 (Ofwat, 2019).

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 from the tight relationship between the human and natural systems in urban environments, where more than half of the population currently resides (House-Peters & Chang, 2011), as

well as the many time- and space-dependent factors that can influence water consumption (Parker & Wilby, 2013). 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-scale water use is likely to be associated with increased noise in the data, leading to a higher uncertainty and thus increased errors.

This study explores in detail and quantifies the relationship between spatial scale and demand forecasting accuracy while identifying the respective importance of different input variables. For this purpose, a Gradient Boosting Machine (GBM) is built that uses a variety of input factors as explanatory variables, to predict consumption 1–7 days into the future for different household group sizes. Overall, it aims to answer two main questions:

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

The current paper is organized 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 data that was used in the study. The methodology section provides an overview of the model-building process, in terms of the different spatial scales, input variable selection, and description of the modeling technique. The next part presents the results of the study, in terms of the accuracy and influencing factors. Finally, the paper concludes with a discussion of the messages derived from this paper and a brief summary of key results and conclusions.

## 2. Background

Several studies attempted to predict water consumption, using a great variety of data, models, methods, as well as explanatory variables (Adamowski et al., 2012; Anele et al., 2017; Brentan et al., 2017; Herrera et al., 2010; Hutton & Kapelan, 2015; Matos et al., 2014; Romano & Kapelan, 2014; Prescott & Ulanicki, 2008; Tiwari & Adamowski, 2013; Xenochristou et al., 2018; Zubaidi et al., 2018). Some studies in the literature even accounted for the spatial variability of water demand (Balling et al., 2008; Chen & Boccelli, 2018; House-Peters et al., 2010; House-Peters & Chang, 2011; Lee et al., 2010; Maheepala et al., 2011; Polebitski & Palmer, 2010; Rathnayaka et al., 2017a). 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 estimates in the forecasting methodology improved the forecasting accuracy, Lee et al. (2010) argued that additional input factors (other than population density) could improve the forecasting accuracy. 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 as predictors, it did not deal with consumption at each scale as a separate problem. Instead, the total consumption was constructed by merely adding the individual end-uses of the households in each aggregation of properties. A study by Balling et al. (2008) investigated water consumption among census tracts and the effect that several weather variables have on it. Using a variety of explanatory variables, it concluded that census tracts' sensitivity to drought depends heavily on their socioeconomic 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, although it was for certain census blocks containing large, new, affluent, and well-educated households.

As it becomes apparent, although few 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 or in many cases the level of spatial aggregation of water consumption data not matching the scale of the explanatory variables. In order to overcome this problem, researchers often have to rely on interpolating or extrapolating data (House-Peters & Chang, 2011; Lee et al., 2010), that is, estimating values for locations within the study area or outside the study area, respectively, which can be a challenging process (Lee et al., 2010). Even when data are available at the household level, it often lacks spatial coordinates (House-Peters & Chang, 2011), sometimes due to privacy concerns. Another main problem derived from the current literature is the lack of a systematic comparison of

accuracy 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 summarize, although a substantial increase in data availability, computational power, and new technologies over the recent years has contributed in developing spatially explicit demand forecasting models, as well as identifying and quantifying relationships among a variety of weather, social, and water consumption data (House-Peters & Chang, 2011; Rathnayaka et al., 2017b; Xenochristou et al., 2018), there is still the need to develop methodologies that incorporate this information at multiple spatial scales (House-Peters & Chang, 2011).

This study aims to address this gap by making use of a very rich data set comprising of a variety of household characteristics, weather data, temporal characteristics, and past consumption. The aim is to use these data 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.

### 3. Data

#### 3.1. Data Description

The consumption data comes from a region in the southwest of England and relates to 1,793 properties. These were monitored by the water company using smart meters at 15–30 min intervals, over a period of almost 3 years (October 2014 to September 2017). The raw data set was carefully cleaned in order to exclude incorrect and missing data, empty properties, and leakage. This process removed readings of more than 450 L/hr, as well as readings that remain unchanged for more than 24 hr. In addition, in order to exclude small, constant leakages, the days and households with less than 10% and the months with less than 20% of readings that are equal to 0 were also removed from the data set. A detailed description of the cleaning process can be found in Xenochristou, Kapelan, and Hutton (2020).

The water company also collected data related to the households' characteristics and partial postcodes. Information regarding the garden size, occupancy rate, metering status, rateable value of the property, residents' socioeconomic status (ACORN), and council tax band became available at the household level. The occupancy rate of the household refers to the number of people living in the property, whereas the metering status reflects if the property is billed based on their meter reading or not. In the United Kingdom, approximately half of the properties are unmetered (Xenochristou, Kapelan, & Hutton, 2020) and their water bill is calculated based on an estimation, partly dependent on the property's rateable value. The higher the rateable value of the property, the higher the water bill (for unmetered properties). ACORN is a geodemographic segmentation of the U.K.'s population in customer types, based on social factors and population behavior (CACI Limited, 2014). According to the ACORN guide, customers are divided into Groups A to Q, with Groups A to E classified as affluent, F to J as comfortable, and K to Q as financially stretched. The council tax band reflects the council tax rate the property belongs to, based on its location. Council tax bands vary from A to H, from the lowest (A) to the highest (H) paying band. The garden size is the size in m<sup>2</sup> of the property's garden. Finally, postcodes in the United Kingdom are composed 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 and used 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 data set into different categories, depending on the individual attributes of each household in the data set. 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 presented in Table 1. Out of all six characteristics, two of them (garden size and metering status) were organized into categories by the water company, whereas the rest of them (rateable value, acorn group, occupancy rate, and council tax band) were divided by the authors. The aim in forming these categories was to create groups that were large enough to be representative, while at the same time being distinct enough from the rest of the groups to offer a certain explanatory value. A *z* statistic was used here to assess the similarity between the groups. For example, the similarity between the distributions of daily consumption values over the 3 years

**Table 1**  
*Household Characteristics and Their Corresponding Categories*

Garden size	Rateable value	Metering status
Large (>165 m <sup>2</sup> )	High (top 30%)	Metered (billed on meter reading)
Medium (61–165 m <sup>2</sup> )	Medium (mid 40%)	Unmetered (billed on an estimation)
Small (<60 m <sup>2</sup> )	Low (bottom 30%)	
ACORN group	Occupancy rate	Council tax band
Affluent (ACORN Groups A–E)	High (3+ occupants)	High (Tax Groups A–C)
Comfortable (ACORN Groups F–J)	Medium (2–3 occupants)	Medium (Tax Groups D–E)
Financially stretched (ACORN Groups K–Q)	Low (1 occupant)	Low (Tax Groups F–H)

in the data between council tax Bands A, B, and C was assessed using a  $z$  statistic and was deemed similar enough to group them together into Categories A–C.

