Open Issues and Challenges on Time Series Forecasting for Water Consumption

Pantelis Chronis
IMIS Institute, Athena RC
pchronis@imis.athena-innovation.gr

Giorgos Giannopoulos
IMIS Institute, Athena RC
giann@imis.athena-innovation.gr

Spiros Athanasiou
IMIS Institute, Athena RC
spathan@imis.athena-innovation.gr

ABSTRACT

In this paper we study the problem of water consumption forecasting, an instance of the general time series forecasting problem, that has not been explored adequately. We base our analysis on two types of data: aggregate and individual consumptions measured by Smart Water Meters. We evaluate a series of state of the art forecasting algorithms and showcase that these models are not suitable for every instance of the forecasting problem: while they work effectively on aggregated data that contain strong seasonal patterns, their performance drops dramatically on individual user consumption time series, where such patterns are weaker. To this end, we identify open issues and challenges on the problem and, also, demonstrate that a simpler model we propose can outperform several of the aforementioned algorithms, although still needing significant improvements.

1. INTRODUCTION

Time series forecasting methods aim at predicting the progress of a time varying phenomenon in the future. The phenomenon is described by a variable or a set of variables, measurements of which are taken in different points in time. In our case, the observed phenomenon is the consumption of water. Given observations of the consumption in the past we wish to predict the future consumption.

The literature for water consumption forecasting is not very extensive [1, 8]. More work has been done on the relevant field of electricity consumption forecasting, which, however, presents significant differences [9, 10, 14, 11, 19, 20]. First of all, energy metering tools can provide measurements of much higher accuracy and granularity compared to water monitoring devices. Also, energy consumption patterns are much more canonical and, sometimes even fixed, than the respective water patterns. For example, it is straightforward to calculate the energy consumption of most household devices through their technical characteristics, while this is not the case for the showers even of the same user.

Further, on both fields, the major focus of approaches in the literature is on the problem of predicting the aggregate consumption of a large population (e.g a city). However there are cases where there is need for prediction per user or even per specific activity (e.g. showering). Water sustainability and saving efforts have lately focused on increasing user-consumer awareness by devising interventions that aim at educating users about their *individual* consumption behavior and guiding them into reducing their consumptions. To do so, R&D projects such as DAIAD[1] and smartH2O[2] research on machine learning techniques for short-term water consumption forecasting and pattern recognition and on intervention methods that exploit these techniques in order to inform consumers and induce behavioral changes. Moreover, companies such as Amphiro[3] invest on water monitoring devices that are installed on household faucets and measure real-time water consumption, providing online statistics and alerts to the users. These interventions and alerts require that individual short-term consumption is forecasted as accurately as possible and in real time, so that it can be compared with the upcoming actual consumptions.

In this work we study the problem of short-term time series forecasting, specializing on water consumption data. We consider two different scenarios/data cases: (a) *Individual consumption data*, where *hourly* measurements are produced by Smart Water Meters (SWM) and each time series is handled separately. (b) *Aggregate consumption data*, where *hourly* measurements are produced by SWM and aggregated into a single time series. We evaluate several state of the art algorithms, including Regression, ARIMA, Exponential Smoothing, and we demonstrate the significant variations in their performance in each case. Specifically, we point out the poor performance of all models in case (a) that is non-aggregate consumption time series. Finally, we demonstrate that a simpler model can outperform several of the aforementioned algorithms. However, since still the prediction performance remains rather poor, we analyse the results of all evaluated algorithms and we discuss directions for improving the precision of the prediction.

2. RELATED WORK

Extensive work is done in the field of time series analysis [15], that includes forecasting, pattern recognition, classification, discretization and event detection. Several works handle the specific cases of water/energy consumption [8, 9, 10, 14, 11, 19, 20]. There are also a few projects studying the field of water timeseries forecasting and analysis [18, 17].

