An empirical investigation of Water Consumption Forecasting methods

Karamaziotis, Panagiotis I.; Raptis, Achilleas ; Nikolopoulos, Konstantinos; Litsiou, Konstantina; Assimakopoulos, Vassilis

International Journal of Forecasting

DOI:
10.1016/j.ijforecast.2019.07.009

Published: 01/06/2020

Peer reviewed version

An empirical investigation of

Water Consumption Forecasting methods[1]

Mr. Panagiotis I. Karamaziotis is an Electrical and Computer Engineer from National Technical University of Athens, and a Visiting Researcher in the Forecasting and Strategy Unit (www.fsu.gr)

Mr Achilleas Raptis is an Electrical and Computer Engineer from National Technical University of Athens, and a Doctoral Researcher in the Forecasting and Strategy Unit (www.fsu.gr)

Professor Konstantinos (Kostas) Nikolopoulos is the Director of forLAB www.forlab.eu. He is an expert in Time Series Analysis & Forecasting, Decision Support Systems and Management Judgment. Professor Nikolopoulos received his Engineering Doctorate in 2002 from National Technical University of Athens. He has worked in the past for the University of Manchester and other prestigious institutions in U.K. and Greece, and now holds the Chair in Business Analytics in Bangor Business School. His work has appeared in JOM, EJOR, IJF, JoF, JORS, Omega, IJPE, IJPR, IMA JMM, JBR, JCIS, IMDS, AE among other journals, and he is an Associate Editor of Oxford IMA Journal of Management Mathematics and Supply Chain Forum: An International Journal. He is co-originator of the Theta forecasting method and the ADIDA temporal aggregation method-improving framework.

Mrs Konstantia Litsiou , BA (Athens), PGCE (Athens) , MA (Lancaster) is a Senior Lecturer in Retail Management in Manchester Metropolitan University Business School. Konstantia is also pursuing a PhD in "Forecasting with Judgmental Methods the success of Megaprojects" in Salford Business School. Konstantia has a background in arts and has worked in big scale art projects and events in Greece for many years. Her work has appeared in international journals and conferences.

Professor Vassilis Assimakopoulos is a professor at the School of Electrical and Computer Engineering of the National Technical University of Athens. He has worked extensively on applications of Decision Systems for business design and he has conducted research on innovative tools for management support in an important number of projects. He specializes in various fields of Strategic Management, Design and Development of Information systems, Statistical and Forecasting Techniques using time series. He is co-originator of the Theta forecasting method and the ADIDA temporal aggregation method-improving framework.

[1] Accepted for publication in the *International Journal of Forecasting* in July 2019

An empirical investigation of

Water Consumption Forecasting methods

Many regions on earth everyday face limitations in the quantity and quality of available water resources. To that end, it is necessary to implement reliable methodologies for water consumption forecasting, that will lead to better management and planning of water resources. In this research, we analyse a first-time used large database containing data from 2 million water meters in 274 unique postal codes, in one of the most densely populated areas in Europe, which faces instances of droughts and overconsumption in hot summer months. With the assistance of R programming language, we built and tested three alternative forecasting methodologies, employing univariate forecasting techniques including a machine-learning algorithm, with very promising results.

Key words Water consumption; Water management; Time-series forecasting; Prediction Intervals; Neural Networks;

1. Introduction

Sustainability is defined as the state where a society meets the needs of the present without compromising the well-being of future generations (Holden, et al., 2013). For that matter, the optimal use of precious natural resources (e.g. water, gas, oil, etc.) is a one of the major issues for today's decision making and policy design (Guerry, et al., 2015). Particularly for water, due to the significant role it plays for the existence of every earthy form of life, the sustainable exploitation of its finite sources is a subject of ongoing regulations, policies and extensive research in the recent decades, where the problem of droughts and overconsumption threatens the ecosystem's equilibrium. While Europe is by large considered as having adequate water resources, water scarcity and drought is an increasingly frequent and widespread phenomenon across EU lands, especially in the southern areas. Since 1980, the number of droughts in Europe has increased, and according to latest findings, water stress around Mediterranean will continue to be intense (Gassert, et al., 2013).

(Figure 1)

Although some researchers openly doubt about the validity of similar approaches (Rijsberman, 2006) it is widely suggested among water scientists that water utilities, in order to ensure future water availability in scenarios of increased water stress, should focus on the improvement of the overall productivity of water rather than endlessly seeking new supplies, as the appropriate water management strategy that will tone down the effects of limited water sources (Gleick, 2003). Keeping these suggestions in mind, we believe that a robust water consumption forecasting methodology can be an inextricable part of a wider strategy for improving operational efficiency of the utility company and consequently, for conserving valuable resources in the long run. Depending on the periodicity of the forecasted variable, water utilities need to predict water demand because, for example, in short term they can better optimize day to day operations whereas in long term horizon, they can efficiently design and arrange capacity expansion for the distribution infrastructure (Billings & Jones, 2008). Additionally, on a more general perspective, having handy a reliable forecast methodology, utilities can predict possible strong upward trends and fluctuations in demand and adjust their policies keeping, in effect, the demand of the predicted variable under control, contributing to its conservation (Billings & Jones, 2008).

Our research can be summarized in the following points:

1. Aim of this paper is the examination of several methodologies and forecasting techniques and propose a reliable water consumption forecasting methodology for one of the largest and most densely populated capital cities in Europe.

2. The research was based on the official monthly data for water consumption in the time interval from 2005 to 2013. The forecast horizon was set to 17 months ahead, thus is classified as mid-term forecast (Gardiner & Herrington, 1986).

3. The examined time series include consumption patterns that come from a rare mixture of domestic, industrial and agricultural end-users. Our data cover domestic consumers that exist not only in the core area of the city, but extend within a radius of 35km and include consumers that live in the more affluent and underpopulated northern and southern suburbs, in touristic and waterside areas or in municipalities built in the foothills of a mountain at the eastern and northern fringes of the city. There also areas where water demand occurs mainly due to industrial and/or agricultural activities, although in volume are significantly smaller compared to the domestic users.

4. Three wider methodologies that use seven popular time series forecasting methods have been developed and compared with each other. In our experiments traditional statistical methods such as ARIMA and Exponential Smoothing as well as Multilayer Perceptrons are developed. Additionally, we developed four techniques that statistically combine the aforementioned individual models.

5. The final forecasts are depicted in a map utilizing ArcGIS software. Using such techniques, utilities can detect possible overcharging in demand in certain areas and take actions to prevent any unwanted local stress in the distribution system.

It will become clear in the next chapters that the proposed methodologies gave very satisfying results. Provided that the scientific approach has been successfully adapted to different consumption patterns, the

methodologies presented in this paper can be utilized unedited by other water utilities who want to optimize their operations in mid-term horizon. Our research serves as an approach in water consumption forecasting using time series forecasting methods. Since water and energy sector are linked through co-dependences (Raucher, et al., 2008), these methods and the findings of this research can be applied by operators of other utilities to upgrade mid-term operations and validate their decision making process when dealing with emergencies, network expansions or other issues. Some examples of time series forecasting techniques used in this paper on behalf of other utilities, are Hahn et al.(2009), Taylor (2003), Taylor et al.(2006) and Amjady (2001).

2. Literature Review

Urban water demand forecasting is a very active field of research. Early methodologies used relatively simple econometric and time series models that required a limited number of data sets and could be performed with modest computing power. In recent years, different methods and approaches have been utilized. These models vary from simple historical extrapolation to sophisticated analytical models employing the extensive fields of Artificial and Computational Intelligence. But there is no general census that imposes certain methodologies for a variety of forecasting variables. Thus, the choice of the appropriate forecasting model depends on the purpose of the forecast and the quality and volume of the available data the utility operator possesses.

Perhaps two of the most popular methods used in urban water demand forecasting have been time series analysis and regression models. Regarding univariate time series models, their appeal has been reduced recently, mostly because they fail to take into consideration the effects of changes in demographic, economic and technological variables as well as water demand management strategies but they are still useful nonetheless. Alhumoud (2008) uses correlation to assess the relationship between water consumption and its determinants and provides a descriptive model of annual water consumption using AR(1) model in order to assist government subsidy reform policy and capacity planning. In the 1980's Maidment et al. (1985) used short-term ARIMA models as a function of rainfall and air temperature for daily municipal water use and one year later, Maidment & Miaou (1986) applied this model to the water consumption of nine US cities.

Smith (1988) developed time series models to forecast daily municipal water demand, which included day-of-week effects and a randomly varying mean, as these two factors were not included in the Maidment's models. For regression approaches, Polebitski & Palmer (2010) used three regression models to explain observed temporal and spatial variation in residential water demand for regional infrastructure expansion, water resources management and understanding of the determinants of water demand. Other examples of water demand forecast modeling which use regression analysis include Hughes (1980) and Cassuto and Ryan (1979).

Machine learning methods, such as artificial neural networks (ANN) have been a robust modeling approach for water demand forecasting in complicated water systems because of their ability to adapt better to nonlinearities (Jain et al.(2001); Adya and Collopy (1998); Adamowski (2008); Ghiassi et al.(2008); Firat et al.(2009); Li and Huicheng (2010)). The principles of ANNs are not new in AI theory, but in the recent decades they have returned to the foreground and can be found in numerous applications. ANN is basically a computational approach that tries to mimic how the human brain works. In the water literature, ANNs have been proposed mostly as an improved method for short-term forecasting of peak daily (Bougadis et al.2005; Adamowski,2008) and hourly (Herrera, et al., 2010) water demand. Jain et al.(2001) developed a gradient-descent ANN models and compared them with time series and regression models for short term water demand forecasting. The ANN approach performed better than the other models. Similarly, Firat et al.(2009) compares the performance of generalized regression neural networks, feed forward neural networks, radial basis neural networks and multiple linear regression models and concludes that the generalized regression neural network is the best, while Behboudian et al.(2014) compared a regression model and an ANN model for the estimation of long-term per capita water demand and proved that the ANN model performed better. Once again, Adamowski et al.(2012) compared multiple linear regression (MLR), multiple nonlinear regression (MNLR), autoregressive integrated moving average (ARIMA), ANN and wavelet neural networks (WA-ANN) models for urban water demand forecasting at lead times of one day for the summer months (May to August) and found that the WA-ANN models were more accurate than the other models. Also points out that wavelet-neural network models are a potentially promising new method of urban water demand forecasting that deserve better research in the near future. Past researches have also employed ANNs with different learning algorithms with promising results. Adamowski & Karapataki (2010) found that the ANN

model utilizing the Levenberg-Marquardt learning algorithm was the most accurate for forecasting of peak weekly water demand in Nicosia, Cyprus. The DAN2 neural network model employs a different architecture than the traditional Feed Forward Back Propagation (FFBP) model and is developed to forecast monthly demand values. It was found that the DAN2 performance for modeling future water demand at multiple temporal scales performed significantly better than the ARIMA method (Ghiassi & Nangoy, 2009).

Finally, another approach in water demand forecasting is classified as hybrid models. These models are either a combination of different models or a combination of methods to estimate components (trend, cyclic, seasonality) of time series (Caiado, 2010; Wang et al.(2009); Zhou et al.(2000)). The concept of combining individual models is almost five decades old when Crane & Crotty (1976) and Ramanathan & Granger (1984) first conceived the idea of model ensembling. Wang et al.(2009) measured the performance of forecasts obtained with the weighted average of a regression model and a back-propagation ANN model and concluded that the combined forecasts outperform the individual ones. In the same way, Li & Huicheng (2010) deduced that the combination of fuzzy neural network and multiple linear regression based on the Hodrick-Prescott filter is better than any other standalone traditional method annual water demand prediction. Also, Kofinas et al.(2014) tested four popular methods for monthly averaged data and concludes that a hybrid method, where an ARIMA model is used to forecast water demand and subsequently a feed-forward ANN is used to simulate the residuals from the ARIMA model, is the best methodology. In most cases, it has been proved that the combination of individual models is preferred leading to more robust results, a conclusion that was also confirmed by the recent M4 competition where out of the 17 most accurate proposed methods, 12 were combinations of individual models (Makridakis, et al., 2018).

3. Research Area, Data and Tools

3.1 Research Area

Attica, the area for which this research is conducted for, is one of the most densely populated territories in Europe with maximum density of 1000 dwellers/km^2 and encompasses the capital city of Greece, Athens. It is 2.5% of the total area of the country but accounts for more than 35% of its population. The total water

consumption in this area is equal to 8% of the whole country's water consumption. Moreover, the area is serviced by a large water distribution network with a length of 7000km. It consists of 1500km main feeder pipeline and 5500km secondary distribution lines. The utility operator company supplies water in areas with ground elevation between 0 and 650 meters from the sea level through a total number of nearly 2 million water meters. Generally, in the research area the water crisis is already noticeable, especially during the summer months when the rainfalls are ceased almost completely. The situation is expected to worsen in the area, in the medium term (10-25 years), following the declining trends in precipitation showcased by many earlier studies (Zanis et al.2008; Feidas et al.2007).

Water Consumption

In recent years a stabilization is observed in water consumption in the research area, as can be seen in figure 2, below. There was also a rapid fall of consumption in the early 90s, because of a water saving campaign that took place (in conjunction with the increase in water rates) in order to address the reduced water supplies. Water consumption reached again the levels of 1991 in 1997, four years after the end of public campaigns for water saving. Using an exponential smoothing method with additive error and trend and no seasonality, we can see that demand will remain mostly constant in the next decade until 2024.

(Figure 2)

Climate

Located in a transitional point between the Mediterranean and the Alpine climatic zones, Attica enjoys a typical Mediterranean climate with the greatest amounts of rain occurring from mid-October to mid-April. The rest of the year remains largely rainless, making it one of the sunniest and driest areas on the European continent. Winters are generally mild, with comfortable daytime temperatures and cool nights, although light frosts may occur on infrequent occasions. It must be noted that the Northern suburbs, which stand at a higher elevation, have a somewhat different micro-climate with cooler summers and colder winters as well as higher average precipitation.

3.2 Data

After defining the recipient and the aim of the modeling process, the next important step is the collection and analysis of the data. This step helps scientists obtain the necessary information in order to draw a clear picture of the necessary features to use in the modeling phase. The main dataset that was used is the monthly observations of more than 2 million water meters (a mix of domestic, agricultural and industrial customers) depicting monthly water consumption related to each available Zip Code. Based on end-customers' water demand data and by the usage of automated procedures, a pool of 274 monthly time series have been created, corresponding to the monthly consumption rates of each of the 274 Zip Codes included in our dataset for the period from 1/2007 to 12/2013.

A secondary dataset that has been used, was the monthly observations of the number of active and connected to the grid water meters for every Zip Code. This dataset was particularly useful because it helped us to standardize the data and perform cluster analysis. In this point, we have to note that in our original data, there were also available monthly water demand rates and water meters data for the years 2005 and 2006 but they weren't used in this research. The decision to exclude these datapoints was based on the data inconsistency for this period. Not only statistical tests for outliers detection but also experts from the water management company pointed out that the specific data were incomplete and unreliable. As a result, we preferred to set as starting point of our observations the year 2007 which had fully recorded demand values in the vast majority of the study areas.

In figure 3 some indicative examples of the consumption patterns existing in our dataset are depicted. The seasonality in figures 3a as well as 3c is obvious but the trend is changing while progressing in time. For example, in the top right time series depicting consumption rates in Aghios Anargiros municipality, there is a downward trend but in 3b the trend is upward till the middle of observations and then gradually wears off. On the other hand, figures 3c and 3d depict more volatile patterns with no apparent seasonality.

(Figure 3)

3.3 Tools

As we have already mentioned, the initial dataset contains millions of datapoints that can be handled only by using automated process during all stages of our research. For this reason and to simplify the replicability of

our experiments, the development of a system architecture was one of our priorities. During the code development process, we used solely the R programming language, a very powerful statistical programming language with a growing supporting community.

(Figure 4)

In Figure 4 the architecture of our system and its components is presented. The end-user of the system can import data through two different ways. The first is by importing data files (such as .csv or .txt) directly to the working space of Rstudio. The second way is via a database server assisted by the RMySQL package for R. In order to efficiently develop our methodologies, the following list of packages were used:

- RMySQL: This package enables R to connect continuously with a MySQL DB providing to the user the appropriate data (Ooms, et al., 2018).
- tsoutliers: tsoutliers package enables the detection of outliers in time series following the Chen & Liu (1993) procedure. Innovational and additive outliers, level shifts, etc. are considered (Lopez-de-Lacalle, 2017).
- forecast: This package provides algorithmic tools for data analysis and forecasting with exponential smoothing and ARIMA models (Hyndman, et al., 2018).
- MAPA: This package implements the Multiple Temporal Aggregation method for time series forecasting (Kourentzes & Petropoulos, 2018).
- forecTheta: Both classic Theta method and its Optimized version were implemented through this package (Fiorucci, et al., 2016).
- nnfor: This package contains methods for the automatic configuration of multilayer perceptrons for time series forecasting (Kourentzes, 2017).
- stringi: A package that contains functions for string manipulation (Gagolewski , 2017).
- Metrics: This package includes some popular evaluation metrics like RMSE, sMAPE, etc. (Hamner & Frasco, 2017).

- stats: This package is part of R's core modules and includes functions for statistical analysis (R Core Team, 2017).

For visualization purposes, two additional tools appropriately interconnected with Rstudio; ArcGIS and Power BI have been used.

4. Methodology

In this chapter, a step-by-step analysis of our forecasting process is presented. Initially, data pre-processing and clustering is introduced and then the design of methodologies and forecasting methods are being discussed. Finally, the estimation of prediction intervals as well as the evaluation procedure is presented too.

4.1 Preprocessing

Tuckey (1962) defined data analysis as "Procedures for analyzing data, techniques for interpreting the results of such procedures, ways of planning the gathering of data to make its analysis easier, more precise or more accurate". After having grouped and aggregated the time series for every Zip Code, the next step would be the data cleaning process due to the existence of extreme, zero or missing values scattered in our original data, the presence of which could not be justified by the analysis and nature of the data. For the adjustment of zero or missing values, the method we used is the following:

- The rate to replace the missing value is defined as the half-sum of previous and the next observation, when the time series characterized by stationarity and has no seasonal behavior.

- If the time series exhibits seasonality, the missing value is replaced with the average of previous and/or following values of the respective periods.

Different treatment is required for zero values which are a different category of "peculiar" values of a time series. These values are divided into two categories. First, there are the zero values recorded as zero due to an error in the log information system and do not correspond to reality. These values are essentially no different from the aforementioned missing values and their treatment is similar. The second category of zero values are the actual zeroes which refer to intermittent demand time series. These values obviously cannot

be filled with any of the previous methods as they are actual recorded zero values and they require an approach based on different methodology that does not involve a change in the value, but the exploitation of the information taken from the zero value. In our data, zero values were of low frequency and after consulting the experts in the water management company we concluded that these values should not be altered. For example, there were few cases of remote areas with positive number of active water meters but zero consumption rates. A valid explanation of this could have been a case of grid expansion with temporary absence of connected end-users to the grid.

The detection and correction of the influence of extreme values is equally important because they affect the selection of the model, the estimation of the parameters and other results pursued by the analysis, such as seasonal adjustment. In this study, we considered various types of outliers (i.e. additive outliers, level shifts and temporary changes) and their detection and adjustment was performed using the "tsoutliers" statistical package for R and specifically the function tso() which is based on the work of Chen & Liu (1993). It is a three-stage iterative process where in the first stage, the potential outliers are identified and in the second stage, joint estimates of the model parameters and outlier effects are obtained using the accumulated outlier information of Stage 1. The adjusted series are obtained by removing only the statistically significant effects. In the final stage, the outliers are identified and their effects estimated again based on the less-contaminated estimates of model parameters obtained in Stage 2. The figure below highlights the effect of applying the tso() function to the original data of a specific time series of our data. It is noticeable that after the removal of the outliers, the time series shows obvious seasonal behavior with a downward trend.

(Figure 5)

4.2 Data Clustering

For the purpose of this research, we had to cluster our Zip Code-level time series data in homogenous groups. In order to group together data that have the same progress in time irrespective of their magnitude, we scaled the data based on the number of water meters existing in each Zip Code. Overall, for each Zip Code the number of water meters remained mostly steady for our examination period, unaffected by potential fluctuations in water consumption. Additionally, the average monthly number of water meters in each Zip Code is significantly positively correlated with the average monthly water consumption rate observed in the

corresponding Zip Codes. Thus, one way to scale the time series for each Zip Code as an input for the cluster algorithm, is to divide the monthly consumption rates with the respective number of water meters.

In our analysis, the agglomerative hierarchical clustering has been used. It is an algorithm that creates a hierarchy of groups in which clusters are created by merging the clusters from the previous, lower level. Contrast to the divisive clustering, agglomerative clustering has a bottom-up approach where at the beginning, each data-point (or time-series in our case) is a standalone cluster. The process continues to merge clusters that are closest until there is only one, big cluster and a dendrogram can be used to depict the arrangement of clusters. In order to define how the merging is performed, a similarity measure between groups (linkage) must be calculated along with the one that is used to calculate pairwise similarities (Hastie, et al., 2009). For the purpose of this research the dissimilarity between individual time-series was measured using Euclidean distance and the between-clusters distances were calculated using the Ward's method that minimizes the total within-cluster variance (Ward Jr., 1963). These two methods were selected due to their vast usage in the literature and because of their sensible results.

For the determination of the number of clusters, one can inspect the dendrogram and "cut the tree" in the height that the largest change in dissimilarity occurs. Apart from this approach, there are several indices that can be used; the elbow method adjusted for hierarchical clustering is one of them and is shown in the following picture.

(Figure 6)

The output results of the elbow method suggest that two clusters might be the optimal choice. However this metric gives indicative results and additional experimentation is required. Therefore, we decided to run forecasting experiments for various number of clusters and scaling methods and the results are summarized in Table 5. In Figure 13 the cluster segmentation is drawn. It seems that the algorithm divided the study area in four main groups: cluster 1 - North, cluster 2 - Central, cluster 3 - Inner South/ Coastal and cluster 4 - Outer South. For the implementation of cluster analysis, we used the function hclust() in the core package "stats" in R.

4.3 Time Series Forecasting Methods

For this research, seven forecasting methods were selected. We used methods used in the literature of urban water demand forecasting and some other methods derived from the time series forecasting literature. The methods are:

Naive

The Naive method is the simplest statistical method, whose forecast is measured as the actual observation of just the previous period. It has been established to use the results of this method as a point reference for the accuracy of other methods. In our research we used the function naive() or snaive(), in the case of seasonal data, for the implementation of the naive method. These functions are part of the "forecast" package in R (Hyndman, et al., 2018).

Exponential Smoothing

The methods of exponential smoothing are particularly popular from the very beginning of their appearance (Gardner, 1985). They give greater weight to more recent data of the time series because they assume that they provide more information compared to the older data. The constant level model (SES) assumes the absence of voltage data and produces forecasts with horizontal extension of the data. The linear trend model (Holt) is an extension of the SES model, which now manages the trend component observed in many time series data. This method involves a forecast equation and two smoothing equations (one for the level and one for the trend). A variation from Holt's linear trend method is achieved by allowing the level and the slope to be multiplied rather than added. Also, the declining trend model (Damped) gradually reduces the effect of trend on the predicted results and controls the growth of their values. Lastly, for seasonal data the Holt's method is expanded by adding a component to account for seasonal time series (Winters, 1960). For this research we built exponential smoothing models using the ets() function in the "forecast" package in R (Hyndman, et al., 2018) and the appropriate components of the model are selected by stepwise estimation of the Akaike Information Criterion (AIC) in the training sample (Hyndman & Khandakar, 2008).

Auto-Regressive Integrated Moving Average

The method of auto-regressive moving average (ARIMA) is a popular prediction method which can be used for one-dimensional or multivariate analysis. The autoregressive part (AR) of the method indicates that every

next value of the time series can be approximated by a regression of past values. The moving average part (MA) uses the past errors to model future values of the variable of interest. The "I" in the name of the method stands for "Integrated" and indicates that the data values have been replaced by the differencing of the previous values with the current values and is a way of making the time series stationary. Therefore, the order of an ARIMA model is defined by three letters; p, d, q denoting the order of autoregressive lags, differencing and moving-average model respectively. For seasonal time series, the seasonal component has the same structure as the non-seasonal model and it may have an AR factor, an MA factor, and/or an order of differencing. Box-Jenkins methodology is a popular way for selecting the appropriate order values of ARIMA (Box & Jenkins, 1970). The implementation of the ARIMA was built utilizing the auto.arima() function in the "forecast" package in R (Hyndman, et al., 2018) for seasonal and non-seasonal configurations of the model and the order selection is based on the stepwise estimation of the Akaike Information Criterion (AIC) of the calculated model in the training set (Hyndman & Khandakar, 2008).

Theta

The Theta method is a one-dimensional method based on the change of local curvature of a time series via the parameter θ (Theta), which is applied multiplicatively to the second class differences of the data. The resulting series that are created maintain the mean and the slope of the original data but not their curvatures. These new time series are named Theta-lines. Their primary qualitative characteristic is the improvement of the approximation of the long-term behavior of the data or the augmentation of the short-term features, depending on the value of the Theta coefficient. The proposed method decomposes the original time series into two or more different Theta-lines. These are extrapolated separately and the subsequent forecasts are combined. The simple combination of two Theta-lines, the line with $\theta=0$ (straight line) and the line with $\theta=2$ (double local curves) was adopted in order to produce forecasts for the 3003 series of the M3 competition with good performance overall (Assimakopoulos & Nikolopoulos, 2000). The implementation of the model was made through the function stheta() in the "forecTheta" package in R (Fiorucci, et al., 2016).

Optimized Theta

Several papers propose methods for optimizing the classic Theta method presented above. In our research we used the model as described by (Fiorucci, et al., 2016). This work extends the classic method Theta, by

selecting the line θ that best describes the short-term behavior of the time series, maintaining unchanged the long-term component (for θ=0). For the selection of this theta line an error function, which is based on error forecasts in a rolling validation sample, is minimized. The combination of forecasts from these two Theta lines happens through weights (not necessarily equal as in the classic Theta method) ensuring the reconstruction of the original time series. The optimized method was built through the function otm() in the "forecTheta" package in R (Fiorucci, et al., 2016).

Multiple Aggregation Prediction Algorithm

The Multiple Aggregation Prediction Algorithm (MAPA) is a forecasting methodology where each time series is broken down into individual time series by time aggregation (for example, the conversion of monthly time series to bi-monthly, quarterly, etc.), in each of these time series an exponential smoothing model is applied and forecasts are made for the respective characteristics of the time series (i.e. level, trend, seasonality). Then, these characteristics from each aggregation level are combined (either using simple/weighted average or median operator) to form the final prediction. The novelty of this algorithm is that it proposes the combination of the individual characteristics of the forecasts for each level and for each type of characteristic. Thus, separate components are combined and then they are aggregated to express the final forecasts (Kourentzes, et al., 2014). The model was built using the function mapa() in the "MAPA" package in R (Kourentzes & Petropoulos, 2018) with weighted average combination since this configuration gave better out-of-sample results overall.

Multilayer Perceptron

A multilayer perceptron (MLP) is a type of feed-forward artificial neural network. A typical MLP has one hidden layer with non-linear activation function and linear activation function on the output neuron. In our research, we used an MLP modeled with the function mlp() of the package "nnfor" in R (Kourentzes, 2017). This method combines filter and wrapper approaches for feature evaluation, construction and transformation and can be fully automated making it ideal for batch forecasting (Crone & Kourentzes, 2010). Specifically, we considered architectures with one hidden layer and a hyperbolic tangent or logistic function as the activation function in every neuron of the hidden layer. The optimal number of neurons and activation types was determined using a 20% validation set randomly sampled and a back-propagation algorithm is used to

adjust the weight of each neuron by calculating the gradient of the loss function. The method trains 20 networks which are used to produce an ensemble forecast using either the average, median or mode operator (Kourentzes, et al., 2014). In our experiments, mode operator produced superior results and was the default choice during the experiments. Moreover, the inputs of the MLP are autoregressive lags of order between 1 to 12 with binary dummies to cope with seasonality if present.

Ensemble Methods

Driven by the extended usage of combination of individual time series forecasting models in the literature (Jose & Winkler, 2008 ; Stock & Watson, 2004) we considered and thoroughly tested the following ensemble models in our dataset:

- <u>Simple Average</u>

  In this approach, all individual methods are combined with equal weights and each point forecast is the average of the respective point forecasts of all the individual models described above.

- <u>Weighted Average</u>

  In this slightly different approach, we combined all individual methods introduced above based on their inverse out-of-sample MASE error for every time series, namely:

$$f_{comb} = \frac{\frac{1}{MASE_i}}{\sum_{i=1}^{n} \frac{1}{MASE_i}} f_{indiv}^{i}$$

  This approach has been used in several papers using primarily a MSE metric (Stock & Watson, 2004) but in our approach, we utilized the MASE error with comparable results.

- <u>Trimmed Average</u>

  Trimmed Average is the method that calculates the mean of the available data after the removal of a percentage of the smallest and largest values in the data. In our experiments, for the production of an overall forecast, we trimmed 20% of the forecasts of the aforementioned individual methods and averaged the remaining.

- <u>Median</u>

  Here, the final forecasts are computed through the median of the individual point forecasts.

4.4 Defining Methodologies

The previously defined time series forecasting models will be utilized by the methodologies that will be described in this section. These methodologies borrowed ideas from hierarchical time series forecasting theory (Hyndman, et al., 2016) and define the way every data and forecasting technique is used in order to produce the final forecasts.

Methodology No1

From the 274 time series for water consumption for each Zip Code, 4 clusters were created based on the cluster analysis presented in section 4.2 and the time series belonging to each cluster were aggregated in order to form 4 cluster-level time series. Then each of the forecasting techniques described in section 4.3 were applied to each of the 4 clustered time series (in total 11*4 = 44 out-of-sample estimations were produced). The model that gave the best performance in a hold-out sample (i.e. has the smallest error in this sample) was selected to produce the optimal forecast for the particular time series. Finally, the optimal forecasts taken for each of the 4 clustered time series were simply aggregated in order to obtain the final estimates of the total consumption of the research area.

Methodology No2

In this approach, we obtained the total estimates of water consumption by aggregating twice the estimations of the 274 time series (2-level hierarchical structure). Having the 274 time series for all Zip Codes, predictions were generated based on the time series forecasting models presented above. In total, 11 forecasting models were fit to 274 time series (i.e. 3014 time series forecasts were produced). The model that gave the best forecasts in a hold-out sample was selected to produce the optimal forecasts for each of the 274 time series. Next, the final forecasts for the total water consumption were produced by aggregating bottom-up these base-forecasts, firstly on a cluster level (middle level), and then to the final series which depict the total consumption.

Methodology No3

Here, each of the 274 time series were added up to form the total water consumption of the area between 1/2007 and 12/2013. We then generated forecasts for the total time series, using the forecasting models

described in the previous section (in total, 11 time series forecasts were produced). The model that has the best performance in a hold-out sample will be selected to produce the final forecasts for this methodology.

4.5 Evaluation

The training of the forecasting algorithms was made by splitting each time series in two parts of 80% and 20% respectively. The first part is termed as in-sample and is used to train the model. In our experiments, the in-sample data of the time series were from 1/2007 till 7/2012 (67 months). The second part is termed as out-of-sample and is used to evaluate the generalization ability of each forecasting model trained in the in-sample data. In our experiments, the out-of-sample part of the time series were between 8/2012 to 12/2013 (17 months).

Error Metrics

In order to select which time series forecasting model performs best in a hold-out sample, MASE error was the only metric considered. For the evaluation of overall performance of the three methodologies presented in 4.4 section, the following metrics summarized below were measured;

1. **Mean Absolute Error (MAE)** is the simplest computationally and easiest to understand from the metrics considered in our experiments. The metric can be calculated as follows:

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |y_i - \bar{y}_i|}{n}$$

   The main downside of the metric is that it is scale-dependent and can't be used to assess the performance of forecasting models in multiple time series.

2. **Mean Absolute Scaled Error (MASE)** was the primary error metric for our experiments and it is defined by the following equation:

$$\mathrm{MASE} = \begin{cases} \dfrac{(n-1)\ \mathrm{MAE}}{\sum_{i=2}^{n} |y_i - y_{i-1}|}, & \text{for non-seasonal data} \\[2ex] \dfrac{(n-s)\ \mathrm{MAE}}{\sum_{i=s+1}^{n} |y_i - y_{i-s}|}, & \text{for data with seasonality s} \end{cases}$$

MASE is a scale-independent error since the regular MAE in the nominator is scaled by the absolute in-sample MAE from the naive forecast method. If MASE is smaller than one then it arises from a better forecast than the average one-step ahead, naive forecast computed in-sample (Hyndman & Koehlerb, 2006). Compared to MAPE which is a popular scale-free metric, MASE can give reliable results in time series that contain zero values while MAPE is unable to handle times series with many zeroes. Furthermore, MASE can penalize positive and negative forecast errors equally whilst MAPE fails on both of these criteria.

3. **symmetric Mean Absolute Percentage Error (sMAPE)** is calculated by the following formula:

$$\mathrm{sMAPE} = \frac{100\%}{\mathrm{n}} \sum_{i=1}^{n} \frac{2\,|\mathrm{y_i} - \bar{\mathrm{y}}_\mathrm{i}|}{\mathrm{y_i} + \bar{\mathrm{y}}_\mathrm{i}}$$

This metric was proposed instead of MAPE by Armstrong (1985) because unlike MAPE, it puts equal penalties in positive and negative errors. The main disadvantage of this metric, is that in cases where y_t is zero, $\bar{y}_t$ is likely to be close to zero which causes division by a number very close to zero. Also, sMAPE values can be negative which causes ambiguity.

4. **Root Mean Square Error (RMSE)** or Root Mean Square Deviation (RMSD) is the most frequently used performance metric for the evaluation of regression forecasts. RMSE can be calculated as:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} (\bar{\mathrm{y}}_\mathrm{i} - \mathrm{y_i})^2}{\mathrm{n}}}$$

RMSE is always non-negative and a value of zero indicates a perfect fit to the data. Generally, a lower RMSE is better than a higher one. However, since RMSE is scaled-dependent, comparisons across different types of data would be invalid. Also, since each error is squared, possible large deviations have a large effect on the RMSE value and thus, making RMSE sensitive to outliers (Armstrong & Collopy, 1992).

4.6 Prediction Intervals

Predictions usually are expressed as single numbers, termed as "point forecasts" which give no information regarding their probability of actually occurring. Computing prediction intervals (PI) is an important part of the forecasting process intended to indicate the likely uncertainty in point forecasts (Chatfield, 1993). For this research, we produced h=17-step ahead forecasts by aggregating point forecasts from a variety of individual or ensemble forecasting models. To construct the PI of the three methodologies, bootstrapped time series were used in order to account for the many uncertainties in time series forecasting (Hyndman & Athanasopoulos, 2018). Specifically, we created 100 bootstrapped time series that are similar to the original data using the Box-Cox and Loess-based decomposition bootstrap (Bergmeir, et al., 2016) and then ARIMA models were used for each of the simulated time series to obtain the prediction intervals[2]. For example, in the second methodology, 100 simulated time series were created for each of the 274 bottom-level time series and the final prediction intervals of the Methodology No2 were derived from the simple bottom-up aggregation of the prediction intervals obtained in each bottom-level time series. The only assumption needed for the validity of this approach is that the residuals are uncorrelated and indeed, this assumption holds true as can be seen in the next figure for the residuals of each methodology.

(Figure 7)

5. Results

Methodology No1

The next table summarizes the out-of-sample MASE error of all the time series forecasting models when applied to the clustered time series of consumption. It is clear that ARIMA performs best in 3 out of the 4 clustered time series and in 1 time series has the eighth smallest MASE error. Also, MAPA model performs best in 1 time series and in 2 others has the third best performance of all the predictive models considered. Additionally, Optimized Theta has marginally better results compared to the classic Theta. On the other hand, MLP model despite exhibiting good fit to the training data, it seems that they cannot maintain the momentum

[2] In the instances where e.g. the upper prediction intervals were smaller than the point forecasts or the actual values, the upper prediction interval was set equal to the respective point forecast.

for longer forecast horizons. Regarding Ensemble methods, trimmed mean has a slight performance gain here over the median combination whereas weighted average performs poorly.

(Table 1)

Aggregating the best forecasts from each clustered time series, we have the total forecast for the first methodology. The following line graph summarizes the predicted values obtained from this methodology along with the actual values and the prediction intervals.

(Figure 8)

Methodology No2

Table 2 summarizes the results of the out-of-sample MASE error of all the time series forecasting models when applied to the Zip Code-level time series of consumption. For this methodology it is impossible to demonstrate here the errors of the methods for each of the 274 time series. Therefore, Table 2 and Figure 9 presents a summary of this extended table of errors. Inspecting this information we can see that again, ARIMA performs better in more time series than any other algorithm. In particular, ARIMA has the smallest and the second smallest MASE out-of-sample error in 120 out of the total 274 time series forecasts. The closest individual model to ARIMA's performance is the Theta method which takes the first and second spot in 59 out of 154 time series (excluding the time series that ARIMA has the first and second best performance). Among Ensemble models, again, trimmed mean performs best and has a slightly better overall performance than ARIMA, making it the best performing model for the second methodology.

(Table 2)

(Figure 9)

Aggregating the best forecasts from each of the 274 time series, we have the total forecast for the second methodology. The next graph summarizes the predicted values obtained from this methodology along with the actual values and the prediction intervals.

(Figure 10)

Methodology No3

In this approach, we fit all the time series models in one total time series depicting overall water consumption in the research area. Table 3 summarizes the out-of-sample MASE errors. Once more, ARIMA has the smallest error, followed by Optimised and classic Theta. For Ensemble models, median here has a marginal improvement over trimmed mean with simple average and weighted average trailing.

(Table 3)

The next graph summarizes the predicted values obtained from this methodology along with the actual values and the prediction intervals.

(Figure 11)

6. Conclusion and Further Research

Summarizing all the forecast errors for all the three methodologies, we have the following Table 4.

(Table 4)

To select the best forecasting methodology, one should consider not just the accuracy of the point forecasts, but the accuracy of the prediction intervals too. The first proposed methodology produces the best point forecasts in the hold-out sample but at the same time consistently fails to include the actual values of the predicted variable of interest within the upper and lower bounds of the prediction intervals. Conversely, the other two methodologies showcase lesser accuracy compared to Methodology No1 in the out-of-sample data and the range of prediction intervals include the actual values of the variable of interest. Concretely, Methodology No3 has respectable forecasting ability and produces tighter prediction intervals than Methodology No3. Therefore, the selection of the best strategy for water consumption forecasting for the study area depends on the specific minimum needs and priorities of the stakeholders. If forecasting accuracy for the unknown future is crucial then Methodology No1 is the frontrunner. On the other hand, if someone needs a methodology that produces solid point forecasts as well as reliable prediction intervals to have a good understanding of future variations of the consumption, then Methodology No3 is preferred.

Additionally, some interesting findings have occurred by inspecting the performance of each individual and ensemble model. The following figure summarizes the overall rankings of each forecasting model in every methodology separately as well as overall. The ranks for each model in each methodology were calculated

as the weighted average of their out-of sample performance (i.e. MASE error), based on the number of time series forecasts considered for each methodology. In the first methodology, the first rank is shared among ARIMA and MAPA forecasting models. In the second methodology the Ensemble model with trimmed mean has the best performance and in the third methodology ARIMA is again the best performing model. The Ensemble model with trimmed mean and ARIMA are sharing the top spot in the overall rankings. In the second and third overall position we see the Ensemble model with median and mean combination respectively. Optimised Theta and classic Theta complete the first five ranks, in the fourth and fifth position respectively.

(Figure 12)

Interestingly, despite their vast utilization in the literature for regression problems, MLPs trained on this dataset failed to deliver as expected since their performance were among the least reliable. In Figure 12 we illustrate the 100% stacked average percentage change in MASE error between in-sample and out-of-sample measurements for all the time series models considered when trained on the dataset for the three methodologies. MLPs have by far the largest increase in error accounting for almost 50% of the total percentage change for all the models considered. Perhaps, one of the reasons for their subpar performance could be attributed to the fact that the training sample was not big enough for the neural network to learn the patterns of the data. On the other hand, classic Theta and its optimised version have the smallest average percentage increase among individual models with less than 2% change each.

(Figure 13)

Moreover, because our water consumption rates are accompanied with Zip Code information, we can spatially visualize the forecasted water demand. Consequently, water utility operators can predict possible water stress in the water distribution system and act accordingly to prevent failures. In Figure 13, the average forecasted consumption rates for every Zip Code is depicted. It seems that the average consumption rate is higher on areas located in the northern and southern coastal suburbs. The cluster segmentation derived from cluster analysis is also highlighted and we can observe consumption patterns from consumers in central, northern and coastal southern and outer southern parts of the city.

(Figure 14)

The cluster-level aggregated consumption rates are depicted in Figure 14. These plots showcase strong seasonality with downward trend. Notice that cluster 4 has almost no trend, especially in the last couple of years. Also, for the second cluster, the downward trend is interrupted around the years between 2011 and 2013.

(Figure 15)

Last but not least, due to the variety of patterns that our data demonstrated for this research, we concluded that the best performing forecasting methodology can be utilized not only for data regarding our research area, but for other regions too. Because our data include consumption patterns such as mixed industrial and domestic areas or mixed industrial and agricultural areas etc, the proposed Methodology No1 can be equally successful in other areas that have one or more of the aforementioned water consumption characteristics.

Nevertheless, there are a few things that can be tested in future iterations of this research that may provide further insights and improvements. For example, the incorporation of other time series models like Support Vector Regression, Random Forests or other architectures of Neural Networks (i.e. recurrent neural networks) may improve the overall performance of the existing methodologies. In addition, exploring further the hierarchical structure of the data and applying novel forecast reconciliation approaches to the dataset may improve the accuracy further (Wickramasuriya, et al., 2018). Also, a more feature-rich dataset could be used along with the main dataset describing the dependent variable. The main factors affecting the demand for water is the weather, the price of the water, and other socio-economic factors (like annual change of GDP or employment rate) (Brekke, et al., 2002). The integration of these factors to the forecasting model can reveal co-dependencies that might boost the performance of the proposed forecasting model. The addition of weather factors is very interesting scientifically, but it has certain drawbacks. The weather conditions of a region are not available at the time of forecast production and one should first make forecasts to predict weather conditions in order to successfully incorporate these factors in the model. So, a thorough research should be done prior to embedding this information to the model. The integration of economic factors is a promising extension too, as the external economic environment greatly affects the behaviour of consumers (Billings & Jones, 2008). Good economic conditions lead to increased water demand and vice versa. It has

been observed that in families with unemployed members (and therefore less income), adjust their habits in terms of water consumption.

Furthermore, the exploitation of the spatial distribution of the Zip Codes is an interesting addition worth exploring and testing. Despite the fact that the exact location of each water meter is unknown, the position of each Zip Code is very much known and this information could be embedded in the modelling process. There are many studies for various applications incorporating the spatial as well as the temporal dimension of the data in time series forecasting and the current dataset could reproduce some of these ideas (Kamarianakis & Prastacos (2005); Giacomini & Granger (2004)). For example, an interesting iteration would be the use of Spatio-Temporal ARIMA (STARIMA) with the help of an adjacency matrix that will incorporate the spatial correlations derived by the specific positioning of the Zip Codes in the study area.

With regards to the clustering process, other distances can be tested for the construction of the dissimilarity matrix between time series. In this research we tested the Euclidean distance transforming the original data with two ways: 1) by scaling each time series with the number of clusters and 2) by applying z-score transformation. As Table 5 suggests, scaling with the number of water meters and grouping the time series in 4 clusters, gave the best performance in the out-of-sample data. However, for time series clustering other measures of distance might be tested such as Dynamic Time Warping and Longest Common Subsequence (Ralanamahatana, et al., 2005).

(Table 5)

References

Adamowski, J. et al., 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(1).

Adamowski, J. F., 2008. Peak Daily Water Demand Forecast Modeling Using Artificial Neural Networks. *Journal of Water Resources Planning and Management*, 134(2), pp. 119-128.

Adamowski, J. & Karapataki, C., 2010. Comparison of Multivariate Regression and Artificial Neural Networks for Peak Urban Water-Demand Forecasting: Evaluation of Different ANN Learning Algorithms. *Journal of Hydrologic Engineering*, 15(10), pp. 729-743.

Adya, M. & Collopy, F., 1998. How effective are neural networks at forecasting and prediction? A review and evaluation. *Journal of Forecasting*, 17(5-6), pp. 481-495.

Alhumoud, J. M., 2008. Freshwater consumption in Kuwait: analysis and forecasting. *Journal of Water Supply: Research and Technology- Aqua*, 57(4), pp. 279-288.

Amjady, N., 2001. Short-term hourly load forecasting using time-series modeling with peak load estimation capability. *IEEE Transactions on Power Systems,* 16(3), pp. 498-505.

Armstrong , J. S., 1985. *Long-Range Forecasting: From Crystal Ball to Computer.* s.l.:s.n.

Armstrong, J. S. & Collopy, F., 1992. Error Measures For Generalizing About Forecasting Methods: Empirical Comparisons. *International Journal of Forecasting*, Volume 8, pp. 69-80.

Assimakopoulos, V. & Nikolopoulos, K., 2000. The theta model: a decomposition approach to forecasting. *International Journal of Forecasting*, 16(4), pp. 521-530.

Behboudian, S., Tabesh, M., Falahnezhad, M. & Ghavanini, F. A., 2014. *Journal of Water Supply: Research and Technology—AQUA*, 63(1), pp. 31-41.

Bergmeir, C., Hyndman, R. J. & Benítez, J. M., 2016. Bagging exponential smoothing methods using STL decomposition and Box–Cox transformation. *International Journal of Forecasting*, 32(2), pp. 303-312.

Billings, B. R. & Jones, C. V., 2008. *Forecasting Urban Water Demand.* 2 ed. s.l.:American Water Works Association.

Bougadis, J., Adamowski, K. & Diduch, R., 2005. Short-term municipal water demand forecasting. *Hydrological Processes*, 19(1), p. 137– 148.

Box, G. & Jenkins, G., 1970. *Time Series Analysis: Forecasting and Control.* San Francisco: San Francisco: Holden-Day.

Brekke, L., Larsen, M. D., Ausburn , M. & Takaichi, L., 2002. Suburban water demand modeling using stepwise regression. *American Water Works Association*, 94(10), pp. 65-75.

Caiado, J., 2010. Performance of combined double seasonal univariate time series models for forecasting water demand. *Journal Of Hydrologic Engineering*, pp. 215-222.

Cassuto, . A. E. & Ryan, S., 1979. Effect Of Price On The Residential Demand For Water Within An Agency. *JAWRA Journal of the American Water Resources Association*, 15(2), pp. 345-353.

Chatfield, C., 1993. Calculating Interval Forecasts. *Journal of Business and Economic Statistics*, Issue 11, pp. 121-135.

Chen, C. & Liu, L.-M., 1993. Joint Estimation of Model Parameters and Outlier Effects in Time Series. *Journal of the American Statistical Association*, 88(421), pp. 284-297.

Crane, D. B. & Crotty, J. R., 1976. A two-stage forecasting model: Exponential smoothing. *Readings in Managerial Economics*, pp. 128-136.

Crone, S. F. & Kourentzes, N., 2010. Feature selection for time series prediction - a combined filter and wrapper approach for neural networks. *Neurocomputing*, 73(10-12), pp. 1923-1936.

Feidas, H., Noulopoulou, C., Makrogiannis, T. & Bora-Senta, E., 2007. Trend analysis of precipitation time series in Greece and their relationship with circulation using surface and satellite data: 1955–2001. *Theoretical and Applied Climatology*, 87(1), p. 155–177.

Fiorucci, J. A., Louzada, F. & Yiqi, B., 2016. *forecTheta: Forecasting Time Series by Theta Models. R package version 2.2.* s.l.:s.n.

Fiorucci, J. A. et al., 2016. Models for optimising the theta method and their relationship to state space models. *International Journal of Forecasting*, 32(4), pp. 1151-1161.

Firat, M., Yurdusev, M. A. & Turan, M. E., 2008. Evaluation of Artificial Neural Network Techniques for Municipal Water Consumption Modeling. *Water Resources Management*, 23(4), p. 617–632.

Gagolewski , M., 2017. *R package stringi: Character string processing facilities..* s.l.:s.n.

Gardiner, V. & Herrington, P., 1986. *Water Demand Forecasting.* s.l.:CRC Press .

Gassert, F., Reig, P., Luo, T. & Maddocks, A., 2013. *Aqueduct Country and River Basin Rankings: A Weighted Aggregation of Spatially Distinct Hydrological Indicators,* s.l.: World Resources Institute.

Ghiassi, M., Zimbra, D. & Saidane, . H., 2008. Urban Water Demand Forecasting with a Dynamic Artificial Neural Network Model. *Journal of Water Resources Planning and Management*, 134(2), pp. 138-146.

Ghiassi, M. & Nangoy, S., 2009. A dynamic artificial neural network model for forecasting nonlinear processes. *Computers & Industrial Engineering*, 57(1), pp. 287-297.

Giacomini, R. & Granger, C., 2004. Aggregation of Space-Time Processes. *Journal of Econometrics*, Volume 118, pp. 7-26.

Gleick, P., 2003. Global Freshwater Resources: Soft-Path Solutions for the 21st Century. *Science,* 302(5650), pp. 1524-1528.

Guerry, A. et al., 2015. Natural capital and ecosystem services informing decisions: From promise to practice. *Proceeding of the national academy of sciences of the United States of America,* Volume 112, p. 7348–7355.

Hahn, H., Meyer-Nieberg, S. & Pickl, S., 2009. Electric load forecasting methods: Tools for decision making. *European Journal of Operational Research,* 199(3), pp. 902-907.

Hamner, B. & Frasco, M., 2017. *Metrics: Evaluation Metrics for Machine Learning. R package version 0.1.3.* s.l.:s.n.

Hastie, T., Tibshirani, R. & Friedman, J., 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Second ed. s.l.:Springer.

Herrera, M., Torgo, L., Izquierdo, J. & Pérez-García, R., 2010. Predictive models for forecasting hourly urban water demand. *Journal of Hydrology*, 387(1-2), pp. 141-150.

Holden, E., Linnerud, K. & Banister, D., 2013. Sustainable development: Our Common Future revisited. *Global Environmental Change*, Volume 26, pp. 130-139.

Hughes, T. C., 1980. Peak Period Design Standards For Small Western U.S. Water Supply Systems. *JAWRA Journal of the American Water Resources Association*, 16(4), pp. 661-667.

Hyndman , R. & Khandakar, Y., 2008. Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software*, Volume 27, pp. 1-22.

Hyndman, R. et al., 2018. *forecast: Forecasting functions for time series and linear models.* s.l.:s.n.

Hyndman, R. J. & Athanasopoulos, G., 2018. *Forecasting: Principles and Practice.* 2nd ed. Melbourne: Otexts.

Hyndman, R. J. & Koehlerb, A. B., 2006. Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), pp. 679-688.

Hyndman, R. J., Wang, E., Lee, A. & Wickramasuriya, S., 2016. *Hierarchical and Grouped Time Series,* s.l.: s.n.

Jain, A., Varshney, A. & Joshi, U., 2001. Short-term water demand forecast modeling at IIT Kanpar using artificial neural networks.. *Water resources management,* 15(5), pp. 291-321.

Jose, V. R. R. & Winkler, R. L., 2008. Simple robust averages of forecasts: Some empirical results. *International Journal of Forecasting*, 24(1), pp. 163-169.

Kamarianakis, Y. & Prastacos, P., 2005. Space–time modeling of traffic flow. *Computers & Geosciences*, Volume 31, pp. 119-133.

Kofinas, D., Mellios, N., Papageorgiou, E. & Laspidou, C., 2014. Urban Water Demand Forecasting For The Island of Skiathos. *Procedia Engineering*, pp. 1023-1034.

Kourentzes , N. & Petropoulos, F., 2018. *MAPA: Multiple Aggregation Prediction Algorithm. R package version 2.0.4.* s.l.:s.n.

Kourentzes, N., 2017. *nnfor: Time Series Forecasting with Neural Networks. R package version 0.9.2.* s.l.:s.n.

Kourentzes, N., Barrow, D. K. & Crone, S. F., 2014. Neural network ensemble operators for time series forecasting. *Expert Systems with Applications*, 41(9), pp. 4235-4244.

Kourentzes, N., Petropoulos, F. & Trapero, J. R., 2014. Improving forecasting by estimating time series structural components across multiple frequencies. *International Journal of Forecasting*, 30(2), pp. 291-302.

Lertpalangsunti, N., Chan, C. W., Mason, R. & Tontiwachwuthikul, P., 1999. A toolset for construction of hybrid intelligent forecasting systems: application for water demand prediction. *Artificial Intelligence in Engineering*, 13(1), pp. 21-42.

Li, W. & Huicheng, Z., 2010. Urban water demand forecasting based on HP filter and fuzzy neural network. *Journal of Hydroinformatics*, 12(2), pp. 172-184.

Lopez-de-Lacalle, J., 2017. *tsoutliers: Detection of Outliers in Time Series. R package version.* s.l.:s.n.

Maidment, D. R. & Miaou, S.-P., 1986. Daily Water Use in Nine Cities. *Water Resources Research*, 22(6), pp. 845-851.

Maidment, D. R., Miaou, S.-P. & Crawford, M. M., 1985. Transfer Function Models of Daily Urban Water Use. *Water Resources Research*, 21(4), pp. 425-432.

Makridakis, S., Spiliotis, E. & Assimakopoulos, V., 2018. The M4 Competition: Results, findings, conclusion and way forward. *International Journal of Forecasting*, 34(4), pp. 802-808.

Ooms, J. et al., 2018. *RMySQL: Database Interface and 'MySQL' Driver for R. R package version 0.10.15.* s.l.:s.n.

Polebitski, A. S. & Palmer, R. N., 2010. Seasonal Residential Water Demand Forecasting for Census Tracts. *Journal of Water Resources Planning and Management*, 136(1), pp. 27-36.

R Core Team, 2017. *R: A language and environment for statistical computing..* s.l.:s.n.

Ralanamahatana, C. A. et al., 2005. Mining Time Series Data. In: s.l.:Springer, Boston, MA, pp. 1069-1103.

Ramanathan, R. & Granger, C. W. J., 1984. Improved methods of combining forecasts. *Journal of Forecasting*, 3(2), pp. 197-204.

Raucher, R. et al., 2008. *Risk and Benefits of Energy Management for Drinking Water Utilities.* Denver: AWWA.

Rijsberman, F. R., 2006. Water scarcity: Fact or fiction?. *Agricultural Water Management,* 80(1–3), pp. 5-22.

Smith, J. A., 1988. A model of daily municipal water use for short-term forecasting. *Water Resources Research*, 24(2), pp. 201-206.

Stock, J. H. & Watson, M. W., 2004. Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, 23(6), pp. 405-430.

Taylor, J., De Menezes, L. & McSharry, P., 2006. A comparison of univariate methods for forecasting electricity demand up to a day ahead. *International Journal of Forecasting,* 22(1), p. 1–16.

Taylor, J. W., 2003. Short-Term Electricity Demand Forecasting Using. *Journal of Operational Research Society,* 54(8), pp. 799-805.

Tuckey, J., 1992. The Future of Data Analysis. In: *Breakthroughs in Statistics.* s.l.:Springer New York, pp. 408-452.

Wang, X., Sun, Y., Song, L. & Mei, C., 2009. An eco-environmental water demand based model for optimising water resources using hybrid genetic simulated annealing algorithms. Part I. Model development. *Journal Of Environmental Management*, 90(8), pp. 2612-2619.

Ward Jr., J. H., 1963. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, 58(301).

Wickramasuriya, S. L., Athanasopoulos, G. & Hyndman , R. J., 2018. Optimal forecast reconciliation for hierarchical and grouped time series through trace minimization. *Journal of the American Statistical Association*.

Winters, P. R., 1960. Forecasting Sales by Exponentially Weighted Moving Averages. *Management Science*, April, pp. 324 - 342.

Zanis, P. et al., 2008. Analysis of an ensemble of present day and future regional climate simulations for Greece. *International Journal Of Climatology*, 29(11), p. 1525–1703.

Zhou, S., McMahon, T., Walton, A. & Lewis, J., 2000. Forecasting daily urban water demand: a case study of Melbourne. *Journal of Hydrology*, 236(3–4), pp. 153-164.

	ETS	ARIMA	Theta	Opt. Theta	MLP	Naive	MAPA	Ensemble Models			
								Mean	Weighted mean	Median	Trimmed mean
Cluster 1-North	1,5964	**0,8976**	1,3457	1,3366	3,2715	2,4146	1,1461	1,2232	7,3989	1,1848	1,1116
Cluster 2-Central	0,4056	**0,2651**	0,7036	0,7028	1,6922	0,8228	0,5305	0,6594	2,7530	0,5635	0,5474
Cluster 3-Inner South / Coastal	0,7596	**0,4394**	1,0209	1,0208	0,8473	1,0490	0,7757	0,7633	1,9882	0,8246	0,8258
Cluster 4-Outer South	0,9448	1,1496	0,8871	0,9069	3,6079	0,8972	**0,8492**	1,3387	1,8935	0,9463	0,9868

Table 1: Out-of-sample MASE errors for each of the four clustered time series forecasts for Methodology No1

	ETS	ARIMA	Theta	Opt. Theta	MLP	Naive	MAPA	Ensemble Models			
								Mean	Weighted mean	Median	Trimmed mean
max	939,28	1227,68	65,88	**58,52**	713,86	878,82	586,83	364,67	105,83	327,14	409,01
average	8,6002	10,7245	2,2771	**2,2409**	8,2818	8,3193	5,9906	4,4153	4,2671	4,1079	4,6854
median	1,1340	**1,0019**	1,1436	1,1374	1,8345	1,1992	1,1079	1,1278	2,6717	1,0837	1,0835
min	0,3510	**0,2254**	0,2662	0,2667	0,3174	0,4065	0,3778	0,2700	0,2554	0,3641	0,3337
IQR	0.2212	0.1594	0.1979	0.2004	0.2766	0.1654	0.3142	**0.1336**	2.459	0.1643	0.1411

Table 2: Summary of out-of-sample MASE errors for each of the 274 zipcode-level time series forecasts for Methodology No2

	ETS	ARIMA	Theta	Opt. Theta	MLP	Naive	MAPA	Ensemble Models			
								Mean	Weighted mean	Median	Trimmed mean
Total water consumption	1,1673	**0,6005**	1,0791	1,0776	2,2575	1,2863	1,1530	1,1820	2,1401	1,0965	1,0973

Table 3: Out-of-sample MASE errors for the one aggregated time series forecasts of Methodology No3

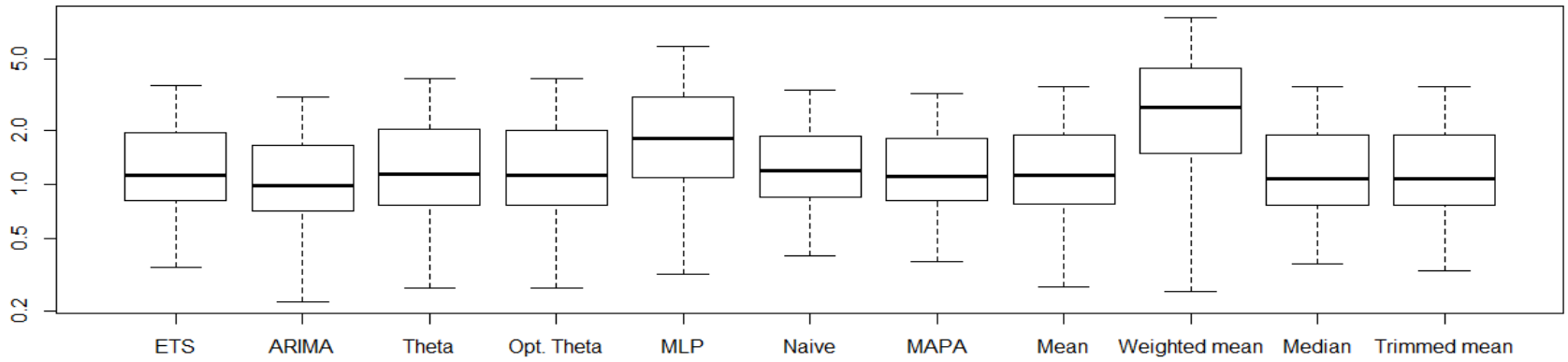

Figure 9: Boxplot of the out-of-sample logarithmic MASE errors for Methodology No2 (omitting large values)

	MASE	sMAPE	MAE	RMSE
Methodology No1	**0,2504**	**0,0093**	**155501,9**	**207147,8**
Methodology No2	0,7854	0,0339	562824,7	622051,3
Methodology No3	0,6005	0,0257	430318,3	510262,2

Table 4: Summary of error metrics for the overall out-of-sample error of the three methodologies. MASE errors for the first methodology for various number of clusters are also displayed

	# clusters = 2	# clusters = 3	# clusters = 4	# clusters = 5	# clusters = 6
Z – score	0,2741	0,2520	0,3023	0,3181	0,4766
# water meters	0,5221	0,2532	**0,2504**	0,3995	0,4892

Table 5: Summary of out-of-sample MASE errors for Methodology 1 for different number of clusters and scaling method

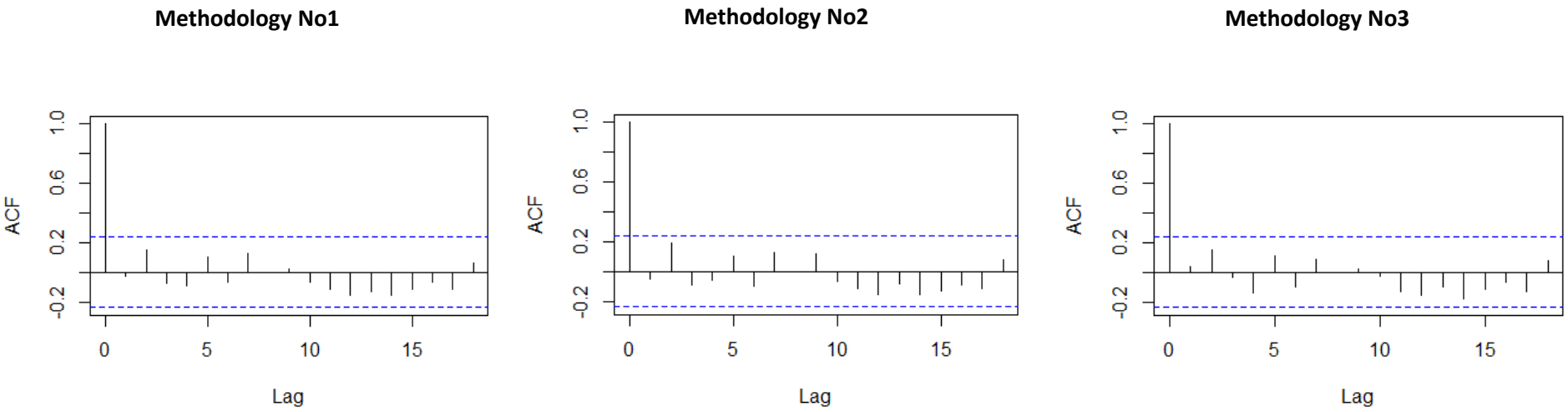

Figure 7: Residual autocorrelation function for each of the three methodologies considered.

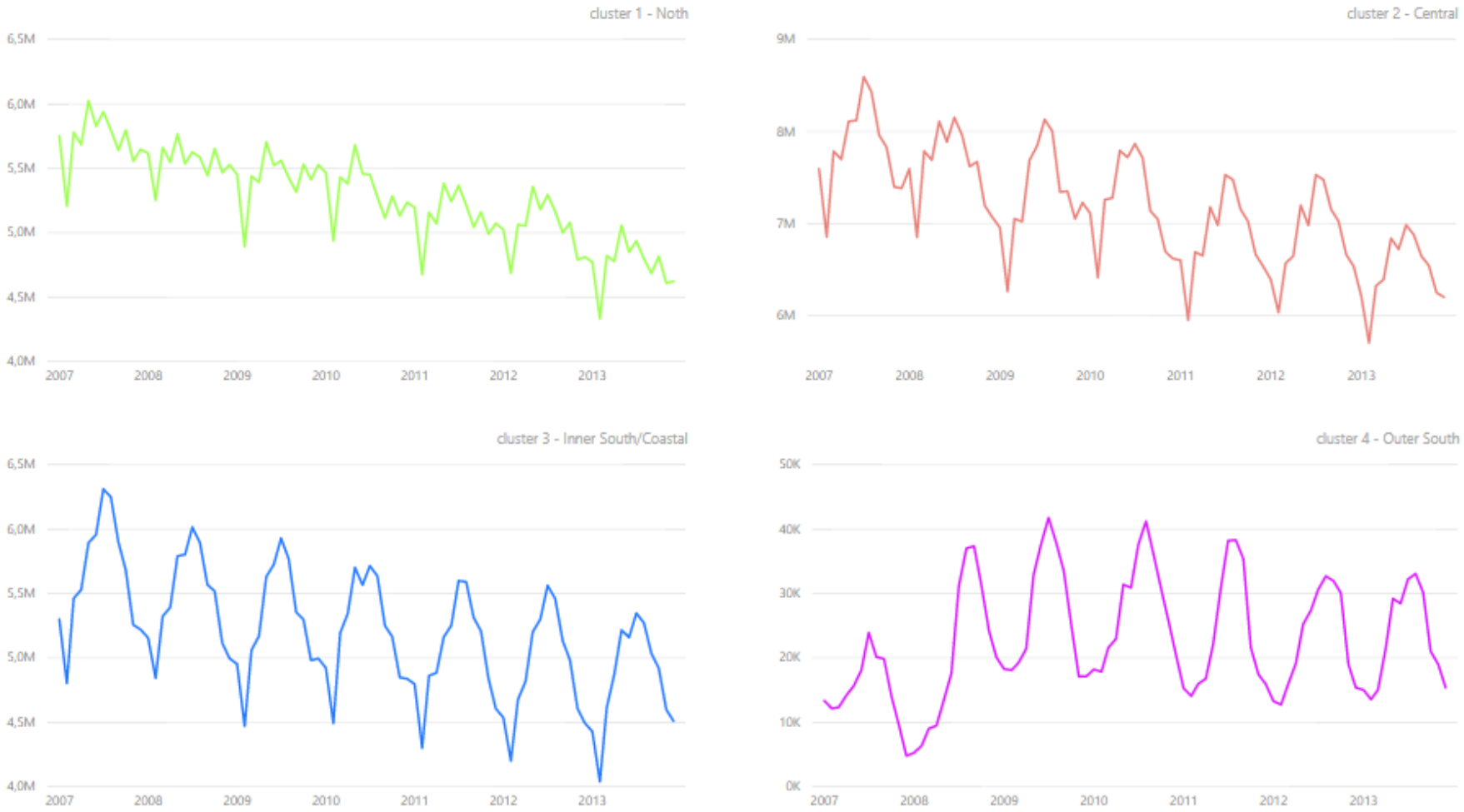

Figure 15: Line graphs of the clustered time series

Figure 3: Example of water consumption patterns existing in our dataset

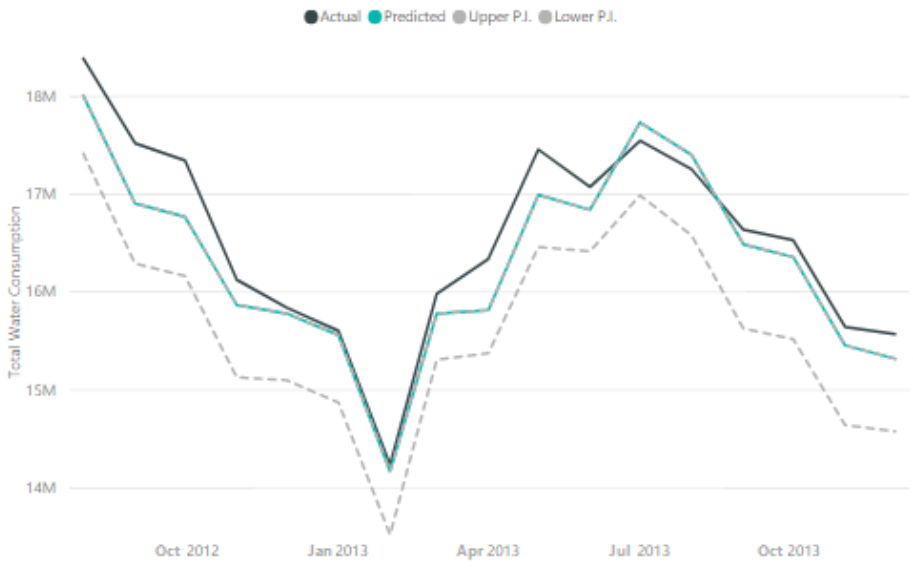

Figure 8: Future estimations for Methodology No1 and actual out-of-sample data. Prediction intervals are highlighted with dashed lines

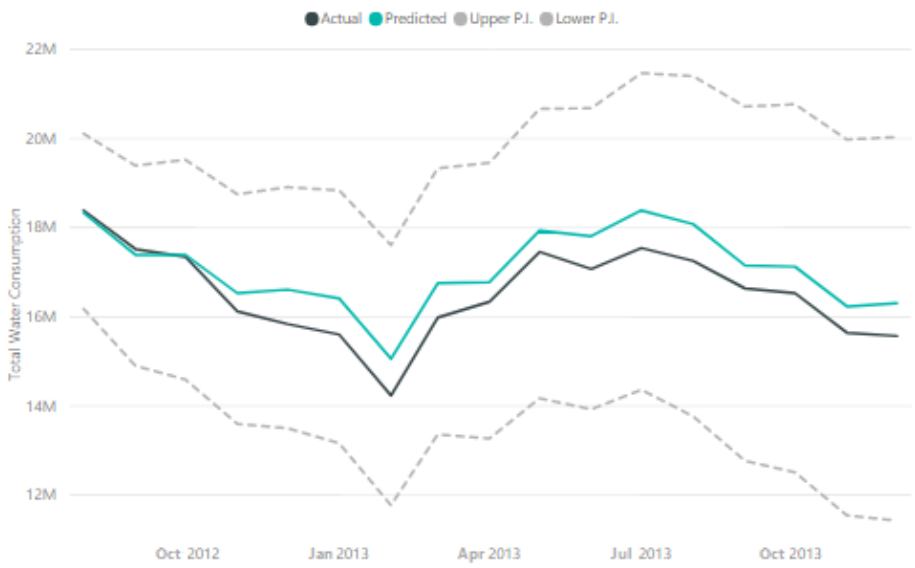

Figure 10: Future estimations for Methodology No2 and actual out-of-sample data. Prediction intervals are highlighted with dashed lines

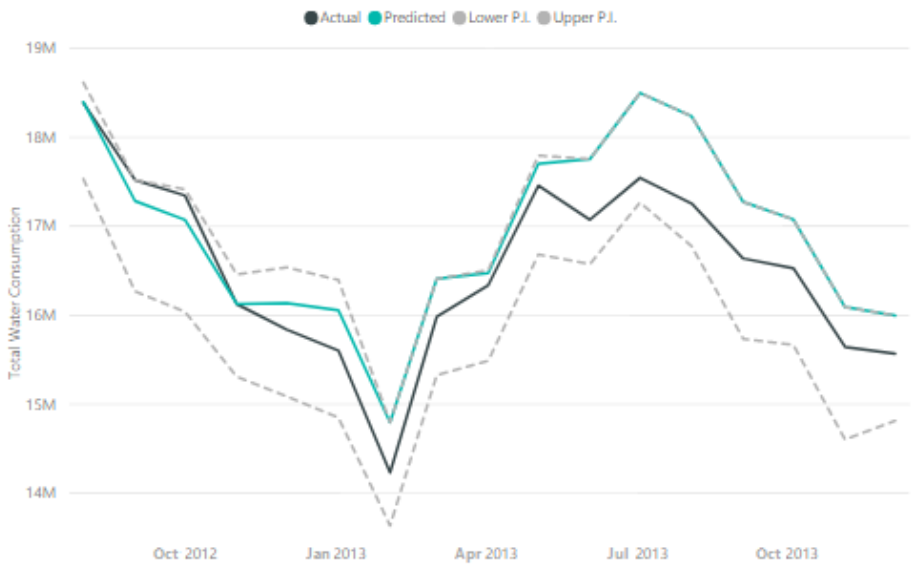

Figure 11: Future estimations for Methodology No3 and actual out-of-sample data. Prediction intervals are highlighted with dashed lines

Figure 2: Annual water demand in the research area and 10-year ahead estimation with prediction intervals

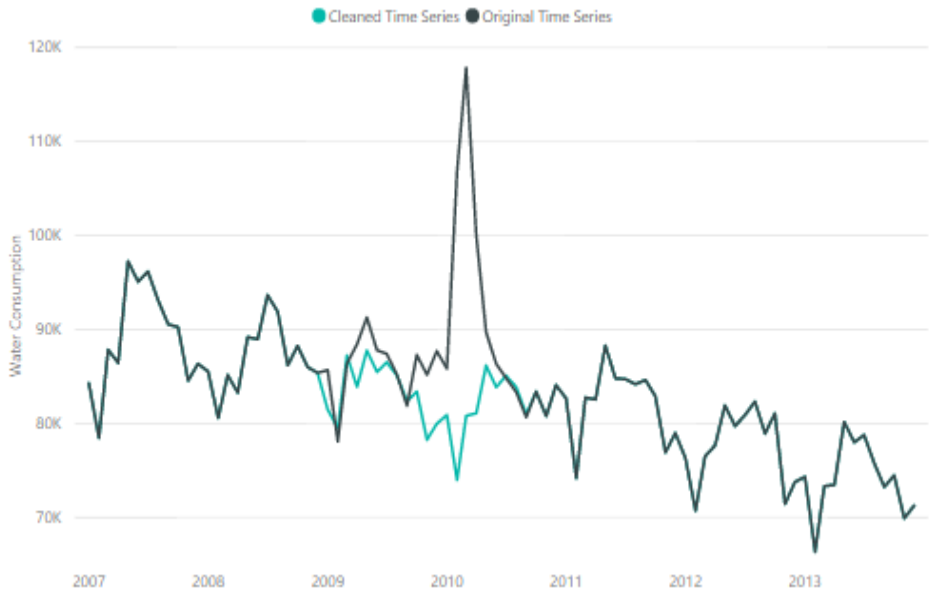

Figure 5: Example of extreme values adjustment

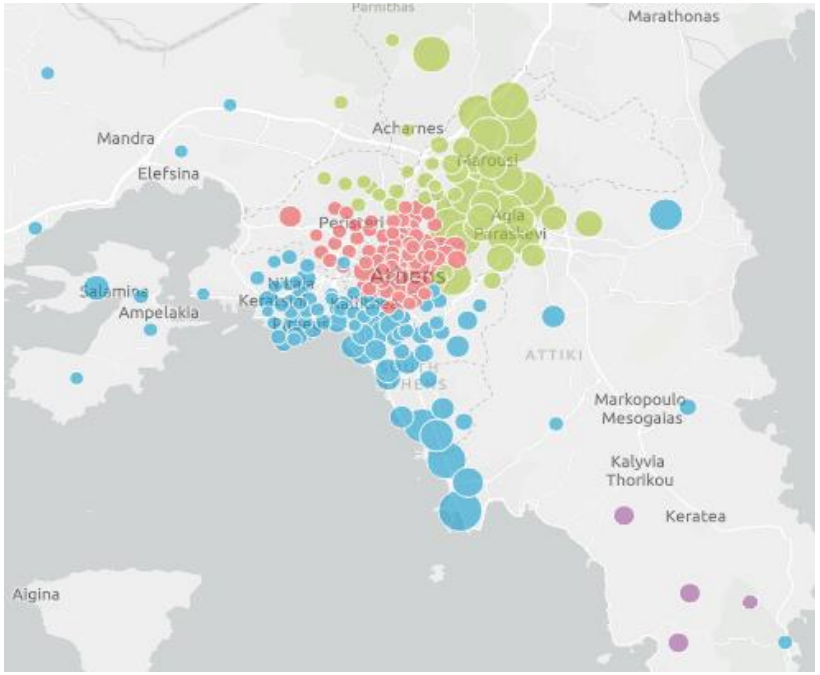

Figure 14: Spatial representation of water consumption forecasts per zipcode and cluster

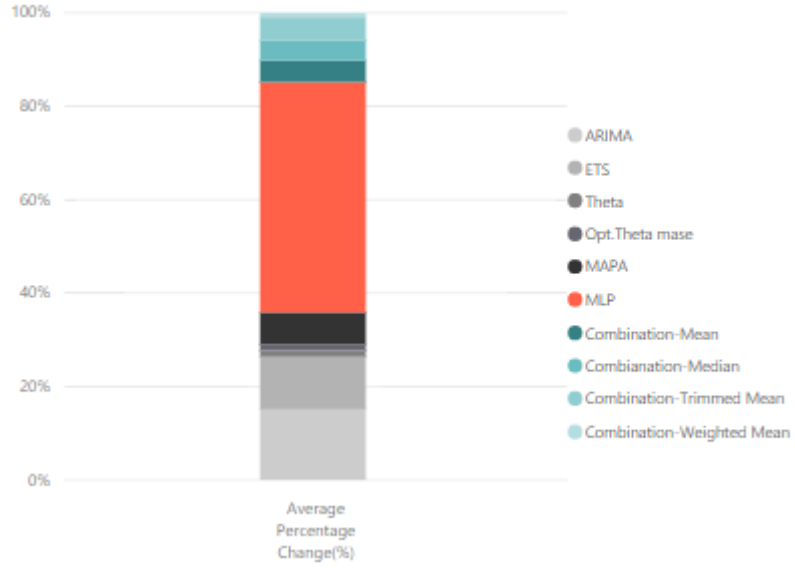

Figure 13: 100% stacked column of the average percentage change between in-sample and out-of-sample MASE error of all methodologies for each forecasting model

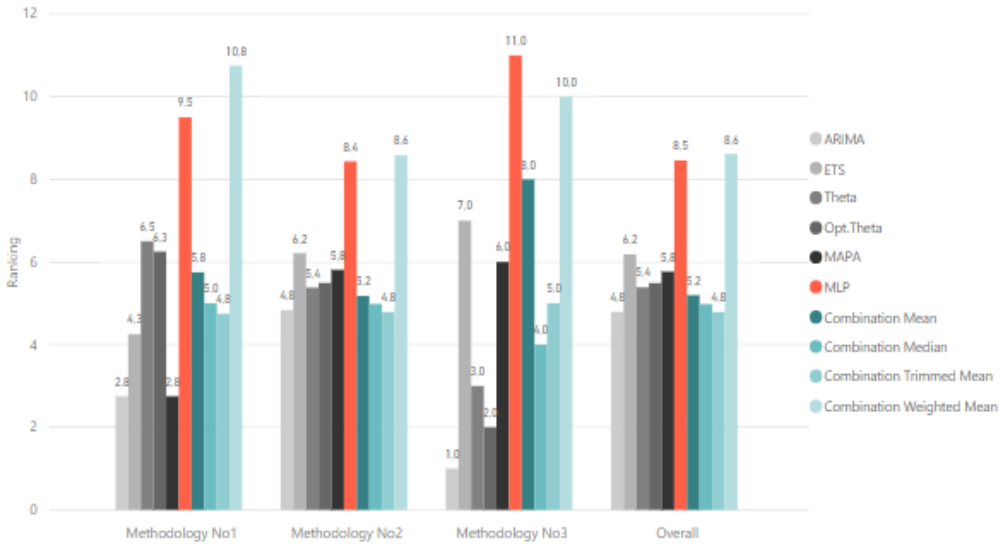

Figure 12: Performance ranking of forecasting models in each methodology and overall

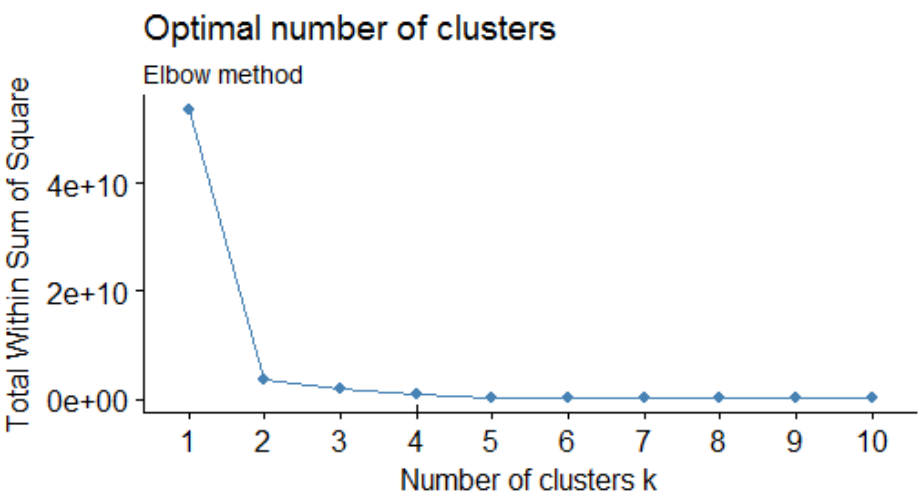

Figure 6: Elbow method for cluster analysis

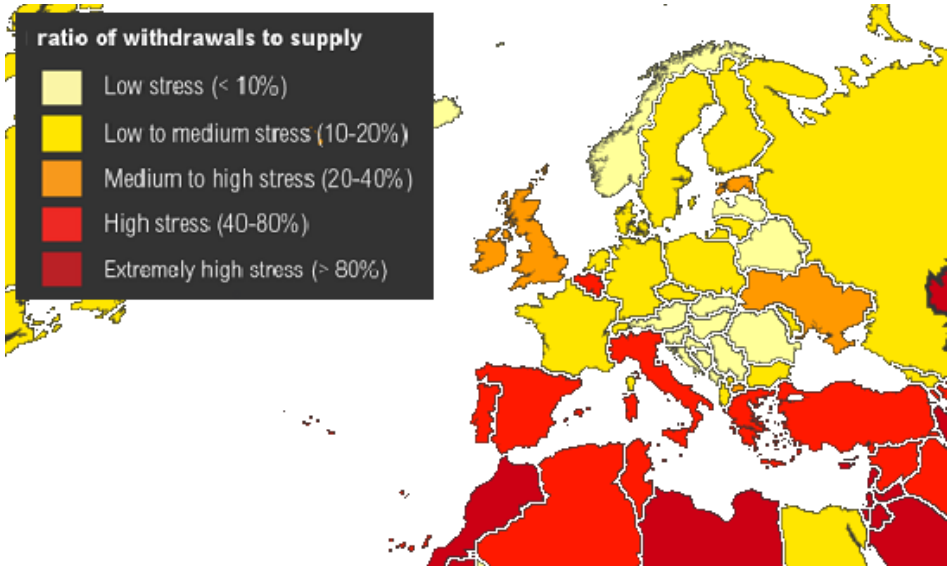

Figure 1: Water stress levels in EU territory (source: World Resources Institute)

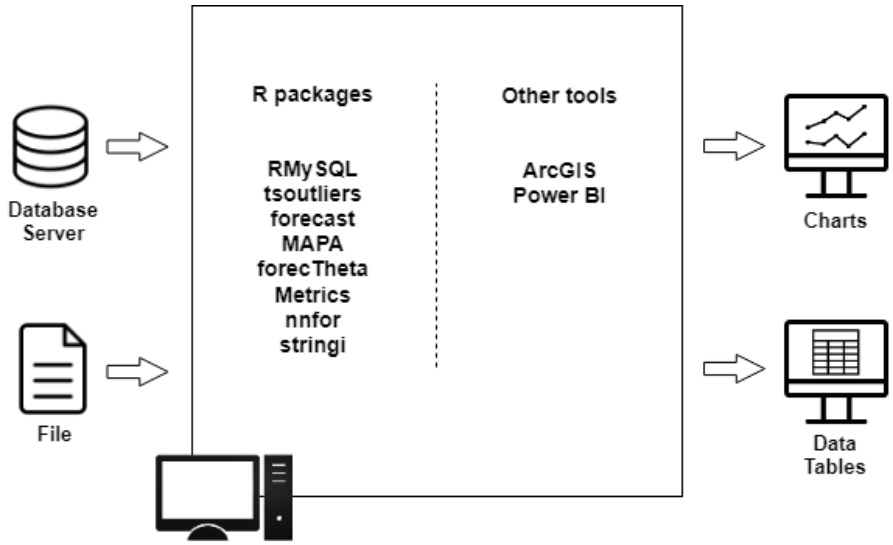

Figure 4: The system architecture