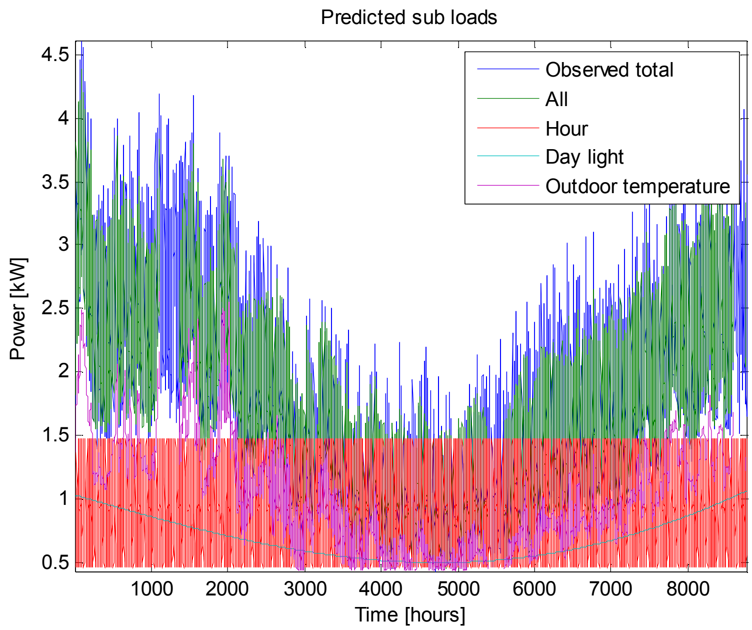
SGEM Research Report D4.2.19

Harri Niska
Jukka Saarenpää

UNIVERSITY OF EASTERN FINLAND

CI/DDM approaches for analysing and modelling loads using smart metering data and other data

Summary of WP4 Task 4.2.1 activities



Abstract

In this report, the research activities related to **SGEM FP2-3 WP4 Task 4.2.1 "Novel load modelling methods"** are briefly reported and summarised. The main objective has been on the investigation and testing of novel data-driven modelling approaches, which combine smart metering data to available external information. Specific focus has been placed on the following research tasks: (i) recognition of heating systems, (ii) disaggregation of subloads (particularly electrical heating), (iii) short-term forecasting of loads and (iv) spatial load modelling. The research results are/will be reported in more details in the international conference papers and journal papers.

Table of contents

Preface

The work presented in this report is part of the project Smart Grids and Energy Marketing (SGEM). The SGEM project belongs to Cluster of Energy and Environment (CLEEN), financed by Finnish Funding Agency for Technology and Innovation, TEKES industrial partners, universities, and research institutes.

We wish to thank the funding organizations and the steering group of SGEM programme for making this work possible to carry out.

The authors

Abbreviations

AMR	Automated metering data
ANN	Artificial Neural Network
CI	Computational Intelligence
DDM	Data Driven Modelling
DR	Demand Response
DSM	Demand Side Management
DSO	Distribution System Operator
GIS	Geographic Information System
LDA	Linear Discriminant Analysis
MLP	Multi-Layer Perceptron
SOM	Self-Organizing Map
SVR	Support Vector Regression

1 Background

Modelling and forecasting of loads is an essential function required by the planning and operation of Smart Grids. From the demand side management (DSM) and load control point of view, reliable methods are required for predicting controllable loads in varying behavioural and environmental conditions. In Scandinavian countries, residential heating loads have shown significant control potential compared with other domestic appliances (e.g. Koponen 2005) and could be shifted without significant disturbance to the comfort of customers. On the other hand, from the perspective of the planning of the grids, the spatial load modelling is required to assess loads and their long-term development in different geographical areas and suggested long-term scenarios.

Different approaches have been developed varying from statistical regression and load profiles to dynamic physical-based models. Many examples can be found in the literature (e.g. Willis et al 2002, Hahn et al. 2009, Räsänen et al 2010, Mutanen et al 2011, Koponen 2011). However, the possibilities and added value of a combined use/enrichment of opening databases and other metering data have not been extensively investigated so far. Thus ideating new types of “data-driven” modelling approaches and functions based on utilisation of automated meter reading (AMR) data together with other available datasets is highly relevant and can result in development of new, useful Smart Grid functionalities and services.

