Electricity Consumption Time Series Profiling: A Data Mining Application in Energy Industry

Hongyan Liu, Zhiyuan Yao, Tomas Eklund, and Barbro Back

Turku Centre for Computer Science,
Åbo Akademi University, Department of Information Technologies
Joukahaisenkatu 3-5, 20520 Turku, Finland
{Hongyan.Liu,Zhiyuan.Yao,Tomas.Eklund,Barbro.Back}@abo.fi

Abstract. The ongoing deployment of Automated Meter Reading systems (AMR) in the European electricity industry has created new challenges for electricity utilities in terms of how to fully utilise the wealth of timely measured AMR data, not only to enhance day-to-day operations, but also to facilitate demand response programs. In this study we investigate a visual data mining approach for decision-making support with respect to pricing differentiation or designing demand response tariffs. We cluster the customers in our sample according to the customers' actual consumption behaviour in 2009, and profile their electricity consumption with a focus on the comparison of two sets of seasonal and time based variables. The results suggest that such an analytical approach can visualise deviations and granular information in consumption patterns, allowing the electricity companies to gain better knowledge about the customers' electricity usage. The investigated electricity consumption time series profiling approach will add empirical understanding of the problem domain to the related research community and to the future practice of the energy industry.

Keywords: Visual Data Mining, Clustering, Business Intelligence, Electricity Consumption Profiling, Self-Organizing Maps, Deviation Detection.

1 Introduction

Within the electricity industry, the deployment of Automated Meter Reading (AMR, i.e., remotely-readable, two-way communication smart meters) has been a topical issue for some time, especially in Europe. The progress of such deployment varies across EU countries. While Italy and Sweden have completed their nation-wide smart meter installations, and Finland is due to finish its large scale rollout to both commercial and household customers by 2013, other countries such as the UK and Belgium are still in the trial or cost-benefit analysis stage. It is well-acknowledged by the electricity industry that the deployment of smart meters and smart metering will benefit the electricity distribution business in several ways. On the one hand, short term benefits will include more efficient and accurate billing, customer services, fault detection and automated healing, just to name a few, while in the long run, it could facilitate the development of

P. Perner (Ed.): ICDM 2012, LNAI 7377, pp. 52–66, 2012.

smart grids, the integration of renewable energy resources (in particular, distributed generation), and ultimately the improvement of energy efficiency. On the other hand, the sheer amount of half-hourly or hourly measured electricity consumption data also introduces both opportunities and challenges for the electricity distribution system operators (DSOs) and /or the electricity retailers, in terms of how to manage and fully utilise such a wealth of data. So far, despite that there are successful business cases from Enel in Italy and Vattenfall Networks in Finland (Cotti and Millan 2011; Garpetun 2011), the utilisation of smart meter data or smart metering is limited to either enhancing distribution operation (e.g., automated fault detection and healing) or cost-saving from manual customer meter reading. For example, Mutanen et al. (2008) presented a method for AMR data to be used to enhance distribution state estimation. Moreover, the utilisation of AMR measurements in improving the accuracy of load modelling has been studied (Mutanen et al. 2011). Several similar studies (Abdel-Aal 2004; Charytoniuk and Chen 2000; Valtonen et al. 2010) have focused on AMR-based short-term load forecasting.

Nonetheless, according to a recent report by CEER (Council of European Energy Regulators), among the three European countries who have made decision to roll out smart meters (i.e., Italy, Sweden, and Finland), none have a demand response scheme based on smart metering. According to CEER's definition, demand response is about "*Changes in electric usage by end-use customers/micro generators from their current/normal consumption/injection patterns in response to changes in the price of electricity over time, or to incentive payments designed to adjust electricity usage at times of high wholesale market prices or when system reliability is jeopardized. This change in electric usage can impact the spot market prices directly as well as over time*" (CEER 2011). This implies that the establishment of a demand response electricity retail market not only requires the electricity end users' active engagement, but also the electricity utilities' capability for incentive pricing is crucial. To this end, we believe that in order to fully utilise the business potential enabled by smart metering technologies, it requires that the DSOs or the electricity retailers have good knowledge about their customers' timely electricity consumption patterns. Therefore, it is necessary to explore the smart meter data deeper for more gold nuggets.

In this paper, we investigate a visual data mining approach in the form of Self-Organizing Maps (SOM), namely electricity consumption time series profiling. We analyse quasi-daily smart meter data for approximately 12,000 customers in a Finnish region in 2009. We compare two sets of variables in terms of seasons and time bands partition. The objective is to study (1) what insightful knowledge can be detected by such a visual data mining approach; and (2) what is the added value for the business practice in applying such an analytical method for decision-making support, with respect to pricing differentiation or designing demand response tariffs. The results indicate that this analytical approach is capable of visualising deviations and more detailed information regarding customer's consumption patterns, which could support the case company in pricing decision-making. As will be illustrated in the following paper, this study will contribute with empirical understanding of the problem domain to the related research community and to the future practice of the electricity industry.

