Available online at www.sciencedirect.com

ScienceDirect

Energy Procedia 70 (2015) 379 – 387

Energy Procedia

International Conference on Solar Heating and Cooling for Buildings and Industry, SHC 2014

# Statistical modeling for real domestic hot water consumption forecasting

Aurore Lomet[a], Frédéric Suard[a], David Chèze[b]

[a]CEA, LIST, F-91191 Gif-sur-Yvette, France
[b]CEA, LITEN, Le Bourget du Lac, France

**Abstract**

In this paper, we study the real domestic hot water (DHW) consumptions from single family houses equipped with solar hot water tank. We model it to understand and forecast the daily needs of inhabitants. Thus, the forecasts can be integrated in a control strategy to optimize the energy cost by heating only the necessary DHW volume.
At first, we realize a data analysis from real uses of several dwellings to lay assumptions of the statistical model. This study highlights a weekly periodicity, random fluctuations and the different profiles of consumption following the residence, the season, and the day of the week. Otherwise, having no prior information like the location or the number of residents, we propose an adaptive time series model which does not require strong a priori and computational time. Then, we develop an ARMA model to forecast the daily DHW volume and we apply it on each individual installation. This model allows to take into account the periodicity of one week, the consumption of the previous days and random fluctuations. The results obtained on real data show that this approach is very promising.


Peer-review by the scientific conference committee of SHC 2014 under responsibility of PSE AG

*Keywords: Domestic hot water consumption, ARMA model, statistical forecasting*

## 1. Introduction

The domestic hot water consumption represents a large part of energy cost of residential end-users. The Domestic Hot Water (DHW) is commonly stored in a water tank of variable size (from 100 to 200L) whose total capacity is heated up to a temperature of 60°C. However the whole heated volume is rarely tapped by end-users and the required energy to keep the whole volume at the desired temperature is wasted through heat losses. A solution to overcome this problem is to heat only the DHW volume that residents need in order to adjust and control only the


Peer-review by the scientific conference committee of SHC 2014 under responsibility of PSE AG
doi:10.1016/j.egypro.2015.02.138

necessary energy. Thus, estimating the DHW consumption becomes an issue for decreasing energy consumption without impact on the comfort of the residents.

For this purpose, different approaches have been proposed to model the energy and DHW consumption as neural networks [1,8], grey-box modeling approaches [2,5] or more generally in the energy field by others models like regressions [9]. However, such models require strong computational time and detailed information about the residents like their number, age or social profile that is often not available. Other methods are based on probability models [4] or moving average models [7]; nevertheless, they need strong assumptions of the DHW load profiles which do not take into account isolated fluctuations of the consumption.

In this work, we propose a model to estimate the daily DHW consumption which does not require strong computation time and information about the residents. Thus, this model can be embedded and integrated in a planning and control strategy. This adaptive model is designed from a data mining study realized on real uses of 8 residences equipped with solar water tank. This study highlights a weekly periodicity, random fluctuations and the different profiles of consumption following the residence, the season, and the day of the week. Therefore, to take into account these results, we propose a time series model whose form is fixed for all residences and whose parameters are tuned for each residence. The proposed ARMA, *Auto Regressive Moving Average*, process considers the DHW consumption of the previous days and of the day of the previous week.

The paper starts with the description of experimental data. Then, the results of the analysis study are presented to lay the assumptions of model. This is followed by a presentation of the modeling approach where the ARMA process and its form are defined. Finally, the results are presented and compared to the model proposed by [7] (named here PG model).

## 2. Data set presentation

The data are gathered trough the measurement of the real uses of the DHW consumption for 8 residential homes in France over two years between 2009 and 2011. The timestamp, the inlet and outlet temperatures of the tank and the consumed volume are given per day but the location and the number of residents is unknown. The measurements are realized every thirty minutes. An analysis with this timestamp shows that the presence of DHW consumption early in the day involves a significant one at the evening. Indeed, for each residence, a day for which a consumption is observed in morning (even lower than 10 liters) presents almost every time a non-negligible one at the end of day. Otherwise, the consumption is short time and irregular in consumed volume and when it is drawn as illustrated in Figure 2. These promptness and irregularity in time due to this granularity make the modeling and the prediction difficult. To solve this problem, we study and model the daily DHW consumption for which an example is given in Figure 1.

