DTU Library

Non-parametric method for separating domestic hot water heating spikes and space heating

Bacher, Peder; de Saint-Aubain, Philip Anton; Christiansen, Lasse Engbo; Madsen, Henrik

Non-parametric method for separating domestic hot water heating spikes and space heating

Peder Bacher, Philip Anton de Saint-Aubain, Lasse Engbo Christiansen, Henrik Madsen

DTU Compute,
Technical University of Denmark,
Matematiktorvet, Building 303B
2800 Kgs. Lyngby Denmark.
pbac@dtu.dk

Abstract

In this paper a method for separating spikes from a noisy data series, where the data change and evolve over time, is presented. The method is applied on measurements of the total heat load for a single family house. It relies on the fact that the domestic hot water heating is a process generating short-lived spikes in the time series, while the space heating changes in slower patterns during the day dependent on the climate and user behavior. The challenge is to separate the domestic hot water heating spikes from the space heating without affecting the natural noise in the space heating measurements. The assumption behind the developed method is that the space heating can be estimated by a non-parametric kernel smoother, such that every value significantly above this kernel smoother estimate is identified as a domestic hot water heating spike. First, it is showed how a basic kernel smoothing approach is too simple to deliver reliable results. Therefore the problem is generalized to a local least squares problem, which makes it possible to design a robust kernel smoother, which estimate is not affected by the spikes. Furthermore, the generalized model makes it possible to estimate higher order local polynomials. Finally, the results are evaluated and it is found that the method is capable of calculating a reliable separation of the total heat load into the two components.

Keywords:
Separation of total heat load, kernel smoother, robust estimation, statistical modeling, time series analyses, smart grid, smart metering, heat metering

1. Introduction

The energy performance of buildings can be assessed in many ways. As more and more data is collected on the actual energy consumption data driven assessment becomes feasible. When analyzing the total energy consumption of a building it is important to acknowledge that it is composed of several components. In many cases the two most important components are related to the building envelope and the occupancy behavior, and for labeling the energy performance of a building it is important to consider which components to include - often the interest is in the envelope.

In this paper a method for separating the total heat load into domestic hot water (DHW) heating and space heating is presented. Data from an individual residential building located in Denmark is used. It consists of a time series of 10 minute values of total heat load, which is the sum of DHW and space heating. DHW is used by the inhabitants for showering, dish washing, etc., and the space heating is used to heat the house. The DHW heating generates spikes added to the (or on top of) the space heating. This is due to the fact that showering and dish washing use intense amount of energy in a short period. A commercial opportunity for this study is that the number of sensors needed can be reduced, since the DHW and space heating don't need to be measured separately, i.e. the same information can be retrieved with a single heat flow meter instead of two. Thus system costs can be decreased and often only the total heat load is available, thus the method can be useful in such cases for data analysis purposes.

The described method for separating the total heat load is quite generic and can therefore easily be used for other applications, where spikes must be separated from other signals. The separation can be useful for building energy performance estimation based on data ([1] and [2]) and for load forecasting where the presented method was actually used [3]. The separated DHW load can be used for example for constructing load profiles for DHW ([4] and [5]), the latter using in-homogeneous Markov chain models providing a fully data-driven stochastic modelling approach. Another important application is in control for heating systems enabling demand response for integration of renewables, for example by using a hot water tank [6] or the building structures [7] for energy storage.

Separating consumption signals into sub-components has been studied quite in-

tensively the last decades, mainly for electrical appliance load monitoring ([8] and [9]), where the electrical load is dis-aggregated into event categories. Also residential water consumption dis-aggregation into end-use categories has been studied [10], where high resolution readings (5 sec.) were used. Such methods are event based, where patterns are matched and related to the properties of appliances or events. Methods for spike detection has been studied in-depth in for example medical applications such as electroencephalography (EEG). Several approaches for determining amplitude thresholds are used. For example different statistical approaches [11], as well as filtering and wavelet approaches [12]. A comprehensive review of different techniques is given by [13].

