Statistical properties of domestic hot water consumption

Nathanael Beeker[#*1], Paul Malisani[#2], Nicolas Petit[*3]

[#]*EnerBaT, EDF Lab*
Av. des Renardières – Ecuelles, 77818 Moret-sur-Loing Cedex, FRANCE
[1]nathanael.beeker-adda@edf.fr, [2]paul.malisani@edf.fr
[*]*CAS, MINES ParisTech, PSL Research University*
60 boulevard Saint-Michel, 75272 Paris Cedex, FRANCE
[3]nicolas.petit@mines-paristech.fr

Abstract

This article studies some of the statistical properties of domestic hot water (DHW) consumption. The study is based on experimental data obtained for a group of households, over a 1 year period. To describe the consumptions, we consider three representative statistical properties of the drain sequence associated to each household: i) the distribution of the magnitude of the drains, ii) a daily pattern of the start times of consumptions, iii) the time between two successive drains. A remarkable outcome of the study is that the time between two successive drains follows a bimodal Weibull distribution. This opens perspectives for piloting applications.

Keywords – energy storage, hot water consumption, stochastic, smart operator

1. Introduction

The increasing share of intermittent renewable electricity sources in the energy mix [2, 3] raises new difficulties in management of the electricity production and equilibrium in distribution networks. Demand Side Management (DSM), which is a portfolio of smart piloting techniques aiming at modifying consumers' demand, is a promising solution for such concerns [3]. A key factor in developing DSM is the availability of energy storage capacities. In this context, the large pools of electric hot water tanks (EHWT) found in numerous countries appear as a very large potential.

An EHWT is a domestic electric appliance which heats a volume of water with a controllable heating element. The user drains hot water from the EHWT at various times of the day, for its comfort. The state of charge of the tank is a function of heating and domestic hot water consumption. Delivery of hot water at all times is a constraint that must be satisfied at best. To design advanced heating control strategies (in view of various objectives such as reducing operation cost for a single tank, or reaching a desirable load profile for the total consumption of a pool of tanks), a model of the demand is a key element. For this reason, we perform some investigations on the dynamics of domestic hot water (DHW) consumption, which has a random nature.

The water drains appear as a sequence of quasi-instantaneous drains in the scale of the day (see Fig. 1). As can be observed on experimental data, the time of occurrence of these drains is not fixed, but is stochastic. In the literature, studies have focused on describing weekly or seasonal consumption patterns, or have presented hour-per-hour mean consumptions [4, 5, 6]. One such typical approach can be found in [7], in which an approach based on aggregation of types of uses allows to generate minute-per-minute load profiles, and in [8], in which forecast over 2 days are generated using the Kalman filter. Nevertheless, more advanced stochastic modeling of the magnitude of these drains and of the temporal correlation in the sequence, based on data analysis, would represent some valuable ways of improvement to design heating strategy for the tanks. This article aims at establishing such statistical properties.

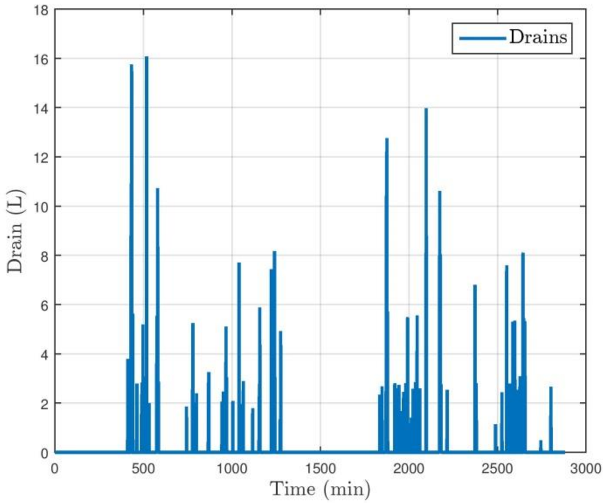

Fig. 1 An example of sequence of drains for an household over 48h.

