Short-Term Forecasting of Household Water Demand in the UK Using an Interpretable Machine-Learning Approach

Xenochristou, Maria; Hutton, Chris; Hofman, Jan; Kapelan, Zoran

DOI
10.1061/(ASCE)WR.1943-5452.0001325

Publication date
2021

Document Version
Accepted author manuscript

Published in
Journal of Water Resources Planning and Management

Citation (APA)
Xenochristou, M., Hutton, C., Hofman, J., & Kapelan, Z. (2021). Short-Term Forecasting of Household Water Demand in the UK Using an Interpretable Machine-Learning Approach. *Journal of Water Resources Planning and Management*, *147*(4), [1325]. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001325

Short-term forecasting of household water demand in the UK: An interpretable machine learning approach

Maria Xenochristou[1], Chris Hutton[2], Jan Hofman[3] and Zoran Kapelan[4]

[1, 4] Centre for Water Systems, University of Exeter, North Park Road, EX4 4QF Exeter, UK

[2] Wessex Water, Claverton Down Road, BA2 7WW Bath, UK

[3] Water Innovation and Research Centre, University of Bath, BA2 7AY Bath Avon, UK

[4] Delft University of Technology, Stevinweg 1, 2628CN Delft, Netherlands

[1]corresponding author: mx220@exeter.ac.uk

Abstract

This study utilises a rich UK dataset of smart demand metering data, household characteristics, and weather data to develop a demand forecasting methodology that combines the high accuracy of machine learning models with the transparency of regression methods. For this reason, a Random Forest model is used to predict daily demands one day ahead for groups of properties (mean of 3.8 households/group) with homogenous characteristics. A variety of interpretable machine learning techniques (variable permutation, Accumulated Local Effects plots – ALE, Individual Conditional Expectation curves – ICE) are used to quantify the influence of these predictors (temporal, weather, and household characteristics) on water consumption. Results show that when past consumption data are available, they are the most important explanatory factor. However, when they are not, a combination of household and temporal characteristics can be used to produce a credible model with similar forecasting accuracy. Weather input has overall a mild to no effect on the model's output, although this effect can become significant under certain conditions.

Keywords: *water demand forecasting, smart demand metering, Random Forest.*

[1] Ph.D. student, Centre for Water Systems, University of Exeter, North Park Road, EX4 4QF Exeter, U.K.
[2] Ph.D., Water Resources Planning Manager, Wessex Water, Claverton Down Road, BA2 7WW Bath, U.K.
[3] Professor, Water Innovation and Research Centre, University of Bath, BA2 7AY Bath Avon, U.K.
[4] Professor, Delft University of Technology, Stevinweg 1, 2628CN Delft, Netherlands

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) and privacy concerns.

Jorgensen et al. (2009) reviewed several studies that used social variables to model and predict water consumption and concluded that most of them found different variables to be the most important explanatory factors of consumption. In addition, the explanatory potential of these models was limited, with R^2 (coefficient of determination) values reaching a maximum of ~30% (Jorgensen et al., 2009). This inability of the models to accurately represent consumption might be the reason for the high deviations between them.

Williamson et al. (2002) used a number of property characteristics (e.g. number of residents, appliance ownership and property type) to predict monthly individual household consumption using a regression-based function. Using the above household characteristics, this method could distinguish between millions of households and explained 44% of the variance (R^2 = 44%) in water demand. The rest was attributed to factors that were not included in the model, such as the garden size. However, aggregating consumption at the monthly scale means that temporal variables such as the day of the week cannot be used as explanatory factors. This might limit the amount of variance explained by the model, as well as the opportunity to understand how these variables influence consumption. In addition, for certain applications (e.g. operational requirements for water distribution systems), predictions with higher temporal resolution might be required.

Jorgensen et al. (2014) used a latent growth curve to predict consumption for single-person households over four quarters in 2009 and 2010. In this case, the maximum variance explained (R^2) in the rate of change of water consumption was 31%. This was achieved using three predictors, the income, type of irrigation system, and beliefs relating to own consumption. However, accuracy could be improved if more variables were included in the analysis.

Duerr et al. (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 for their ability to forecast monthly individual household consumption, such as machine learning, linear regression and time series models. The one that performed best was the time series model, with a minimum Root Mean Square Error (RMSE) of 1,246 gallons/month (the equivalent of an average of 155 litres/day), for predictions one month ahead. Similarly to previous studies, the level of accuracy is problematic, while consumption is aggregated at the monthly scale.

Overview and Aim

The benefit of explanatory variables depends on the model's capability to capture the complicated relationships between them and water consumption. In most cases, even when explanatory variables (e.g. household and climatic variables) are deployed to produce water demand forecasts, this is often done using linear regression analysis or geodemographic profiling based on census data (Parker and Wilby, 2013). These techniques have traditionally been used because they are simple and able to capture the relationships between the predictors and water demand in a transparent way (Goodchild, 2003; Wong et al., 2010). However, their ability to model the complicated relationships between a set of predictors and water consumption may be limited. At the same time, the non-linear and non-univariate effect of some weather variables on water demand, as well as their interactions with other variables that were observed in previous studies (Parker and Wilby, 2013; Parker, 2014; Xenochristou et al., 2018; Xenochristou et al., 2019) require further attention.