The number of daily measurements varies between 632 and 738 following the residence due to the missing values as illustrated in Figure 1 where the zero values are the absence of residents as holidays and not the missing values. Due to the application of the final model and to avoid introducing bias in this study, no interpolation or estimation is realized to replace these values. Each data file (for each residence) contains less than 5% of missing values. Then, the proposed model has to be robust of missing values, since the model is learned from the observed data.

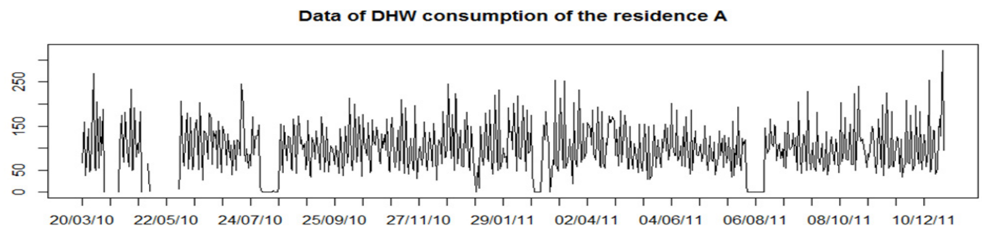


Fig. 1. Consumed volume in liters per day for the residence A.

## 3. Data analysis

In the literature, authors make various assumptions about the factors which impact the volume of the DHW consumption in order to explain it. For example, [4] proposed a probabilistic model to generate data of DHW consumption which depends on the season, the week, the day of the week, a probability for the summer holidays and the type of consumption (bath, shower, short load and medium load). In our study, since the data come from real uses, we study the behavior of the residents about their consumption of the eight residences to highlight the different phenomena which impact the daily consumed volume. For this purpose, we analyze the consumption by month, by week and the auto-correlation of the data.

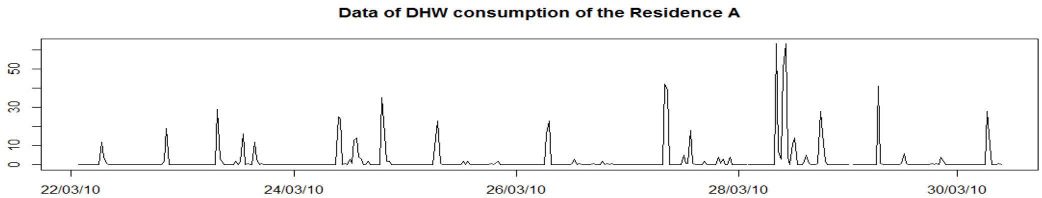


Fig. 2. Consumed volume in liters per 30 minutes for the residence A.

The statistical analysis of these data highlights that different factors have an influence on the DHW consumed volume. Thus, the season has a strong influence since the needs are lower in summer than in winter. Moreover the consumed volume depends on the month which decreases also during February and August as illustrated in Figure 3. These graphs represent the average of consumed volume of DHW per month of the year and the average of the consumed DHW volume per day of occupancy for the residence A. This last one highlights that the decreases observed during February and August on the left graph are due to the lower number of days in February and to the absence of the residents since these decreases do not appear when we consider the average of consumption per days of occupancy. The same observation could also be pointed out in Figure 4 which draws the boxplot of consumed volume for each month in liters per day of occupancy. Indeed, the median values of the DHW consumption in February and August are not lower than the others months. However, the dispersion and the maximum value during summer are generally lower than others months. In this graph the month of January seems uncharacteristic with its weak dispersion. This fact is due to the lack of data for this month (only 22 days against 60 or 62 for the others) and then, it is not considered in the analysis.

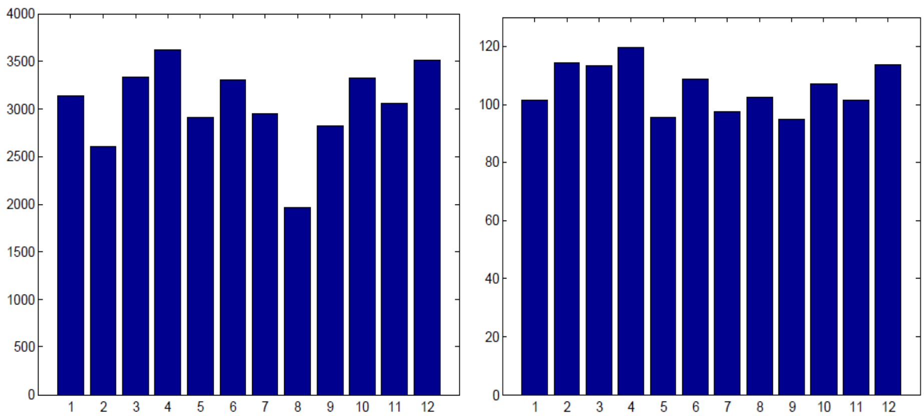


Fig. 3. Average of consumed volume in liters per month of the year for the residence A: on the left, the graph represents the average of the consumed volume per month; on the right, the average of the consumed DHW volume per day of occupancy.