In the present study a statistical time series approach [14] based on kernel smoothing techniques for time series ([15] and [16]) is used and combined with robust estimation (see [17] and [18]). The proposed method enables separation of very high spikes to be carried out without interfering with the remaining signal. The method is based on the idea of using a non-parametric model (i.e. the kernel smoother) to estimate the space heating. The space heating changes over time as it is a low-pass filtered response of mainly the outdoor temperature and the solar radiation. Consequently, the space heating changes at frequencies related to those variables, thus in a slow moving and rather smooth pattern. The DHW generates spikes (a noisy and high frequency component) added to the slower changing space heating signal. The method is therefore designed such that the non-parametric estimate follows the slower changes only, without being influenced by the spikes - which is possible by using a robust estimation scheme. Since the estimate is not affected by the spikes they can be identified and separated from the space heating signal.

Outline
In Section 2 the data used for the study is presented and in Section 3.1 a simple kernel smoother model applied to separate the DHW spikes from the data is described. In Section 3.2 a robust estimation scheme is presented, and in Section 3.3 it is described how the method is extended further by using a second order polynomial model. Finally, in Section 4 the results are analyzed and in Section 5 the conclusions are drawn.

2. Data

The data used in the study consists of the total heat load of a single-family free standing residential building with two occupants. The total heat load is the sum of DHW heating used for heating water for showering, dish washing, etc., and space heating used for heating the building. Sønderborg Fjernvarme (consumer owned district heating company) located in the southern part of Denmark delivered the data. The period used is covering one month from 1st of March to 2st of April 2010. The data was logged every 10th minute. The total heat load is represented with the time series

$$\left\{ Q_t, \quad t = 1, \ldots, N \right\} \tag{1}$$

where Q_t is the value at time t and $N = 4607$ is the number of observations in the times series, i.e. equidistant sample points. Per convention $\dot{Q}$ is used for denoting the rate of heat, however the dot is left out to simplify the notation, such that Q_t is the heat transferred in the time interval from $t-1$ to t and the unit of the heat series is kept in megajoule per hour [MJ/h]. The upper plot of Figure 1 shows the raw data from this period. Some of the spikes are as high as 160 MJ/h and have been cut off by the frame of the plot in order to make the lower variations visible. It can be seen that in a two week period from Friday 12th until Friday 26th there are no spikes and the total heat load has very little variation. It is assumed that the inhabitants were on holiday and left the house with the usual space heating during these two weeks. The inclusion of this holiday period in the evaluation provides a further opportunity to evaluate the separation, since the models should predict that no DHW heating is used in the period. The lower plot of Figure 1 shows four days covering the transition to the period with no spikes, and in this plot also the characteristics of the spikes can be seen in more detail.