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. The interpretability of machine learning models is a topic with increasing popularity as more methods are developed (Doshi-Velez and Kim, 2017; Adadi and Berrada, 2018; Carvalho et al., 2019; Molnar, 2019a) and find use in different fields, particularly in medical applications (Berk et al., 2016; Choi, 2018; Cremona et al., 2018; Carmichael et al., 2019; Chang et al., 2019). However, machine learning interpretability methods have not been applied and tested in the field of water demand forecasting. As a result, the ability to use these models to provide guidance to water utilities has been limited.

The overall aim of this paper is to present a novel approach towards water demand forecasting that combines the high accuracy of machine learning models with the interpretability of simpler methods. 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. Water demand modelling that reconstructs detailed household, temporal, and weather variables would enable planners to predict small area demands and test new tariffs (Clarke, 1997). In addition, these variables can enhance the understanding of water use behaviours and thus support improved demand management practices (Duerr, 2018). This is particularly important when the distribution of customer demand is highly skewed, particularly on peak demand days, when a small number of customers are responsible for a high percentage of the total water use. Results of this study would allow understanding and targeting particular household types (i.e. the types that use the most water) to reduce peak demands, which can be valuable during drought periods, as well as improve the understanding of the complicated relationships between weather and water consumption.

In order to achieve this, a machine learning model based on Random Forests is implemented to predict daily demands for small household groups with homogenous characteristics, with and without past consumption data. Next, three interpretability techniques (variable permutation, Accumulated Local Effects - ALE plots, Individual Conditional Expectation - ICE curves) are used to assess the influence of a variety of household, temporal, and weather variables, as well as their interactions, on the model's outcome.

Data

The dataset comprises of water demand data and household characteristics from the southwest of England, collected by Wessex Water, one of the UK water companies, as well as weather data provided by the Met Office. A detailed description of each data type is available in this section.

Consumption Data

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. More details about this process can be found in Xenochristou et al. (2019). After the pre-processing of the data, 1,793 properties are included in the dataset. Recordings for each property correspond to a maximum duration of 1,019 days, although this number is reduced for most properties due to gaps in the data.

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 occupancy rate), available at the household level. Information about garden sizes and occupancy rates were collected by questionnaires that customers fill in when they want to switch to a smart water meter. The rest of the household properties were collected by their respective agencies.

In order to limit the processing time and reduce complexity, the properties in the dataset are grouped in two to three segmentation categories for each household characteristic (Figure 1). 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 in the study area 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, 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 (e.g. households in council tax bands A, B, C, etc.). 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).
Figure 1 demonstrates the percentage of properties in each segmentation category among all properties in the study area, for each one of the six household characteristics.

Weather Data

The weather dataset includes Met Office data on air and soil temperature at 10 cm depth, humidity, sunshine duration, and rainfall. These data are 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 preceding consecutive days without rain is also calculated based on the rainfall data.

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, as a weighted average of all 56. 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). For example, if weather station A is the nearest weather station to 100 properties and weather station B is the nearest weather station to 160 properties, weather station B is assigned a higher weight. Weather stations that have no properties in the nearest proximity are assigned a zero weight. The above methodology is adopted in order to account for the location of the weather stations. Instead of calculating a mean value among all stations in the area, the proximity of the stations to the properties in the dataset is taken into account. This is likely to result in more accurate estimates of weather values, especially for the weather variables that demonstrate a higher spatial variability.

Methodology

This section outlines the steps of the methodology adopted here, in terms of the model variables, the household grouping, the modelling technique (Random Forests), the model and variable assessment methodologies, and finally the model's technical implementation.

Model Input variables

The first step towards model building is to define the pool of variables that will be included in the analysis. All available variables are investigated for their influence on the model's results, for forecasts one day into the future, 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. Figure 2 demonstrates an example of how water consumption, averaged across all properties in the dataset, follows a weekly pattern over two consecutive weeks, from 18th April till 2nd May 2016. In Figure 2, the 2nd May, which is a Monday, corresponds to unusually high consumption, which is typically characteristic of weekends. This is due to the fact that this day is also a bank holiday in the UK;

- 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 regularly suspected to influence demand;
- Weather variables: six variables relating to daily air and soil temperature at 10 cm depth, relative humidity, total sunshine hours and rainfall amount as well as the total number of preceding days without rain are used to account for the weather induced variance in water consumption.

Household Grouping

In order to maintain the heterogeneity of the original dataset, six household characteristics are used in order to create homogenous groups of properties. For example, one group comprises of properties with large gardens, high rateable value, metered 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 household groups:

HG (3,072) = GS (4) * RV (4) * MS (3) * Acorn (4) * CT (4) * OR (4),

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

Even though the theoretical number of groups is 3,072, some of the above household characteristics combinations contain no houses, for all or part of the days in the dataset (1,019 days in total), while others contain only one household. For this analysis, the minimum amount of households allowed in each group is set to two. Each data point represents consumption for a given group and a given day, resulting in 56,020 data points, containing 2-24 households each, or a mean of 3.8 households.