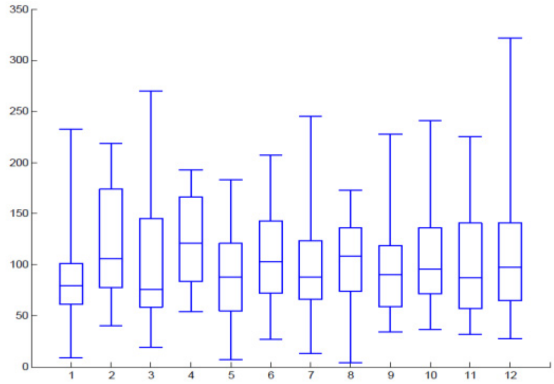


Fig. 4. Boxplot of the consumed volume for each month in liters per day of occupancy for the residence A.

Another result is the definition of different consumption profiles which vary following the residence and the day of the week as illustrated in Figure 5. These boxplots represent the consumed volume in liters per day of the week for the residences A (left subfigure) and C (right subfigure). Thus we can observe that the consumed volume on Wednesday is generally up to two times compared to the average volume of Monday for the residence A because the school is closed (in France) while this is not the case for the residence C. More generally, the profile of a typical week depends on the considered residence: for example, the mean of the DHW consumption varies between 62.1 and 109.4 liters. Therefore the model used to forecast the DHW consumption has to be adaptable to the different daily profiles of each residence.

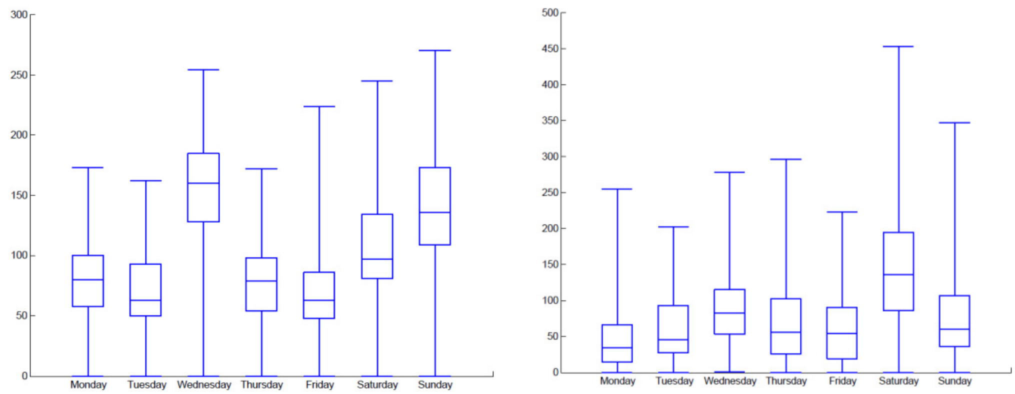


Fig. 5. Boxplot of the consumed volume in liters per day of the week for the residences A (in the left) and C (in the right).

Even if the consumption profiles are different for each residence, all of them present a periodic pattern of the consumption of one week as shown in Figure 6. These two autocorrelograms represent the similarity between observations as a function of the time lag between them in order to find repeating patterns. They highlight significant peaks (above the blue line) each seven days which means that the values are correlated with a periodicity of a week.