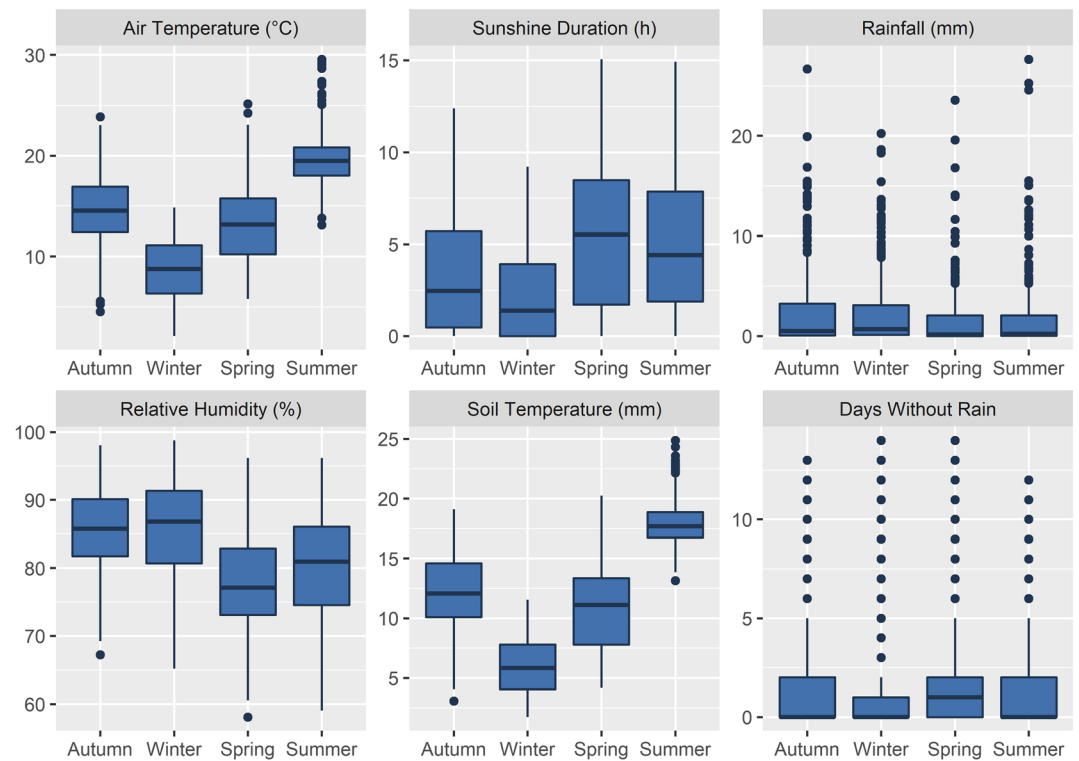
Furthermore, weather data on air temperature, soil temperature at 10 cm depth, humidity, sunshine duration, and rainfall became available by the U.K.'s Meteorological Office (Met Office). These data were recorded at the hourly or daily scale over the same period (October 2014 to September 2017), from 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, 2006b, 2006c, 2006d, 2006e). When recorded hourly, the values were transformed to either mean or total daily values. One additional weather variable was created based on the rainfall data, indicating the number of consecutive days without rain. Since weather data was gathered from hundreds of weather stations across the southwest, one value for each weather variable was calculated as a weighted sum of the recorded values among all weather stations. Each property was assigned to the weather station in the closest proximity and the weight of each weather station was based on the number of properties assigned to it. The more properties a weather station was the closest to (more than any other station), the higher the weight of its recordings (Xenochristou, Kapelan, & Hutton, 2020).

Figure 1 gives a brief overview of the distribution of the six weather variables over the period of the study. Weather in England is characterized by mild temperatures and consistent rainfall all year round. Generally, maximum air temperatures vary between 5°C and 25°C, with very few exceptions, mostly over the winter and summer months (Figure 1). Springs and summers are generally characterized by higher temperatures, increased sunshine hours and lower humidity, although seasonality is not as prominent as in continental climates. Finally, the total amount of rainfall seems to be reduced over the spring and summer months. The presence of rainfall, however, which is often found to be the determining factor in water demand forecasting studies, is consistent over all seasons, although it appears to be lower over the winter months.

Previous analysis explored the interactions and correlations between all available explanatory variables (Xenochristou, 2019). Although results showed that most household variables are weakly to moderately correlated, these interactions were not strong enough to justify excluding any one of them from the model (Xenochristou, 2019). Out of the six household variables examined here, the council tax band is the most interrelated one, as it correlates with the acorn group as well as the property's rateable value and garden size. In terms of the weather variables, a strong correlation was identified between air and soil temperature, as well as between rainfall and days without rain ( $R^2 > 0.90$ ). Since the rainfall amount has a limited impact on water consumption in the United Kingdom and soil temperature has a lesser effect than air temperature, these two variables were excluded from further analysis (Xenochristou, Kapelan, & Hutton, 2020). Only four weather variables, the maximum air temperature, total sunshine duration, relative humidity, and number of days without rain are included in the following.