This grouping is adopted in order to reduce the number of data points and the noise in 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 l/p/d to ~390 l/p/d, with a mean consumption of 127.4 l/p/d.

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 features tested 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. Reducing the mtry increases the randomness of the trees and reduces processing time while reducing the nodesize cause the trees to grow deeper, with the danger of overfitting.

It is commonly believed that default values of these hyperparameters (e.g. mtry = number of variables/3 in regression) can produce good results, although there is no theoretical framework that supports this assumption (Scornet, 2017). A search for the optimum set of hyperparameters (mtry, nodesize, ntrees) confirmed the belief that RFs are fairly robust to changes in hyperparameter values, at least when these are varied within reasonable limits. Thus, the hyperparameter nodesize for the models is set to 200 and the number of trees at 300, although all models are tuned for the optimum value of the mtry parameter.

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., 2018). 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.

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 sensitive to outliers; the MAPE is weighted more towards smaller values and is independent of units and therefore system capacity (Xenochristou, 2019); the R^2 indicates the agreement between observed and predicted values.

Each one of the above metrics is calculated as follows:

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

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

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

where n is the total number of values. The observed values are symbolised with O and predicted values with P. Thus, O_i and P_i are the i_{th} observed and predicted value, respectively, while $\hat{O}$ and $\hat{P}$ are the observed and predicted means, respectively (Xenochristou, 2019).

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, therefore it destroys its predictive capability. 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. This process is affected by variable interactions for two reasons. First, correlated predictors mask 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 result when varying the values of the predictor within a small window. 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 assess 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). In other words, an ICE plot shows the response of the dependent variable (the daily water consumption), for a change in the independent variable (the weather), for each data instance. 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). The ICE plots are used to capture the varying effect of the weather variables, across different types of households and days in the data (Xenochristou et al., 2019).

More details and explanations regarding these three methods (variable permutation, ICE curves, and ALE plots) can be found in Molnar (2019a). All of the above analysis is performed using the R programming language, particularly the RandomForest (Liaw, 2018) and iml (Molnar, 2019b) packages.

Technical Implementation

As the methods described earlier (variable permutation and ICE curves) are affected by variable interactions, the correlations between the predictors need to be assessed. Many household variables 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 1). The x^2 varies between 1 and -1, indicating a perfect positive or negative correlation, respectively. According to Table 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. Properties with larger gardens have a higher rateable value and are occupied by residents in higher acorn groups (Table 1). Although there are clear relationships between the household variables, these were not considered strong enough in order to remove one of them as input.

An investigation into weather variable interactions (Xenochristou et al., 2019) showed that sunshine hours and humidity, rainfall and days without rain, as well as air and soil temperature are correlated. 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 are also auto-correlated from one day to the next one.

Based on the above, two groups of RF models are developed for daily predictions one day into the future (Table 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 mask the effect of other predictors.

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 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 2).

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 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.

Results and Discussion

Preliminary Analysis

The preliminary data analysis demonstrates 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 test if these conclusions align with the modelling results.

Figure 3 shows the distribution of consumption for each 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 per capita compared to high occupancy ones (3+ residents) (Figure 3(a)). Differences also appear between households in different council tax bands (Figure 3(b)), with houses in bands A-C (lower council tax bands) consuming less water per capita than houses in bands F-H (higher council tax bands).

Figure 3 also shows that 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. This is likely because lower consumption constitutes basic consumption, i.e. water used in order to perform essential day to day activities such as toilet flushing, showering, and cooking. On the other hand, higher demand values and variance, typically found in higher council tax band households, are due to additional, non-basic water consumption activities such as garden watering that occur on some days but not on 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 per capita consumption (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 4 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 4, (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 4, (b) and (c)).

Prediction Accuracy

A summary of the modelling results for the training and test datasets are shown in Table 3. Model 6 has the best performance (MAPE = 17.9%, R^2 = 54.9%). 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%). For comparison, the model that assumes water demand for each day in the data is equal to mean demand, averaged across all days, has a MAPE = 29.3% and R^2 = 0. The model that predicts consumption for each household group to be equal to the previous day has a MAPE = 23.7% and R^2 = 34.9%.

Clearly, two benchmark models are relatively simple hence unsurprisingly the RF model is performing the best. Regardless of this, the fact that the RF model is able to predict significant portion of the variance in the household consumption (R^2 = 54.9%) despite the relatively high amount of randomness associated with this level of aggregation speaks for itself. It is believed that this is due to proven ability of the RF-based models to generalize well the underlying patterns/trends in observed data whilst avoiding overfitting, even in the presence of noise. This ability of RF comes from the fact that RF model is an ensemble model comprised of multiple decision trees with different trees generalising slightly different patterns/trends in data hence ensuring that all patters/trends of interest are captured.

In addition, note that previous studies that have specifically analysed the effect of spatial (i.e. household aggregation) scale on demand prediction accuracy have highlighted the reduction in predictive performance associated with small scale household consumption (Xenochristou et al., 2020). When predicting household consumption at the monthly scale and household level, previous studies achieved a maximum R^2 of only 44% (Williamson, 2002), while in many other cases the maximum R^2 was limited to a maximum of ~30% (Jorgensen et al., 2009). Therefore, the RF model developed here (R^2 = 54.9%) performs significantly better at the challenging, single household scale.

Further, according to Table 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, Models 3-5), whereas excluding six days of past consumption from model 1 leads to increased forecasting errors (Table 3, Model 2). Model 7, which includes only six household variables and the type of day as input, performs better than model 3, which has additional temporal and weather variables. Removing all variables other than past consumption and the type of day from model 1 also slightly increases the prediction accuracy (Table 3, Model 6). In both cases, this is likely due to overfitting problems, i.e. the model learning patterns from the variables that do not influence consumption.

Based on the above, for the purposes of demand prediction, water utilities do not necessarily need to rely heavily on extensive smart metering programs over the whole network, although there are potential benefits of smart metering data beyond demand forecasting. These benefits include reduced consumption, leakage detection and deriving a greater understanding of household water consumption for individual water users. In terms of demand forecasting, smaller scale metering 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 unmetered, and overall smart meter penetration is significantly lower.

Variable Permutation

One variable is permutated at a time for each model and results appear in Figure 5 (models with past consumption) and Figure 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 5, when seven days of past consumption are included as model input, they are by far the most important predictors (Figure 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 day of the week is the only other important variable, whereas the other predictors have a mild to zero 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 individually does not have a major effect on the output.

However, things are different for model 2 (Figure 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 (i.e. by 50-53%), 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 5, Model 2).

Figure 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 6). Similarly to Figure 5, all other variables, including the weather and the rest of the temporal characteristics, are very close to a factor of one. This means that even when past consumption is not included as model input, excluding these variables from the model does not influence accuracy.

Although there are slight differences among models 3-5 (Figure 6), the importance factors relating to each predictor are very similar. Removing correlated predictors (e.g. the season, month, and various weather variables) from models 4 and 5 in this case did not increase their importance.

Notably, there is a large difference in the scale of feature importance between Figure 5 (with past consumption) and 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 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 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 or 130%.

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.

Influence of Household Variables

Next, the effect that different household characteristics have on the predictions is explored using the ALE plots (Figure 7). The y axis shows different categories of each explanatory variable, while the x axis demonstrates the deviation from the mean predicted consumption for each household category (Figure 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.

As it can be seen from Figure 7, the results are in agreement with previous analysis that explored the distribution of consumption for each household category (Figure 3). Occupancy has by far the highest influence on predicted consumption, as properties with low occupancy rate (1 resident) are predicted to consume ~75 l/p/d of water more than properties with high occupancy (3 or more residents) (Figure 7(a)). The next most influential variable is the council tax band (Figure 7(b)). Higher paying bands (F-H) have a predicted consumption of ~26.5 l/p/d more than lower bands (A-C), while unmetered customers are also on the higher end, with ~19.5 l/p/d more than metered customers (Figure 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 l/p/d more than customers in the comfortable acorn group (Figure 7(f)). Properties with large gardens are predicted to consume ~5 l/p/d more than the ones with small gardens (Figure 3.7(e)), whereas properties with high rateable values are predicted to consume ~3.5 l/p/d more than the low ones (Figure 7(d)).

The above results are in general agreement with studies in the literature (Russac et al., 1991; Edwards and Martin, 1995; Bellfield, 2001; Butler and Memon, 2006). Edwards and Martin (1995) concluded that lower acorn classes are associated with lower per capita consumption (PCC), whereas other studies (Russac, 1991; Bellfield, 2001) found no strong relationship between the acorn group and water use. Although some studies (Russac, 1991) observed that as the rateable value increases, so does water consumption, others (Bellfield, 2001) did not find any relationship between the two. Finally, the relationship between the garden size and water consumption has been so far difficult to establish (Bellfield, 2001; Gato, 2006).

Influence of Temporal Variables

The effect of four temporal characteristics on the model's result is also investigated using the ALE plots (Figure 8). According to Figure 8, the type of day and the day of the week have the highest impact on predicted water demand, whereas the month and season have almost no influence. Overall, water consumption on weekends and holidays is predicted to be ~11 **l/p/d** higher than on working days (Figure 8(c)). Water demand gradually declines from Monday to Friday, to then increase again on Saturday and Sunday. Sundays claim almost 8 **l/p/d** more on average compared to Fridays, the day with the lowest predicted consumption (Figure 8(a)). Although the month and season have almost no influence on the model's result, summers cause a slight increase in consumption (<1 **l/p/d**). An even smaller influence is observed for December (<0.5 **l/p/d**), 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.

Time variations in water use patterns are widely recorded in the literature (Edwards and Martin, 1995; Hartley, 1995; Kowalski and Marshallsay, 2005; Gato, 2006; Billings and Jones, 2008; Parker and Wilby, 2013). Water use is higher in the weekends, as this is when people tend to be more regularly at home (Edwards and Martin, 1995; Hartley, 1995; Bellfield, 2001; Gato, 2006; Parker and Wilby, 2013). Typically, water use peaks over the summer months, although lower peaks have also been observed over the winter (Billings and Jones, 2008; Parker and Wilby, 2013). However, in a temperate climate like the UK with lack of strong seasonality and rainfall well distributed over the year, it is expected that the seasonal pattern is going to be weaker than in other countries.