The paper is organised as follows: in the next two sections, the data mining method used in this study will be described first, followed by a brief introduction to the business case area. Thereafter, the experiment, results, and the analysis will be presented, and in the last part of this paper, the conclusion will be drawn and limitations and future research will be addressed.

2 Methodology

The Self-Organising Map (SOM) is one type of data mining technique based upon Artificial Neural Networks (ANNs). ANNs are designed to mimic the basic learning and association patterns of the human nervous system, and consist of a number of neurons (simple processors) connected by weighted connections. ANNs learn by adjusting the weight of each connection, increasing or decreasing the importance of the input (information) being transferred, until a desired output is achieved. Essentially, they are non-linear, multivariate regression techniques, better able to handle erroneous and noisy data than parametric statistical tools (Bishop1995).

The SOM is a widely used unsupervised neural network, particularly suitable for clustering and visualisation tasks (Han and Kamber 2000; Kohonen 1997). It is capable of projecting the relationships between high-dimensional data onto a two-dimensional display (or map), where similar input records are located close to each other (Kohonen 1997). By adopting an unsupervised learning paradigm, the SOM conducts clustering tasks in a completely data-driven way (Kohonen 1997; Kohonen et al. 1996), i.e., no target outputs are required. Because of its robustness, it requires little a priori information or assumptions concerning the input data, and is more tolerant towards difficult data, including non-normal distributions, noise, and outliers, than traditional statistical tools. In other words, the SOM combines the objectives of both data and dimensionality reduction methods, as seen either in the clustering techniques (e.g., K-means) or in the visualisation techniques (e.g., Sammon's mapping) (Sarlin and Peltonen 2011). This capacity of the SOM motivated the authors to apply it in the present study. As the SOM algorithm itself is well-known, we refer readers to Kohonen (2001) for details.

The SOM has been applied as an analytical tool in finance, medicine, customer relationship management, and engineering applications (Back et al. 2001; Deboeck and Kohonen 1998; Eklund et al. 2003; Kaski et al. 1998; Oja et al. 2002; Yao et al. 2010). In particular, the SOM has been used in the energy sector for e.g., power system stability assessment, on-line provision control, load forecasting, as well as electricity distribution regulation and benchmarking (Lendasse et al. 2002; Nababhushana et al. 1998; Rehtanz 1999; Riqueline et al. 2000; Liu et al. 2011).

A SOM is typically composed of two layers: an input and an output layer. Each input field is connected to the input layer by exactly one node, which is fully connected with all the nodes in the output layer (Berry and Linoff 2004; Wiskott and Sejnowski 1998). When the number of nodes in the output layer is large, the adjacent nodes need to be grouped to conduct clustering tasks. Accordingly, Vesanto and Alhoniemi (2000) proposed a two-level approach, e.g., the SOM-Ward clustering, to perform clustering tasks. The dataset is first projected onto a two-dimensional display using the SOM, and the resulting SOM is then clustered. Several studies have shown the effectiveness of the two-level SOM, especially the superiority of the SOM-Ward over some classical clustering algorithms (Lee et al. 2006; Samarasinghe 2007).

As mentioned previously, the SOM-Ward clustering is a two-level clustering approach that combines local ordering of the SOM and Ward's clustering algorithm to determine the clustering result. Ward's clustering is an agglomerative (bottom-up) hierarchical clustering method (Ward 1963). The SOM-Ward starts with a clustering where each node is treated as a separate cluster. The two clusters with the minimum Euclidean distance are merged in each step, until there is only one cluster left on the map. The distance follows the SOM-Ward distance measure, which takes into account not only the Ward distance but also the topological characteristics of the SOM. In other words, the distance between two non-adjacent clusters is considered infinite, which means only adjacent clusters can be merged. A low SOM-Ward distance value represents a more natural clustering for the map, whereas a high value represents a more artificial clustering. In this way, the users can flexibly choose the most appropriate number of clusters for their data mining tasks.

3 The Finnish Business Case

The business case studied in this paper is provided by one DSO in Finland – Ålands Elandelslag (ÅEA, which is a non-profit ownership cooperative). ÅEA's distribution area has distinct geographical features and customer structure. Åland is an autonomous Finnish archipelago region with nearly 300 habitable islands. It is situated between mainland Finland in the east and Sweden in the west. ÅEA is responsible for the electricity distribution to 15 municipalities in Åland. Its distribution area covers 14,097 customers, of which Jomala is the largest (2,290 customers) and Sottunga is the smallest (184 customers). Its distribution power lines totalled 3,217 km in 2009, with high voltage lines (10kV) 1,163 km and low voltage lines (0.4kV) 2,054 km. Åland's geographical features determine that its economy is heavily dominated by shipping, trade, and tourism. The majority of the housing is in the form of summer cottages, detached houses, or town houses, while multi-storeyed buildings only account for a very small portion.