#### 4. Methodology

This section describes the main steps of the model development process. These include the selection of the spatial aggregation levels and candidate input variables, as well as the description of the modeling technique and model technical implementation and assessment.



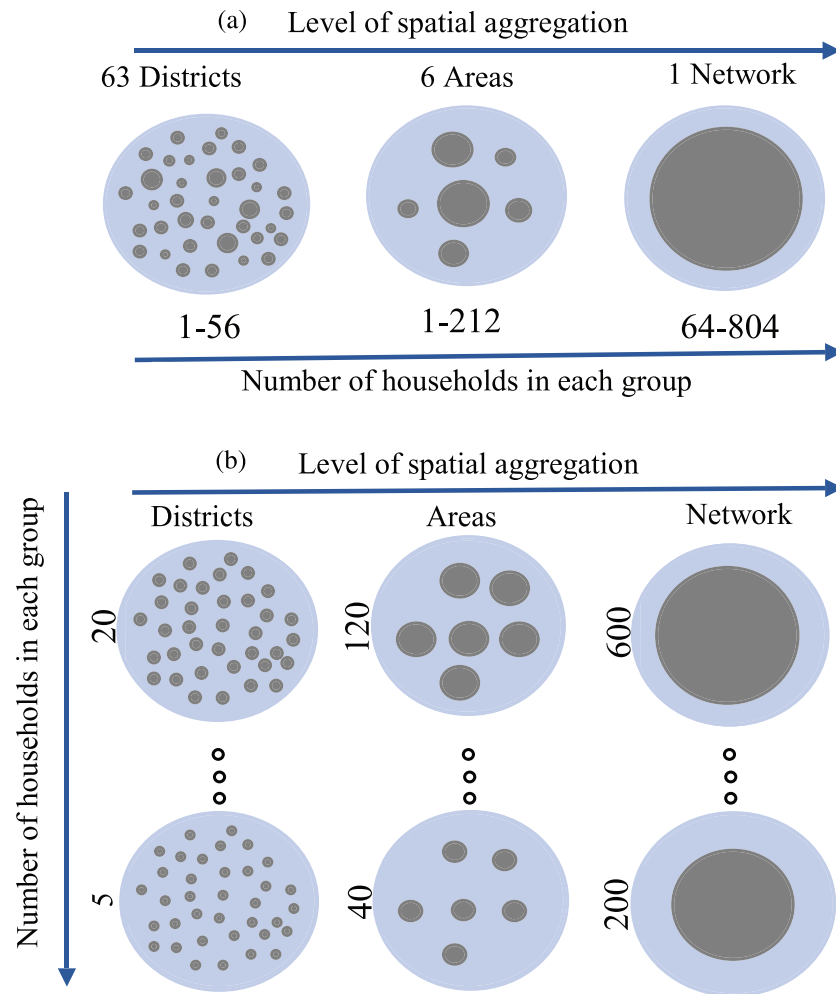
**Figure 1.** Variation of weather variables within each season over the period of the study.

#### 4.1. Spatial Aggregation

Initially, the households are grouped spatially based on their postcodes. This way, it is easy to ensure that properties that are grouped together are in close geographical proximity and each property is counted exactly once. As a result, the following three levels of spatial aggregation are created:

1. **Network grouping:** No grouping criteria are used. Consumption is aggregated among all properties for each day in the data (Network; Figure 2a). Due to errors and inconsistencies, consumption is not available for every property over each day. Therefore this group can vary in composition among different days, that is, include a slightly different collection of properties. The network group consists of 1,056 data points (each data point represents 1 day), with 64–804 properties in each one, depending on data availability for the corresponding day.
2. **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 2a), with 1–212 properties in each one (depending on data availability for the corresponding postcode and day). Each data point represents the consumption of an area for 1 day.
3. **District-based grouping:** The first and second part of the postcode (e.g., BA1) is used to group the properties into districts. This group consists of 76,032 data points (Districts; Figure 2a), 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 1 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 data set, 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 the supporting information.



**Figure 2.** (a) Range of household group sizes for each level of spatial aggregation among days and groups. Each grey dot represents one household group. The number of the dots represents the number of groups in each spatial scale. The size of the dots reflects the number of houses in each group. (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.

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 2b). Each aggregation level contains a number of household groups for each day (this might slightly vary due to missing data), which is 63 for the district level, 6 for the area level, and 1 for the network level. When the household group size is set to a fixed number, the groups that are smaller than the threshold are excluded from the data set, whereas the groups that are larger are reduced to the fixed number of properties. When this threshold increases, the number of data points decreases, as groups with less than the required number of households are removed from the data. The result is nine different spatial scales, comprising of different household group sizes (Figure 2b). The group sizes are set to 5, 10, and 20 for the district groups, 40, 80, and 120 for the area groups, and 200, 400, and 600 for the whole network. The dots in Figure 2b illustrate the number and size of household groups that correspond to each spatial scale, for each day in the data.

#### 4.2. Model Inputs

As it was mentioned in the data section, a variety of input variables became available, including past consumption and weather data as well as postcodes and household characteristics. Based on their nature, the variables were divided into four distinct types:

**Table 2**  
*Model Configurations Tested at Each Level of Spatial Aggregation*

Variable group	Model input variables	Model							
		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	

1. Past consumption data: Past 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. This means that for every day in the data, the mean daily consumption for each one of the 7 days prior to it was used to make predictions.
2. 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 category 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 explanatory values (0.30, 0.60, and 0.10), one for each category. The same applies to the rest of the household variables.
3. 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.
4. 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, and acorn), whereas others need to be collected through questionnaires and inspections (Xenochristou, Kapelan, & Hutton, 2020).

Eight models with different configurations of the above input variables are tested at each level of spatial aggregation (Table 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 day and group of properties corresponds to one training data point. Thus, the smaller the aggregation level, the higher the number of training data points.