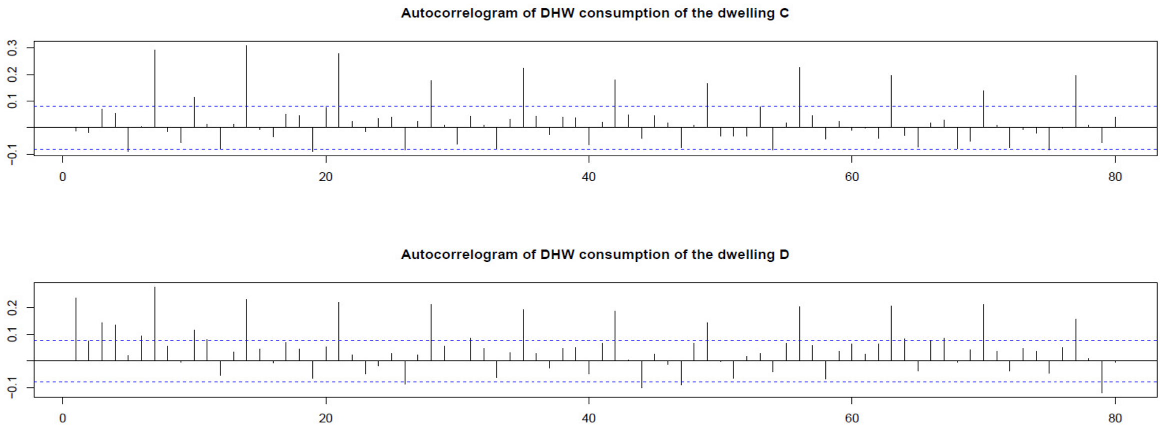


Fig. 6. Autocorrelogram of the DHW consumed volume per day of the residence C and D (lag in days).

Although the DHW consumption of these 8 residences presents various profiles, we can observe similar trends like the decrease in summer and the periodicity of one week. However all these phenomena are not sufficient to explain in an exhaustive way the real consumption which is too irregular to be simply modeled by them. Considering the objective of embedded computation, a constraint is put on the model complexity. Then the predictive model used for the DHW consumption has to present a similar low complexity and takes into account the periodicity. This model also has to be flexible and self-adaptive to each residence for considering the different profiles and the random fluctuations. Moreover having no knowledge about residences excepting the consumed volume, this model has to be able to learn it from the historical data. Accordingly we propose to use a model for time series with an autoregressive part to integrate all the preliminary observations.

## 4. Predictive model for daily DHW consumption

The DHW consumption is given as a volume in liters per day. Since these observations $(X_t)$ are following a temporal dimension in discrete time, they can be defined as time series model like ARIMA [3] (*Auto Regressive Integrated Moving Average*) in order to analyze the previous behavior and predict its future behavior. This model is a generalized form of the stationary ARMA model [6] (*Auto Regressive* Moving *Average*) and allows to take into account the non-stationary and the periodicity of the data. An ARIMA process $(Y_t, t \in \mathbb{Z})$ of order $(p, d, q)$ is defined by three components: the autoregressive model of order $p$, the integration of order $d$, the moving average of order $q$ and given by:

$$X_t = \nabla^d Y_t \ , \qquad (1)$$

where $X_t$ is an ARMA process of order $(p, q)$ given by:

$$X_t = \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t \quad , \qquad (2)$$

with the parameters $\varphi_i$ and $\theta_i$ are linear coefficients and $\epsilon_i$ a white noise. The first sum of this model represents the autoregressive part which expresses linearly the output variable $X_t$ according to its previous values. The second sum of this model represents the moving average defined by a linear regression of the current value of the series against current and previous white noise error terms. Considering a time lapse of 12 weeks for the training set of the time series, the data are considered as stationary. Therefore, an ARMA process could be applied to the DHW consumption.

To set the general form of the model, we consider the results of the data analysis presented before which highlights a periodicity of one week for each residence. Moreover, to take into account promptly the variations of the consumption due to the absence of the residents during the holidays and other sudden changes, we also consider in the model the DHW consumed volume of the previous days. Various forms of model presenting these characteristics have been compared: different numbers of previous days and numbers of weeks have been integrated in the model for example. Finally, we retain the model for which minimizes the forecasting error. Thus, the form of the model is fixed for all residences as:

$$X_t = \varphi_1 X_{t-1} + \varphi_2 X_{t-2} + \varphi_7 X_{t-7} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \theta_7 \epsilon_{t-7} + \epsilon_t \ + Cst \ , \quad (3)$$

where $Cst$ is a constant value (the mean of DHW consumed volume).

Thus only the coefficient values $\varphi_i$ and $\theta_i$ depend on the residence. Practically, these parameters are tuned through a learning phase based on a historical dataset. Since no parameter is involved in the model, the learning dataset should be properly chosen: it is preferred that it contains exhaustive pattern of behaviors and not a wide interval of missing values. After several tries with different training sets and sizes, we set it at 12 weeks and choose for each residence a window, as it is possible, which contains a short absence of the residents, fluctuations, periodic behaviors and which does not contain too many missing values: the model is robust to some missing values.