For forecasting large scale aggregate consumption similar approaches are adopted in most works, which focus on the strong seasonal patterns of the data. [1] apply double seasonal ARIMA, Holt Winters Exponential Smoothing and Garch models in 6 years of aggregate water consumption in Spain. In [2], double seasonal Holt Winters model and double seasonal ARIMA with daily and weekly

(c) 2016, Copyright is with the authors. Published in the Workshop Proceedings of the EDBT/ICDT 2016 Joint Conference (March 15, 2016, Bordeaux, France) on CEUR-WS.org (ISSN 1613-0073). Distribution of this paper is permitted under the terms of the Creative Commons license CC-by-nc-nd 4.0

[1] http://daiad.eu/
[2] http://www.smarth2o-fp7.eu/
[3] http://amphiro.com/

cycles, a simple feedforward Neural Network (NN) and PCA are applied on aggregate hourly electricity demand data. The best performance comes from the Holt Winters model. In [3] the authors propose triple seasonal Holt Winters and ARIMA models to capture the daily, weekly and yearly cycles and demonstrate improvement over the double seasonal models.

For the problem of forecasting on smaller scale of aggregation there exists more variability in the approaches. [4] compare Linear Regression (LR), NNs and Support Vector Regression (SVR), at different scales of aggregation, on hourly electricity demand data. They find that SVR outperforms the other models at higher aggregation but the LR outperforms the SVR on lower aggregation (< 16 households). However, the prediction errors for individual forecasting are much higher than the ones on aggregated data. In [5] the authors compare three NNs with three ARIMA based algorithms on aggregate electricity demand data from 90 and 230 households, with the ARIMA achieving best average performance but with the NNs having better performance at specific parts of the day. [7] propose a Kalman Filter on a single household electricity consumption data. In [6] the SVR and NN algorithms are applied in single household data and are found to have very similar performance.

3. DATASETS

3.1 Per user SWM hourly consumption data

The first dataset consists of time series produced by SWM, installed in 121 households, that provide hourly measurements of the water consumption for a period of one year. In some cases, due to malfunctions, the next measurement is taken sooner or later than one hour. This causes two issues. First, most algorithms require as input fixed time intervals between measurements. Second, in the case we want to aggregate all the time series, we need them to be aligned, i.e. to contain measurements for the same time intervals. Thus, we align the time series using linear interpolation and use the aligned dataset for the rest of our analysis.

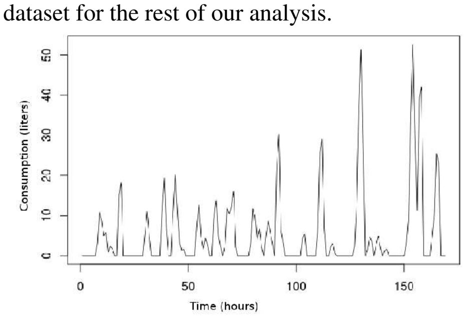

Figure 1: Water Consumption - Per user dataset

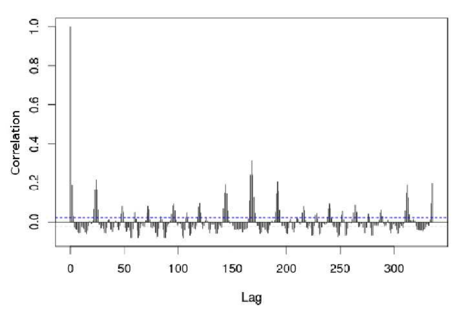

Figure 2: Correlation - Per user dataset

Figure 1 presents a sample time series of the second dataset, where there appear to be some coarse-grained patterns: the consumption is low at night and is followed by some spikes in consumption, usually three, during the day. There is also significant noise that affects both the height and the position of the spikes in time. As shown in the correlogram (Figure 2) there is a relatively large correlation between each measurement and the previous one. Also, there is a clear sign of seasonal structure. The seasonality is daily and weekly, with the weekly component being more important, as seen by the increased values around hours 24 (daily) and 168 (weekly). The correlation is relatively small but well above the statistically significant level, because of the large size of the data ($\sim$8000 measurements). Of course, non-linear relationships or relationships including more than one previous values (that cannot be captured in Figure 2) might exist.

3.2 Aggregate SWM hourly consumption data

The second dataset consists of a single time series representing the aggregate consumption of 121 households, in hourly measurements for a period of one year. It is created by summing the 121 time series of the first dataset. This time series tends to become more regular as more time series are summed, because the noise is cancelled out. Figures 3 and 4 show that the second dataset is particularly regular. It presents a very clear daily seasonal pattern and a less clear weekly pattern. This type of data is the one usually addressed in water/energy consumption forecasting literature.

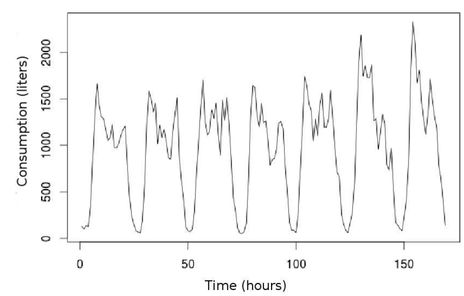

Figure 3: Water Consumption - Aggregate dataset

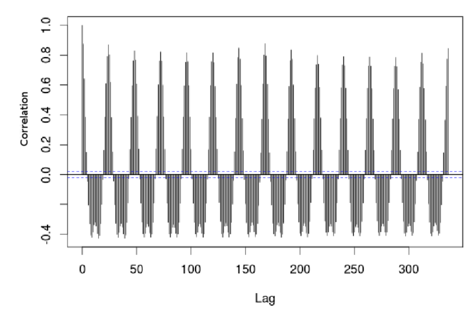

Figure 4: Correlation - Aggregate dataset

4. ALGORITHMS

This section presents the forecasting algorithms that were examined in this work. First, we briefly describe a set of state of the art algorithms and discuss their expected performance on the specific problem, based on the logic behind the algorithms and their empirical performance in the literature. We also present a simpler, first-cut prediction method based on time series discetization and probabilistic prediction of next values, given previous ones. The model is based less on seasonality and more on the premise that the latest consumption values influence the upcoming ones.

4.1 State of the art prediction algorithms

We evaluated the following state of the art algorithms, used in the literature for short term water/energy time series forecasting.

- Linear Regression (LAR). The idea of LAR is to model the values to be predicted as a linear function of the previous values of the time series [13].
- Support Vector Regression (SVR). The algorithm aims at finding the simplest hyperplane so that all the training points (previous time series values) stay within a specified distance from it [13].
- Autoregressive Integrated Moving Average (ARIMA). ARIMA is a class of algorithms for univariate time series forecasting that model the current value as a linear combination of the previous values and the errors of the previous predictions [3].
- Exponential Smoothing (ES). ES [3] is another class of algorithms used for univariate time series forecasting that composes the time series from three parts: level, trend and seasonality.
- Artificial Neural Network (aNN). aNNs [13] are composed of layers of nodes, at each of which the output is a linear combination of the inputs passed through an activation function. Each layer feeds the next one and the last layer gives the dependent variable-prediction.

LAR, which is conceptually the simplest of the algorithms is expected to work well in the aggregate SWM dataset, where the linear correlations are large. However, it is not expected to have good performance in the other datasets where such correlations are not present. SVR has the advantage of using suitable kernels in order to capture non-linear patterns. However, the complexity of most widely used kernels is limited to relatively simple analytical functions; defining an appropriate kernel for the respective datasets is a problem of its own. The ARIMA and ES algorithms are mostly used in the literature of water/energy consumption forecasting on aggregate datasets. They both capture the seasonal patterns and perform well in very regular datasets. ARIMA has the limitations imposed by the model itself, i.e. the data must be generated from a linear combination of the previous values and the previous errors, which may not hold in general. ES can only follow changes in the time series after they have happened, because it calculates the components for its prediction from the previous values, so it is designed for time series that are consistent in terms of level trend and seasonality. Both those models are expected to work well on the SWM aggregate and perhaps the per-user hourly dataset. The aNNs, similarly to SVR, can be made arbitrarily complex. They are expected to perform well in the aggregate SWM dataset but, because of their black-box nature, it is hard to predict their performance on the per-user dataset.