Data have been gathered in 11 distinct households with EHTW equipped with flow meters and temperature sensors, over periods ranging from 292 to 337 days, with a sampling time of one minute. In this paper, using data, we identify statistical characteristics of DHW consumption. This is the main contribution of this article.

The paper is organized as follows. In Section 2, we identify three main statistical features. Then, in Sections 3 to 5, model of these characteristics are presented. Conclusion and perspective are drawn in Section 6.

2. Characteristics of domestic hot water consumption

Domestic Hot water consumption aggregates various uses in a household: bathing, cleaning, cooking, etc. The durations of resulting drains range from a few seconds to a few minutes. They can be represented as quasi-instantaneous events of various magnitudes in the scale of the day. These drains can be described as a volume of hot water or a quantity of energy taken from the tank (the energy contained in hot water being defined with respect to a cold water temperature reference).

To model the drain sequence, one needs to describe when the drain happens, and how much hot water is consumed during those drains (we make the assumption that two drains occurring within a single minute correspond to a single bigger drain).

The following statistical properties are considered: *i)* the distribution of the magnitude of the drains, *ii)* a daily pattern of the start times of consumptions, *iii)* the time between two successive drains.

This defines a stochastic process for the total hot water consumption $HWC(t)$ at time t of the form

$$HWC(t) = \sum_{j=1}^{+\infty} M_j \theta(t - t_j) \quad (1)$$

where t_j and M_j are the time of occurrence and the magnitude of the drain j, resp., and $\theta(t)$ is the Heaviside function: $\theta(t) = 0$ if $t < 0$ and $\theta(t) = 1$ if $t \geq 0$.

Classically, the magnitude of the drains can be characterized through frequentist inference [9], by estimating a probability density function from the frequency of the data. The results are presented in Section 3.

The times of occurrence of the drains are more complex to describe. The consumption start times are related to the number of persons in the household and their domestic habits. As will be shown in Section 4, they are distributed according to an average daily pattern which is reported. Then, the time between two successive drains follows a bimodal Weibull distribution. This is shown in Section 5.

3. Frequentist inference for the water drains distribution law

A simple way to model the diversity of magnitude of the drains is to represent the drains as random variables drawn from the same probability law. Since the drains are positive, this probability law can be represented with a probability density function on $[0,+\infty[$ (see Fig. 2).

The probability distribution reported in Fig. 2 has been estimated with a frequentist approach to inference [9], for all households and all measurements. The displayed distribution law is the frequency of occurrence of each drain for a large number of measurements.

An interesting observation is that four distinct peaks are visible in the filtered distribution function. This is in accordance with the results presented in [7], obtained by considering four types of uses (small and medium drains, shower bath and bath tub).

If needed, this result could be refined by taking into account that the measurements are strongly related to the habits of the households (e.g. in term of bathing), the time of the day and the season. For households equipped with additional sensors, the conditional distribution could be estimated online. However, we do not consider this in this article.

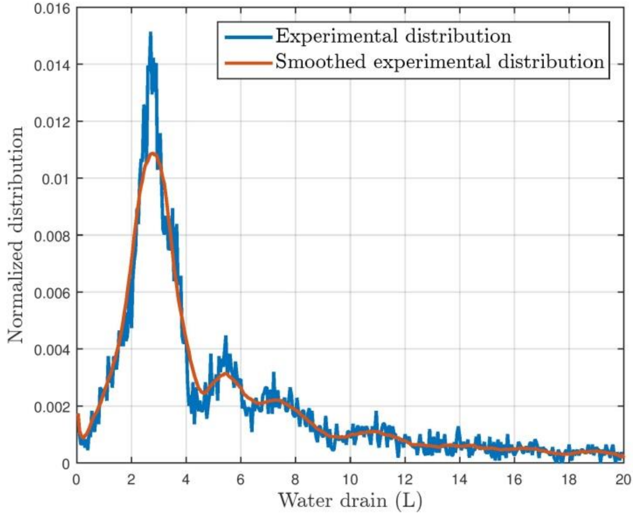

Fig. 2 Experimental probability distribution of drain magnitudes (over the whole recorded sequences).