### 4.3. GBMs

Previous work (Xenochristou & Kapelan, 2020) focused on comparing a selection of machine learning models for water demand forecasting and identifying the one that achieves the best prediction accuracy. In that



case, the models were compared at the postcode area level. This spatial scale was chosen to (a) avoid forming very small groups of properties, as this would have interfered with the accuracy of the results, and (b) allow for sufficient data points to train and test the model. Results showed that the GBM method combines high prediction accuracy with ease of implementation hence was chosen for this work.

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 data set at each branch in a way that maximizes entropy, that is, the homogeneity within each of the split groups. At each branch (node) of the tree, a variable as well as a threshold value are chosen for splitting the data set. The tree will keep dividing until it reaches a limit, typically defined by the user, such as a maximum tree depth or minimum final node size.

The GBM algorithm uses bagging, as well as boosting in order to achieve the best result. Each tree is trained on a subset of the original data, while at each node of the tree, the best variable for splitting is chosen among a random sample of the input variables (bagging). At each step, one regression tree is built on the residual errors of the previous tree with the aim to improve the final result. 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 optimization problem, where the objective is to minimize the error in the objective function (gradient descent).

#### 4.4. Model Implementation and Assessment

In order to build the model, the data set is randomly shuffled and divided into a training (70% of the data) and a test (30% of the data) set. The training set is used to train and tune the model for the optimum set of hyperparameters, whereas the test data set 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. Model training is the process of fitting the model on the training data, whereas the tuning step refers to the selection of a set of hyperparameters that are chosen before the training begins. These are important as they define how closely or loosely the model fits the training data. In order to enhance the robustness of the hyperparameter selection process, the performance of the hyperparameter values is tested on multiple subsets of the training data using a fivefold cross validation process (Zhang, 1993). This means that the training set is divided into five parts and at every iteration, four parts are used for training while one is used to assess the model performance.

The GBM is trained and tuned for the optimum set of hyperparameters using the “h2o” package (LeDell et al., 2019) written for R (R Core Team, 2013), which serves as an interface for the “h2o” machine learning platform (Aiello et al., 2018). Predictions are made for different model configurations, groups of properties, and forecast horizons. The model is retrained and retuned for every change in the input variables, forecast horizon, or spatial aggregation. The automated machine learning capability of “h2o,” called “automl” (H<sub>2</sub>O.ai, 2019), is used to identify the optimum set of hyperparameters in each case, using a random search (Bergstra & 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 counterproductive, especially since the aim is to train, tune, and compare a large number of models.

Nine hyperparameters are tuned in this study for the GBM algorithm: the total number of trees that construct the final model (ntrees); the size of the subsample of the training data set 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 and col\_sample\_rate\_per\_tree, respectively); 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); and the minimum requirements for splitting at each node (min\_split\_improvement and min\_rows). More information regarding the model hyperparameter can be found in the ‘h2o’ documentation (H<sub>2</sub>O.ai, 2019).

After the model is properly trained and tuned, it is used on the test data set to make predictions for daily consumption 1–7 days into the future. The model performance is assessed based on three criteria, the Mean

**Table 3**  
Prediction Accuracy Achieved for Nine Models Trained at Different Spatial Scales

Aggregation level	Data points	Group size	MAPE (%)		MSE (L <sup>2</sup> /day <sup>2</sup> )		R <sup>2</sup> (%)	
			Train	Test	Train	Test	Train	Test
District	43,875	5	16.2	17.0	1,047	1,133	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

Absolute Percentage Error (MAPE), mean square error (MSE), and  $R^2$  correlation coefficient, as each one of these provided slightly different information. 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 per household consumption). The MSE is sensitive to outliers, while the  $R^2$  shows the variance in the dependent variable that can be explained by changes in the independent variable (Xenochristou, 2019).

## 5. Results

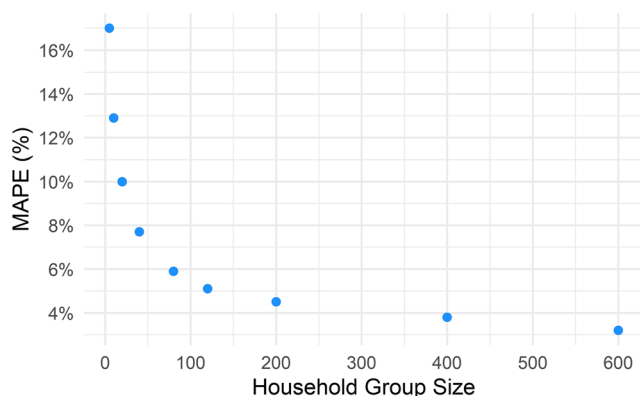
### 5.1. Demand Forecasting Accuracy at Different Spatial Scales

Increasing the level of spatial aggregation consequently 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 and tuned for the optimum set of hyperparameters and consequently assessed for their ability to predict demand for different household group sizes, 1 day into the future. For comparison purposes, each model is trained with the same input, 7 days of past consumption. Table 3 shows the aggregation level, group size, and number of data points that were used to train each model as well as the results acquired from each one based on three assessment criteria, the MAPE, MSE, and  $R^2$ , for the training and test data set. The results of the hyperparameter tuning process are summarized in the supporting information.