4.2 Proposed methods

4.2.1 *Formal Description of the data.*

A time series y is a vector of values (measurements). Value i of y is $y_i, 0 \leq i < n$. The algorithms we propose require discretized data so we discretize our time series values. Given a set of symbols $S, s_j \in S, |S| = k$ and a function $f : \mathbf{R} \rightarrow S$, which corresponds to the chosen discretization method, we transform the continuous time series y to the discrete time series $d, d_i = f(y_i)$. Also, we assign a continuous value to each symbol which is the average of all the values that map to symbol s_j in the training set. This value is used when we want to obtain a continuous prediction from the algorithms that provide discrete outputs. The discretization methods we examined are the following: (a) using equal ranges, (b) using equiprobable ranges and (c) using k-means clustering.

In the method of equal ranges, the axis of consumption volume is divided into equal ranges. In the method of equiprobable, ranges the axis of consumption is divided into ranges such that an equal number of points falls in each range. In the clustering method, the k-means clustering algorithm is used to obtain k centroids. For each centroid, a range is created that includes those points that are closer to the given centroid than any other. After obtaining the ranges, a symbol is assigned to each range and the time series is discretized by substituting each measurement with the symbol of the range it falls into.

4.2.2 *Sequential Regression (SR).*

In order to capture complex non-linear behaviour we propose a simple non-parametric algorithm that first discretizes the values and then searches for the most probable discrete subsequence.

The problem is defined as follows: given the classes of the previous b consumption events, predict the class of the next consumption event. To calculate this, we count the occurrence of each subsequence of size $b + 1$. More formally, we define a mapping data structure C. The elements of C count the occurrences of subsequences; for example, $C[0, 2, 1] = 4$ means that the subsequence $0, 2, 1$ has occurred 4 times. We count all the subsequences of size $b + 1$ by scanning the time series.

We denote the concatenation of symbol s_j to sequence d as $d + s_j$. Our prediction for $d_i, i > b$ is:

$$d'_i = argmax_{s_j} C(d_{i-b:i-1} + s_j)$$

For subsequences that have never been observed before we give as prediction the most probable symbol of the time series.

4.2.3 *Sequential Regression 2 (SR2).*

We also attempt to extend the SR algorithm by modelling the distribution of the time interval between two subsequent occurrences of each symbol (consumption range). The intuition behind this is that events of each class have an expected frequency (e.g. a user takes a bath once every 3 days). Thus, we take into account the probability of each subsequence, *as well as* the probability of the recurrence time.

4.2.4 *Classed Linear Regression (CLR).*

Based on the discretization assumption, we also adjust LAR algorithm to utilize discretized time series input. We perform linear regression considering the classes-ranges of the previous b consumptions as explanatory variables. To indicate the classes we use indicative variable vectors, i.e. vectors where each dimension corresponds to a class. The dimension of the current class is 1 and all others are 0.

5. EVALUATION

Next, we present the detailed configurations we used on the evaluated algorithms, as well as their performance in terms of prediction precision. For the SWM per user dataset, that consists of multiple time series, the reported values are the average ones on all the examined time series. We note that we chose a train/test partition of $70\%/30\%$ in all cases and we did not use a validation set, since our aim was not to perform a strict comparison of the methods, rather than to highlight shortcomings of current state of the art on specific water consumption forecasting settings and identify possible research directions towards improving the forecasting performance on such settings.