According to Statistics Åland, in 2009, Åland's electricity consumption by sector is as follows: Households (45.04%), Agriculture (7.01%), Industry (11.77%), Services (21.22%), and the Public Sector (14.97%), respectively. It shows that households, services, and the public sector constitute the majority in terms of electricity consumption in Åland. This differs from the electricity consumption breakdown on mainland Finland, where industry's electricity consumption amounts to 46%, whereas housing and agriculture, and services and construction, consume 29% and 22% respectively (source: Energiateollisuus).

The data investigated is from ÅEA meter reading registers in 2009. For each meter, the electricity usage is registered with 27 hours 20 minutes time intervals, due to the

AMR and communication technology adopted (Turtle Automated Meter Reading system). The Turtle AMR uses the power line for data transmission. The data is collected by a receiver installed at a substation and held until requested by a computer at the main office, then sent via SMS. Turtle AMR also calculates the highest rate of electricity usage for each meter during each 27hrs20mins interval, i.e., the Peak Load. Therefore, the data from meter reading registers includes Meter ID, Electricity Usage, Reading Time, Peak Load, and Peak Time.

The analysis is carried out with a focus on three types of consumption time series, including (i) weekdays vs. weekends consumption comparison, (ii) consumption seasonality, and (iii) load patterns at various times of the day (i.e., different time bands).

4 The Experiment

Even though the ÅEA smart meter data is not hourly measured, it is still possible to look into customers' electricity consumption patterns in terms of day-of-the-week, seasonal, and time band effects. Based on the meter register data, a great deal of data pre-processing work, including data transformation, aggregation, and normalisation, has to be performed to create customer signatures, with one record per customer and a range of variables capturing customers' demographic and consumption related features. We excluded the customers whose records included less than one year, or whose annual consumption is 0 kWh. There are in total 11,964 customers included in this study. The variables used fall into two types based upon their purpose – one type is used to describe the customer's general consumption and demographic profile, and the other is to investigate customers' weekday-weekend, seasonal, and time-band related consumption patterns. Regarding the second type, we compared two sets of variables – the first set is adopted from ÅEA's partition as weekdays/weekends for the time of week, seasons (i.e., January-April, May-September, and October-December), and day time (7:00-23:00)/night time (23:00-7:00) for every 24hrs, which can be seen in ÅEA's electricity tariff of Time rate; the second set is proposed by the authors, as weekdays/Saturday/Sunday for the day-of-week, seasons (i.e., Summer: March-September, and Winter: October-February), and four time bands (i.e., 6:00-9:00, 9:00-16:00, 16:00-22:00, 22:00-6:00) for every 24hrs. In total, there are 31 variables used in this analysis. The variables are described as follows:

Average Consumption (kWh) – is the customer's average consumption per 27hrs 20mins +/- 8mins.

Average Peak Load (kW) – is the customer's average peak demand in 2009, which is based on the highest load aggregated from three consecutive 20min intervals during each 27hrs 20mins period.

Electricity Rate – is the contractual electricity tariff the customer has chosen among 5 categories: Normal rate, Economic rate, Time rate, Irrigation rate, and Temporary Working rate, which are provided by ÅEA (available at http://www.el.ax/files/tariffhafte_20110101.pdf, in Swedish). Due to the previously mentioned customer selection criteria set in pre-processing, the data records with Irrigation rate and Temporary Working rate are not included1.

Housing Type – is based on historical statistics, provided by ÅEA as a reference variable, including 5 categorical attributes: Summer Cottage, Detached House, Town House, Multi-storeyed Building, and Others. Again, as with Electricity Rate, the data records with Housing Type as Others are not included in the final dataset.

Seasonal and day-of-the-week Consumption (kWh) – includes Weekday Consumption1, Weekend Consumption1, Jan.-Apr. Consumption, May-Sep. Consumption, Oct.-Dec. Consumption, which are adopted from ÅEA's Time-of-Use tariff; and Weekday Consumption2, Saturday Consumption, Sunday Consumption, Winter Consumption, and Summer Consumption, which are proposed by the authors.

Time-based Peak Load (kW) – is the customer's average peak demand at various times of the day, including: Peak Load_Day, Peak Load_Night, which are based on ÅEA's electricity tariff; and Peak Load_6-9, Peak Load_9-16, Peak Load_16-22, Peak Load_22-6, which are proposed by the authors.

Time-based Peak Frequency (%) – is the percentage of peak demand occurring at different times of the day, including: Peak Frequency_Day, Peak Frequency_Night, which are based on ÅEA's electricity tariff; and Peak Frequency_6-9, Peak Frequency_9-16, Peak Frequency_16-22, Peak Frequency_22-6, which are proposed by the authors.

