Prediction of Domestic Warm-Water Consumption

ELENA ŞERBAN1, DANIELA POPESCU2
Department of Computer Science, Department of Fluid Mechanics
Technical University Gheorghe Asachi Iasi
Bd. D. Mangeron nr. 53A, Iasi, 700050
1eserban@cs.tuiasi.ro, 2daniela_popescu@tuiasi.ro, http://www.tuiasi.ro

Abstract: - The paper presents methodologies able to predict dynamic warm water consumption in district heating systems, using time-series analysis. A simulation model according to the day of a week has been chosen for modeling the domestic warm water consumption in a block of flats with 60 apartments. The analysis of the residuals indicates good simulation and prediction models for the cases studied. Double-cross validation was done using data collected by the SCADA system from District Heating Company of Iasi.

Key-Words: - district heating systems, domestic warm water, time series models, autoregressive model, simulation, prediction.

1 Introduction

Increasing energy efficiency is an important issue for District Heating Companies [1]. A source of saving energy is production of thermal energy according to demand. Very good information about future consumption are needed in order to do good production plans. Dynamic simulations enable the ongoing development of operational optimization models, but unfortunately, estimation of heat demands is a complex task. To make full advantages of the district heating network modeling, the systems must have measurement points. Nowadays, SCADA systems were implemented in Romania and historical data both regarding space heat and domestic warm water consumption are available. It is time to use them, not only for monitoring but also for a better management by doing realistic scenarios of future consumption.

For a typical district heating system, four load components shape the total heat-load: space heating for buildings, domestic warm-water, distribution loss, additional work day loads [2]. Werner investigated six Swedish district heating systems using 5-11 years monitoring data. He found out that 60% of the heat was consumed for space heating, 30% for domestic warm-water preparation, 6-8% by distribution losses, the rest representing loads that are dependent on the day of the week [3].

There are computer programs such as CONDOR, EcNetz, RNET, SYSTEM RORNET, TERMIS, BoFiT, ANSYS, DH SIM that can make space heat load prediction [4]. Scientific literature also suggests interesting models. For instance, Dotzauer [5] proposes a simple model based on the insight that the space heat demand is mainly affected by the outdoor temperature and the social behavior of the consumers. Artificial neural networks represent an alternative. The approach used by Ian Beausoleil-Morrison and Moncef Krarti used a multilayer feedforward neural network with the back-propagation learning algorithm [6] and leads to very good results. Prediction of thermal performance of a hot water system is presented in [7]. Time-series analysis is used to model systems and to predict their behavior in different area (alone or combined with other methods) [8], [9], [10].

Even if domestic warm water is used within a house for bath/shower, wash hand basin, dish washing, clothes washing, studies about the thermal energy consumed for preparing it are only a few [11]. The energy used for heating domestic warm water depends on the human behavior most of all, a factor that is difficult to control. Usually, prediction means an average value for the monthly/daily fluid flow rates per person to be taken into consideration. It is not enough for a real-time operation of a district heating system. At least hourly heat load profiles are needed.

The objective of this work is to develop and analyze methodologies able to predict dynamic warm water consumption in district heating systems (DHS), using time-series analysis. Validation of the methods was performed by comparing the modeling results with acquired data via a monitoring system from the District Heating Company of the city of Iasi (Romania).

The theoretical bases are presented in section 2, the simulation algorithm in section 3, section 4 presents the statistical analysis of computational...
results and finally in Section 5 some conclusions are given.

2 Theoretical basis
In order to predict dynamic warm water consumption in district heating systems, a simulation model using experimental data must be done.

The experimental data are sequence of values, each value corresponding to a time moment. Graphical representation of experimental data in time is a time history representation, and the sequence of data is a time series.

A time-series can be model as the output of a system that has as input a white noise signal. With this observation, the general form of a time-series model is [13]:

\[ A(q^{-1})y(t) = C(q^{-1})e(t) + e(t) \]  

(1)

where \( A(q^{-1}) \) and \( C(q^{-1}) \) are polynomials in the delay operator \( q^{-1} \) and they have the following forms:

\[ A(q^{-1}) = 1 + a_1q^{-1} + a_2q^{-2} + \ldots + a_nq^{-na} \]

\[ C(q^{-1}) = 1 + c_1q^{-1} + c_2q^{-2} + \ldots + c_nq^{-nc} \]  

(4)

where \( a_i, i = 1, na \) and \( c_i, i = 1, nc \) are coefficients of model polynomials, \( A(q^{-1}) \) and \( C(q^{-1}) \), respectively. They are referred as parameters of the model. The degrees of the respective polynomials are \( n_a \) and \( n_c \), respectively.