## 5. Results

The previous ARMA model is applied on each residence dataset. The training window of 12 weeks defined previously is used to estimate the values of coefficients during the learning stage. Then, a different set is used to perform the forecast. This function allows to set the periodicity at seven days and the form of the model as presented in section 3. The parameter estimation is performed thanks to the conditional mean square method.

Then, the model is applied to forecast the daily DHW volume over one year with a daily update of data. The output of our model is compared to estimations given by the daily average consumption and by the PG model [7] which is a moving average on the same day of the week during the last two months.

Figures 6 to 9 represent the forecasts and error distribution of the ARMA model for three residences. On the left, the testing data set and the estimations are represented: the upper figure is the forecasts of the mean, the middle figure is the PG model estimations and bottom figure is our ARMA model forecasts. On the right, the error distribution and median for each (in red) are plotted. The negative values (respectively positive values) of the histograms correspond to an underestimation of consumption (respectively an overestimation) which causes a discomfort for the residents (respectively an over-consumption of energy for heating water). Ideal distribution is a Dirac centered in zero.

They highlight that our model corrects the latency problem of PG model in the case of resident absences and large variations of the DHW consumption. The histograms also highlight that the error distribution of our model is generally more centered on zero than the PG model and also show that large errors are reduced. Indeed, the autocorrelation part of our model recovers the trend of the data in two days as illustrated with the summer holidays and in Figure 9 with the large fluctuations of consumption. Although the PG model is efficient for regular data as showed in Figure 7, it needs more than one month to find the correct trend in Figure 9 after the variation of consumption in the end of October while our model takes few days

.

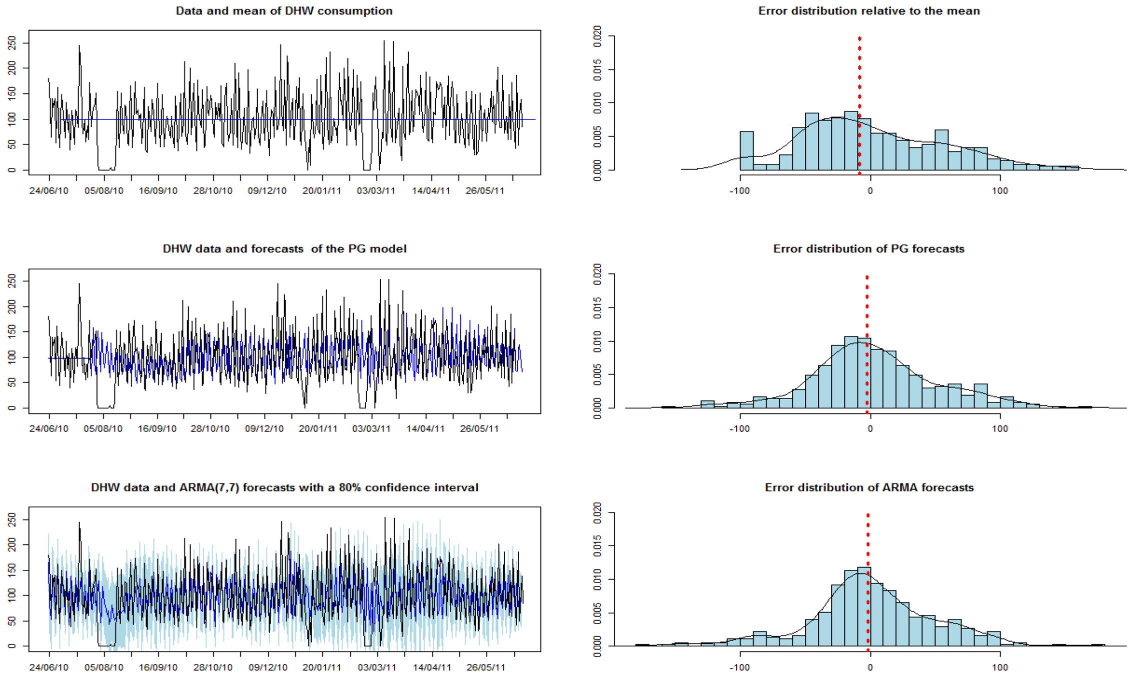


Fig. 7. Forecast and error distribution of the ARMA model for the residence A: on the left, the data (in black), the forecasts of the mean, the PG model and our ARMA model (in blue) are represented; on the right, the error distribution and median (in red) are represented. On the histograms, the left side of zero corresponds to an overestimation of consumption and the right, to an underestimation.