In this study, Viscovery SOMine v.5.0 (http://www.eudaptics.com/) is used to perform the visual data mining task. SOMine uses an expanding map size and the batch training algorithm, allowing for efficient training of maps (Deboeck and Kohonen 1998). The SOM-Ward clustering method is also used to identify clusters based on actual consumption behaviour, which eliminates the need for subjective identification of clusters (Vesanto and Alhomiemi 2000). The two sets of seasonal and time-based variables are normalised according to the respective average value before map training, i.e., each entry in a field is divided by the mean of the entire field (Baragoin et al. 2001; Collica 2007). The purpose is to address the relative significance of the value of a particular variable against the overall mean of that variable. For example, customers exhibiting average consumption patterns are given normalised values of 1, while a normalised value of 2 implies that their consumption amount or peak load is two times more than the average. In addition, all the variables included in the training process were scaled to comparable ranges in order to prevent variables with large values from dominating the result. Viscovery SOMine offers two forms of scaling, linear and variance scaling. Linear scaling is simply a linear scaling based upon the range of the variable, and is suggested as default when the range of the variable is greater than eight times of its standard deviation. Otherwise, variance scaling is used. In this study, range scaling was applied to the variables of Electricity Rate and Housing Type, while variance scaling was applied to the others.

---

[1] Categorical variables, such as Electricity Rate and Housing Type, must be split into binary dummy variables in order to be used with the SOM, as they represent nominal data with no inherent numerical order or distance.

We experimented with different combinations of parameters, and selected the map based on following criteria: average quantization error, normalized distortion measure, the meaningfulness of clusters, the visual interpretability, the smoothness of neighbourhood of each node, and the SOM-Ward cluster indicator. The map was trained using a map size of 279 nodes, a map ratio of 100:49, and a tension of 0.5. During the training process, the priority of categorical variables such as Electricity Rate and Housing Type, as well as the seasonal and time-based variables proposed by the authors, was set to 0. These variables thus have no influence on the training process. However, their distribution in each of the segments can be visualised on the map for comparison and profiling purposes.

In order to evaluate the robustness of the training method, a supervised ten-fold cross-validation was conducted. The entire training dataset was firstly partitioned into 10 subsets, then using 9 out of the 10 subsets each time to reiterate the map training with the same set of training parameters as was described above. The map selecting criteria set above can be held over the ten-fold iteration.

5 Results and Analysis

5.1 Cluster Profiles

The SOM divided the 11,964 customers into four clusters according to their consumption similarity in 2009. The SOM results can be seen in Figures 1-3. Since the warm colour code (e.g., red) in SOM map denotes high values while a cold colour code (e.g., blue) represents low values, the characteristics of each cluster (I-IV) can be easily identified, as summarised in Table 1. A description of each cluster follows:

- *Cluster I: High consumption customers*

Customers in cluster I account for 10% of total customers investigated and stand for 28.9% of the total consumption. They have the highest consumption profile (Average Consumption 63.0 kWh and Average Peak Load 5.1 kW). The proportion of customers using the Economic rate in cluster I (19%) is much higher than that of the other three clusters, although 80% of the customers still prefer the Normal rate. The majority of customers in cluster I live in detached house (88%), while 7%, 4%, and 1% of them are in summer cottages, town houses, or multi-storeyed buildings, respectively.

- *Cluster II: Medium-high consumption customers*

17% customers are in cluster II and they stand for 30.7% of the total consumption. They have the Medium-high consumption profile (Average Consumption 39.3 kWh and Average Peak Load 3.2 kW). Even though the majority housing type is detached house (75%), the proportion of summer cottage (18%) is the second highest after cluster IV. 5% of the customers in this cluster chose Economic rate, while 94% of them went for Normal rate.

- *Cluster III: Medium-low consumption customers*

Customers in cluster III account for 25.9% of the customer base and stand for 24.1% of the total consumption. They have Medium-low consumption profile

(Average Consumption 20.3 kWh and Average Peak Load 2.0 kW). The characteristics of cluster III are very similar to those of cluster II in that most of the customers (96%) use Normal rate and 76% of the customers live in detached houses. But the proportion of town house owners (12%) is the highest comparing to the other three clusters.

- *Cluster IV: Low consumption customers*

47.1% customers belong to cluster IV, which has the lowest consumption profile (Average Consumption 7.5 kWh and Average Peak Load 0.6 kW). They stand for 16.2% of the total consumption. 99% of customers in cluster IV have the Normal tariff contracts. Summer cottage (70%) is the major housing type within cluster IV, while detached house, town house, and multi-storeyed building account for 18%, 8%, and 4%, respectively.