AMR data are now increasingly available and form the basis for the building and verification of load models. In addition, there exist data from various external information sources that can be used in load modelling. Combined with AMR and other data new more accurate and reliable models could be achieved. Potential external information sources include georeferenced data hosted mainly by the public sector. Among the most promising datasets in Finland from load modelling perspective are building, meteorological, and demographic datasets. Some of the public sector data are opening for free use for research and commercial purposes (see e.g EU’s PSI and INSPIRE Directives).

However, there are some challenges and limitations related to the use of above mentioned datasets which should be carefully considered. Firstly, one common problem is that AMR is aggregated representing the total load of user-site instead of the loads caused by separate appliances or sub load components (e.g. heating, ventilation). Secondly, the common one hour resolution of AMR data largely restricts the analysis and modelling of load dynamics which often occur in minute-to-minute level (e.g. ventilation). In addition, when it comes to the integration of AMR with other datasets, missing common key, id or spatial reference, dealing with heterogeneous resolutions, missing data and finding an appropriate generalisation level in modelling are common problems.

2 DDM/CI based load models and related methodology

2.1 DDM/CI models

When dealing with large amount of data, it is often difficult for human to notice the patterns and interrelationships within. However, in ”data-rich” conditions data-driven modeling (DDM) methods provide new possibilities for the analysis and modelling. More sophisticated DDM methods rely often on novel data mining and computational techniques contributed by the field of Computational Intelligence (CI) or Artificial Intelligence (AI), the most well-known techniques including:

- Artificial neural networks (e.g. support vector machines, multi-layer perceptron networks, self-organizing maps, radial basis networks, elman networks)
- Evolutionary and genetic algorithms
- Fuzzy logic
- Clustering methods (e.g. k-means, isodata, fuzzy-c-means)

The DDM methods combined with conventional statistical methods and geocomputing techniques could result substantial enhancements in solving load modelling problems related to the planning and operation of Smart Grids. Some advantages of DDM/CI methods are that they are capable of (Niska and Saarenpää 2011):

- Searching complex spatiotemporal patterns, load curves, in different data presentation levels
- Modelling non-linearity and temporal dynamics of loads, including time-delays and interaction with external variables
- Forecasting future behaviour of load time-series
- Handling measurement errors, noise and missing data

Recently clustering based approaches have been developed to calibrate and update group-specific load curves using AMR data (e.g. Mutanen et al. 2011; Räsänen et al. 2010). Furthermore, regression and artificial neural network (ANN) models have been presented for load modelling and forecasting (e.g. Hahn et al., 2009; Aldo et al. 2010; Hong 2009; Beccali et al. 2004).

2.2 Physical models vs DDM models

DDM methods, which rely solely on the data, are often shown to outperform physical/mechanistic based models in several problem domains. However, the drawbacks of DDM methods are that they are restricted solely on the domain of training data and in some cases it might be difficult to understand the model dynamics between the inputs and outputs (thus often referred to as black box methods). If new situations outside the data range are under modelling and prediction, some physical reality is required to be incorporated to the models.

The benefit of the physical models over data-based models is that, if well-formulated, they can cover also the situations, which are out of the training data. This is particularly important from the perspective of load control and management.

A challenge is however often that loads are caused by complex interactions/dynamics between customer behaviour (use of appliances etc), physical characteristics of building and environmental conditions (particularly outdoor temperature). The description of such system at sufficient level of representation is often complicated.

However, the use of simple physical models could catch major behaviour at a level sufficient for load and response control (Koponen 2011).

2.3 Hybrid models

Developing hybrid approaches, which combine the benefits of data-based models and physical models can yield significant enhancements in the load modeling. Interesting question is how to construct and update site or region-specific models efficiently using the collected AMR data and other supporting background information.

3 Objectives and research tasks during SGEM 2-3 FP

The main objective during SGEM 2-3FP has been on investigating possibilities and added value of an integrated use of public sector originated data and AMR data in load modelling and forecasting. The work has been carried out in a conjunction with WP6, Task 6.10, in which the emphasis has been placed on spatial modelling issues (Figure 2).