5.1 Evaluation Metrics and Baselines

Mean Absolute Percentage Error (MAPE) is a widely used measure for assessing the prediction performance of forecasting algorithms, which we adopt in our evaluation. However, in the case of per user hourly data it is very common that there are zero measurements, so MAPE cannot be defined. Due to this, for the per user data, we use a variation, Normalized Mean Absolute Error (NMAE), which divides, for each measurement, the absolute error of the prediction with the average value of the time series. Given a time series y, with y_i a measured value and y'_i the respective predicted value in time i, NMAE is defined as:

$$NMAE = \frac{1}{n}\frac{1}{mean(y)}\sum_{i=1}^{n}|y_i - y'_i|$$

In order to evaluate the performance of the algorithms we also consider three baseline algorithms. The first baseline (*BR0*) gives as prediction the median value of the training set. The median is chosen instead of the mean, because it minimizes the MAPE metric. The second one (*BR1*) gives as prediction the previous value of the time series, so it takes advantage of possible similarity of successive observations. The third baseline (*BR2*) gives as prediction the value that occurred 24 observations ago, essentially considering that the same hours of each day correspond to exactly the same consumption values.

5.2 Per-user hourly consumption data

In the case of per user SWM consumptions, we considered only the b previous values of the time series, since no metadata were available. We experimented with various b, $1 \leq b \leq 168$. For LAR and SVR, the best performance was for $b = 24$, which corresponds to identifying daily periodicity in consumptions. ARIMA was used with configuration $(3, 0, 3)(2, 0, 2)_{24}$. ES was used with two seasonal cycles, one of 24 (daily) and one of 168 (weekly) hours. The aNN was also used with $b = 24$ in order to capture the daily pattern. SR0 achieves the best performance for $b = 1$. The reason for this is probably the fact that, for large b the model of SR0 becomes sparse, i.e. there are not enough observations for each combination of previous values.

For SR0, SR2 and CLR the number of classes corresponding to the best performance were 18, 3, 15. For SVR we set $C = 0.1$, $\epsilon = 0.0001$ and the linear kernel was used. For the aNN two hidden layers of size 10 and 5 were used and the convergence threshold was set to 0.1.

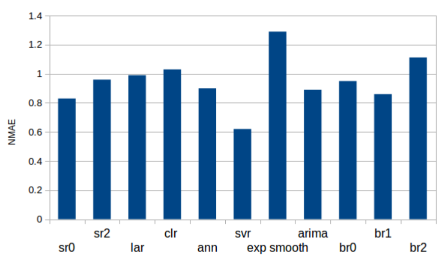

Figure 5: Forecasting NMAE on per user hourly data

In Figure 5, we can see the performance of the algorithms in terms of NMAE. The error is above 60% for all algorithms which constitutes a poor performance. This can be partly justified by the fact that the mean value of the time series, which is at the denominator of the metric, is low and that results to big values for the relative error. Nevertheless, the performance compared to the baselines is disappointing. The best performance comes from SVR with 62%. The performance of SR0 is 83% which is 3% better than the naive baseline (BR1). All other algorithms have worse performance than BR1. This probably happens because the time series do not follow some simple analytical model. However, there may be patterns that can be captured by a more data-driven analysis, like that of SR0.

The best performance of SVR is achieved by using the linear kernel. The benefit of SVR is its ability to avoid overfitting and provide good generalisation performance. However, its performance could possibly be improved by analysing the residual errors (i.e. the actual observation minus the predicted value) for non-linear patterns, possibly using our SR0 algorithm or some other non-parametric algorithm. In Figure 6, we see that (for an exemplary time series) there is significant linear autocorrelation left in the residual errors of the second best performing algorithm, SR0. This can be attributed to the fact that the algorithm is run with b=1 and correlations of previous values are not used.

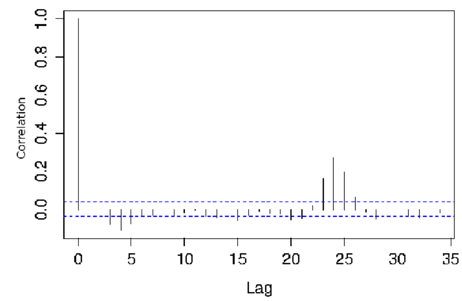

Figure 6: Autocorrelation of the residual errors of SR0 on a single time series on SWM per user data

All the above indicate a possible direction for improvement of forecasting on the specific kind of data. Extending the SR0 algorithm to capture the seasonality relationships and also exploiting the linear correlation in the residuals can lead to a better performing algorithm. Moreover, trying to identify specific recurring patterns by inspecting the time series seems to be more suitable for this kind of data, than trying to identify an analytical model.