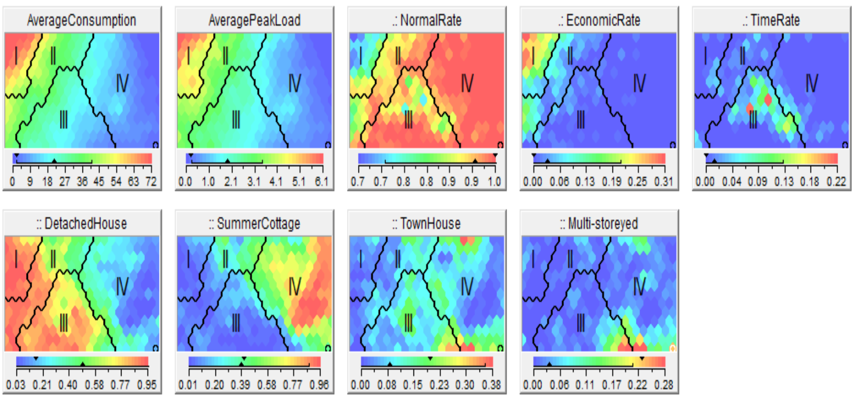

Fig. 1. Cluster profiles

5.2 Consumption Time Series Profiling

5.2.1 Weekdays vs. Weekends Consumption Comparison

Figures 2 and 3, specifically, reveal the patterns of each cluster (i.e., day-of-the-week, seasonal, and different time band consumption), and those of ÅEA's customers in general. For instance, from cluster I through cluster IV, both weekday consumption- and weekend consumption- patterns (see Figure 2) are ranging from high, medium to low, which also are in accordance with the patterns of Average Consumption in Figure 1. In addition, if comparing the consumption during weekdays/weekends (see Figure 2), or weekdays/Saturday/Sunday (see Figure 3), the patterns are nearly identical. This implies that if ÅEA intended to shift customers' demand between weekdays and weekends to mitigate system constrains or when the wholesale market price is high, ÅEA should devise enough incentive in their price signals for customers to adjust their consumption behaviour between weekdays and weekends.

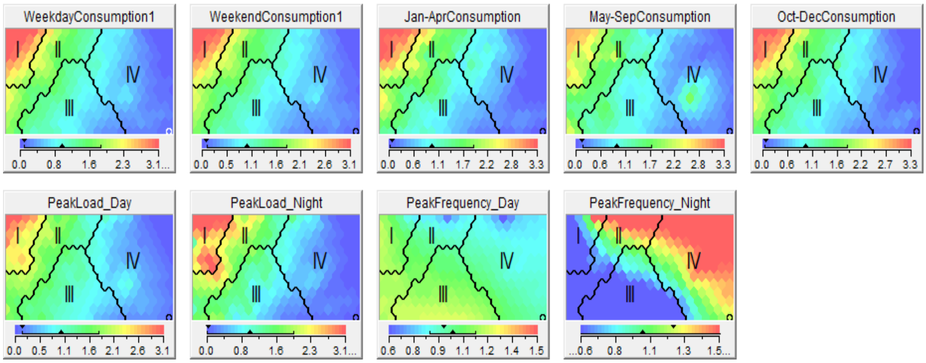

Fig. 2. Consumption patterns with ÅEA variables

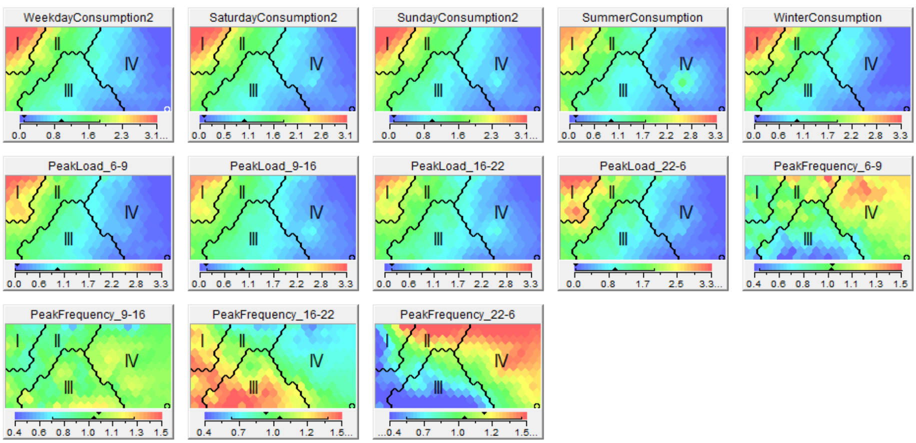

Fig. 3. Consumption patterns with proposed variables

5.2.2 Consumption Seasonality

The customers' seasonal consumption patterns vary. They follow the typical Nordic phenomena: electricity consumption is relatively higher in cold winter months than in summer time. This can be seen from both sets of seasonal consumption variables (see Figure 2 and Figure 3). However, it is important to note that there is a special group of customers in cluster IV (see Figure 4), whose electricity consumption in May-September is higher than the rest of cluster IV. This special group can be identified both from Figure 2 (May-Sep Consumption) and Figure 3 (Summer Consumption), which emphasizes that the consumption pattern deviation of this special group of customers in summer time is without regard to the summer months partition (i.e., May-Sep. as following ÅEA, or March-Sep. as proposed by the authors). At this point, it demonstrates that such a SOM-based data mining approach can visualize latent information for companies to take into account.

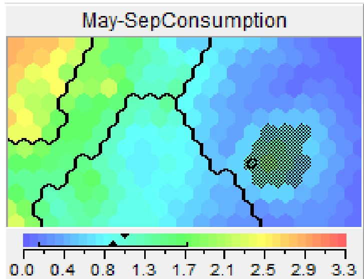

Fig. 4. Special group of customers in cluster IV

Based on the SOM visual clustering results, Figures 5, 6, and 7 summarize the comparison of various time series profiles among clusters. Figure 5 illustrates the consumption profile breakdown of each cluster and the special group within cluster IV, regarding weekday/weekend as well as seasonal consumption patterns. The different clusters have distinct consumption profiles in different seasons. For instance, regarding the Medium-low consumption customers (cluster III), their electricity usage is relatively even across different seasons (Jan-Apr., May-Sep. and Oct.-Dec.) in 2009 (red line in Figure 5). But High and Medium-high consumption customers (green and purple lines in Figure 5) had lower electricity consumption in summer time, compared to their respective cold weather seasons. On the other hand, as was pointed out before, among Low consumption customers, their May-September period consumption is relatively higher than in the rest of the seasons (see two blue lines in Figure 5).

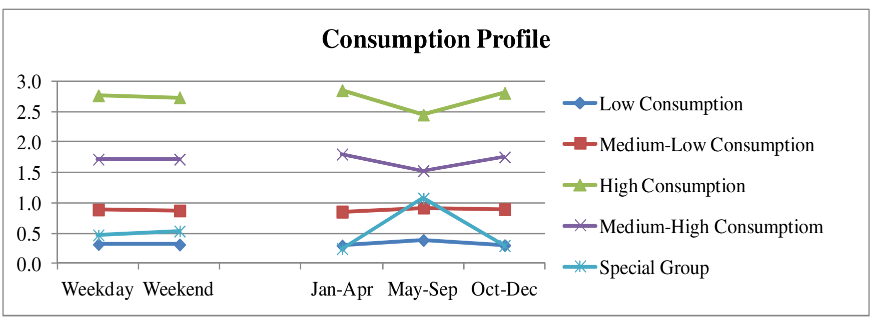

Fig. 5. Consumption profile breakdown

5.2.3 Load Patterns at Various Times of the Day

Accordingly, one can see that the patterns in terms of Peak Load at day time (7:00-23:00) and night time (23:00-7:00) (see Figure 2) are in line with the patterns of Average Peak Load in each cluster (see Figure 1). However, if examining Peak Load in four time bands in Figure 3, instead of the 2 (i.e., Day and Night) in Figure 2, slightly different picture emerges: the customers in cluster I have relatively higher peak demand in the early morning (6:00-9:00) and in the late night (22:00-6:00), compared to usual working hours (9:00-16:00) or usual peak consumption time period

(16:00-22:00). This is also represented in Figure 6, where the green line (High consumption customers of cluster I) bends up towards the ends considerably. It suggests that using the proposed four time bands can reveal more detailed information about the customers' consumption behaviour. And it might be beneficial if the company would consider using more than two time bands in their Time-of-Use pricing. The evidence can also be seen from Peak Frequency, i.e., where time-wisely speaking Peak Frequency at 6-9, 9-16, and 16-22 are equivalent to Peak Frequency_Day, but provide more information about consumption behaviour in different clusters. The comparison regarding how much extra information can be extracted with four time bands partition is shown in Figure 7.

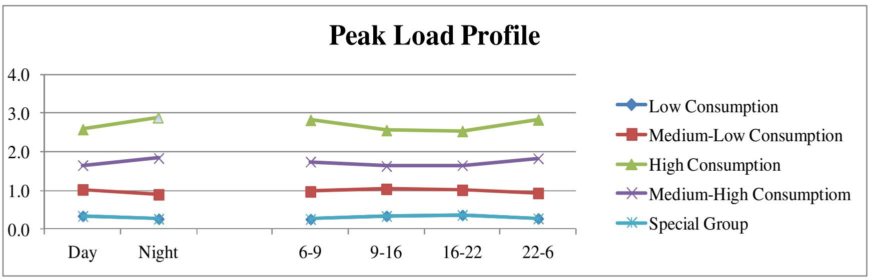

Fig. 6. Peak load profile breakdown

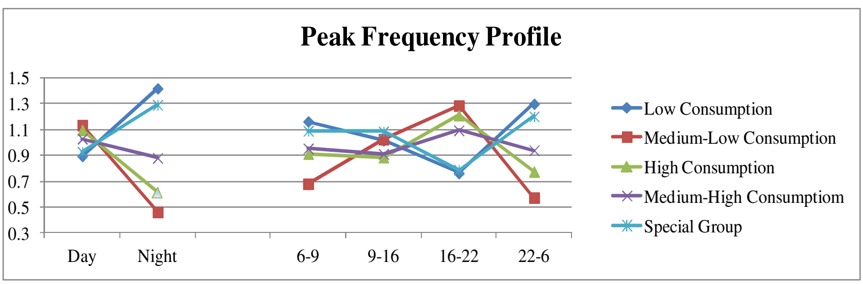

Fig. 7. Time bands partition comparison

6 Conclusion

Traditionally, the electricity utilities have classified customers according to their business nature (i.e., industrial, commercial, and residential) and their consumption bands (e.g., annual consumption < 2,000kWh, >5,000kWh, or > 18,000kWh) and housing types (e.g., detached houses, town houses, and multi-storeyed buildings) for household customers. Even in the same customer class, the consumption patterns may vary considerably due to customers' business nature / life style diversity (Keppo and Räsänen 1999). Additionally, the customer type is usually determined when the electricity connection is contracted, which is highly likely out-dated because of later changes in the customer's profile, for example, occupancy changes in a household. Now, smart meter data provides the opportunity to group and compare the customers according to their actual energy usage, especially taking seasonal variations into account.

Enabled by the smart metering technologies and motivated by the analytical robustness of the SOM in visualisation and data exploration, in this paper we have examined a SOM-based visual data mining approach, in order to investigate how the electricity utilities can fully explore smart meter data to gain better knowledge about their customers' timely electricity consumption patterns, and in turn to support pricing decision-making. We studied a case company from Finland—ÅEA's AMR data in 2009, with the purpose of demonstrating (1) what kind of actionable knowledge the examined electricity consumption time series profiling approach can offer and (2) what is the added value for DSOs or electricity retailers in applying such a visual data mining driven analytical method in decision making support, especially with regard to pricing differentiation or dynamic pricing. First, we used the SOM to cluster 11,964 customers into four groups according to their electricity consumption similarity in 2009. Then, the consumption profile of each cluster was visualized through feature plane analysis. During the analysis we compared different variable sets in day-of-the-week, season, and time band partition, in order to extract more detailed information about the customers' consumption patterns. For instance, the result shows that there is a special customer group within the low consumption cluster IV, whose consumption pattern in summer time deviated from the rest of the cluster. Moreover, the consumption visualisation indicated that the benefit for ÅEA to design different Time-of-Use tariff on weekdays or weekends calls for a review of its pricing differentiation strategy. In addition, there is evidence that the authors' proposed four time bands could provide granular information regarding customer consumption behaviour. These findings are actionable information for the case company to take into account in their future pricing strategy making. To this end, the conclusion can be drawn that this study provides an empirical example with regard to exploring timely measured smart meter data for customer's consumption behaviour analysis. It could induce further scientific interests regarding this emerging problem domain, for example, in terms of the intersection between ubiquitous computing, data mining, and demand response simulation. It also will contribute to the future practice of the energy industry in terms of integrating data mining into their pricing decision-making support.

Nevertheless, there are limitations to this study. Firstly, it would be of great interest to compare the SOM application to using other visualisation and clustering methods such as K-means or multi-dimensional scaling methods. However, it is beyond the scope of this paper. Secondly, the scope of this analysis is determined by the specific data domain. Therefore, the discovered knowledge has its particular locality. On the other hand, such an analytical approach can be applied to other AMR data for further examination and comparison studies.

Table 1. Summary of cluster characteristics

ID	Cluster Profile	Average Daily Consumption (kWh)	Average Peak Demand (kW)	Cluster Size and Percentage of Total Consumption (%)
I	High consumption: 80% Normal rate, 19% Economic rate; 88% detached house, 7% summer cottage.	63.0	5.1	10.0, 28.9
II	Medium to high consumption: 94% Normal rate, 5% Economic rate; 75% detached house, 18% summer cottage, 6% town house.	39.3	3.2	17.0, 30.7
III	Medium to low consumption: 96% Normal rate; 75% detached house, 9% summer cottage, 12% town house.	20.3	2.0	25.9, 24.1
IV	Low consumption: 99% Normal rate; 18% detached house, 70% summer cottage, 8% town house.	7.5	0.6	47.1, 16.2

Acknowledgements. The authors gratefully acknowledge the financial support of the Academy of Finland (grant nos. 127592 and 127656) and the Fortum Foundation. The case organization's cooperation is also gratefully acknowledged.

References

Abdel-Aal, R.E.: Short Term Hourly Load Forecasting Using Abductive Networks. IEEE Transactions on Power Systems 19(1), 164–173 (2004)

Back, B., Toivonen, J., Vanharanta, H., Visa, A.: Comparing numerical data and text information from annual reports using self-organizing maps. International Journal of Accounting Information Systems 2(4), 249–269 (2001)

Baragoin, C., Andersen, C., Bayerl, S., Bent, G., Lee, J., Schommer, C.: Mining Your Own Business in Retail Using DB2 Intelligent Miner for Data. IBM Redbooks (2001)

Berry, M.J.A., Linoff, G.: Data mining techniques: for marketing, sales, and customer relationship management. Wiley Computer Publishing (2004)

Bishop, C.M.: Neural Networks for Pattern Recognition. Oxford University Press, Avon (1995)

CEER Advice on the take-off of a demand response electricity market with smart meters, Ref: C11-RMF-36-03 (December 2011)

Charytoniuk, W., Chen, M.-S.: Very Short Term Load Forecasting Using Artificial Neural Networks. IEEE Transactions on Power Systems 15(1), 363–368 (2000)

Collica, R.S.: CRM Segmentation and Clustering Using SAS Enterprise Miner. SAS Publishing (2007)

Cotti, M., Millan, R.: Cervantes project and 'meters and more': the state of the art of smart metering implementation in Europe. In: Proceeding of 21st International Conference on Electricity Distribution (CIRED 2011), Frankfurt, Germany, paper 0829 (2011)