4. Daily pattern for the start times of drains

The hot water consumption is strongly related to the habits of the house occupants. Also, it is clearly visible in the data (as could be expected) that the drains mostly take place at certain times of the day (see Fig. 1). In fact, a daily pattern can be defined. For this, one can simply consider the mean value of the number of drains at a given time of the day, for each household. The procedure is as follows. This procedure is commonly used in statistics (e.g. to determine the intensity function of a non-homogeneous Poisson process [10]).

Given a household, we represent each of its n set of 24 hours measurements (labeled by k=1…n) with a function $N_k(t)$ that represents the number of drains having occurred in this day over the interval $[0, t]$, for $t \in [0, 24\text{h}]$. In other words, consider any day k and the sorted drain times $t_1^k, \dots, t_m^k$, then $N_k(t)$ is defined as

$$N_k(t) = \sum_{j=1}^{m} \theta\left(t - t_j^k\right). \tag{2}$$

Then, we define the daily pattern as the mean of these functions

$$M(t) = \frac{1}{n} \sum_{k=1}^{n} N_k(t)\ . \tag{3}$$

This function, defined over the time interval $[0, 24\text{h}]$, gives, for any household, the average (expected) number of drains that should have occurred at time t (see Fig. 3): it is representative of the habits of this household in terms of frequency of the drain. The function (3) obtained with the discussed dataset is reported in Fig. 3. The presence of sharp transients between low slope regions shows that the average function (2) is representative of the individual behaviors in each household. High slopes correspond to time periods with frequent drains.

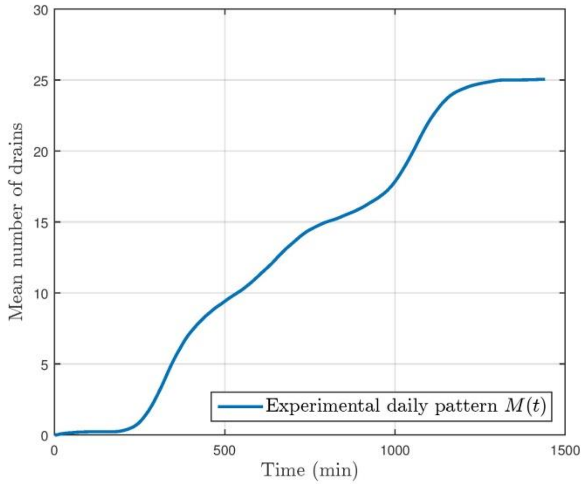

Fig. 3 Averaged daily cumulative number of drains

5. Distribution of the time between two successive drains

The time between two successive drains $y_i^k = t_i^k - t_{i-1}^k$ (during a day k) is also related to the habits of the house occupants. Therefore, we isolate data accordingly.

In the case when the number of drains over time $N_k(t)$ are samples of a non-homogeneous Poisson process (i.e. the drains are not correlated), then it is possible to construct adjusted increments with the mean value function $M(t)$, that must follow an exponential distribution. As will appear, this assumption is not valid in our case, but we still perform the same analysis, yielding different conclusions. The construction is done the following way.

For each period k of 24 hours, a set is constructed by taking the image of each drain time $t_1^k, \dots, t_m^k$, through the function M (see Fig. 4). The $M(t_i^k)$ are then used to define the set of successive increments

$$x_i^k = M\left(t_i^k\right) - M(t_{i-1}^k)\ . \tag{4}$$

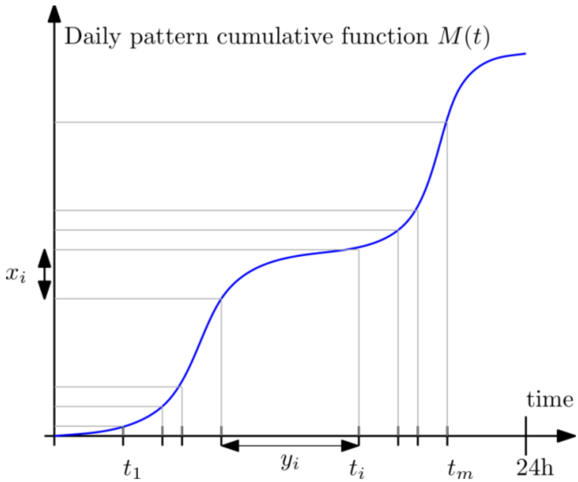

Fig. 4 Construction of the ajusted increments

These variables have an interpretation: for any day k, x_i^k is the average increment of the number of drains from t_{i-1}^k to t_i^k. Its expected value (for all possible k) is 1.

These normalized increments (represented in the ordinate axis in Fig. 4) are assumed to be independent from the choice of the households. They are representative of the correlation between successive drains.

The x_i^k are distributed in the $[0, +\infty[$ interval. Using the whole dataset, an experimental cumulative distribution function (CDF) can be obtained. It is reported in Fig. 5. Remarkably, this function shows a good fit to $f(x) = 1 - exp(x^\alpha / \rho)$, which corresponds to the CDF of the Weibull distribution of shape parameter α and scale parameter ρ [11]. Among other possibilities (exponential distribution, Gamma distribution, Beta distribution), this is (by far) the best fit.

The Weibull distribution is commonly used to model failure rates over time. In the literature, it is often associated with Autoregressive Conditional Duration (ACD) models [12] to forecast the distribution of a succession of duration times (e.g. time lapse between two transactions in stock market [13]). In Weibull models, the parameter α defines the nature of the process. The case $\alpha = 1$ corresponds to an exponential distribution. If $\alpha < 1$, then the occurrence of an event raises the probability of a closely following event (cases of immediately consecutive events are frequent). On the contrary, the case $\alpha > 1$ corresponds to the case when successive events are spaced out.

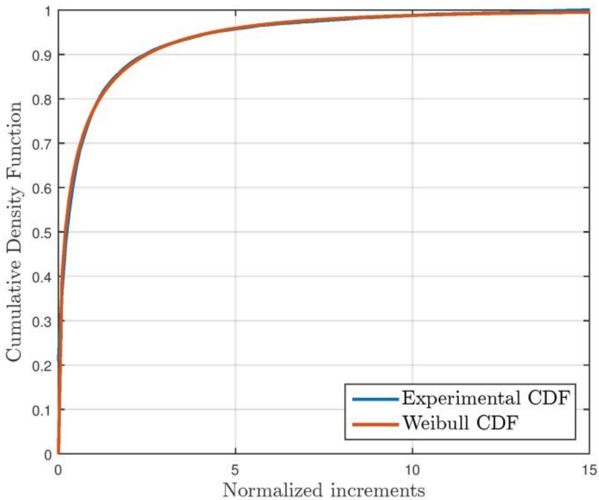

Fig. 5 Experimental CDF for the ajusted increments compared to the Weibull CDF

To establish that the experimental data follow the assumed distribution, a graphical method can be employed. The probability plot consists in plotting two CDF, one against the other: if the two distributions are similar, the points of the probability plot should lie on a straight line. Further, for the Weibull distribution, the shape parameter α can simply be deduced from the slope [11].

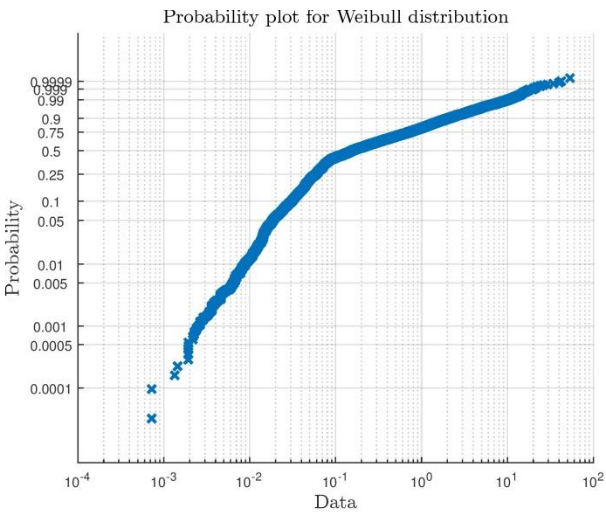

Fig. 6 Probability plot of the data against the Weibull distribution

Interestingly, the probability plot in Fig. 6 displays two distinct lines with different slopes. This is a well-identified characteristic of a multimodal or mixed Weibull distribution, indicative of the fact that several shape parameters coexist [14, 15]. In our case, we propose the following interpretation. It is likely that two different timescales come into play in the time correlation. Those timescales can correspond to the fact that the user is at home or not, doing an activity that requires hot water. The two modes have parameters $\alpha < 1$, i.e. at all times during the day the occurrence a water drain strongly suggests that another water drain should be expected right after.

6. Conclusion and perspectives

In this article, we have identified three distinct statistical characteristics present in a dataset of time series of DHW consumption. According to the investigations conducted on the presented dataset, the conclusions are well grounded but they remain preliminary. If these characteristics remain valid for more various cases (e.g. in different countries or with a wider range in the type of households), then this would pave the way to interesting works on stochastic modelling of hot water consumption.

These perspectives are important for smart piloting applications of EHWT. Indeed, the identification of the time lapse between drains with a Weibull distribution appears as particularly useful, since it stresses that if the user is draining water, he is very likely to drain more water shortly. Considering this likely event, piloting strategies can be updated online to take into account the high probability of future drains. Future works should include the definition, if possible, of a global Weibull ACD model that can be fitted online for each household.

At another scale, for a large number of EHWT, stochastic modelling of the consumption opens the way to the development of models for the distribution of the state of charge of the tanks, using e.g. Fokker-Planck equations.

Acknowledgment

EDF Lab is gratefully acknowledged for making available the dataset on domestic hot water consumption.

References

[1] European Commission. Energy roadmap 2050: Communication from the Commision to the European Parliament, the Council, the European Economic and Social Committee abd the Committee of the Regions (2011).

[2] O. Edenhofer, R. Piches-Madruga, Y. Sokona and K. Seyboth. Report on renewable energy sources and climate change mitigation. Technical report, Intergovernmental Panel on Climate Change (2011).

[3] P. Palensky and D. Dietrich. Demand side management: Demand response, intelligent energy systems, and smart loads. IEEE Transactions on Industrial Informatics (2011).

[4] E. Vine, R. Diamond and R. Szydlowski. Domestic hot water consumption in four low-income apartment buildings. Energy, Vol. 12, No. 6, 459-467 (1987).

[5] J.P. Meyer and M. Tshimankinda. Domestic hot water consumption in South African houses for developed and developing communities. Int. Jour. of energy research, Vol. 21, 667-673 (1997).

[6] K.T. Papakostas, N.E. Papageorgiou andB.A. Sotiropoulos. Residential hot water use patterns in Greece. Solar Energy, Vol. 54, No. 6, 369-374 (1995).

[7] U. Jordan and K. Vajen. Influence of DHW load profile on the fractional energy savings: a case study of a solar combi-system with TRNSYS simulations. Solar Energy, Vol. 69, Nos 1-6, 197-208 (2000).
[8] T. Prud'homme, D. Gillet. Advanced control strategy of a solar domestic hot water system with a segmented auxiliary heater. Energy and Buildings, Vol. 33, 463-475 (2001).
[9] J. Neyman. Outline of a Theory of Statistical Estimation Based on the Classical Theory of Probability. Philosophical Transaction of the Royal Society of London A, 236, 333-380 (1937).
[10] C. Ruwet. Processus de Poisson. Université de Liège (2007).
[11] D.N. Prabhakar Murthy, M.Xie, R. Jian. Weibull models. Wiley (2004).
[12] S. Crowley. Point Process Models for Multivariate High-Frequency Irregularly Spaced Data (2013).
[13] R. Tsay. Autoregressive Conditional Duration Models. In: T. Milles and K. Patterson. Handbook of Econometrics II, Palgrave Publishing Company (2007).
[14] N. Doganaksoy, G.J. Hahn and W.Q. Meeker. Reliability Analysis by Failure Mode. Quality Progress of June 2002, 47-52 (2002).
[15] D. Kececioglu. Reliability Engineering Handbook : Volume 1. DEStech Publications (2002).