Our ongoing work on improving the prediction accuracy on individual consumption time series involves attempting to extract features in an unsupervised way, like in [16]. This can be coupled with the problem of disaggregation which has been studied for the case of energy consumption but is entirely open for the case of water consumption. What we intend to do is extract features, such as reoccurring patterns of the time series and consumption events, and then use them as input in a prediction algorithm. A more direct modification that could be tested is to treat different parts of the time series independently (e.g. the nights), since the consumption then has different behaviour as described in [5]. Relaxing the time constraints of the prediction, i.e disregarding small time shifts in prediction, can also be useful on this dataset.

5.3 Aggregate SWM hourly consumption data

Again, in the case of aggregate SWM consumptions, we considered only the previous consumption values as features. The number of previous values b that were eventually selected were 1, 1, 168, 24, 24 and 24 for SR0, SR2, LAR, CLR, SVR and aNN respectively. The configuration of ARIMA was $(3, 0, 3)(2, 0, 2)_{24}$ and the cycles for exponential smoothing were 24 and 168. The number of classes were 18, 3, 24 for SR0, SR2 and CLR. For SVR we set $C = 0.5$, $\epsilon = 0.001$ and the polynomial kernel was used. For the ANN two hidden layers of size 10 and 5 were used and the convergence threshold was set to 0.1.

In Figure 7 we can see the performance of the algorithms on the aggregate SWM hourly data. The performance is much better than in the other setting. This was expected because of the very strong seasonal patterns of the time series. The best performance comes from the aNN. SVR abd LAR are the second and third best approaches, respectively. This is expected because of the high linear autocorrelation of the time series. However, we would expect more problem-specific algorithms like ARIMA and ES to outperform, e.g., LAR which did not happen. This could be attributed to the relatively high irregularity of the time series, which is explained by the fact that it aggregates only 121 individual consumption time series. In the literature, the errors reported for the same algorithms on energy consumption datasets for one step ahead prediction are even lower ($< 5\%$). The datasets in those cases are aggregates of much larger samples and, thus, even more regular. Also the time series of energy consumption tend to be more canonical.

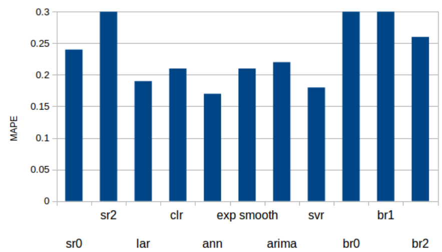

Figure 7: Forecasting MAPE on aggregate hourly data

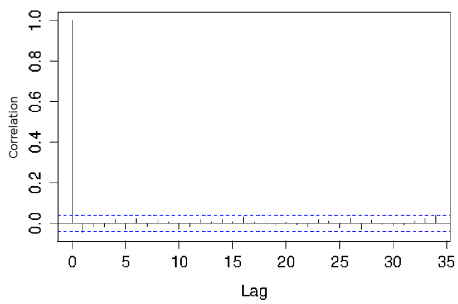

Figure 8: Autocorrelation of the residual errors of the aNN on SWM aggregate data

The path that we are considering for future research on this kind of data is to attempt prediction on the residual errors of several of the examined algorithms, and of the aNN specifically, which has the higher performance, and use it to improve the prediction precision. The algorithm should, of course, take into account non-linear relations because, as we see in Figure 8, there is no linear autocorrelation left in the residuals. In this direction, we are currently attempting to apply the Nearest Neighbour algorithm on the residual errors. More specifically, we search for instances with similar previous errors to those observed at the current time and try to find whether the next error follows some pattern.