Deboeck, G., Kohonen, T.: Visual explorations in finance using self-organizing maps. Springer, Berlin (1998)

Eklund, T., Back, B., Vanharanta, H., Visa, A.: Using the self-organizing map as a visualization tool in financial benchmarking. Information Visualization 2(3), 171–181 (2003)

Garpetun, L.: Experiences from operations after a full-scale smart metering rollout regarding availability and reliability. In: Proceedings of the 21st International Conference on Electricity Distribution (CIRED 2011), Frankfurt, Germany, paper 0415 (2011)

Han, J., Kamber, M.: Data Mining: Concepts and Techniques. The Morgan Kaufmann Series in Data Management Systems. Morgan Kaufmann (2000)

Kaski, S., Kangas, J., Kohonen, T.: Bibliography of Self-Organizing Map (SOM) Papers 1981-1997. Neural Computing Surveys 1, 102–350 (1998)

Keppo, J., Räsänen, M.: Pricing of electricity tariffs in competitive markets. Energy Economics 21, 213–223 (1999)

Kohonen, T.: Self-organizing maps. Springer Series in Information Sciences, vol. 4, pp. 22–25. Springer, Berlin (1997)

Kohonen, T.: Self-organizing maps, 3rd edn. Springer, Berlin (2001)

Kohonen, T., Hynninen, J., Kangas, J., Laaksonen, J.: SOM PAK: The self-organizing map program package. Technical Report A31, Helsinki University of Technology, Laboratory of Computer and Information Science (1996)

Lee, S.-C., Gu, J.-C., Suh, Y.-H.: A Comparative Analysis of Clustering Methodology and Application for Market Segmentation: K-Means, SOM and a Two-Level SOM. Foundations of Intelligent Systems, 435–444 (2006)

Lendasse, A., Lee, J., Wertz, V., Verleysen, M.: Forecasting electricity consumption using nonliner projection and self-organizing maps. Neurocomputing 48, 299–311 (2002)

Liu, H., Eklund, T., Back, B., Vanharanta, H.: Visual Data Mining: Using Self-Organizing Maps for Electricity Distribution Regulation. In: Ariwa, E., El-Qawasmeh, E. (eds.) DEIS 2011. CCIS, vol. 194, pp. 631–645. Springer, Heidelberg (2011)

Mutanen, A., Repo, S., Järventausta, P.: AMR in distribution state estimation. In: Proceedings of Nordic Distribution and Asset Management Conference, Bergen, Norway (2008)

Mutanen, A., Repo, S., Järventausta, P.: Customer Classification and Load Profiling Based on AMR Measurements. In: Proceedings of the 21st International Conference on Electricity Distribution (CIRED 2011), Frankfurt, Germany, paper 0277 (2010)

Nababhushana, T.N., Veeramanju, K.T., Shivanna: Coherency identification using growing self organizing feature maps (power system stability). In: IEEE Proceedings of EMPD 1998. International Conference on Energy Management and Power Delivery, vol. 1, pp. 113–116 (1998)

Oja, M., Kaski, S., Kohonen, T.: Bibliography of Self-Organizing Map (SOM) Papers: 1998-2001 Addendum. Neural Computing Surveys 3, 1–156 (2002)

Rehtanz, C.: Visualisation of voltage stability in large electric power systems. In: IEEE Proceedings Generation, Transmission and Distribution, vol. 146, pp. 573–576 (1999)

Riqueline, J., Martinez, J.L., Gomez, A., Goma, D.C.: Possibilities of artificial neural networks in short-term load forecasting. In: Proceedings of the IASTED International Conference Power and Energy Systems, pp. 165–170. IASTED/ACTA Press, Anaheim (2000)

Sarlin, P., Peltonen, T.: Mapping the State of Financial Stability. ECB WP No. 1382/2011 (2011)

Samarasinghe, S.: Neural networks for applied sciences and engineering: from fundamentals to complex pattern recognition. CRC Press (2007)

Valtonen, P., Honkapuro, S., Partanen, J.: Improving Short-Term Load Forecasting Accuracy by Utilizing Smart Metering. In: Proceedings of the 20 th International Conference on Electricity Distribution (CIRED 2010), Lyon, France, paper 0056 (2010a)

Vesanto, J., Alhoniemi, E.: Clustering of the self-organizing map. IEEE Transactions on Neural Networks 11(3), 586–600 (2000)

Ward, J.H.: Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association 58(301), 236–244 (1963)

Wiskott, L., Sejnowski, T.: Constrained optimization for neural map formation: A unifying framework for weight growth and normalization. Neural Computation 10(3), 671–716 (1998)

Yao, Z., Holmbom, A.H., Eklund, T., Back, B.: Combining Unsupervised and Supervised Data Mining Techniques for Conducting Customer Portfolio Analysis. In: Perner, P. (ed.) ICDM 2010. LNCS, vol. 6171, pp. 292–307. Springer, Heidelberg (2010)