According to Table 3, the prediction error (MAPE and MSE) decreases (i.e., improves) 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 data set (Group size = 600; Table 3). The largest MAPE on the other hand (MAPE = 17%) relates to the smallest aggregation scale, at the district level, with a group size of five households (Group size = 5; Table 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 the best balance between prediction accuracy and household group size, that is, at which spatial scale a further increase in group size does not offer a significant reduction in prediction errors. This is depicted in Figure 3, which represents the balance between the MAPE and spatial scale, for the test data set. According to Figure 3, the model error increases exponentially as the household group size decreases. When everything else remains the same (model structure and input variables), increasing the prediction group size from 40 to 120 households reduces the MAPE by 2.6% (Figure 3). 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



**Figure 3.** Model accuracy achieved for each household group size for the test data set.

**Table 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 per capita per day)
Network	100	657	992	117–175
Areas	60	114	5,592	100–195
Districts	20	29	8,537	80–250

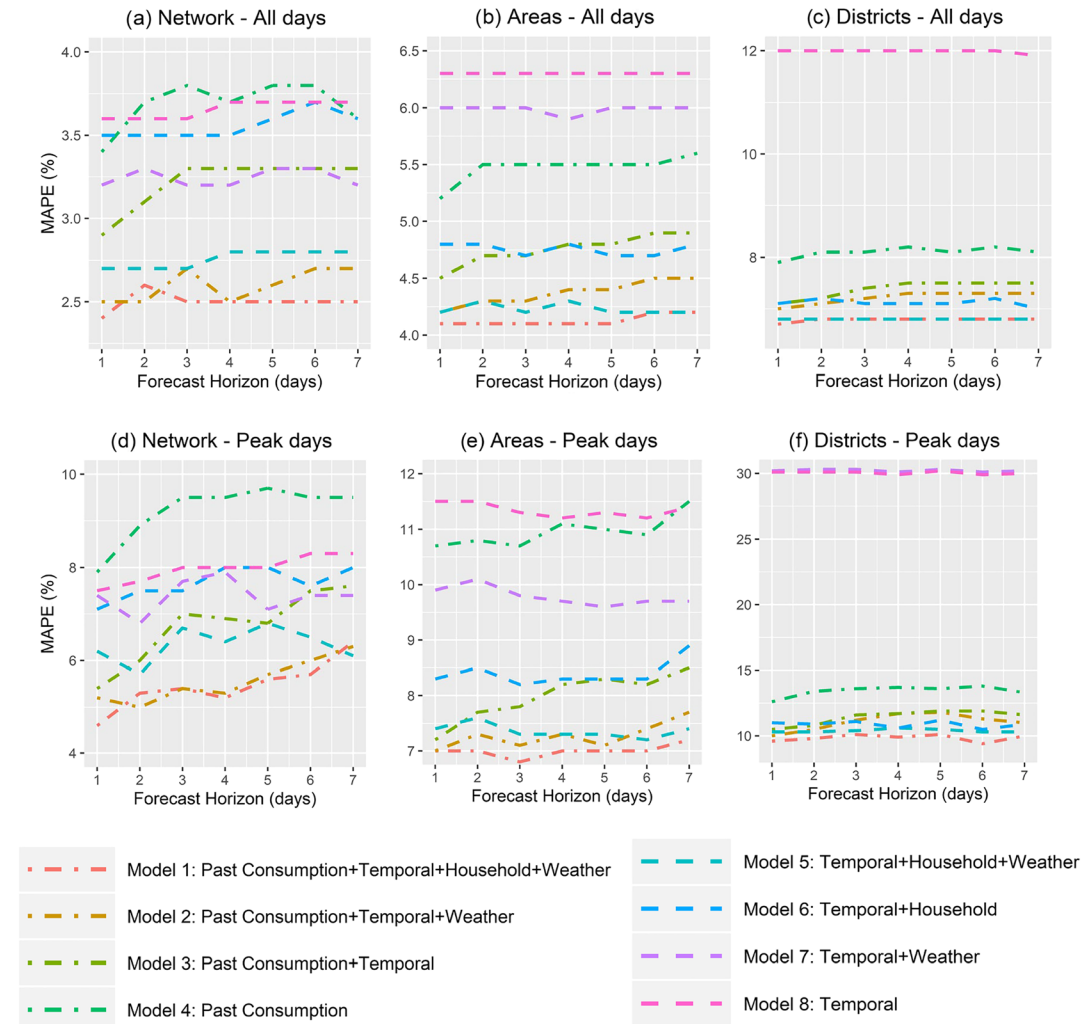
households, the MAPE decreases marginally for a high increase in group size (Figure 3).

**5.2. Variable Importance at Different Spatial Scales**

The three aggregation levels contain different household group sizes, with different ranges in their daily consumption and different amounts of data points (Table 4). In order to avoid increased prediction errors associated with very small groups (<20 households), while allowing to create distinct enough group sizes to allow for a meaningful comparison, 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 and variable for smaller household groups, the range in daily consumption also increases (Table 4).

Results are summarized in Figure 4 and Table 5. Figure 4 shows the prediction accuracy, in terms of MAPE, for predictions 1–7 days ahead, over all days in the data (Figures 4a–4c), as well as peak days, that is, 10% of the days with the highest consumption (Figures 4d–4f). Each plot represents one aggregation level (network, area, and district) and eight model configurations, with each configuration corresponding to a different set of input variables (Table 2). Table 5 shows the MAPE for each model and each aggregation level, for 1 as well as

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 and variable for smaller household groups, the range in daily consumption also increases (Table 4).



**Figure 4.** Mean Absolute Percentage Error (MAPE) for different model configurations and different spatial aggregations.

**Table 5**  
*MAPE Achieved for Eight Different Model Configurations for Predictions 1 and 7 days into the future, for Three Spatial Aggregations of Properties*

Model	NETWORK – MAPE (%)				AREAS – MAPE (%)				DISTRICTS – MAPE (%)			
	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

7 days into the future, for all days and peak days. The hyperparameter values selected for each model are available in the supporting information.

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).

Although the MAPE value and variance increase for peak days, results are overall very similar. The best performing model (MAPE = 4.6%), for 1 day lead time, is the one that uses all predictors (Model 1). However, for predictions 7 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). Temporal characteristics, on top of past consumption, improve the MAPE by 2.5% (Model 3), for 1 day lead time. Weather input further reduces errors by 0.2% (Model 2) and household characteristics by 0.6% (Model 1). For Models 5–8 however (the ones excluding past consumption data), household and weather input reduce errors by 0.4% (Model 6) and 0.1% (Model 7), for predictions 1 day ahead. The combined effect of both of the above reduces 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).

