

Forecasting Domestic Water Consumption Using Bayesian Model

Wojciech Froelich and Ewa Magiera

Abstract In this paper, we address the problem of forecasting domestic water consumption. A specific feature of the forecasted time series is that water consumption occurs at random time steps. This substantially limits the application of the standard state-of-the-art forecasting methods. The other existing forecasting models dedicated to predicting water consumption in households rely on data collected from questionnaires or diaries, requiring additional effort for gathering data. To overcome those limitations, we propose in this paper a Bayesian model to be applied for the forecasting of the domestic water consumption time series. The proposed theoretical approach has been tested using real-world data gathered from an anonymous household.

Keywords Forecasting time series · Domestic water consumption · Bayesian networks

1 Introduction

Forecasting water consumption at the household level is a problem that can be addressed from different perspectives. First, on the basis of the forecasted water consumption, a corresponding system of forecasted bills/charges can be applied. In this case, the resident pays her bills calculated on the basis of forecasted water consumption. This results in more accurate forecasting, less frequent reading of water meters, and the charges paid by the consumer are fair. For this purpose, the monthly time scale is usually applied.

W. Froelich (✉) · E. Magiera
Institute of Computer Science, University of Silesia, Sosnowiec, Poland
e-mail: wojciech.froelich@us.edu.pl

E. Magiera
e-mail: ewa.magiera@us.edu.pl

From the perspective of water conservation, the forecasted values of water consumption can be compared with actual values. In the cases when the actual value is lower than the forecasted value, the water conservation process can be recognized; otherwise, the consumer may be warned of excessive water usage. For the purposes of water conservation, the considered time scale in which forecasts are made should be detailed. The households should be equipped with smart water meters that are enabled to perform readings at daily, hourly, or even more detailed temporal scales. Furthermore, to assess water conservation with respect to the type of consumption (e.g., showering, dish-washing, drinking or other activities), water metering should be made separately for each of the corresponding areas of consumption. The role of water efficient technology such as smart meters in reducing water consumption throughout the day is crucial [4].

Forecasting of water consumption is subject to two variables as described in the literature: (a) the analysis of domestic water consumption [5, 15] and (b) the forecasting of time series [7, 8].

A detailed analysis of water consumption at households was made in [16] and more recently in [17]. Water consumption in the kitchen was investigated [20], where water withdrawal was related to different take events, e.g. drinking, cooking, etc. In [4] the patterns of water consumption in the household were analyzed on an hourly scale. A household water saving model based on Bayesian network was proposed [13]. In [9] a general architecture of the decision support system based on forecasting water consumption was described. More recently, the potential of water conservation during dish washing has been investigated in a dissertation [10].

Knowledge about physical properties of households, e.g., number of rooms, type of house/flat, garden, was used for the analysis of residential water consumption [5]. Also in [5] the problem of missing data in water meter readings was investigated. A practical decision support system for water conservation at household has been developed by IBM. On the basis of the available report [12] it can be noted that one of the applied solutions was based on the recognition of trend of water consumption. An overview of the literature on the analysis of domestic water consumption was made [5, 15].

The problem of forecasting domestic water consumption was investigated in several publications. In [11] the daily and hourly patterns of water consumption in households were analyzed. A general review of selected forecasting models used at the urban and household levels was made in [11]. Recently, it has been shown that the bottom-up approach to the forecasting of urban water demand is superior to the top-down approach based on bulk water meters [2].

In particular, Bayesian networks have been used for the forecasting of time series [1, 22]. Recently, an overview of Bayesian forecasting approaches has been reported [3]. Our previous studies [6, 18] applied Bayesian models for the forecasting of water demand at the city level. Due to the different characteristics of the data and the domain of application, those publications are not directly related to the approach presented in this paper.

Most of the existing forecasting models reported in the literature rely on information gathered from households using questionnaires, including information about household occupancy, household income, or even from daily diaries prepared by residents. Significant additional effort is required to gather such data and these data are usually not anonymous. By contrast, our approach utilizes only the near real-time readings of water meters.

For the purpose of forecasting domestic water consumption, we design a structure of the Bayesian network. In addition, we investigate a related problem, the dependence between the equal-length discretization technique and its influence on the accuracy of forecasting.

The paper is organized as follows. Section 2 provides basic notions on Bayesian networks. In Sect. 4, the forecasting accuracy measures selected for this research are reviewed. Section 3 describes our contribution: the adaptation of Bayesian network to the forecasting of domestic water consumption. The results of experiments are presented in Sect. 5. Section 6 concludes the paper.

2 Bayesian Networks—Basic Notions

Bayesian network (BN) is a knowledge representation tool that is proposed in this paper for the application of forecasting of domestic water consumption. The BN is able to relate the symbolic values of explanatory variables to the discretized numerical values of the forecasted time series. BN is a transparent graphical model that explicitly depicts the probabilistic dependencies discovered in data.

A Bayesian network is a triple $BN = (X, DAG, P)$, where X is a set of random variables, $DAG = (V, E)$ is a directed acyclic graph, and P is a set of conditional probability distributions [14]. Each node of the graph $v_i \in V$ is related to the discrete random variable $X_i \in X$. The edges from the set $E \subseteq V \times V$ of DAG correspond to conditional dependence between random variables. $P(X_i | X_{pa(i)})$ is the conditional probability distribution for each $X_i \in X$, where: $pa(i)$ is the set of conditioning variables of the X_i . The posterior probability of unknown X_i given the set of evidence variables $XE \subset X$ is calculated using the BN model.

The learning of the Bayesian network consists of two steps: (1) structure learning and (2) learning of parameters, i.e., the probability distributions P .

To learn the structure of BN we selected already implemented state-of-the-art algorithms: Hill-Climbing—a greedy search on the space of the directed graphs; Tabu Search—a Hill-Climbing with the additional mechanism of escaping from local optima; Max-Min Hill-Climbing—a hybrid algorithm combining the Max-Min algorithm and the Hill-Climbing algorithm; and Restricted Maximization—a generalized implementation of Max-Min Hill-Climbing.

To learn the parameters, Maximum Likelihood Estimation (MLE) has been selected. The details of the applied algorithms are available in the documentation of 'bnlearn' library [21] of the R package [19].

3 Bayesian Approach to the Forecasting of Domestic Water Consumption

Let $\{W(t)\}$ denote the considered water consumption time series, where $W(t) \in \mathfrak{R}$ is a real-valued variable and $t \in [1, 2, \dots, n]$ is a discrete time scale of the length $n \in \mathfrak{N}$. We define a set of the random variables related to the considered time series. Let $Hour(t) \in [1, 24]$ be a variable indicating the hour of day. Similarly, $Day(t) \in [1, 7]$ is a variable indicating the day of week.

The analysis of water consumption data revealed that the variations in water consumption are related to the part of the day. It is possible to distinguish 5 intervals in which the water consumption can be analyzed: morning, day, afternoon, evening and night. The determination of these particular periods with respect to people's behaviour is a complex problem. For this paper, we decided to make an arbitrary mapping of those intervals to the hours of the day, according to observations of the behavior of the inhabitants in the single household that was considered in our research. The assumed mapping is given in Table 1. Further investigation of the problem of that mapping is left for future research.

The partitioning of the day led to the introduction of the related variable $POD(t) \subset \{morning, day, afternoon, evening, night\}$ assuming symbolic values related to the parts of the day.

To enable the application of the standard Bayesian network for the forecasting of water consumption we discretize the domain of $W(t)$. For the purpose of this paper we decided to use the simplest, equal-length discretization method. The discretized version of $W(t)$ is denoted as $W_d(t) \in \mathfrak{N}$, where the values of $W_d(t)$ are the positive integers specifying the intervals to which the actual values of $W(t)$ belong. The parameter $k \in \mathfrak{N}$ determines the number of the discretization intervals applied.

The lagged value $W_d(t - 1)$ was assumed as the last of the explanatory variables. Due to a very short forecasting interval of 30 s, the lack of short-term seasonality confirmed by the plot of correlation function (ACF), there was no reason for using higher order lags. The final set of the variables is given in Table 2.

Table 1 Temporal partitioning of day

Period	Start[h]	End[h]
Morning	4.00	9.00
Day	9.00	14.00
Afternoon	14.00	18.00
Evening	18.00	23.00
Night	23.00	4.00

Table 2 The set of variables

Variable	Description
$Hour(t)$	Hour
$Day(t)$	Day of week
$POD(t)$	Part of day
$W_d(t - 1)$	Lagged water consumption
$W_d(t)$	Water consumption

4 Forecasting Accuracy Measures

To evaluate forecasting accuracy of the discretized version of time series, i.e., $W_d(t)$ we use two measures already applied in our previous study [18]. To assess the ratio of perfect forecasts PFR (related to all forecasts made) the formula (1) is used:

$$PFR = \frac{\sum_{t=1}^n 1|W'_d(t) = W_d(t)}{n}, \quad (1)$$

where $W'_d(t)$, $W_d(t)$ denote the forecasted and actual values of water consumption respectively; $n \in \mathfrak{N}$ is the final time step of the forecasted time series.

To calculate the accumulated forecasting error the discrete mean absolute error (DMAE) accuracy measure is selected (2):

$$DMAE = \frac{\sum_{t=1}^n |W'_d(t) - W_d(t) + 1| \cdot d}{n}, \quad (2)$$

where d denotes the length of the applied discretization interval.

To evaluate the forecasting accuracy, the time series was partitioned into learning and testing periods. We applied the concept of a growing window. Following this assumption, the learning period began at the first time step and finished at time $t - 1$ of the historical time series. This way, the length of the growing window increases as time continues. The BN model was retrained at every time step and the prediction was made 1-step ahead, i.e., for the step t . The minimum length of the growing window was set to 20 steps (days).

5 Experiments

For the purpose of this study we have collected data readings from smart water meters installed in the kitchen of the anonymous household. All experiments described in this section were performed using the 'bnlearn' library [21] of the R package [19].

5.1 Source Data

Source data were collected from the initial period of the project lifetime, from 28 November 2014 to 21 February 2015. Due to the short period of gathering data, the collected time series did not cover monthly variations of water consumption. The data were gathered from two smart meters measuring the water intake of the dishwasher and that used by the sink. The readings were made in time intervals of 30 s.

Figure 1a depicts the mean water usage of the dishwasher for every day of the week. As can be noted, the dishwasher was used only on Fridays, Saturdays and Sundays with a very low variation of water usage. For that reason we focused our attention only on forecasting water consumption at the sink. As Fig. 1b shows, the most water consumption occurred just after and before weekends, on Mondays and Fridays, respectively. As Fig. 1c shows, the highest mean water consumption can be recognized in the afternoon with very low consumption at night. When considering short-term water usage, a peak can be observed at about 5 AM, see Fig. 1d.

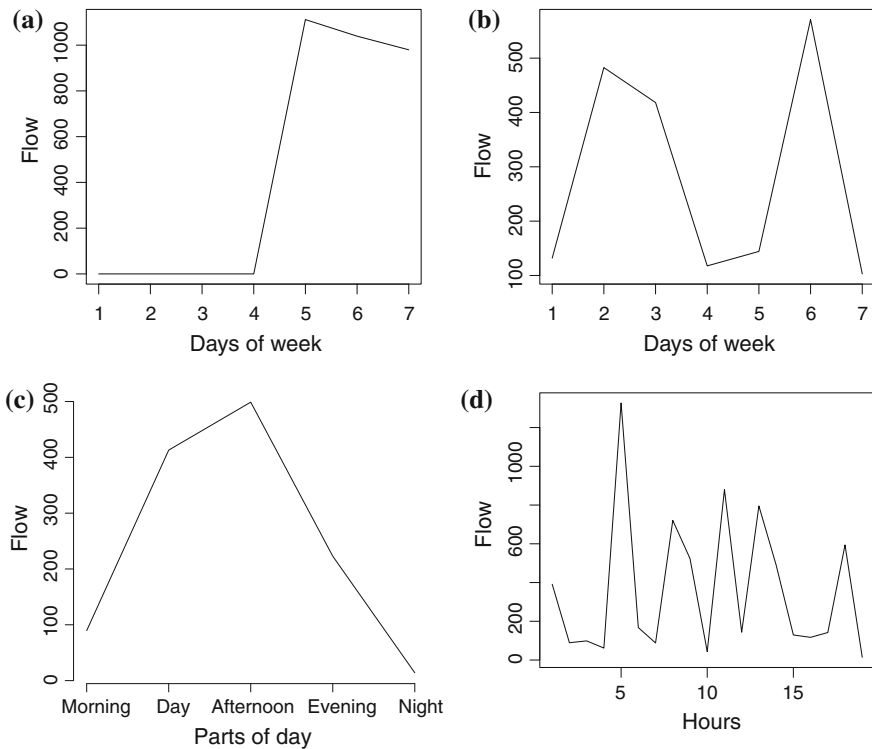


Fig. 1 Mean water consumption in kitchen. **a** Dishwasher—mean daily consumption, **b** Sink—mean daily consumption, **c** Sink—mean daily pattern, **d** Sink—mean hourly consumption

Table 3 Distribution of water consumption during the week

	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Morning	66.32	97.85	29.40	178.78	0.0	35.625	101.54
Day	172.29	1702.16	696.66	49.0	7.0	7.0	193.16
Afternoon	45.16	430.40	689.12	19.85	277.25	1166.3	53.5
Evening	212.99	273.99	394.92	109.76	15.5	17.5	81.73
Night	0.0	0.0	14.0	0.0	0.0	0.0	0.0

Table 3 presents the distribution of water consumption with respect to the days of the week and the parts of the day. Peaks for water usage were recorded on Monday during the day and on Friday afternoon. The content of Table 3 can be interpreted as an individual, weekly consumption profile of the individual water consumer. Further analysis of these type of profiles for other points of water intake placed in kitchen is planned. After gathering data from the other households, the analysis of data will be performed with respect to other users.

5.2 Experimental Results

First, the structure of the BN was learned automatically using all of the algorithms given in Sect. 2. However, the obtained values of DMAE were high. An attempt was undertaken to design the structure of the BN. Following our previous experiences with the forecasting of time series using Bayesian networks [6, 18], we assumed $W_d(t)$ as a single dependent variable. The rest of the variables that played the role of evidence variables directly influenced $W_d(t)$. This way, only forward inference within the BN was assumed. Table 4 describes the investigated BNs with the related evidence variables.

We started the investigations with a simple BN1 containing only a single evidence variable and the conditional distribution $W_d(t)|W_d(t - 1)$ assigned to the child node. In the second step we complemented the network by adding other variables. This

Table 4 Forecasting accuracy for the designed structures of BNs

Bayesian network	Variable
BN1	$W_d(t - 1)$
BN2	$W_d(t - 1), Day(t)$
BN3	$W_d(t - 1), POD(t)$
BN4	$W_d(t - 1), Hour(t)$
BN5	$Hour(t), POD(t), Day(t)$
BN6	$W_d(t - 1), POD(t), Day(t), Hour(t)$

Table 5 Forecasting accuracy DMAE(PF) for the designed structures of BNs

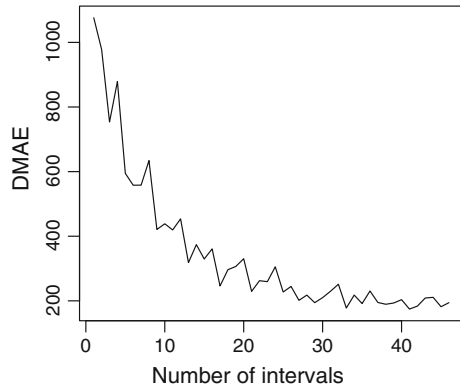
k	BN1	BN2	BN3	BN4	BN5	BN6
10	227.05(0.85)	233.47(0.83)	210.20(0.88)	265.56(0.77)	292.84(0.70)	309.69(0.67)
20	124.35(0.82)	132.38(0.78)	112.32(0.86)	142.00(0.75)	168.88(0.69)	186.93(0.60)
30	110.98(0.70)	115.26(0.67)	93.33(0.74)	126.76(0.64)	147.35(0.55)	175.70(0.49)
40	91.66(0.63)	88.65(0.61)	75.41(0.68)	107.10(0.55)	117.93(0.51)	137.79(0.42)
50	96.59(0.55)	97.23(0.52)	75.89(0.62)	111.03(0.47)	124.35(0.44)	148.10(0.35)

way BN2, BN3, and BN4 were constructed. Because the addition of $POD(t)$ was shown to be the most beneficial in terms of DMAE, we complemented BN3 further by adding first the variable $Day(t)$ and then, $Hour(t)$. In this way we obtained BN5 and BN6. The obtained results of DMAE with the PFR given in brackets are shown in Table 5. The parameter k denotes the number of equal-length intervals used for the discretization.

Table 5 shows that the BN3 is the best model and that further addition of evidence variables does not help.

As expected, the ratio of perfect forecasts decreases while increasing the number of intervals and thus making those intervals shorter. On the other hand, the value of DMAE increases reaching it's lowest values for $k > 30$. To further investigate this dependency, we repeated the experiment with BN3 for all values of $k \in [5, 50]$. The results are shown in Fig. 2. This way it has been confirmed that the lowest value of DMAE are obtained for $k > 30$. Then the DMAE exhibits random variation. On the other hand, to obtain the PFR as high as possible, for the considered data, we selected $k = 30$ for future research.

Fig. 2 Discrete mean absolute error



6 Conclusions

In the presented study, we investigated the application of standard Bayesian networks to the forecasting of water consumption. We designed a Bayesian network model that is able to efficiently forecast a discretized time series of domestic water consumption. We have discovered that the application of the evidence variable related to the parts of day augments good forecasting accuracy. Further work is required to compare the obtained results with the other forecasting methods. After collecting more data covering seasonal and monthly variations of water consumption, theoretical and experimental research will be extended. The planned research will also include the analysis of user specific water consumption profiles.

Acknowledgments The work was supported by ISS-EWATUS project which has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement no. 619228.

References

1. Abramson, B., Brown, J., Edwards, W., Murphy, A., Winkler, R.L.: Hailfinder: a Bayesian system for forecasting severe weather. *Int. J. Forecast.* **12**(1), 57–71 (1996)
2. Bennett, C., Stewart, R.A., Beal, C.D.: Ann-based residential water end-use demand forecasting model. *Expert Syst. Appl.* **40**(4), 1014–1023 (2013)
3. Biondi, D., Luca, D.D.: Performance assessment of a bayesian forecasting system (bfs) for real-time flood forecasting. *J. Hydrol.* **479**, 51–63 (2013)
4. Carragher, B.J., Stewart, R.A., Beal, C.D.: Quantifying the influence of residential water appliance efficiency on average day diurnal demand patterns at an end use level: A precursor to optimised water service infrastructure planning. *Resources, Conservation and Recycling*, pp. 81–90 (2012)
5. Fox, C., McIntosh, B., Jeffrey, P.: Classifying households for water demand forecasting using physical property characteristics. *Land Use Policy* **26**(3), 558–568 (2009)
6. Froelich, W.: Forecasting daily urban water demand using dynamic gaussian bayesian network. In: *Proceedings of the BDAS 2015*, pp. NA–NA. Springer, Ustro, Poland (2015)
7. Froelich, W., Papageorgiou, E.I.: Extended evolutionary learning of fuzzy cognitive maps for the prediction of multivariate time-series. In: *Fuzzy Cognitive Maps for Applied Sciences and Engineering—From Fundamentals to Extensions and Learning Algorithms*, pp. 121–131. Springer (2014)
8. Froelich, W., Salmeron, J.L.: Evolutionary learning of fuzzy grey cognitive maps for the forecasting of multivariate, interval-valued time series. *Int. J. Approx. Reason.* **55**(6), 1319–1335 (2014)
9. Froukh, M.: Decision-support system for domestic water demand forecasting and management. *Water Res. Manag.* **15**(6), 363–382 (2001)
10. Fuss, N.A.: Determination and verification of possible resource savings in manual dishwashing, Ph.D. thesis. Rheinische Friedrich-Wilhelms-Universitt Bonn, 1st edn. (2011)
11. Gato, S.: Forecasting Urban Residential Water Demand, Ph.D. thesis. School of Civil, Environmental and Chemical Engineering Science, RMIT University, 1st edn. (2006)
12. IBM: Smart Water Pilot Study Report. IBM Research, 1st edn. (2011)
13. Inman, D.: The use of Bayesian networks to facilitate implementation of water demand management strategies, Ph.D. thesis. Cranfield University, School of Applied Sciences (2008)

14. Jensen, F.V.: Bayesian Networks and Decision Graphs. Springer (2001)
15. Jorgensen, B., Graymore, M., O'Toole, K.: Household water use behavior: an integrated model. *J. Env. Manag.* **91**(1), 227–236 (2009)
16. Krants, H.: Matter that Matters. A study of household routines in a process of changing water and sanitation arrangements. Department of Water and Environmental Studies, 1st edn. (2005)
17. Linkola, L.: Behaviorally based modeling of domestic water use. Masters thesis. Leiden University, Delft University of Technology, 1st edn. (2011)
18. Magiera, E., Froelich, W.: Application of bayesian networks to the forecasting of daily water demand. In: Proceedings of the KES IDT-15, Sorrento, Italy. pp. NA–NA. Springer, 17–19 July 2015
19. Package, R.: <http://www.r-project.org>
20. Richter, C., Stamminger, R.: Water consumption in the kitchen a case study in four european countries. *Water Res. Manag.* **26**(6), 1639–1649 (2012)
21. Scutari, M.: Bayesian network structure learning, parameter learning and inference (2014). <http://www.bnlearn.com/>
22. Vlachopoulou, M., Chin, G., Fuller, J.C., Lu, S., Kalsi, K.: Model for aggregated water heater load using dynamic bayesian networks. In: Proceedings of the DMIN'12 International Conference on Data Mining, pp. 1–7 (2012)