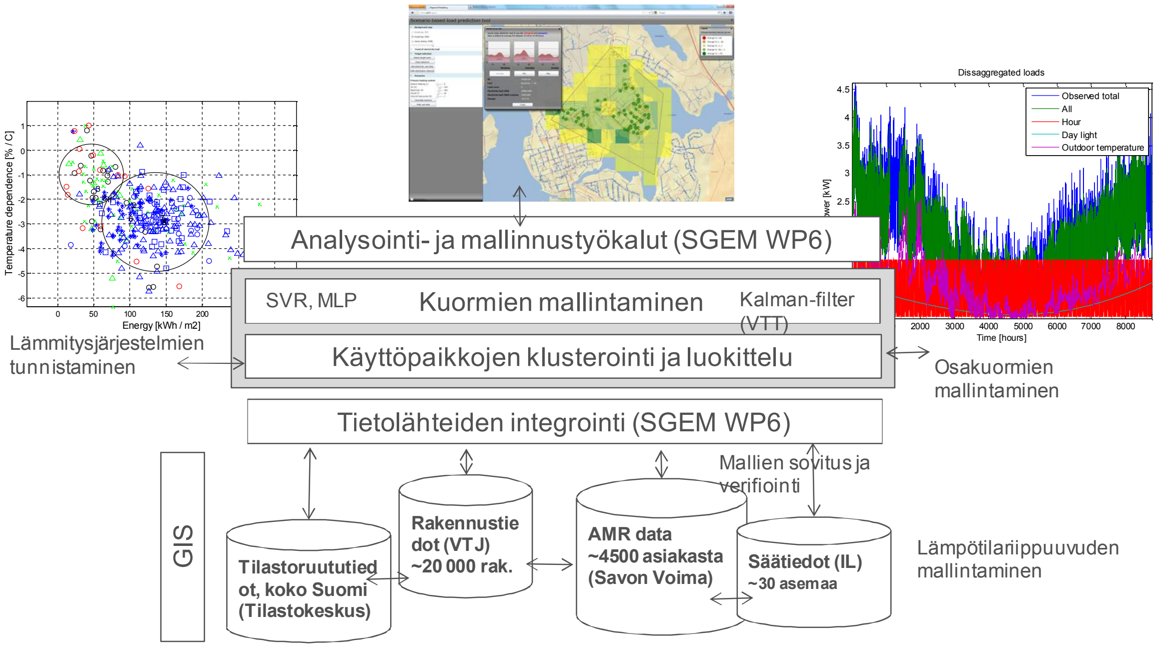


Figure 1: SGEM 2-3FP research tasks in WP4, Task 4.2.1.

The following specific research questions and tasks have been studied:

- How to combine heterogeneous, region-level data to user-site specific smart meter data (i.e. building information <-> AMR, population data <-> AMR)
- How to automate the recognition of heating systems from the AMR data, and how to disaggregate/model controllable heating loads from the aggregated AMR data
- What is the accuracy of different methods (such as ANNs, load profiles and physically based methods) in load modelling
- Is it possible to develop a model based on spatial units, which can be used to support strategic planning of distribution networks in different scenarios concerning EV, DER, DG, population, climate, etc.

In the SGEM project, UEF has been collecting various datasets from the area of the electricity distribution network of Savon Voima Verkko Oyj for the basis of model development and testing. The main datasets collected are as follows:

- AMR data, about 4000 customer (from Savon Voima Verkko)
- Building information (VTJ/RHR/KTJ)
- Socio-economic grid data (Statistics Finland)
- Weather data, 13 met. statiotions (FMI)
- Land use data, CORINE (MML/SYKE)
- HEV registration data (Trafi)

In addition UEF has been developing energy and indoor air quality monitoring system in order to research energy efficiency and indoor air quality in several research projects (e.g. Asteka, Asko, Insulavo). The system was first presented in the Kuopio housing fair 2010. The developed monitoring system is composed of sensors, data transfer unit (ZigBee), database and user-interface (see Appendices 1-2). The more detailed information of the system can be found in Skön et al. (2011).

The monitoring system is installed for the pilot houses (currently covers approx. 8 apartment houses, 10 schools, 5 detached houses) and can be used to develop and test models e.g. an interaction between loads and indoor air quality with customer specific high-resolution data.