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 as well as peak days (Figures 4b and 4e). 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 errors by 0.3% (Model 7), for 1 day lead time, whereas household characteristics reduce it by 1.5% (Model 6), for all days (Table 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). 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). 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.

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 4c). 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 7 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). 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 7 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).

It is worth noting the upward trend of all models that include past consumption as an explanatory factor (Models 1–4), as predictions move further into the future. Since water consumption is highly autocorrelated from 1 day to the next one, predictions for 1 day ahead are more accurate than 7 days ahead. However, adding weather and household input does reduce 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 output (Figure 4). The result of this is that the best model sometimes shifts depending on the forecast horizon, as models that include past consumption often perform best for 1 day lead time but are outperformed by the ones that have temporal, household, and weather input for increased lead times (e.g., 7 days).

## 6. Discussion

This paper shows that if everything else stays the same, water demand prediction errors improve for larger aggregations of households, reaching constant prediction accuracy for groups larger than ~200 houses. This is likely due to the fact that as the household group size decreases, water demand becomes more variable as well as more random/erratic and therefore more difficult to predict. This is illustrated by the level of water demand variability, which is clearly associated with the level of spatial aggregation; smaller groups have a much higher daily water consumption range (80–250 L per capita per day for the district groups) compared to larger ones (115–175 L per capita per day for the network grouping). As errors reduce for larger group sizes, the  $R^2$  value increases, but only within the same aggregation level. While the variance in the response variable (i.e., the water consumption) decreases as the group size increases, moving to a higher aggregation level (e.g., from districts to areas) also has a negative effect; grouping together houses that are further away from each other potentially creates less homogenous groups and thus reduces the explanatory value of the predictor variables, in this case past consumption.

This demand variability in smaller household groups can be largely explained by different behaviors and habits, and thus, results can be improved by adding the right explanatory factors as model inputs. Smaller groups show a much higher range of MAPE, as they are associated not only with higher variations in water demand but also higher variations in corresponding household characteristics. This means that the variance in consumption among the smaller groups is largely explainable and prediction results can be significantly improved by identifying the right explanatory variables to include in the model (Figure 4, Districts). However, when groups are larger, hence more homogenous, the potential for error reduction are significantly smaller (Figure 4, Network). This is the reason why explanatory variables that improve predictions also vary between different spatial scales, with household characteristics becoming more important as groups become smaller (Figure 4, Districts). The higher the variation of household composition between groups, the higher the importance of household characteristics as explanatory factors of consumption (Figure 4). On the other hand, when predicting water consumption for groups of households with similar characteristics (e.g., when aggregating all households in the network for each day in the data), these characteristics cannot explain the variance in consumption (Figure 4, Network). In order for a variable to be useful input to a forecasting model, it needs to influence water consumption, as well as have a wide enough range of values among the households or groups in the data set (Figure 4, Districts).

Past consumption data also became more important as the household group size reduced (Figure 4). Household characteristics are embedded in past consumption, in addition to other factors that can define the consumption behavior of a certain property or group of properties. Therefore, using past consumption

data can be particularly valuable for smaller groups, since it can capture the individual behavior that relates to the variability in their individual characteristics. This is demonstrated by examining the influence of the explanatory variables for the district areas (Figure 4). When past consumption data is available, household characteristics do not further improve the prediction accuracy of the model. However, when past consumption is not used as model input, a combination of household, weather, and temporal characteristics can adequately be used to characterize and thus predict water demand with the same accuracy. For example, adding weather and household variables on top of past consumption reduced the MAPE a maximum of 1.6% for peak days and district areas whereas for the model that did not include past consumption, adding household and weather characteristics achieved a reduction of 19.7%, from 30% to 10.3%.

The effect of weather became noticeable only for larger groups of properties (Network and Areas, Figure 4), while it is rather irrelevant when attempting to predict consumption for smaller household groups (Districts, Figure 4). Previous studies found that the effect of weather on water consumption varies between households, days and times in the year (Xenochristou, Kapelan, & Hutton, 2020). Out of all households in the data set, only few of them will alter their consumption behavior based on the weather and therefore using weather input cannot improve predictions at small levels of spatial aggregation. In these cases, the model would “learn” based on the majority of the data points, for which weather does not actually have an influence on consumption. However, when aggregating all properties for each day in the data, the effect of weather can be seen in each data point (each day) used to train the model; therefore, in this case weather is found to have a (slight) impact on consumption. Notably, the combined contribution of household and weather characteristics in the model was in most cases much higher than their individual contributions. This result confirms further what was already concluded from previous studies (Xenochristou, Kapelan, & Hutton, 2020) that the influence of weather on water consumption is dynamic and it strongly depends on the type of property and residents. Therefore, providing additional context in terms of household characteristics on top of weather information can further improve results.

Finally, adding another dimension (or variable) to the problem, such as the temporal scale or model type could provide further insights. Here, a GBM model and daily scale are used to compare the forecasting accuracy and variables of interest at different spatial scales. The daily scale allowed to incorporate additional input variables in the model, such as the day of the week, and thus account for the weekly pattern of water consumption. The GBM model was chosen for its accuracy and ease of implementation, based on previous work that compared the forecasting accuracy of several machine learning models under different scenarios (Xenochristou & Kapelan, 2020). Ideally, different models would have been tested at different spatial scales, in order to determine the best one for each application. In addition, testing demand predictions over a grid of spatial and temporal aggregations of consumption would demonstrate the limitations and opportunities that arise at each scale. However, including each dimension of the water demand forecasting problem (such as the temporal scale and model type) as an unknown variable would increase significantly the dimensionality of the problem. As a result, it would also increase disproportionately the computational and time requirements of the analysis, and equally the processing and understanding of the results. In this case, the model type and temporal scale were chosen by the authors and were not considered in the analysis as another dimension of the problem, as this was out of the scope of the current work. Future work should focus on creating a modeling framework that integrates more unknown aspects of the problem simultaneously, in order to determine how they influence results.

## 7. Summary and Conclusions

This study explored the effect of the spatial scale on water demand forecasting, both in terms of prediction accuracy and influencing factors. In order to achieve this, multiple models with different input variables were trained on real-life U.K. daily consumption records for different aggregations of consumption. Initially, three different levels of spatial aggregation were created using the properties' postcode. One group included all the households in the network (up to 804 properties per group) while the other two aggregated the properties in the data set in 6 areas (up to 262 households per group), or 63 districts (up to 56 households per group). At the same time, three household group sizes were fixed and tested for each aggregation level, varying from 5 (for the districts) to 600 (for the network) properties per group per day. A GBM was trained using each of the above configurations and a prediction was made for the water consumption of the same

groups, for 1 day into the future, using only past consumption as an explanatory factor. The purpose of this was to compare the modeling accuracy among models for different spatial scales. After this, different types of model input variables (temporal characteristics, weather data, household characteristics, and past consumption) were used in order to improve the prediction accuracy at each level of spatial aggregation (Network, Areas, and Districts) and identify the most influential input factors.

The results obtained show the following:

1. The level of spatial aggregation has a direct influence on the demand forecasting accuracy. In general, the higher the spatial scale of household aggregation, the more accurate are demand forecasts. For groups of fewer than 20 households, the prediction error measured via MAPE increases exponentially with a decrease in household group size. On the other hand, for group sizes above approximately 200 households, a further increase in group size only marginally reduces the MAPE.
2. Demand forecasting errors can be reduced by using additional explanatory variables, especially in the case of smaller groups, where the error range varied significantly depending on the input factors used. In this study, the most influential input variables that improved the demand forecasting accuracy varied for different levels of spatial aggregation. Past consumption became more important for smaller aggregations of properties, along with household characteristics, while weather data contributed to the model's accuracy only for larger household groups.

Although the effect of different levels of spatial aggregation was investigated in detail in this paper, this was done within a fixed set of environmental conditions. All of the above analysis reflected the consumption of houses in the southwest of England. In a different setting, with different prominent household and resident characteristics, as well as climate, these results could be very different. Although the above methodology could be replicated anywhere where the related data is available, it is important to note that the results could possibly vary.

### Data Availability Statement

The data for this study was made available by Wessex Water and is protected under a nondisclosure agreement. Interested parties can ask for data access directly from Wessex Water. The weather data used in this study was collected and became available by the Met Office. This data was provided to the author for research purposes only and is available for purchase by the Met Office.

### Acknowledgments

This study was funded as part of the Water Informatics Science and Engineering Centre for Doctoral Training (WISE CDT) under a grant from the Engineering and Physical Sciences Research Council (EPSRC), Grant EP/L016214/1.

### References

- Adamowski, J., Fung Chan, H., Prasher, S., Ozga-Zielinski, B., & Sliusarieva, A. (2012). Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. *Water Resources Research*, *48*, W01528. <https://doi.org/10.1029/2010WR009945>
- Aiello, S., Eckstrand, E., Fu, A., Landry, M., & Aboyoun, P. (2018). Machine Learning with R and H2O. <http://h2o.ai/resources/>
- Anele, A. O., Hamam, Y., Abu-Mahfouz, A. M., & Todini, E. (2017). Overview, comparative assessment and recommendations of forecasting models for short-term water demand prediction. *Water (Switzerland)*, *9*(11), 887. <https://doi.org/10.3390/w9110887>
- Bailey, S. (2019) "Consumption per head in the UK: 2017", (December), pp. 1–9.
- Balling, R. C., Gober, P., & Jones, N. (2008). Sensitivity of residential water consumption to variations in climate: An intraurban analysis of Phoenix, Arizona. *Water Resources Research*, *44*, W10401. <https://doi.org/10.1029/2007WR006722>
- Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, *13*, 281–305
- Brentan, B. M., Meirelles, G., Herrera, M., Luvizotto, E. Jr., & Izquierdo, J. (2017). Correlation analysis of water demand and predictive variables for short-term forecasting models. *Mathematical Problems in Engineering*, *2017*, 1–10. <https://doi.org/10.1155/2017/6343625>
- CACI Limited (2014). The Acorn user guide. The Consumer Classification, p. 108. Available at: <https://acorn.caci.co.uk/downloads/Acorn-User-guide.pdf>
- Chen, J., & Boccelli, D. L. (2018). Forecasting hourly water demands with seasonal autoregressive models for real-time application. *Water Resources Research*, *54*, 879–894. <https://doi.org/10.1002/2017WR022007>
- H2O.ai (2019). AutoML: Automatic machine learning. Available from: <http://docs.h2o.ai/h2o/latest-stable/h2o-docs/automl.html> [Accessed 06/09/2019]
- Herrera, M., Garcia-Diaz, J. C., Izquierdo, J., & Perez-Garcia, R. (2010). Predictive models for forecasting hourly urban water demand. *Journal of Hydrology*, *387*(1–2), 141–150. <https://doi.org/10.1016/j.jhydrol.2010.04.005>
- House-Peters, L., Pratt, B., & 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), 461–472. <https://doi.org/10.1111/j.1752-1688.2009.00415.x>
- House-Peters, L. A., & Chang, H. (2011). Urban water demand modeling: Review of concepts, methods, and organizing principles. *Water Resources Research*, *47*, W05401. <https://doi.org/10.1029/2010WR009624>
- Hutton, C. J., & Kapelan, Z. (2015). A probabilistic methodology for quantifying, diagnosing and reducing model structural and predictive errors in short term water demand forecasting. In *Environmental Modelling and Software* (Vol. 66, pp. 87–97). New York: Elsevier Ltd. <https://doi.org/10.1016/j.envsoft.2014.12.021>