Influence of Weather Variables

The influence of four weather variables on the model's response, i.e. the daily water consumption, is assessed using the ICE plots (Figure 9). Previous work (Xenochristou et al., 2019) 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 here. To avoid significant 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 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). The percentage associated with each curve represents the percentage of data points that belong to each cluster.

According to Figure 9, the weather variable that causes the biggest spike in water consumption is air temperature (Figure 9(a)). This effect is non-linear and becomes significant when temperature exceeds approximately 18°C and to a lesser extent for near-freezing temperatures, which is likely due to water used to prevent pipes from freezing (Billings and Jones, 2008), or leakages between the meter and the property. Although water consumption starts increasing for temperatures over the 18°C threshold, the rate of increase varies significantly (Figure 9(a)). 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 group), the model predicts an increase in water use of up to 15 l/p/d, 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 l/p/d (Figure 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 l/p/d, 6 l/p/d 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 l/p/d. 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 9(c)). For the rest 24% of data points, water consumption drops by 4 l/p/d, 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 l/p/d, for 16% of data points. This could potentially cause problems in the future, if the length of droughts increase. For the rest of the days and households, the number of preceding days without rain has no effect on consumption.

In previous studies, the temperature (Bellfiield, 2001; Parker and Wilby, 2013; Dos Santos and Pereira, 2014), sunshine hours (Bellfield, 2001), and humidity (Dos Santos and Pereira, 2014), have all been found to influence water demand, whereas the rainfall amount had a lesser effect on water consumption (Bellfield, 2001; Schleich and Hillenbrand, 2009).

One reason that could explain this low impact of weather on prediction accuracy could relate to the mild UK climate, which lacks seasonal extremes. In this region, household demand uplifts associated with the weather are typically in the order of 5% during hot summer periods, thus weather induced demand is overall limited. Even more so, the years included in this study did not capture a particularly hot dry summer. During the record summer temperatures of 2018, the non-linear influence between weather and demand was seen at a broader aggregation – e.g. from DMA to company level. Therefore, stronger weather effects could have been observed if the analysis included 2018 data.

Another reason for the limited weather effect could be the small size of household groups (a mean of 3.8 properties/group). At this level, the noise in the consumption signal might be too strong to allow for the subtle changes due to weather to show. Previous work showed that the effect of weather becomes noticeable only for certain households, days, and times (Xenochristou et al., 2019). Therefore, when looking at the overall influence of the weather over all customer types and days, it is averaged and thus diminished.

Summary and conclusions

This study demonstrates a novel approach that combines the high accuracy of machine learning models with the interpretability of regression methods. As part of this work, a RF model is developed that predicts daily water consumption one day ahead for homogenous groups of properties (~3.8 households/group). A variety of interpretable machine learning techniques (variable permutation, ALE and ICE curves) is used in order to assess the contribution of the predictors on the forecasting accuracy and predicted water consumption.

Based on the results obtained the following conclusions can be drawn:

- The RF based short-term demand forecasting model is able to accurately capture the complex and non-linear dependencies between water consumption and different explanatory variables such as temporal, household, and weather characteristics.
- When past consumption is not available, credible forecasting models can be developed using household and temporal characteristics, while weather input does not further improve results. The best performing forecasting model in this case is the one including six household variables (occupancy rate, council tax band, metering status, rateable value, acorn, and garden size) as well as the type of day as inputs.
- When past consumption is not available, 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. The weekly pattern of consumption also becomes evident as weekends and holidays have a higher predicted consumption compared to working days, although the monthly and seasonal patterns are very weak.
- When past consumption data are included in the demand forecasting model, no other variable can significantly improve the prediction results. The best performing model in this case is the one using seven days of past consumption and the type of day as inputs.
- Although weather input does not improve the forecasting 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. This non-linearity is important to identify and is relevant to help understand and predict changes in household consumption under potential changes in the UK climate.

The above results help identify the factors that can explain consumption variability among households. Thus, they may assist with effectively targeting water conservation strategies, testing new tariffs, assessing the impact of population and lifestyle changes, as well as evaluating the effect of potential changes in the climate at the household level. In addition, this methodology can lead to the development of improved water demand forecasting models and enhance the usefulness of machine learning models, even when past consumption is not available.

The same methodology can be adopted and applied in different studies in order to determine the predictors of water demand with respect to the characteristics of each individual case. However, the results of each study are specific to and dependent on its individual characteristics that can relate to environmental factors such as climatic variables, as well as household characteristics, customs and habits, and the interactions between them. Therefore results should always be interpreted within the context of the specific case study.

In addition, this work uses a certain level of temporal (daily) and spatial (~3.8 households/group) aggregation. The small temporal and spatial scales implemented here allow to maintain the heterogeneity of the dataset and account for the influence of the different household, temporal, and weather variables, as well as their interactions, on the model’s output. However, this choice might have influenced the results. Increasing the level of spatial aggregation decreases the range of demand values and thus it reduces forecasting errors, while the variable importance also changes at different aggregation levels (Xenochristou et al., 2020).