The variance of the white noise \( e(t) \) is considered to be \( \sigma^2 \).

The order of the model is the pair \( (n_a, n_c) \).

If \( C(q^{-1}) = 1 \) then we have an autoregressive (AR) model and the model (3) becomes:

\[ A(q^{-1})y(t) = e(t) \]  

(5)

If \( \theta \) is the parameters vector defined in equation (6):

\[ \theta = [a_1 \ldots a_n; c_1 \ldots c_n]^T \]  

(6)

and \( \phi(t) \) is defined in equation (7):

\[ \phi(t) = [y(t-1) \ldots y(t-n_a); e(t-1) \ldots e(t-n_c)] \]  

(7)

an equivalent form of equation (1) is obtained:

\[ y(t) = \phi^T(t)\theta + \epsilon(t) \]  

(8)

Choosing the appropriate order structure for the ARMA model is an iterative procedure which involves data analysis, model parameters estimation, model analysis, selection and validation.

In order to obtain a measure of the model fitness some quantities are calculated in equations (9)

The percentage of the output model variation that is explained by the model is represented by the fit function described in equation (9)

\[ Fit = \left[ 1 - \frac{\|y - \bar{y}\|}{\|y - \bar{y}\|} \right] \cdot 100 \]  

(9)

where \( y \) is the observed (real) time-series, \( \bar{y} \) is the simulated time-series and \( \bar{y} \) is the mean value of time series \( y \) calculated applying the formula:

\[ \bar{y} = \frac{1}{N} \sum_{i=1}^{N} y_i \]  

(10)
where $N$ is the number of samples from time series under consideration and $y_i$ is $i$-th sample of the time-series $y$.

In equation (9) $\|w\|$ is the Euclidean norm of the time-series $w$ calculated with the formula:

$$\|w\| = \sqrt{\sum_{i=1}^{n} w_i^2}$$

(11)

where $N$ is the number of samples of the time-series $w$ and $w_i$ is the value of the $i$-th sample of the time-series $w$.

Another criterion for model evaluation is the loss function. Loss function is calculated as the determinant of the covariance matrix of the prediction errors (residuals of the model). The loss function $V$ is calculated with the following relation:

$$V = \det \left( \frac{1}{N} \sum_{i=1}^{N} \varepsilon_i \varepsilon_i^T (\hat{\theta}) \right)$$

(12)

where $\hat{\theta}$ is the estimate of the parameters vector $\theta$ and $\varepsilon_i$ is the $i$-th value of the model residuals time-series.

The Akaike Information Criterion (AIC) for an estimated model is defined as the value of the negative log-likelihood function at the estimated parameters plus the number of estimated parameters [12]. If the disturbance source is Gaussian with covariance matrix $\Lambda$, the logarithm of likehood function is:

$$L(\theta, \Lambda) = \frac{1}{2} \sum_{i=1}^{N} \varepsilon_i (\theta)^T \Lambda^{-1} \varepsilon_i (\theta) - \frac{N}{2} \log \det \Lambda + C$$

(13)

where $C$ is a constant. Maximizing this analytically with regard to $\Lambda$ and then maximizing the results with regard to $\theta$, gives

$$L(\theta, \Lambda) = C + \frac{N \cdot p}{2} + \frac{N}{2} \log(V)$$

(14)

where $p$ is the number of outputs and $V$ is the loss function defined in equation (12). After removing constants and suitable normalization, the following expression is reached:

$$AIC = \log(V) + \frac{2d}{N}$$

(15)

where $d$ is the number of estimated parameters.

Akaike Final Prediction Error (FPE) [12] for an estimated model is calculated with the formula:

$$FPE = V \cdot \frac{1 + \frac{d}{N}}{1 - \frac{d}{N}}$$

(16)

It is technically possible for FPE to become negative if the number of estimated parameters exceed the number of data. In such a case it is better to use AIC.

