# MINING SMART METER DATA – CASE FINLAND

Antti MUTANEN
Tampere University of Technology
Finland
antti.mutanen@tut.fi

Harri NISKA
University of Eastern Finland
Finland
harri.niska@uef.fi

Pertti JÄRVENTAUSTA
Tampere University of Technology
Finland
pertti.jarventausta@tut.fi

## ABSTRACT

*Smart meters collect a lot of data on customer level electricity consumption and this, together with other data sources e.g. environmental information and public open data, provides an excellent basis for data mining. As a part of a recent smart grid project conducted in Finland, several different ways of mining smart meter data were studied. The project brought advances in customer classification and clustering, load profiling, spatial load analytics, behaviour change detection and load forecasting.*

## INTRODUCTION

Smart metering is spreading quickly. Many distribution system operators (DSOs) in western countries have installed, are installing, or are planning to install smart meters. The drivers for smart meter implementation are multiple. In some countries legal and regulatory requirements are driving the change as governments are enforcing the transition to liberated energy markets where consumers can freely choose their energy retailer. Smart metering is also seen as an essential enabler for smart grids that can host large amounts of $CO_2$ free distributed generation. They also enable new kind of dynamic tariffs that support energy efficiency targets and operation of the electricity markets. Sometimes, smart meter investments are made on economic and financial grounds, with the assumption that on the long run smart meters and the new services they enable, provide more revenue.

The primary role of smart meters is to provide fine-grained interval data on end customer energy consumption to the utility, but the cost of installing smart meters may not be justified if the meters are used only for reading energy consumption data. In addition to simple meter reading, other functions with added value are needed to make the smart meter investments profitable. Already today, several possibilities to utilize smart meters in electricity distribution and energy retail businesses exist, for example: remote disconnection and reconnection of electricity supply, interruption and power quality statistics collection, alarms based on exceptional events (i.e. network faults and voltage violations), web services for reporting customers electricity consumption, and demand side management.

In addition to the above mentioned rather direct ways of using smart meters, some added value can be extracted from the huge amount of historical electricity consumption data stored in meter data management (MDM) systems. Data mining can be used to extract useful information that is normally hidden within this vast amount of measurements.

This paper reviews the Finnish smart metering background and presents some smart meter data mining applications developed in Finland during the recent years.

## CASE FINLAND

In 2009, the Finnish Government passed a new act [1], which states that at least 80% of the customers of each DSO must have a smart meter by 31 December 2013. The law requires that the installed smart meters are read at least once a day, provide hourly energy measurements and that these measurements are used in balance settlement. If requested by the customer, the smart meter must have a standardised interface for real-time consumption monitoring. Other requirements include; registrations of long interruptions (>3 min), possibility to execute or relay load control commands, and storing the smart meter measurements at least six years. Also, the metering system must encourage customers to use electricity in efficient and economical manner.

Another law [2], that has been passed to fulfil the European Union energy efficiency directive, requires that all energy retailers must provide their customers, at least once a year, with a report that shows the energy consumption on the reporting period and on three previous years. This report must include comparison to other similar energy end users and information how to improve the energy efficiency.

Although only 80 % smart meter penetration is required by law, in practice almost every customer (98 %) in Finland is supplied with a smart meter. This makes Finland an excellent place for developing smart meter based functionalities and electricity consumption data mining methods. Between the years 2010 and 2015, several Finnish universities, research institutes, DSOs, and industry partners co-operated in a 52 million euro Smart Grids & Energy Markets -program (SGEM). Smart metering was one of the subtopics in SGEM and the research work done during this project resulted in several advances in the field of smart meter data mining, some of which are presented in this paper.

## SMART METER DATA MINING

When combined with external information sources, electricity consumption interval data provides useful insights into customer behaviour. The external information sources can be weather observations and forecasts (outdoor temperature, solar irradiance etc.), socio-economic data, building information, customer location, contractual information, or information collected directly from the end users. With data mining methods the above mentioned information sources and smart meter interval data can be combined and useful information can be extracted.

### Clustering and load profiling

In Finland, customer class load profiles are widely used in distribution network calculation and analysis. Distribution network operation, operation planning, and network planning could all benefit from better customer classification and load profiles.

During the SGEM project, we developed a load profiling method that clusters customers into similarly behaving groups and calculates statistical load profiles [3]. In addition to cluster profiles, individual load profiles are calculated for those large customers that exhibit unique consumption characteristics. Seasonal temperature dependencies are calculated for every cluster and individual profile and when combined with outdoor temperature forecasts, the new load profiles can be used also for load forecasting [4].

The developed clustering procedure starts with customer wise temperature dependency calculation. Next, temperature normalization to long term monthly temperatures is done so that measurements are comparable between different years. Pattern vectors describing the average hourly consumption every day of the week in a certain month are calculated. Analysis of variance (ANOVA) is used to determine whether weekdays should be modelled separately or with a common model. A two-stage clustering procedure utilizing weighted K-means clustering is performed. After the first stage, outliers (customers with the highest distances from cluster centers) and candidates for individual load profiling (customers with the highest weighted distances from cluster centers) are separated from others. Weighting is done according to yearly energies. After the second clustering stage, cluster wise temperature dependencies are calculated and load profiles are formed using the existing Finnish load profile format where each hour of the year has an expectation value and a standard deviation. Fig. 1 shows the diversity of cluster center shapes typically found among Finnish electricity consumers.

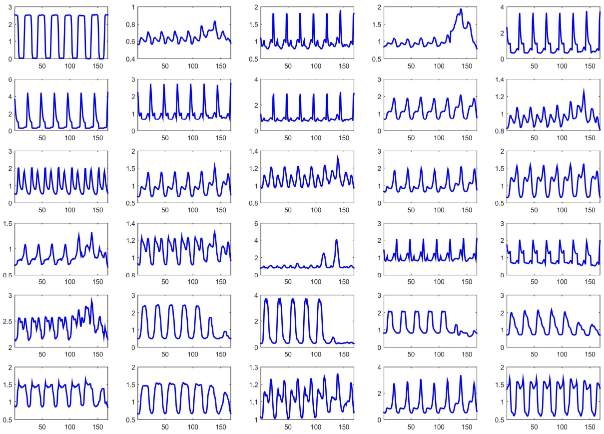

**Fig. 1.** Example of clustering results (type weeks for october).

### Customer behaviour change detection

Sometimes customers change their electricity consumption habits, for example due to a change in the heating solution or purchase of a micro generator or electric vehicle. Possibility to detect these kind of abrupt changes makes load profiling more dynamic and guarantees that the customer classification and load profiles are always up-to-date. Once a change is detected, the old pre-change measurements should be omitted and only the post-change measurements should be used in load profiling.

The newly developed change detection method [5] is based on studying weekly segments of smart meter data. Changes are detected based on weekly energies and weekly profile shapes. In the shape based detection, all customers are first clustered and then changes in the individual consumer's weekly cluster memberships are compared with previous year memberships. If the sum of differences is larger than a predefined detection threshold, a change is declared. Fig. 2 shows an example. The upper subfigures show the hourly electricity consumptions during the years 2009 and 2010. The subfigures in the middle show the weekly cluster memberships. The sum of membership differences (subfigure in bottom right) confirms that there has been a change in the intra-week behaviour from week 30 onward, but the weekly energies (subfigure in bottom left) have not changed.

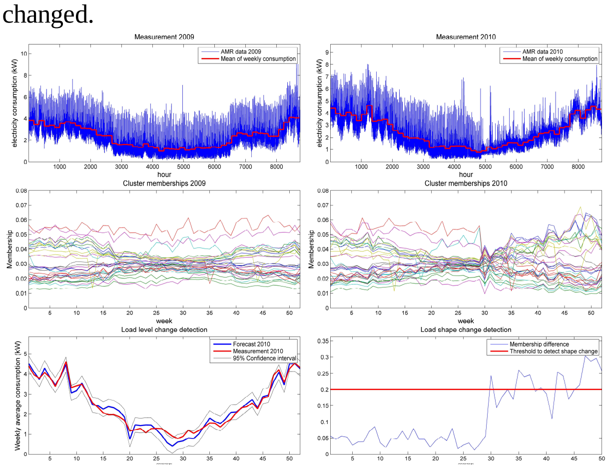


**Fig. 2.** Change detection example.

### Comparison curves

In order to fulfil the requirements given in [2], the most Finnish electricity retailers have provided their customers with access to an on-line portal where they can view their electricity consumption history and compare their consumption with other similar end users through comparison curves. The calculation of comparison curves has its own challenges and to overcome these challenges a new comparison curve formation methods have been presented in [6] and [7].

The method proposed in [6] groups customers based on the available customer type information and then conducts interval data based clustering to reduce the effect of classification errors occurred in the first grouping phase. The method proposed in [7] starts with a classification based on the two most significant attributes; the type of housing and the heating system, and then continues with weighted K-means clustering. Clustering is applied to a set of vectors that contained information on daily load profiles and selected customer background variables (in this case gross floor area and the number of residents in different age groups).

In these studies, the information characterizing customers is collected directly from customers either through the on-line portal [6] or through a separate questionnaire [7]. This information could be extracted also from open databases (e.g. national building information database), but the information might not be up-to-date. Fig. 3 presents a prototype program where the customer can select the information used in the comparison curve calculation. Finally, the difference with the comparison group is shown graphically and in euro.

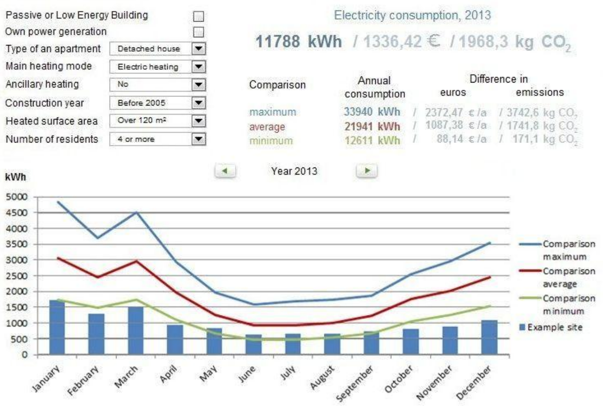


**Fig. 3.** Comparison curve with selectable customer type attributes. *Source* [5]

## Load forecasting

Short-term load forecasting is required by dynamic optimization of control actions, electricity market transactions, and distribution network operation. However, forecasting errors cause high balancing and network costs, thus cancelling the financial benefits from active demand. Extensive smart metering data can be utilized to improve the accuracy of the existing models as well as developing new enhanced modelling schemes for load forecasting. In this context, data mining techniques, including machine and statistical learning, are of interest in enriching data-based models, as well as hybrid models based on their combinations with other techniques such as physically based modelling.

During the SGEM project, several different models utilizing hourly metered consumption data were evaluated for short-term load forecasting. The studied models were; a cluster profile based predictor, a Kalman-filter based predictor with input nonlinearities and physically based main structure, a neural network (NN) model [4], and a support vector machine (SVM) model [8]. The NN and SVM models were the most accurate, but also the other methods had their relative merits. The work is now continued in the Academy funded RESPONSE project (2015–2018), where the aim is to develop a hybrid model, which combines the benefits of physical and purely data-driven modelling schemes in forecasting of loads and their load control responses.

## Spatial data analytics and modelling

Combining smart metering data with available geographic information will open totally new possibilities for data analytics in the planning and operation of future electricity networks and energy systems. In Finland, as well as in EU-level, different registries of the societies are opening in parallel with legislative frameworks (INSPIRE, PSI) that drive for the release and use of public sector data. Potential datasets include e.g. socioeconomic grid data, building and population information, land-use data, meteorological observations, and weather forecasts.

During the SGEM project, several GIS based data analytics and modelling tool demonstrations were carried out by combining smart metering data with other available geographic information [e.g. 9, 10]. Those experiments were mainly focused on the use of socioeconomic grid data (Statistics Finland) and building information (Population Register Centre) in spatial load modelling and what-if type load analytics. Data mining techniques such as unsupervised and supervised neural networks were used to model interactions between smart metering data and complex spatial information. Fig. 4 shows a screenshot from the developed scenario based load prediction tool [10]. This

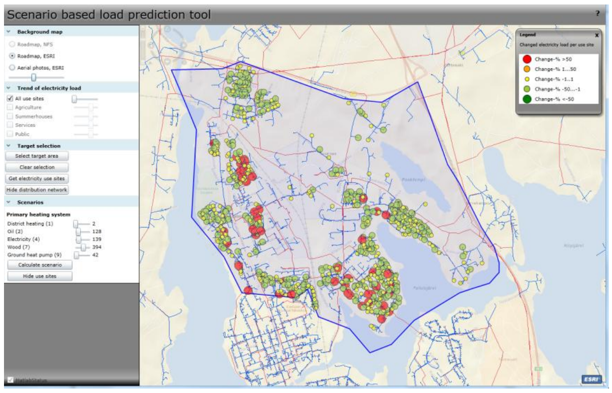


**Fig. 4.** Spatial load analytics tool based on smart metering data and building information.

tool can calculate, for example, how changes in heating solutions affect the regional peak loads and shows on the map where the load increases and where it decreases.

## Future development needs and opportunities

The SGEM project provided many answers, but also new research questions arose. For example, the developed change detection method is able to detect changes and point out when the change happened, but it cannot specify what caused the change. There is a need to develop more specific methods for detecting and identifying customers with photovoltaic panels, heat pumps and electric vehicles.

In future, the utilization of demand response in a larger and more dynamic scale requires development of new load models that can forecast load responses accurately in different situations, can adapt to changes in load behaviour and are easy to maintain and update. In this context, smart metering data and its enrichment using data mining techniques play an important role and can enable data-driven model creation and updating.

The Finnish transmission network operator Fingrid has been assigned a task of implementing a national data hub that will centralize all information exchange in electricity markets. All smart meter measurements normally exchanged between customers, DNOs and electricity retailers will be stored in a single database. Expected to finish in 2019, the data hub will offer unprecedented possibilities for smart meter data mining.

## CONCLUSIONS

This paper presented smart meter data mining related studies conducted during the SGEM-project. During this project, it was proven that clustering can improve the accuracy of customer class load profiles and a methodology for smart meter data based clustering was developed. In order to make the load profiling more dynamic, a method for detecting changes in customer behaviour was developed. In comparison curve calculation, smart meter measurements were combined with information characterizing consumers and consumption sites so that customers can compare their electricity consumption habits with other similar customers. Neural network and support vector machine based forecasters were successfully applied to short term load forecasting. Geographical grid data available in open databases was combined with smart meter data in order to create spatial load analyses.

To sum up, smart metering data and its mining together with opening public sector data will enable the development of new powerful data analytics and services for enabling reliable, and cost-efficient operation and planning of future flexible energy systems as well as providing tools for managing customer side energy efficiency and flexibility of prosumer demand and generation.

## REFERENCES

[1] REG 1.3.2009/66. Valtioneuvoston asetus sähköntoimitusten selvityksestä ja mittauksesta (Finnish council of state act on electricity settlement and metering, in Finnish).

[2] REG 1211/2009. Laki energiamarkkinoilla toimivien yritysten energiatehokkuuspalveluista (Law on energy efficiency services for companies operating in energy markets, in Finnish).

[3] A. Mutanen, P. Järventausta, M. Kärenlampi & P. Juuti, “Improving Distribution Network Analysis with New AMR-based Load Profiles,” *International Conference and Exhibition on Electricity Distribution (CIRED)*, Stockholm, Sweden, June 10–13, 2013.

[4] P. Koponen, A. Mutanen & H. Niska, “Assessment of Some Methods for Short-Term Load Forecasting,” *IEEE PES Innovative Smart Grid Technologies (ISGT) European Conference*, Istanbul, Turkey, Oct. 12–15, 2014.

[5] T. Chen, A. Mutanen, P. Järventausta & H. Koivisto, “Change detection of electricity Customer behavior based on AMR measurements,” *IEEE PowerTech conference*, Eindhoven, Netherlands, June 29 – July 2, 2015.

[6] V. Välipirtti, “Development of comparison feedback for private customers in electricity consumption reporting service,” M.Sc. Thesis, Tampere University of Technology, Tampere, Finland, 2014.

[7] M. Mononen, J. Saarenpää, M. Johansson & H. Niska, “Data-driven method for providing feedback to households on electricity consumption,” *International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP) conference*, Singapore, April 21–24, 2014.

[8] H. Niska, P. Koponen & A. Mutanen, ”Evolving smart meter data driven model for short-term forecasting of electric loads,” *IEEE Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, Singapore, April 7–9, 2015.

[9] J. Saarenpää, M. Kolehmainen, M. Mononen & H. Niska, “A data mining approach for producing small area statistics-based load profiles for distribution network planning,” *IEEE ICIT conference*, Seville, Spain, March 17–19, 2015.

[10] H. Niska, J. Saarenpää, T. Räsänen, J. Tiirikainen & M. Kolehmainen, “Scenario based load prediction tool for distribution planning and management,” *International Conference on Electricity Distribution (CIRED)*, Frankfurt, Germany, 6–9 June, 2011.