Finally, the RF model was selected for this analysis due to its accuracy and ease of implementation. However, forecasting accuracy may further improve if a different model is used instead. The performance of RFs with respect to the characteristics of the problem, such as the temporal and spatial scale, forecast horizon, and data availability, compared to other machine learning models, has been the topic of future work (Xenochristou and Kapelan, 2020).

Data availability statement

Some or all data, models, or code used during the study were provided by a third party. Direct requests for these materials may be made to the provider as indicated in the Acknowledgments.

Acknowledgements

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 number EP/L016214/1. The data for this study was made available by Wessex Water.

References

Adadi, A. & Berrada, M. (2018). 'Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)' *IEEE Access*, 6, pp. 52138-52160. doi: 10.1109/ACCESS.2018.2870052

Akoglu, H. (2018). User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine*, *18*(3), 91–93. https://doi.org/10.1016/j.tjem.2018.08.001

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*, *9*(11). https://doi.org/10.3390/w9110887

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

Bellfield (2001). Short-term domestic water demand: estimation, forecasting and management. PhD Thesis. University of Leeds.

Berk, R.A., Sorenson, S.B. and Barnes, G. (2016), Forecasting Domestic Violence: A Machine Learning Approach to Help Inform Arraignment Decisions. Journal of Empirical Legal Studies, 13: 94-115. doi:10.1111/jels.12098

Billings, R.B. and Jones, C.V. (2008). Forecasting Urban Water Demand. 2nd Edition.

AWWA Publishing.

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

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

Butler D. and Memon F.A. (2006). Water Demand Management. IWA Publishing.

CACI Limited. (2014). The Acorn User Guide. *The Consumer Classification*. Retrieved from https://acorn.caci.co.uk/downloads/Acorn-User-guide.pdf

Calaway, R., Microsoft Corporation, Weston, S., & Tenenbaum, D. (2018). doParallel: Foreach Parallel Adaptor for the 'parallel' Package. R package version 1.0.14

Calaway, R., Microsoft Corporation, & Weston, S. (2017). foreach: Provides Foreach Looping Construct for R. R package version 1.4.4

Carmichael, I., Calhoun, B.C., Hoadley, K.A., Troester, M.A., Geradts, J., Couture, H.D., Olsson, L., Perou, C.M., Niethammer, M. Hannig, J., and Marron, J. S. (2019). Joint and individual analysis of breast cancer histologic images and genomic covariates. Retrieved from: https://arxiv.org/abs/1912.00434 [Accessed 15/01/2019].

Carvalho, D.V., Pereira, E.M., and Cardoso, J.S. (2019). Machine Learning Interpretability: A Survey on Methods and Metrics. *Electronics*, 8(8), 832. https://doi.org/10.3390/electronics8080832

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

Choi (2018). Doctor AI: Interpretable Deep Learning For Modeling Electronic Health Records. PhD Thesis. Georgia Institute of Technology. Retrieved from: https://smartech.gatech.edu/bitstream/handle/1853/60226/CHOI-DISSERTATION-2018.pdf?sequence=1&isAllowed=y [Accessed 15/01/2019].

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

Cominola, A., Giuliani, M., Piga, D., Castelletti, A., & 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*, 198–214. https://doi.org/10.1016/j.envsoft.2015.07.012

Cremona, F., Tuvikene, L., Haberman, J., Nõges, P., Nõges, T. (2018). Factors controlling the three-decade long rise in cyanobacteria biomass in a eutrophic shallow lake *Sci. Total Environ.*, 621, pp. 352-359

DEFRA. (2011). Water for life. Department of Environment, Food and Rural Affairs. Retrieved from http://static.ow.ly/docs/WaterisLife_1hi.pdf

Donkor, E. A., Mazzuchi, T. A., Soyer, R., & Roberson, J. A. (2014). *Urban Water Demand Forecasting: Review of Methods and Models*. *140*, 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.

Doshi-Velez, F., & Kim, B. (2017). Towards A Rigorous Science of Interpretable Machine Learning.

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

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

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.

Goodchild, C. (2003). Modelling the impact of climate change on domestic water demand. Water Environ. J., 17(1), 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 166 Expectation. Journal of Computational and Graphical Statistics, 24(1), pp. 44- 65, https://doi.org/10.1080/10618600.2014.907095

Hartley, J.A. (1995). A Neural Network and Rule Based System Application in Water Demand Forecasting. PhD Thesis. Brunel University.

Herrera, M., Torgo, L., Izquierdo, J., & Pérez-García, 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

Hope, R., Foster, T., Money, A., Rouse, M., & Thomas, M. (2011). *Smart Water Systems*.

Jorgensen, B.; Graymore, M.; O'Toole, K. Household water use behavior: An integrated model. J. Environ. Manage. 2009, 91, 227-236.

Jorgensen, B.S., Martin, J.F., Pearce, M.W., and Willis, E. (2014). Predicting Household Water Consumption with Individual-Level Variables. *Environment and Behavior.* 46(7), pp. 872-897. DOI: 10.1177/0013916513482462

Kowalski M, Marshallsay D (2005) Using measured microcomponent data to model the impact of water conservation strategies on the diurnal consumption profile. Water Sci Technol Water Supply 5(3–4):145–150

Liaw, A. (2018). randomForest: Breiman and Cutler's Random Forests for Classification and Regression. R package version 4.6-14.