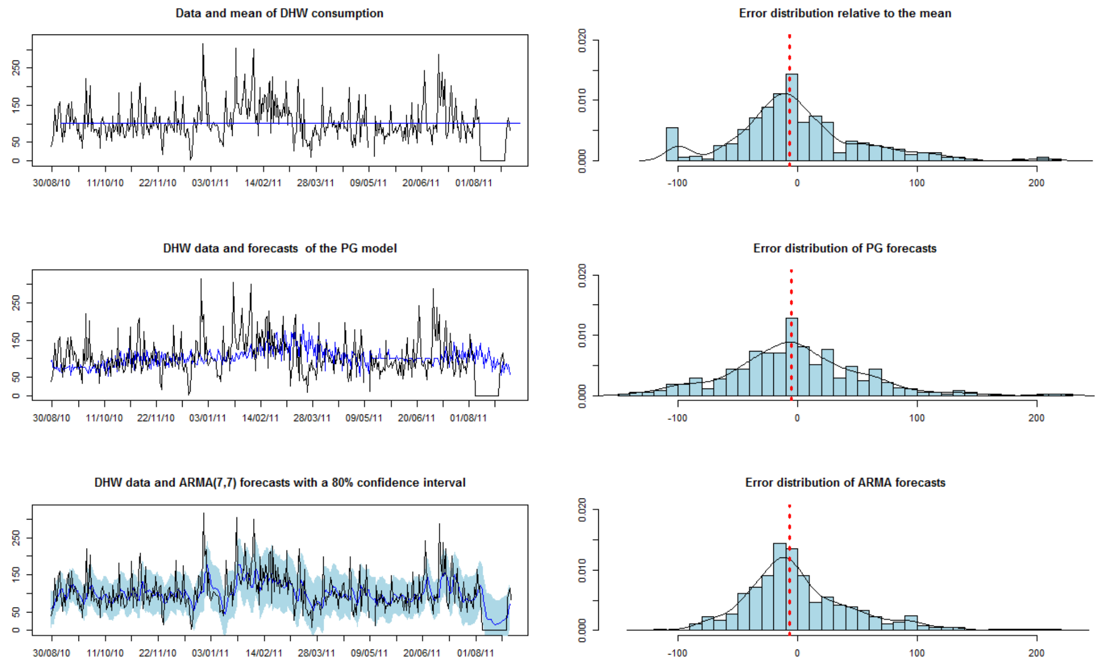


Fig.8. Forecast and error distribution of the ARMA model for the residence C: on the left, the data (in black), the forecasts of the mean, the PG model and our ARMA model (in blue) are represented; on the right, the error distribution and median (in red) are represented. On the histograms, the left side of zero corresponds to an overestimation of consumption and the right, to an underestimation.