- DeDell, E., Gill, N., Aiello, S., Fu, A., Candel, A., Click, C., et al. (2019). Package “h2o”, R package version 3.26.0.2
- Lee, S. J., Wentz, E. A., & 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), 283–295. <https://doi.org/10.1007/s00477-009-0317-z>
- Maheepala, S., Blackmore, J., Diaper, C., Moglia, M., Sharma, A., & Kenway, S. (2011). Integrated urban water management planning manual.
- Matos, C., Teixeira, C. A., Bento, R., Varajao, R., & Bentes, I. (2014). An exploratory study on the influence of socio-demographic characteristics on water end uses inside buildings. *Science of the Total Environment*. Elsevier B.V., 466–467, 467–474. <https://doi.org/10.1016/j.scitotenv.2013.07.036>
- Met Office (2006a). MIDAS: UK Soil Temperature Data. NCAS British Atmospheric Data Centre, 26/07/2020. <https://catalogue.ceda.ac.uk/uuid/8dc05f6ecc6065a5d10fc7b8829589ec>
- Met Office (2006b). MIDAS: UK Daily Rainfall Data. NCAS British Atmospheric Data Centre, 26/07/2020. <https://catalogue.ceda.ac.uk/uuid/c732716511d3442f05cdecbe99b8f90>
- Met Office (2006c). MIDAS: UK Daily Temperature Data. NCAS British Atmospheric Data Centre, 26/07/2020. <https://catalogue.ceda.ac.uk/uuid/1bb479d3b1e38c339adb9c82c15579d8>
- Met Office (2006d). MIDAS: UK Daily Weather Observation Data. NCAS British Atmospheric Data Centre, 26/07/2020. <https://catalogue.ceda.ac.uk/uuid/954d743d1c07d1dd034c131935db54e0>
- Met Office (2006e). MIDAS: UK Hourly Weather Observation Data. NCAS British Atmospheric Data Centre, 26/07/2020. <https://catalogue.ceda.ac.uk/uuid/916ac4bbc46f7685ae9a5e10451bae7c>
- Ofwat (2017). PR19. Appendix 2: Delivering outcomes for customers.
- Ofwat (2018). Reporting guidance leakage, pp. 1–44. Available at: <https://www.ofwat.gov.uk/publication/reporting-guidance-leakage/>
- Ofwat (2019). Ofwat’s forward programme 2019–20.
- Parker, J. M., & 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>
- Polebitski, A. S., & Palmer, R. N. (2010). Seasonal residential water demand forecasting for census tracts. *Journal of Water Resources Planning and Management*, 136(1), 27–36. [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0000003](https://doi.org/10.1061/(ASCE)WR.1943-5452.0000003)
- Prescott, S., & Ulanicki, B. (2008). Improved control of pressure reducing valves in water distribution networks. *Journal of Hydraulic Engineering*, 134(1), 56–65. [https://doi.org/10.1061/\(ASCE\)0733-9429\(2008\)134:1\(56\)](https://doi.org/10.1061/(ASCE)0733-9429(2008)134:1(56))
- 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., & 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*. Elsevier B.V., 117, 85–92. <https://doi.org/10.1016/j.resconrec.2016.11.014>
- Rathnayaka, K., Malano, H., Arora, M., George, B., Maheepala, S., & Nawarathna, B. (2017b). Prediction of urban residential end-use water demands by integrating known and unknown water demand drivers at multiple scales II: Model application and validation. *Resources, Conservation and Recycling*. Elsevier B.V., 117, 85–92. <https://doi.org/10.1016/j.resconrec.2016.11.014>
- Romano, M., & Kapelan, Z. (2014). Adaptive water demand forecasting for near real-time management of smart water distribution systems. *Environmental Modelling and Software*, Elsevier Ltd, 60, 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.royal-mail.com/delivery/mail-advice/postcodes-addresses-explained> [Accessed 05/09/2019].
- Tiwari, M. K., & Adamowski, J. (2013). Urban water demand forecasting and uncertainty assessment using ensemble wavelet-bootstrap-neural network models. *Water Resources Research*, 49, 6486–6507. <https://doi.org/10.1002/wrcr.20517>
- Ulanicki, B., AbdelMeguid, H., Bounds, P., & Patel, R. (2009). Pressure control in district metering areas with boundary and internal pressure reducing valves. *Geotechnical Special Publication*, 187, 691–703. [https://doi.org/10.1061/41023\(337\)58](https://doi.org/10.1061/41023(337)58)
- Xenochristou, M. (2019). ‘Water demand forecasting using machine learning on weather and smart metering data.’ PhD thesis. University of Exeter. Retrieved from: <https://ore.exeter.ac.uk/repository/handle/10871/39792> [Accessed 15 December 2019]
- Xenochristou, M., & Kapelan, Z. (2020). An ensemble stacked model with bias correction for water demand forecasting. *Urban Water Journal*, 17(3), 212–223. <https://doi.org/10.1080/1573062X.2020.1758164>
- Xenochristou, M., Kapelan, Z., & Hutton, C. (2020). Using smart demand-metering data and customer characteristics to investigate the influence of weather on water consumption in the UK. *Water Resources Planning and Management*, 146(2). [https://doi.org/10.1061/\(ASCE\)WR.1943-5452.0001148](https://doi.org/10.1061/(ASCE)WR.1943-5452.0001148)
- Xenochristou, M., Kapelan, Z., Hutton, C., & Hofman, J. (2018). *Smart water demand forecasting: Learning from the data*. Paper presented at 13th International Conference on Hydroinformatics: EPIC Series in Engineering 3. EasyChair Publications.
- Zhang, P. (1993). Model selection via multifold cross validation. *The Annals of Statistics*, 21(1), 299–313. URL: <http://www.jstor.org/stable/303559>
- Zubaidi, S. L., Gharghan, S. K., Dooley, J., Alkhaddar, R. M., & Abdellatif, M. (2018). Short-term urban water demand prediction considering weather factors. *Water Resources Management*, 32(14), 4527–4542. <https://doi.org/10.1007/s11269-018-2061-y>