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

Met Office, 2012. *England Averages.* [Online] 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, N. B. A. C., 2006e. *MIDAS: U.K. Soil Temperature Data,* s.l.: s.n.

Met Office, N. B. A. D. C., 2006a. *MIDAS: U.K. Daily Rainfall Data,* s.l.: s.n.

Met Office, N. B. A. D. C., 2006b. *MIDAS: U.K. Daily Temperature Data,* s.l.: s.n.

Met Office, N. B. A. D. C., 2006c. *MIDAS: U.K. Daily Weather Observation Data,* s.l.: s.n.

Met Office, N. B. A. D. C., 2006d. *MIDAS: U.K. Hourly Weather Observation Data,* s.l.: s.n.

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

Molnar, C. (2019b). iml: Interpretable Machine Learning. R package version 0.9.0.

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

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

Russac, D. A. V, Rushton, K. R., & 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), 144–162. https://doi.org/10.1051/proc/201760144.

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

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/ThamesWater/Corporate/AboutUs/Our-strategies-and-plans/Waterresources/Document-library/External-reports/demandforecastingjan16.pdf

Wong, J., Zhang, Q., & 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/10.1029/2009WR008147

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/wdsaccw/article/view/12048/7605.

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 (under revision).*

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.* https://doi.org/10.1061/(ASCE)WR.1943-5452.0001148.

Xenochristou, M., Kapelan, Z., Hutton, C., & Hofman, J. (2020). Water demand forecasting accuracy and influencing factors at different spatial scales using a Gradient Boosting Machine. *Water Resources Research (under revision).*

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

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.

Tables

Table 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

Table 2: Input variables for Models 1-7.

Variable Group	Model Input Variables	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
	Sunshine hours	X	X	X	X			
	Soil Temperature	X		X		X		
Weather	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

Table 3: Model configuration and prediction accuracy for models 1-7. The two best models, with and without past consumption data, are highlighted in bold.

	Model Parameters				Training			Testing		
Models	Cons Data	mtry	nodesize	ntrees	MAPE (%)	MSE (l^2/day^2)	R^2 (%)	MAPE (%)	MSE (l^2/day^2)	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**

Figure Captions

Figure 1: Percentage of properties in each segmentation category of the six household variables.

Figure 2: Mean PCC among all properties in the study area for two consecutive weeks, between 18th April and 2nd May 2016.

Figure 3. Distribution of consumption for different categories of six household characteristics. Each distribution comprises of mean daily per capita consumption (PCC) among all properties with the corresponding characteristic, for each day in the data.

Figure 4. Distribution of consumption for different categories of temporal characteristics. Each distribution comprises of mean daily per capita consumption (PCC) among all properties for each day in the data, for different (a) weekdays, (b) months, (c) seasons, and (d) day types.

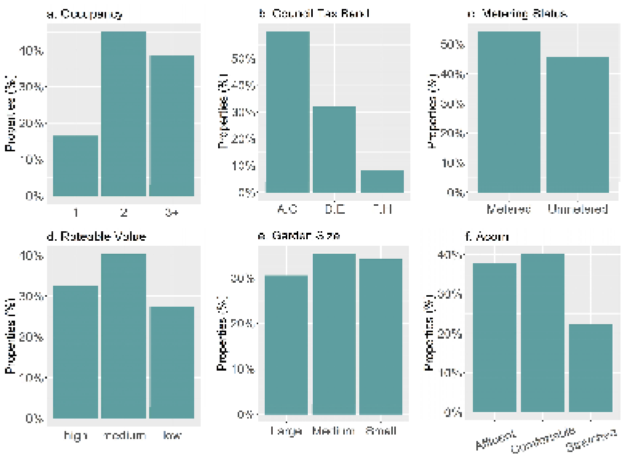
Figure 5. Factor by which the MSE increases when each feature is permutated for models 1 and 2.

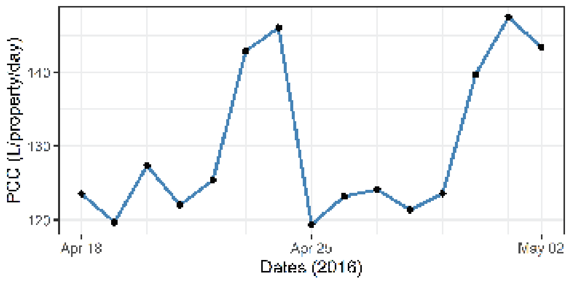
Figure 6. Factor by which the MSE increases when a feature is permutated for models 3 - 5.

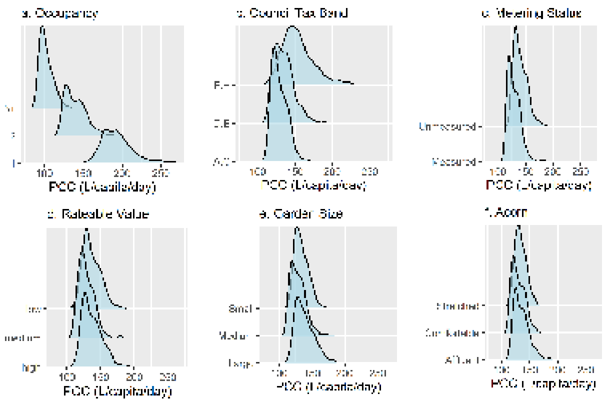
Figure 7. Influence of six household characteristics on predicted water consumption – ALE plots.