3. Models

In the following the applied kernel smoother models are presented. It is assumed that the spikes represent DHW heating and the remaining signal represents space heating. The kernel smoother estimates the space heating and thus spikes significantly higher than the kernel smoother estimate are identified as DHW heating. First, a simple zero order kernel smoother is presented and thereafter, in Section 3.2 and 3.3, improvements of this simple model are suggested.
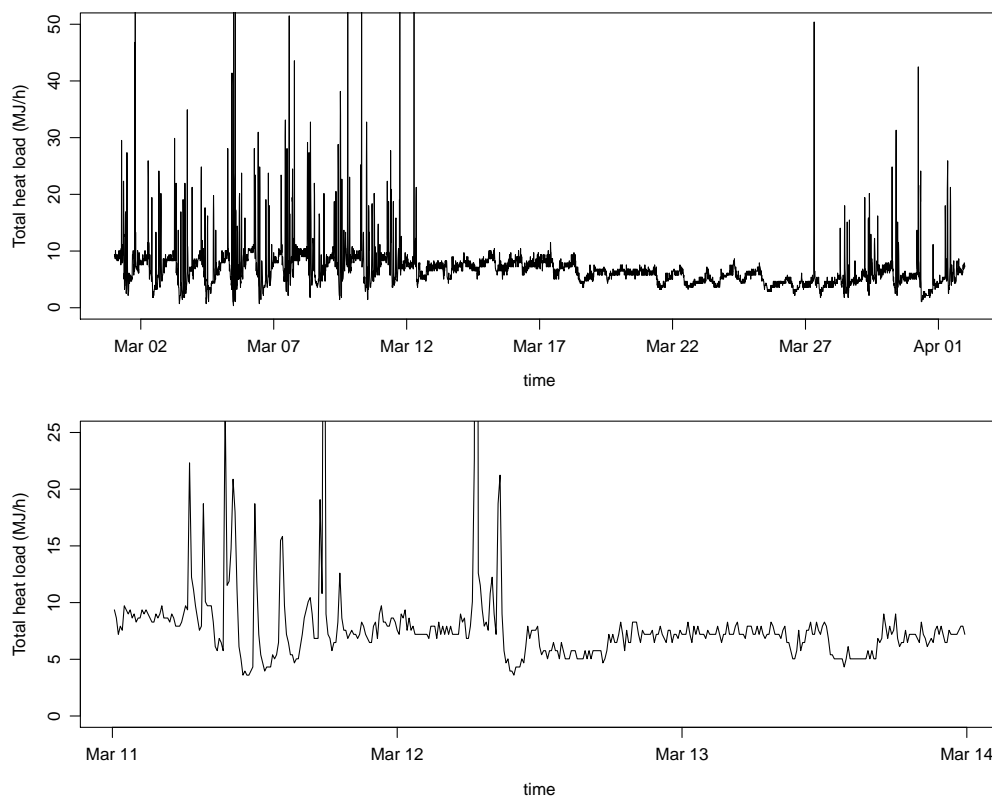
Figure 1: The upper plot is of the total heat for March 2010 and the lower plot is for the four days period, where the spikes end and it is assumed that the inhabitants leave the house.

3.1. Zero order kernel smoother

A kernel smoother is a method to estimate the underlying function of some given noisy measurements. Kernel estimation is a non-parametric estimation technique, where no explicit description of the true function is needed and only a bandwidth parameter needs to be set [16]. The kernel smoother is

$$\hat{g}(t) = \sum_{i=1}^{N} \frac{Q_i k\left(\frac{t-i}{h}\right)}{\sum_{i=1}^{N} k\left(\frac{t-i}{h}\right)} \tag{2}$$

where $\hat{g}(t)$ is the kernel estimate for a given time t, $k(\cdot)$ is the kernel function and h is the bandwidth parameter. From the formula it is seen that the kernel smoother is a local weighted average around the given time t, hence a zero order local estimate. The kernel function $k(\cdot)$ determines how the weight should be put on the neighboring data points. The Gaussian kernel $k(u) = \frac{1}{2\pi} \exp\{-\frac{u^2}{2}\}$ is used. The bandwidth h is a smoothing parameter which determine the width of the kernel. As $h \to \infty$ the estimate will go towards the average $\frac{1}{N} \sum_{t=1}^{N} Q_t$. Therefore for too large values of h the kernel estimate will be biased. As $h \to 0$ the kernel estimate would just be equal to the nearest data points and there will be no bias, but a large variance of the estimate. Hence the bandwidth needs to be tuned for the particular data. In the present case the value were set according to a manual tuning, which gave a good result when visually inspecting the separation, however for applications - where manual tuning is not an option - a scheme for tuning the bandwidth (and the two other parameters presented below) should be developed in further studies. In the present case a bandwidth equal to $h = 12$ (which is 2 hours) is found adequate. This results in the kernel seen in Figure 2.

The separation is carried out by identifying the DHW heating by

$$\hat{Q}_t^{\text{water}} = I(Q_t > q_{\text{thres}} \, \hat{g}(t)) \, (Q_t - \hat{g}(t)) \tag{3}$$

where $I(\cdot)$ is the indicator function. Hence spikes above $q_{\text{thres}} \cdot \hat{g}(t)$ are identified as DHW heating and the value of them are found by subtracting the kernel estimate. The separation threshold q_{thres} needs to be tuned and it should be set related to the local variance of noise in the space heating signal, such that the spikes are significantly higher than this noise level. In the present case q_{thres} is set to 1.3. Since only one time series is available a scheme for tuning of q_{thres} is left for future studies where many different series can be included. The space heating is found simply by subtracting DHW heating from the total heat load

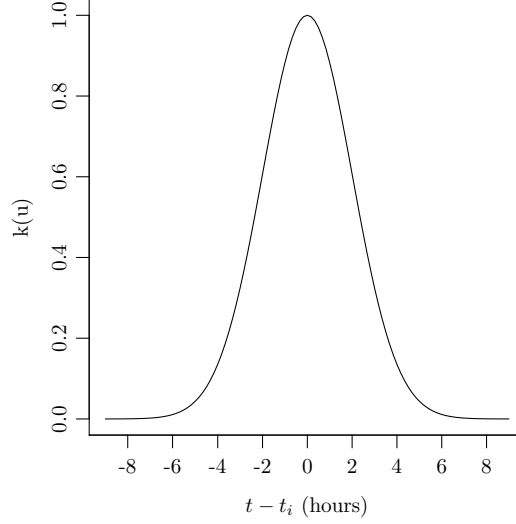$$\hat{Q}_t^{\text{space}} = Q_t - \hat{Q}_t^{\text{water}} \tag{4}$$

Figure 2: The normal kernel with $h = 12$ used for the smoothing.

The result of the separation is shown in Figure 3. The uppermost plot shows the original data together with $1.3 \cdot \hat{g}(t)$. The middle plot shows the DHW heating. The lower plot shows the space heating. Therefore, the sum of the two lower plots is equivalent to the uppermost plot. It is seen that the estimated space heating has many short-lived spikes especially during periods with many spikes, this is not appropriate. The reason is that the kernel estimate is too affected by the spikes, which leads to smaller spikes being below $1.3 \cdot \hat{g}(t)$. The next sections are dedicated to show how a modification by using a suggested robust kernel smoother can prevent the estimate from being affected by the high spikes.

3.2. Robust zero order kernel smoother

The idea behind robust estimation is to make the estimation method robust against outliers or extremes [18]. Optimization methods generally try to minimize some function $\rho(\varepsilon)$ of the residuals ε. In this case the simple zero order kernel estimator in Equation (2) is a zero order local regression model [19] and can be formulated as

$$\hat{g}(t) = \arg\min_{\theta} \frac{1}{N} \sum_{i=1}^{N} k\left(\frac{t-i}{h}\right)(Q_i - \theta)^2 \tag{5}$$
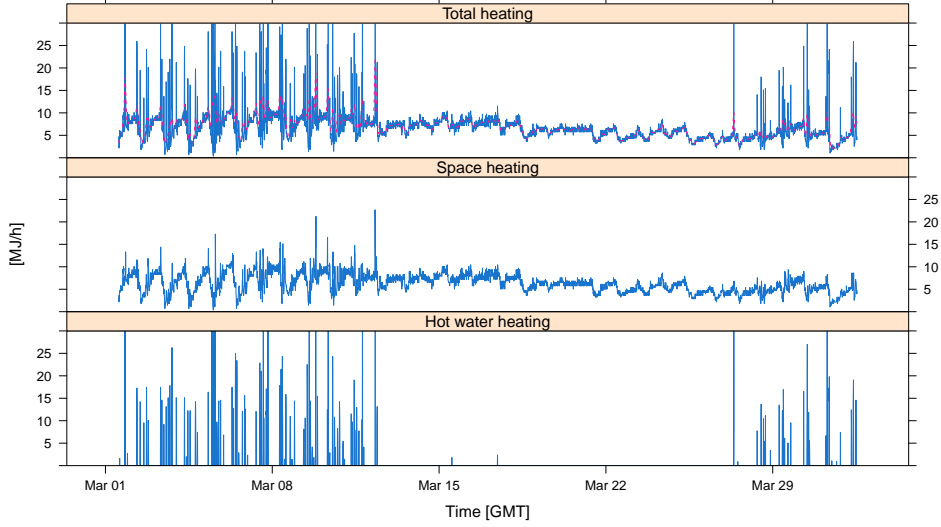
7

Figure 3: Result of the separation with the zero-order kernel smoother. The red dashed line is $1.3 \cdot$ kernel estimate.

where the residuals minimized are $\varepsilon_i = Q_i - \theta$, hence the least squares method minimizes the sum of the squared residuals $\rho_{\mathrm{LS}}(\varepsilon) = \varepsilon^2$. The estimation is made robust by replacing the quadratic function with Tukey's biweight function, also known as the bisquare function, see [17]. The biweight estimation minimizes the following function

$$\rho_{\mathrm{biweight}}(\varepsilon) = \begin{cases} \frac{1}{6} \frac{\varepsilon^2 \left(\varepsilon^4 - 3\,\varepsilon^2 \gamma^2 + 3\,\gamma^4 \right)}{\gamma^4} & \text{if } |\varepsilon| \leq \gamma \\ \frac{1}{6}\,\gamma^2 & \text{if } |\varepsilon| > \gamma \end{cases} \tag{6}$$

The biweight function is approximately quadratic for small residuals and constant for residuals larger than γ. A plot of $\rho_{\mathrm{biweight}}(\varepsilon)$ and a scaled version of $\rho_{\mathrm{LS}}(\varepsilon)$, together with their derivatives are shown in Figure 4. The derivative is also known as the influence function. The biweight function induces that outliers do not cause displacement of the resulting estimate. For residuals further away than the γ limit the influence function $\rho'(\varepsilon)$ is zero. The parameter γ is a selected threshold determining when residuals are considered large. For the actual heating data a reasonable value found to be $\gamma = 7$ Mj/h.

For a given time t the robust zero order estimate is found by solving the opti-
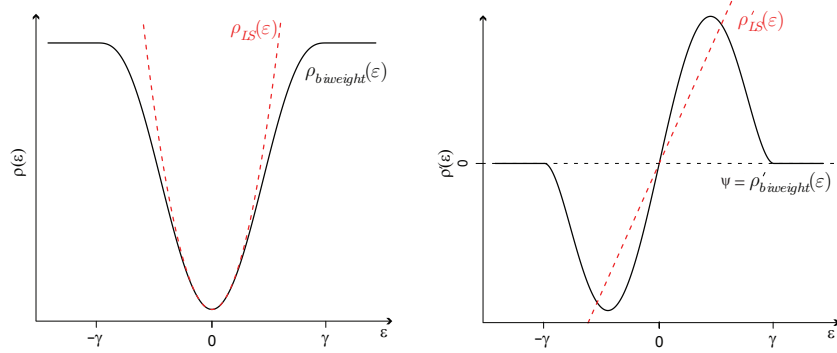
Figure 4: Left: Tukey's biweight and a square function. Right: The derivatives also known as the influence function.

mization problem

$$\hat{g}(t) = \arg\min_{\theta} \frac{1}{N} \sum_{i=1}^{N} k\left(\frac{t-i}{h}\right) \cdot \rho_{\text{biweight}}(Q_i - \theta) \tag{7}$$

Hence the residuals are $\varepsilon_i = Q_i - \theta$. The result of the separation using this model is shown in Figure 5. It is seen that almost all of the spikes in the heating are removed compared to the separation using the simple kernel, and thus by using the robust kernel the problem of the estimate being too affected by the large spikes is solved.

3.3. Robust second order kernel smoother

The previous models were based on a zero order local estimate (i.e. the locally weighted average). Applying a higher order polynomial kernel smoother might improve the performance of the estimator, since then the higher order moments of the local estimation functions are included in the model. This can be important when for load profiles with fast changes e.g. caused by a nightly set-back resulting in a significant diurnal profile. A robust second order local model, which includes the curvature of the locally estimated function, is applied by using the cost function

$$\theta(t) = \arg\min_{\boldsymbol{\theta}} \frac{1}{N} \sum_{i=1}^{N} k\left(\frac{t-i}{h}\right) \cdot \rho_{\text{biweight}}(Q_i - \theta_0 + \theta_1(t_i - t) + \theta_2(t_i - t)^2) \tag{8}$$
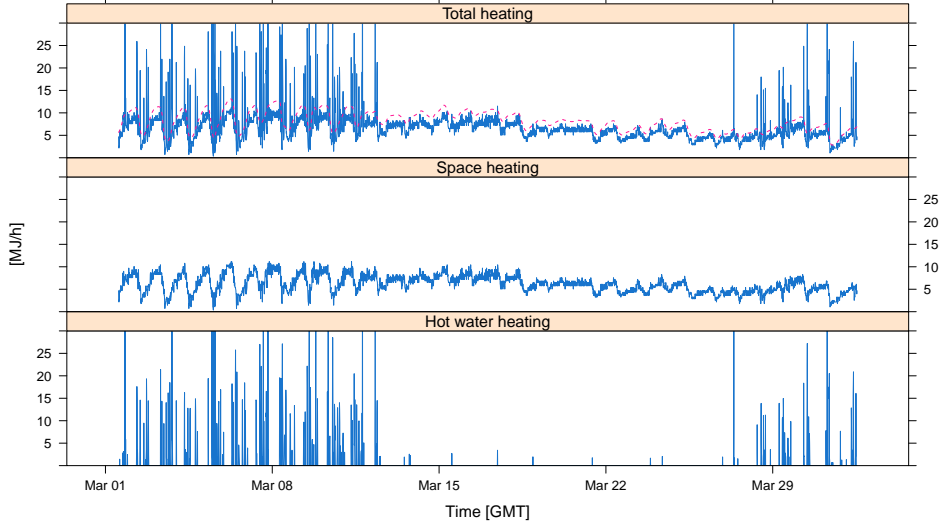
$$\hat{g}(t) = \theta_0(t) \tag{9}$$

9

Figure 5: Result of the separation with robust zero order kernel smoother. The red dashed line is $1.3 \cdot$ kernel estimate.

Hence for any given time t the minimization of the cost function is carried out for $\boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2)$ and the intercept θ_0 of the locally fitted polynomial is the estimate. This model has less tendency to get biased in high curvature points. Therefore a higher bandwidth can be used making it even less affected by the spikes. The bandwidth is increased to $h = 18$. The result of separating the total heat load into DHW and space heating using the robust second order kernel smoother is shown in Figure 6. Comparing the plots for the separation using the robust zero order and the robust second order smoother only marginal differences are seen. Hence for the present heating series it is not found necessary to use an higher order estimator and the robust zero order smoother is found to be a sufficient model.

4. Discussion

Certainly, it is needed to apply the method on many series with different patterns, and where DHW and space heating are measured separately in order to validate the performance more thoroughly. This will also enable the construction of a scheme for automatically tuning the parameters: the kernel bandwidth h, the separation threshold q_{thres} and the γ threshold for the biweight function in the robust estimation scheme.
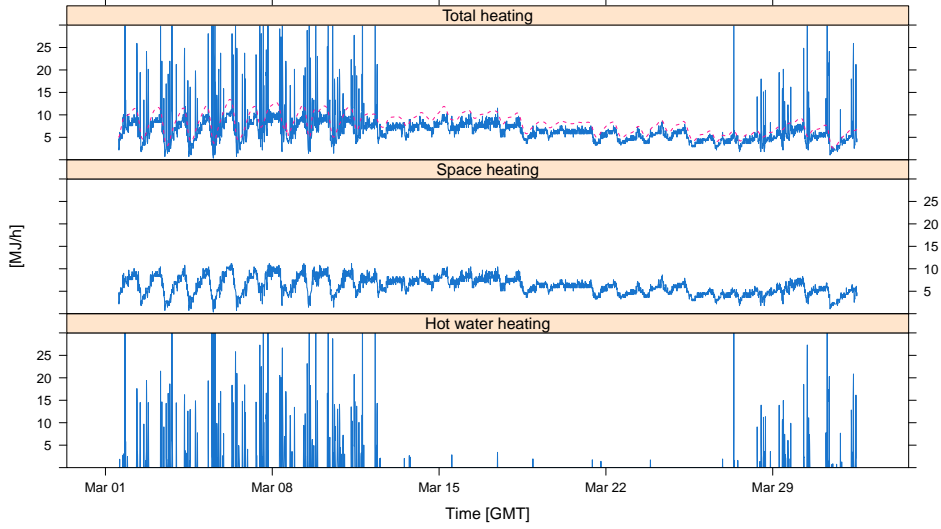
10

Figure 6: Results of the separation with the robust second order kernel smoother. The red dashed line is $1.3 \cdot$ kernel estimate.

Only marginal differences was found between the results from the zero order and the second order robust kernel for the present series, however it could be expected that for series with significant diurnal variation, e.g. from a nightly setback, that a higher order model is better.

Other explanatory variables can also be included as inputs to the model, e.g. ambient temperature, solar radiation and electrical consumption. However, one should not expect to explain all the variation in the data in this way as most houses represent a relative large thermal mass that combined with thermostats creates dynamics in the response, hence these slower dynamics are modelled quite well with the suggested model. Also it is mentioned, that a quantile regression model could be used, since this is also robust to outliers. For example the median (50% quantile) or a lower quantile could be estimated. Quantile regression models can be fitted in a similar non-parametric approach using a kernel smoother [20] and such models can also be fitted using a recursive scheme [21], which can enable faster computation times in online applications.

The proportion of the total heat load that is used for space heating obviously depends on the outdoor climate and therefore on the season. This will lead to changes in signal to noise ratio and it would also be of interest to investigate how

11

this approach works for different periods of the year. Another important consideration is the sampling frequency. In this study a sample is recorded every 10 minutes which is on the same time scale as typical DHW usage for dish washing and showering while at the same time being frequent relative to the dynamics of the space heating.

5. Conclusion

A method for separating spikes from a signal is presented using a case study of the total heat load measured in a single family residential house, where the DHW heating is separated from the space heating. It is assumed that the DHW heating are spikes added to the space heating in the total heat load. First a local average kernel is applied, however the spikes affect the estimates and the results were not reliable. Therefore a robust estimation is suggested and it is shown that this enables a much better separation of the DHW and space heating. It is found that the method is very promising and that it can be concluded that the method is useful for separating the DHW heating from space heating. However, it is emphasized that further studies are needed, where the method is applied on many different heating series and where the DHW and space heating are measured separately, for a more in-depth development and verification of the method and its accuracy.

The described method can be used in other fields of operation as well. In any problem where noisy data includes spikes the described method could possibly provide a technique for removing the spikes.

6. Acknowledgments

7. References

[1] A. Rabl, Parameter-estimation in buildings - methods for dynamic analysis of measured energy use, Journal of Solar Energy Engineering-transactions of the ASME 110 (1) (1988) 52–66.

[2] M. Jiménez, H. Madsen, K. Andersen, Identification of the main thermal characteristics of building components using MATLAB, Building and Environment 43 (2) (2008) 170–180. `doi:10.1016/j.buildenv.2006.10.030`.

[3] P. Bacher, H. Madsen, H. A. Nielsen, B. Perers, Short-term heat load forecasting for single family houses, Energy and Buildings 65 (0) (2013) 101–112. `doi:http://dx.doi.org/10.1016/j.enbuild.2013.04.022`.
URL `http://www.sciencedirect.com/science/article/pii/S0378778813002752`

[4] J. Widén, M. Lundh, I. Vassileva, E. Dahlquist, K. Ellegård, E. Wäckelgård, Constructing load profiles for household electricity and hot water from time-use data—modelling approach and validation, Energy and Buildings 41 (7) (2009) 753–768. `doi:http://dx.doi.org/10.1016/j.enbuild.2009.02.013`.
URL `http://www.sciencedirect.com/science/article/pii/S0378778809000413`

[5] P. D. Andersen, A. Iversen, H. Madsen, C. Rode, Dynamic modeling of presence of occupants using inhomogeneous markov chains, Energy and Buildings (0) (2013) . `doi:http://dx.doi.org/10.1016/j.enbuild.2013.10.001`.
URL `http://www.sciencedirect.com/science/article/pii/S0378778813006397`

[6] R. Halvgaard, P. Bacher, B. Perers, E. Andersen, S. Furbo, J. B. Jørgensen, N. K. Poulsen, H. Madsen, Model predictive control for a smart solar tank based on weather and consumption forecasts, Energy Procedia 30 (2012) 270–278. `doi:10.1016/j.egypro.2012.11.032`.

[7] S. Prívara, J. Cigler, Z. Váňa, F. Oldewurtel, E. Žáčeková, Use of partial least squares within the control relevant identification for buildings, Control Engineering Practice 21 (1) (2013) 113–121. `doi:http://dx.doi.org/10.1016/j.conengprac.2012.09.017`.
URL `http://www.sciencedirect.com/science/article/pii/S0967066112002006`

[8] G. Hart, Nonintrusive appliance load monitoring, Proceedings of the IEEE 80 (12) (1992) 1870–1891. `doi:10.1109/5.192069`.

[9] L. Farinaccio, R. Zmeureanu, Using a pattern recognition approach to disaggregate the total electricity consumption in a house into the major end-uses, Energy and Buildings 30 (3) (1999) 245–259. `doi:http://dx.doi.org/10.1016/S0378-7788(99)00007-9`.
URL `http://www.sciencedirect.com/science/article/pii/S0378778899000079`

[10] K. Nguyen, R. Stewart, H. Zhang, An intelligent pattern recognition model to automate the categorisation of residential water end-use events, Environmental Modelling & Software 47 (0) (2013) 108–127. `doi:http://dx.doi.org/10.1016/j.envsoft.2013.05.002`.
URL `http://www.sciencedirect.com/science/article/pii/S1364815213001084`

[11] L. E. Christian, D. O. Everson, S. L. Davis, A statistical method for detection of hormone secretory spikes, Journal of Animal Science 46 (3) (1978) 699–706. `arXiv:http://www.journalofanimalscience.org/content/46/3/699.full.pdf+html`.
URL `http://www.journalofanimalscience.org/content/46/3/699.short`

[12] R. Q. Quiroga, Z. Nadasdy, Y. Ben-Shaul, Unsupervised spike detection and sorting with

wavelets and superparamagnetic clustering, Neural Computation 16 (8) (2004) 1661–1687. `doi:10.1162/089976604774201631`.
URL `http://dx.doi.org/10.1162/089976604774201631`

[13] S. B. Wilson, R. Emerson, Spike detection: a review and comparison of algorithms, Clinical Neurophysiology 113 (12) (2002) 1873–1881. `doi:http://dx.doi.org/10.1016/S1388-2457(02)00297-3`.
URL `http://www.sciencedirect.com/science/article/pii/S1388245702002973`

[14] H. Madsen, Time Series Analysis, Chapman & Hall/CRC, 2007.

[15] V. A. Epanechnikov, Non-parametric estimation of a multivariate probability density, Theory of Probability & Its Applications 14 (1) (1969) 153–158.

[16] P. M. Robinson, Nonparametric estimators for time series, Journal of Time Series Analysis 4 (3) (1983) 185–207. `doi:10.1111/j.1467-9892.1983.tb00368.x`.
URL `http://dx.doi.org/10.1111/j.1467-9892.1983.tb00368.x`

[17] P. J. Huber, Robust Statistics, John Wiley & Sons, 2003.

[18] P. J. Rousseeuw, A. M. Leroy, Robust regression and outlier detection, Vol. 589, Wiley.com, 2005.

[19] J. Friedman, T. Hastie, R. Tibshirani, The elements of statistical learning, Vol. 1, Springer Series in Statistics, 2001.

[20] R. Koenker, Quantile Regression, Cambridge University Press, 2005.

[21] J. K. Møller, H. A. Nielsen, H. Madsen, Time-adaptive quantile regression, Computational Statistics and Data Analysis 52 (3) (2008) 1292–1303.