The estimation $\hat{\theta}$ of the parameters vector is calculated using the condition [14]:

$$\hat{\theta} = \arg \min_{\theta} V(\theta)$$

(17)

where the criterion function $V(\theta)$ is defined in equation (18):

$$V(\theta) = \sum_{i=1}^{N} e_i^2 (t) = \sum_{i=1}^{N} \left[ y_i(t) - \phi^T(t) \theta \right]^2$$

(18)

The solution of equation (17) is:

$$\hat{\theta} = \left[ \sum_{i=1}^{N} \phi(t) \phi^T(t) \right]^{-1} \left[ \sum_{i=1}^{N} \phi(t) y(t) \right]$$

(19)

In equation (19) the following matrix:

$$\sum_{i=1}^{N} \phi(t) \phi^T(t)$$

(20)

has to admit an inverse and to be semi-positive defined.

Another form of the solution of equation (17), equivalent with solution (19) is:

$$\hat{\theta} = \left[ \Phi^T \Phi \right]^{-1} \Phi^T Y$$

(21)

where

$$Y = [y(1) \ldots y(N)]^T$$

$$\Phi = [\phi(1) \ldots \phi(N)]^T$$

(22)

The matrix (21) admits inverse if the elements of the vector $\phi(t)$ are linear independent.
In order to predict the behavior of the system an on-line determination of the systems parameters can be made. This procedure is based on the recursive algorithm for the parameters estimation [14].

If \( \hat{\theta}(t) \) is the estimation of the parameters of the system (8), the estimation based on the first \( t \) measurements, then:

\[
\hat{\theta}(t) = \left[ \sum_{s=1}^{t} \varphi(s) \varphi^T(s) \right]^{-1} \left[ \sum_{s=1}^{t} \varphi(s) y(s) \right] \tag{23}
\]

Estimation (23) satisfies, by definition, the equation

\[
\sum_{s=1}^{t} \epsilon(s) = y(s) - \varphi^T(s) \theta(s) \tag{24}
\]

The weighted form of the criterion (18) is considered:

\[
V(\theta) = \sum_{s=1}^{t} \lambda^{2-s} \epsilon^2(s) \tag{25}
\]

In equation (25) \( \lambda \) is the forgetting factor. This way the criterion function uses, for parameters estimation, the last measurements. The other measurements, before these, are almost ignored.

The recursive algorithm is described by the following equations:

\[
\hat{\theta}(t) = \hat{\theta}(t-1) + K(t) \epsilon(t) \\
\epsilon(t) = y(t) - \varphi^T(t) \hat{\theta}(t-1) \\
K(t) = \frac{P(t-1) \varphi(t)}{[\lambda + \varphi^T(t) P(t-1) \varphi(t)]} \\
P(t) = \left[ P(t-1) - K(t) \varphi^T(t) P(t-1) \right] / \lambda
\]

where \( P(t) \) is of form (28), depending on the covariance matrix of the estimator:

\[
P(t) = \left[ \sum_{s=1}^{t} \lambda^{2-s} \varphi(s) \varphi^T(s) \right]^{-1} \tag{28}
\]

If there are no a-priori information about \( \theta \), the following initializations are made:

\[
\hat{\theta}(0) = 0 \quad \text{and} \quad P(0) = \alpha I \tag{29}
\]

where \( I \) is the unity matrix of appropriate order and \( \alpha \) is constant. A value too small for \( \alpha \) leads to a slow convergence of the estimates, and a value too big for \( \alpha \) has as consequence oscillation with high amplitude of the parameters values (which are elements of vector \( \theta \)). If the analyzed system has time-varying parameters, then the chosen value for \( \lambda \) must be less then 1 (smaller if the system parameters have a fast variation in time). Usual, \( \lambda \in [0.95, 0.995] \).

3 Numerical Simulation

The simulation modeling assumes measurements of data for a given period. Data collected during three months by a datalogger were used in this study. The conversion of the pulse signal to a voltage signal appeared not to be as constant in time as desired and to give meaningful results the heat meter signal had to be filtered.