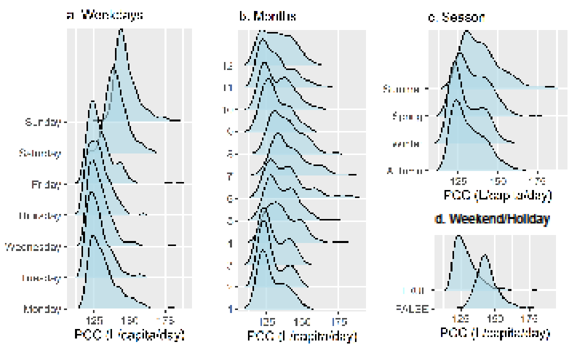
Figure 8. Influence of four temporal characteristics on predicted water consumption – ALE plots.

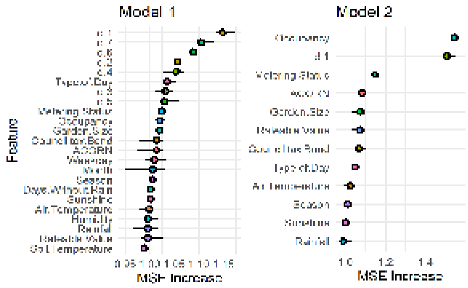
Figure 9. Influence of four weather variables (air temperature, sunshine duration, relative humidity, and days without rain) on predicted water consumption – ICE plots. The percentage associated with each curve represents the percentage of data points that belong to each cluster.

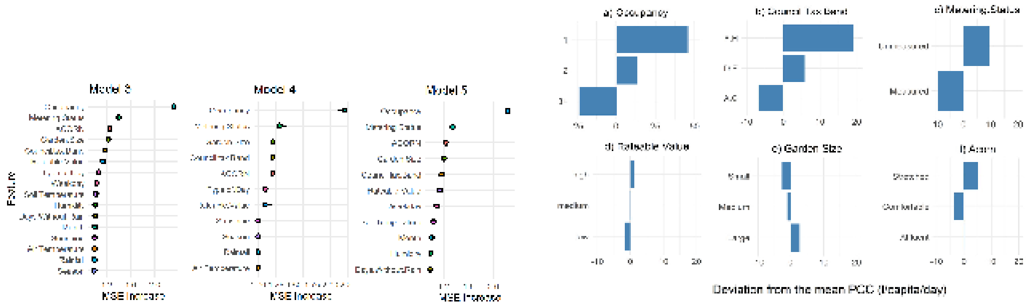


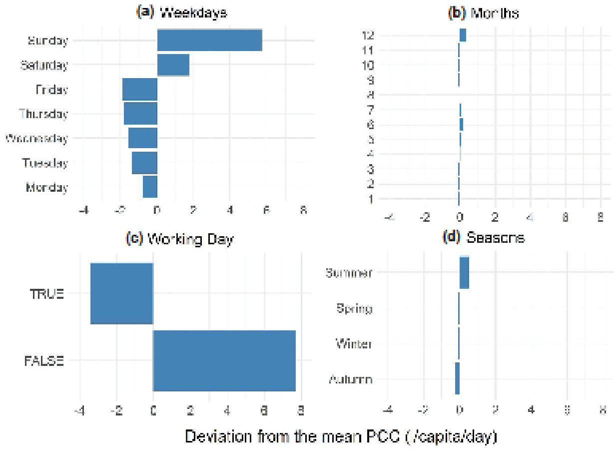


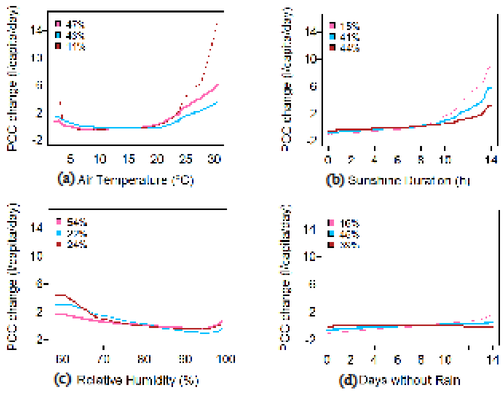







Model 3
Model 4
Model 5
MSE Increase
a) Occupancy
d) Rateable Value
c) Garden Size
Deviation from the mean PCC (l/capita/day)


(a) Weekdays
Sunday
Saturday
Friday
Thursday
Wednesday
Tuesday
Monday
(b) Months
(c) Working Day
TRUE
FALSE
(d) Seasons
Summer
Spring
Winter
Autumn
Deviation from the mean PCC (/capita/day)


PCC change (l/capita/day)
(a) Air Temperature (°C)
(b) Sunshine Duration (h)
(c) Relative Humidity (%)
(d) Days without Rain