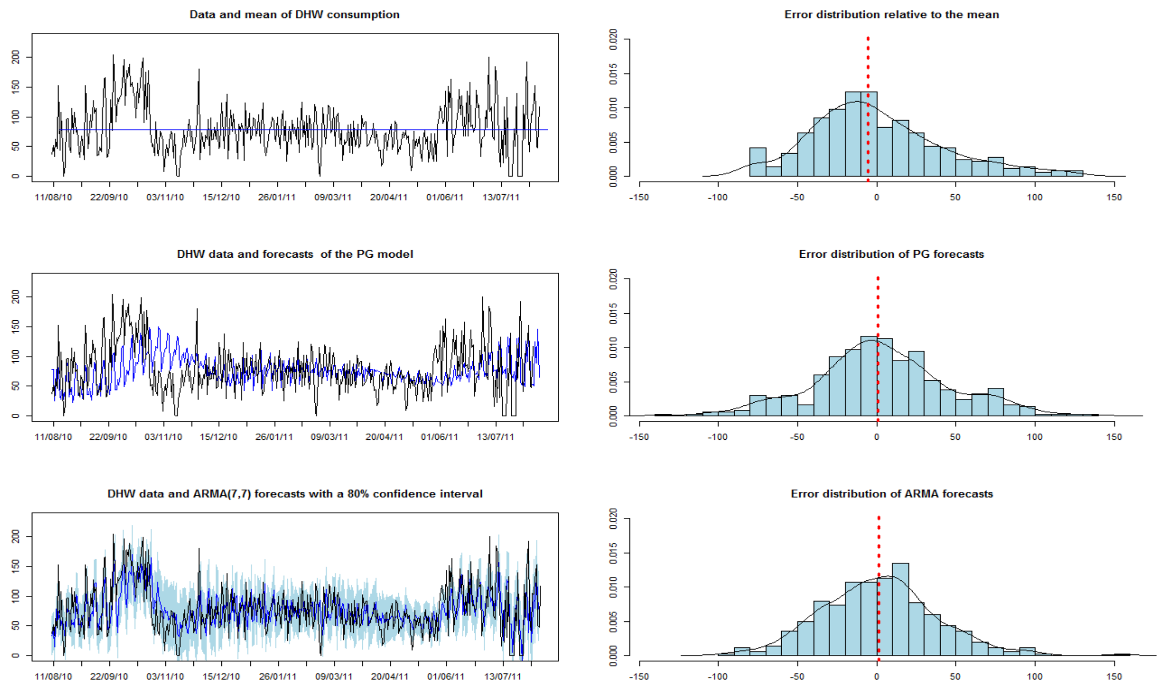


Fig. 9. Forecast and error distribution of the ARMA model for the residence E: on the left, the data (in black), the forecasts of the mean, the PG model and our ARMA model (in blue) are represented; on the right, the error distribution and median (in red) are represented. On the histograms, the left side of zero corresponds to an overestimation of consumption and the right, to an underestimation.

## 6. Conclusion

The domestic hot water consumption is an important part of energy demand. The forecast of the consumed DHW volume allows to control and reduce the cost by heating only the needed volume. In this context, we study the consumption profiles of various individual residences. In terms of data, only the inlet and outlet temperatures of the water tank and the consumed volume are known.

In a first step, we performed a data analysis on 8 residences which highlights some common characteristics as a periodicity of one week or a decrease of consumption in summer, different profiles defined by the mean or the type of consumption during the week and large variations.

Based on this preliminary result we propose an ARMA model that takes into account the periodicity, the consumption of the two previous days and the consumption of the corresponding day of the previous week. The parameter values are estimated over a training window of 12 weeks for which the data are considered stationary. The forecast of the daily DHW volume is performed on a period of one year.

The results on real data are compared to the mean and the PG model, a moving average on the same day of the week during the last two months. They showed that our model performs the forecast of the DHW daily consumed volume with higher precision and recovers quickly the trend of consumption when large variations like the summer holidays.

Since this model has shown its efficiency on a daily forecast basis, a perspective is to combine it with a model of the DHW profiles within the day. Thus the DHW heating management could be realized each half-hour in order to heat the desired water volume at the right time only what would save energy without discomfort for the residents.

## References

[1] M. Aydinalp, V. Ismet Ugursal, and A. S. Fung. Modeling of the space and domestic hot-water heating energy-consumption in the residential sector using neural networks. *Applied Energy* 2004, 79(2) :159-178.
[2] P. Bacher, H. Madsen, H. A. Nielsen, and B. Perers. Short-term heat load forecasting for single family houses. Energy and Buildings, 2013.
[3] G. E. Box, G. M. Jenkins, and G. C. Reinsel. Time series analysis: forecasting and control. John Wiley & Sons, 2013.
[4] U. Jordan and K. Vajen. Influence of the dhw load profile on the fractional energy savings: A case study of a solar combi-system with trnsys simulations. *Solar Energy* 2001, 69: 197-208.
[5] H. A. Nielsen and H. Madsen. Modelling the heat consumption in district heating systems using a grey-box approach. *Energy and Buildings*, 38(1) :63-71, 2006.
[6] S. Makridakis andM. Hibon. Arma models and the box-jenkins methodology. *Journal of Forecasting*, 16(3) :147-163, 1997.
[7] T. Prud homme and D. Gillet. Advanced control strategy of a solar domestic hot water system with a segmented auxiliary heater. *Energy and buildings* 2001, 33(5): 463-475
[8] L. G. Swan, V. I. Ugursal, and I. Beausoleil-Morrison. Occupant related household energy consumption in Canada : Estimation using a bottom-up neural-network technique. *Energy and Buildings*, 43(2) :326-337, 2011.
[9] L. Suganthi and A. A. Samuel. Energy models for demand forecasting - a review. *Renewable and Sustainable Energy Reviews*, 16(2) :1223-1240,2012.