Figure 1 shows the consumption of domestic warm water for a block of flats. Like in other studies described by scientific literature, variation is large and unsystematic. In the present case, the monthly quantity of heat consumed for domestic warm water preparation was 243 KWh/month per apartment. Other studies took into consideration different values. For instance, Bohm B. investigated consumption of domestic warm water in Copenhagen and found out that the thermal energy used was 106 kWh/month per apartment. Yang used data from an apartment having a thermal energy consumption for preparing warm water of 200-222kWh/month, and Lavaetz used values in the domain 108-233kWh/month per apartment [15]. Yao R. and Steemers K. analyzed typically hourly domestic warm water load profile for an average size domestic house hold from U.K [11]. They took into consideration five types of consumers according to different occupancy scenarios: unoccupied from 9.00-13.00, unoccupied from 9.00-18.00, unoccupied from 9.00-16.00, all day occupied, unoccupied from 13.00-18.00. They found out a typical consumption profile and used it in a computer program for prediction of daily load profile for households from U.K.

The building studied in the present paper has 60 apartments; therefore it is difficult to choose a
model specific for all the inhabitants. A simulation model according to the day of a week has been chosen. Usually, studies assume that the behavior of consumers depends most of all on the fact that the day is a working one or not. What about the same type of day? Is really the consumption similar in non-working days?

This work studies consumption of domestic warm water during two non-working days: Saturday and Sunday. The simulation model was built using data for one single day and another day was used for validation. Also cross validation was done checking out the accuracy of predictions.

In order to choose a simulation model for modeling the warm water consumption many ARMA model were taken into consideration. The criteria for choosing one model or another were the value of fit function (9), the Akaike Final Prediction Error (16) and Akaike Information Criterion (15). For fit function larger values are better, while for FPE and AIC smaller values are better.

In Table 1 are presented the fit functions values for analyzed models, Table 2 contains values for FPE and Table 3 contains values for AIC. In the model name the first number represent the value for $n_a$ and the second the value for $n_e$.

<table>
<thead>
<tr>
<th>Model</th>
<th>Fit Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMX(2,2)</td>
<td>75.86 67.81</td>
</tr>
<tr>
<td>AMX(2,4)</td>
<td>81.59 75.41</td>
</tr>
<tr>
<td>AMX(2,6)</td>
<td>81.48 76.34</td>
</tr>
<tr>
<td>AMX(4,4)</td>
<td>81.65 75.52</td>
</tr>
<tr>
<td>AMX(4,10)</td>
<td>82.06 76.64</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>FPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMX(2,2)</td>
<td>2.03443 2.0763</td>
</tr>
<tr>
<td>AMX(2,4)</td>
<td>1.21434 1.2456</td>
</tr>
<tr>
<td>AMX(2,6)</td>
<td>1.26239 1.1708</td>
</tr>
<tr>
<td>AMX(4,4)</td>
<td>1.23486 1.2605</td>
</tr>
<tr>
<td>AMX(4,10)</td>
<td>1.29019 1.2339</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMX(2,2)</td>
<td>0.7102 0.7306</td>
</tr>
<tr>
<td>AMX(2,4)</td>
<td>0.1942 0.2196</td>
</tr>
<tr>
<td>AMX(2,6)</td>
<td>0.2330 0.1577</td>
</tr>
<tr>
<td>AMX(4,4)</td>
<td>0.2109 0.2315</td>
</tr>
<tr>
<td>AMX(4,10)</td>
<td>0.2544 0.2101</td>
</tr>
</tbody>
</table>

Figure 1. Daily thermal power consumption of domestic warm water during a month.
The chosen model for Saturday is described by equation (3) and the polynomials of the models are:

\[ A(q) = 1 - 1.006 q^{-1} - 0.2217 q^{-2} \]
+0.1405 q^{-3} + 0.09357 q^{-4} \tag{30}

C(q) = 1 + 0.5169 q^{-1} + 0.3459 q^{-2} +
+ 0.1216 q^{-3} + 0.1592 q^{-4} - 0.8185 q^{-5} -
- 0.3647 q^{-6} - 0.293 q^{-7} - 0.09375 q^{-8} -
- 0.2134 q^{-9} - 0.1543 q^{-10} \tag{31}

Estimated using PEM from data set zi1s.
Loss function 1.0917 and FPE 1.29019.
Sampling interval: 300 s.