6. CONCLUSIONS

In this work, we investigated the performance of several state of the art forecasting algorithms on time series of two different water consumption datasets. The results indicated that the performance of each algorithm depends on the dataset. Another important issue is that the performance of all algorithms becomes poor as the scale of aggregation decreases from 121 households, to 1 household. Driven by this fact, we proposed a first-cut approach which shows relatively better results in the latter case compared to most soa algorithms and presents an interesting direction for future research in the case of individual consumption forecasting.

7. ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Union Seventh Framework Programme - Collaborative Project - DAIAD (http://www.daiad.eu/) under grant agreement No. 619186.

8. REFERENCES

[1] J. Caiado, *Forecasting water consumption in Spain using univariate time series models*, Proceedings of IEEE Spanish Computational Intelligence Society, 2007.

[2] J. W. Taylor, L. M. Menezes, P. E. McSharry, *A comparison of Univariate Methods for Forecasting Electricity Demand Up to a Day Ahead*, International Journal of Forecasting, 2006.

[3] J. Taylor, *Triple Seasonal Methods for Short-Term Electricity Demand Forecasting*, European Journal of Operational Research, 2010.

[4] S. Humeau, T. K. Wijaya, M. Vasirani, K. Aberer, *Electricity Load Forecasting for Reisdential Customers: Exploiting Aggregation an Correlation between Households*, IFIP, 2013.

[5] A. Marinescu, C. Harris, I. Dusparic, S. Clarke, V. Cahill, *Residential Electrical Demand Forecasting in Very Small Scale: An Evaluation of Forecasting Methods*, 2nd International Workshop on Software Engineering Challenges for the Smart Grid, 2013.

[6] K. Gajowniczek, T. Zabkowski, *Short term electricity forecasting using individual smart meter data*, 18th International Conference on Knowledge-Based and Intelligent Information and Engineering Systems, 2014.

[7] M. Ghofrani, M. Hassanzadeh, M. Etezadi-Amoli, M. S. Fadali, *Smart Meter Based Short-Term Load Forecasting for Residential Customers*, N. A. Power Symposium, 2011.

[8] S. Alvisi, M. Franchini, A. Marinelli, *A short-term, pattern-based model for water-demand forecasting*, Journal of Hydroinformatics, 2007.

[9] F. Martinez Alvarez, A. Troncoso, J. Riquelme, J. Aguilar-Ruiz, *Energy Time Series Forecasting Based on Pattern Sequence Similarity*, TKDE, 2011.

[10] R. Weron, A. Misiorek, *Forecasting spot electricity prices: A comparison of parametric and semiparametric time series models*, TKDE, 2011.

[11] A. Livera, R. Hyndman and R. Snyder, *Forecasting time series with complex seasonal patterns using exponential smoothing*, J. of the American Statistical Association, 2011.

[12] R. Kato, T. Shiohama, *Model and Variable Selection Procedures for Semiparametric Time Series Regression*, Journal of Probability and Statistics, 2009.

[13] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer 2007.

[14] R. Weron, *Modeling and Forecasting Electricity Loads and Prices*, Wiley 2006.

[15] J. Lin, S. Williamson, K. Borne, D. DeBarr, *Pattern Recognition in Time Series*, Advances in Machine Learning and Data Mining for Astronomy, Chapman and Hall, 2012.

[16] M. LÃď'ngkvist, L. Karlsson, A. Loutfi *A review of unsupervised feature learning and deep learning for time-series modeling*, Pattern Recognition Letters, 2014.

[17] *DAIAD Deliverable: State of the Art Report*, available at http://www.daiad.eu.

[18] *ICeWater Deliverable: State of the Art Analysis*, available at http://www.icewater-project.eu.

[19] B. Chen, M. Chang, and C. Lin *Load Forecasting Using Support Vector Machines: A Study on EUNITE Competition 2001*, IEEE Transactions on Power Systems, 2001.

[20] J. Chen, W. Wang, C. Huang *Analysis of an adaptive time-series autoregressive moving-average(ARMA) model for short-term load forecasting*, Electric Power Systems Research, 1995.