4 Main results achieved

Main results achieved during 2-3FP can be divided roughly into the following categories:

- Recognition of heating systems
- Disaggregation of subloads (particularly electrical heating)
- Short-term forecasting of loads
- Spatial load modelling

The main outputs and results in these research domains are next briefly presented and summarised. More detailed description of methods and results can be found in the resulted publications (e.g. ISSNIP 2013, CIRED 2013)

4.1 Recognition of heating systems

The recognition of heating systems is highly necessary function since DSOs have no up-to-date information about customer’s heating systems. This is also partly due to that there are no requirements for collecting information on all the installations (e.g. heat pumps, solar panels etc). Therefore, classification approach based on the combination of linear discriminant analysis (LDA) and multi-objective genetic algorithm was tested for the recognition of heating systems using the AMR and building information (Niska, 2013).

Approaches were based on customer-specific feature variables extracted from AMR data, namely:

- Temperature delay (hours), analysed using the linear correlation between the delayed 24h average outdoor temperature and daily energy consumption for each customer.
- Temperature dependency %/C
- Daily load characteristics (peak hour, minimum load hour, standard deviation, range)
- Daily average energy kWh
- Time series characteristics (curtosis, skewness and variance)

The accuracy obtained was moderately good in terms of performance indices (Kappa index: 0.62, TPR: 87%, and TNR: 82%). The approach was able to discriminate customers either to non-electrical or electrical heating group with sufficient accuracy. However, when it comes to the discrimination of sub loads inside primary electrical heating systems (e.g. air-to-air source heat pumps, etc), the performance was degenerated. This was obtained to be largely due to insufficient feature variables extracted from the data. As a recommendation, more development work is needed to be placed for developing and extracting discriminative features for the basis of classification. Moreover, more sophisticated classification methods such as SVMs should be tested.

4.2 Disaggregation of heating loads

Disaggregation of sub loads is an essential question when dealing with the aggregated AMR data and DSM. This is needed to analyse and model the controllable parts of loads such as electrical heating, which is a particular issue in Finland (e.g. Koponen 2005). However, the disaggregation of sub loads is difficult task due to various correlating factors (e.g. correlation between seoson/time and temperature).

Support vector regression (SVR) based models have emerged as potential and accurate methods for load modelling and forecasting (Hahn et al. 2009). As a part of the SGEM 2-3FP experiments, a computational approach based on SVR was tested for modelling heating loads from aggregated AMR data (Figure 2). The basic principle of the developed SVR model is on the non-linear regression between timing variables (hour, day of week, day of year), outdoor temperature (˚C) and hourly power (kW). The prediction performance was observed to be high in terms of various performance indices (R2=94%, the index of agreement=98% and RMSE=0.18kW).

The SVR method and results are further presented and discussed in IEEE ISSNIP paper (Niska 2013).

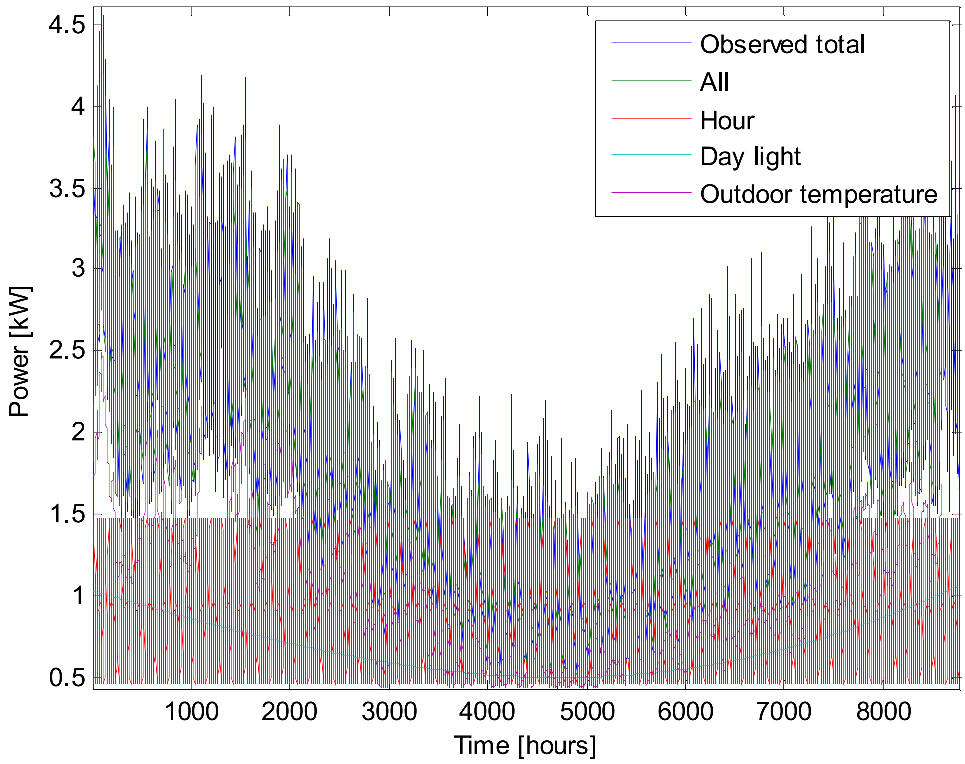


Figure 2. Sub loads modelled using the SVR model (modified from Niska 2013, IEEE ISSNIP 2013).

4.3 Load forecasting

Short term load forecasting is central function required by the load control operations. Sufficient forecasting accuracy is needed in order to ensure cost-efficient control. When it comes to the operational forecasting of temperature-dependent loads (e.g. electrical heating), accurate weather forecasts are needed.

During 2-3FP, the ANN/MLP based approach for the short-term forecasting of average and sum powers of a customer group was developed. The basic principle is close to the SVR model presented earlier, but in addition time-lagged temperature and power input values are proposed. The accuracy (SSE, std of error, etc) of the proposed ANN model approach was compared against the corresponding accuracies obtained using the load profile (TUT) and partly physically based models (VTT). The results will be published during 4-5FP separately in international journal articles and conference papers.

The initial results obtained show high forecasting performance. Also a combination of ANN based modelling and Kalman filtering approach (VTT, Koponen 2011) is under consideration in order to further enhance prediction accuracies.

4.4 Spatial load modelling

The objective has been on developing more generic and computationally powerful tools for spatial modelling of loads, which better consider information (e.g. properties and trends) extracted from the existing GIS data (e.g. population, buildings, climate etc.).

A spatial modelling concept based on the use of spatial grids or spatial planning units is proposed (Figure 3). Without going here to the details (those will be presented in CIRED 2013, Niska et al. 2013), the basic idea of the approach follows the load profile concept (e.g. Mutanen et al. 2011) but is made in spatial unit level. The benefit of such approach is suggested to be better integrity to available geographic information and related models and scenarios, concerning also DER (such as PHEVs) and DG (solar panels, wind etc).

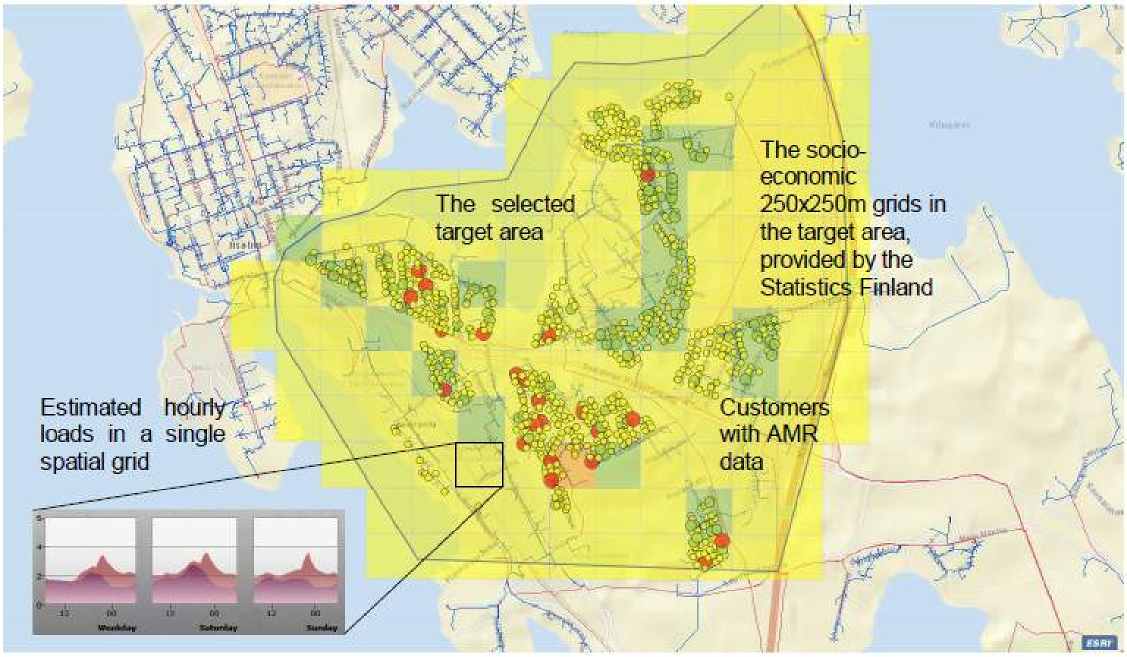


Figure 3. Spatial load modelling concept based on the spatial grids (modified from Niska et al. 2013, CIRED 2013).

5 Conclusions and further SGEM activities

During the SGEM 2-3FP, WP4 Task 4.2, various load modelling activities were performed in a solid collaboration with WP6, Task 6.10 and other SGEM industrial and research partners. Particular viewpoint has been on the requirements of DR, and especially in the control of heating loads. Promising results were obtained using CI/DDM methods, which clearly point out the possibilities and importance of enhanced data analytics and modelling in the field of Smart Grids.

In the next stages, during SGEM 4-5FP, the objective is to further develop and demonstrate the developed load models and concepts in deeper collaboration with the partners. The specific objective is to demonstrate the model functionalities in the real operating environments.

6 References

Aldo et al. 2010. Functional clustering and linear regression for peak load forecasting. International Journal of Forecasting 26, 700-711.

Beccali et al. 2004. Forecasting daily urban electric load profiles using artificial neural networks. Energy Conversion and Management 4, 2879-2900.

Hahn, H., Meyer-Nieberg, S., Pickl, S. 2009. Electric load forecasting methods: Tools for decision making, *European Journal of Operational Research*, vol. 199, 902-907.

Hong 2009. Electric load forecasting by support vector model. Applied Mathematical Modelling 33, 2444-2454.

Koponen, 2005. Real-time pricing project at small customers in Finland, Demand Response Workshop 19 April 2005 in Helsinki.

Koponen 2011. Identification of simple physically based models of the response dynamics of electrical heating loads. In Koponen and Saarenpää (Eds.), Load and response modelling workshop in project SGEM. 10 November 2011.

Mutanen et al. 2011. Customer classification and load profiling based on AMR measurements. CIRED 21st International Conference on Electricity Distribution, Frankfurt 6-9 June 2011. Paper 0277.

Mutanen et al. 2011. Customer classification and load profiling method for distribution systems. IEEE Trans. on Power Delivery 26, 1755-1763.

Niska, Saarenpää, Räsänen, Tiirikainen, and Kolehmainen, 2011. Scenario based load prediction tool for distribution planning and management. CIRED, 6-9 June, 2011, Frankfurt.

Niska and Saarenpää 2011. DDM/CI methods and experiments in load modelling using AMR and other environmental data. In Koponen and Saarenpää (Eds.), Load and response modelling workshop in project SGEM. 10 November 2011.

Niska 2013. Extracting controllable heating loads from aggregated smart meter data using clustering and predictive modelling. Proceedings of ISSNIP 2013, 2-5 April 2013, Melbourne, Australia.

Niska, Saarenpää, Kolehmainen 2013. Computational approach for spatiotemporal modelling of heating loads using AMR and other external data. Accecpted to CIRED 2013.

Räsänen et al. 2010. Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data. Applied Energy 87, 3538-3545

Skön J-P., Kauhanen O. and Kolehmainen M. 2011. Energy Consumption and Air Quality Monitoring System. In: Proceedings of the 7th International Conference on Intelligent Sensors, Sensor Networks and Information Processing '11, Adelaide, Australia, pp. 163-167.

Willis, H. L. Spatial Electric Load Forecasting, 2002. Second Edition Revised and Expanded. New York, Marcel Dekker.