Figure 2 presents the experimental data and the simulation output for Saturday.
The chosen model for Sunday is described by equation (3) and the polynomials of the models are:

A(q) = 1 - 1.431 q^{-1} + 0.0099 q^{-2} +
+ 0.5751 q^{-3} - 0.1506 q^{-4} \tag{32}

C(q) = 1 + 0.08072 q^{-1} - 0.3723 q^{-2} -
- 0.03749 q^{-3} - 0.0645 q^{-4} + 0.9458 q^{-5} -
- 0.0696 q^{-6} + 0.4276 q^{-7} +
+ 0.08007 q^{-8} + 0.05085 q^{-9} - 0.04453 q^{-10} \tag{33}

Estimated using PEM from data set zis2.
Loss function 1.08164 and FPE 1.2783.
Sampling interval: 300 s

It can be noticed that the polynomials of the models and the shape of a graph are different even if both days are non-working days. Concluding, a different simulation model has to be done for every day of the week.

4 Statistical analysis
In order to validate the model, the residual analysis is used. Residuals represent the portion of the validation data not explained by the model. Residual analysis consists of two tests: the whiteness test and the independence test. According to the whiteness test criteria, a good model has the residual autocorrelation command inside the model confidence interval, indicating that the residuals are uncorrelated. According to the independence test criteria, a good model has residuals uncorrelated with past inputs. Evidence of correlation indicates that the model does not describe how part of the output relates to the corresponding input [12].

4.1 Simulation model
This work contains the autocorrelation of residuals for Saturday and Sunday, presented in

Figure 4 and Figure 5, respectively. The horizontal dotted lines represent the limits of the confidence interval set to 99%.

It can be noticed that almost all values of the residuals autocorrelation are inside the confidence interval indicating a good simulation model for both days.

![Figure 4. Autocorrelation of residuals for simulation model. Day Saturday.](image)

![Figure 5. Autocorrelation for residuals for simulation model. Day Sunday.](image)

4.2 Prognosis model
The main purpose of modeling is forecasting consumption in order to have an efficient production. Space heating is obviously the first option of scientists’ works [16], [17]. Forecast of domestic warm water consumption is a different problem depending on other parameters, so different methods are needed [18]. The present work tries to do it using time-series analysis.

Previous paragraphs present two different simulation models: one for a Saturday day and the other for a Sunday day. The next step is to check with experimental data for other days of Saturday and Sunday, if these time-series simulation models are or not appropriate for prognosis.
An analysis of residuals when the simulation model is used for prediction should give the answer to this problem. Each simulation model was used for prognosis of consumption for two different future days. Figure 6 presents the residuals of Saturday’s forecasting and figure 7 presents the residuals of Sundays’ forecasting. The confidence interval was set to 99%.

For both figures prediction for the first future day is represented with solid line, and the second future day is represented with dashed line. The confidence interval limits are the dotted lines.

![Figure 6](image6.png)

**Figure 6. Autocorrelation of residuals for prediction models. Prognosis for Saturday days.**

![Figure 7](image7.png)

**Figure 7. Autocorrelation of residuals for prediction models. Prognosis for Sunday days.**

It can be noticed that the prediction is very good. Even if the confidence interval is tight, practically all the residuals are within it.

The prediction method described above considers that the system's parameters are constant in time. If these parameters vary in time, the recursive method described by equations (27) and (28) has to be applied. The recursive method has also the advantage that in any moment, the next value of the time series can be calculated.

In Figure 7 the evolution in time of the Saturday model parameters are presented. Figure 8 presents the experimental data (the blue line) and predicted data (green line) obtained applying the recursive algorithm for data for Saturday.

![Figure 8](image8.png)

**Figure 8. Experimental and recursive predicted data for Saturday.**

In Figure 9 the evolution in time of the Sunday model parameters are presented. Figure 10 presents the experimental data (the blue line) and predicted data (green line) obtained applying the recursive algorithm for data for Sunday.
Neither of these two models (for Saturday and for Sunday) do not have model parameters, and a recursive method for model parameters evaluation might be more appropriate.

Figure 11 presents the time-history for model parameters for the days of the week-end and Figure 12 presents the observed data and predicted data for the days of the weekend. As one can notice observing Figure 11, in this case the model parameters vary in time, but, after a time, they tend to have constant values.

5 Conclusions
The paper presents a methodology for prognosis of domestic warm water consumption in district heating systems based on time series analysis. Double cross validation was done; the simulation model was used for prediction in other two cases. Recursive algorithm was applied highlighting that the model parameters are not constant in time. The results of the statistical analysis are very good pointing out that daily simulation models using time series analysis are a powerful and appropriate tool for the prognosis of consumption in district heating systems.
References:


