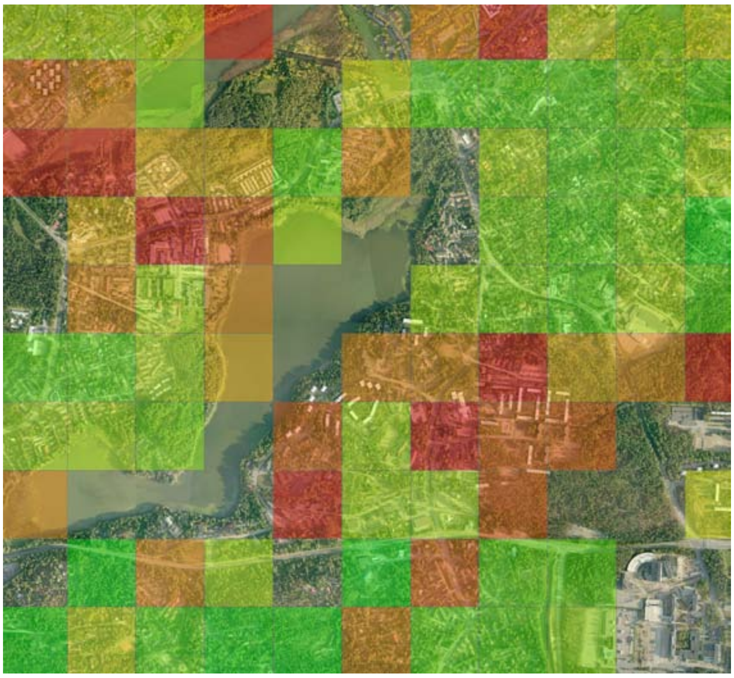
SGEM Research Report D4.2.3

Jukka Saarenpää
Harri Niska

UNIVERSITY OF EASTERN FINLAND

Regional modelling of smart grid related technology adoption - The possibilities and limitations of environmental data



Original title: Regional modelling of loads and DER adoption – The possibilities and limitations of environmental data

Abstract

A huge amount of environmental data has been collected by various public and private sector institutions over the years. Traditionally the data has been difficult to obtain but the recent political decisions have been favourable for opening more of the publicly funded data sources for free use.

The large amount of available data creates new possibilities to improve the planning and management of smart grids. At the same time, the “information overload” requires new methods for efficient management and utilization of the data. In this report the possibilities and limitations of environmental data in planning and management of smart grids are discussed. Moreover, an example of estimating the regional PHEV adoption using external data is presented.

The report is part of the University of Eastern Finland (UEF), SGEM funding period 2, research activities on the usage of environmental data in predicting regional electricity loads. The work is part of the work package 4.2 “Estimation of customer loads, DG and storage”, and more specifically the subtask 4.2.1 “Novel load modelling methods using new available data”.

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

EV	Electric Vehicle
HEV	Hybrid Electric Vehicle
PHEV	Plug-in Hybrid Electric Vehicle
DSO	Distribution System Operator
CHP	Combined Heat and Power
DR	Demand Response
DER	Distributed Energy Resources
SOM	Self-organizing Map
CI	Computational Intelligence
DDM	Data-driven Modelling
SVR	Support Vector Regression

1 Introduction

Some of the Smart Grid's goals are to improve energy efficiency, reliability, economics and sustainability. The realization of the Smart Grid vision highly depends on widespread use of technologies such as distributed energy resources (DER), demand response (DR) and electric vehicles (EVs). Thus, modelling the adoption of Smart Grid related technologies is important and highly relevant especially in distribution network planning and for aggregator and charging station businesses.

Nowadays increasing amount of environmental data is being available for commercial and research purposes. The huge amount of information potentially enables development of new and more accurate models. On the other hand appropriate methods are needed to process the huge amount of data and to extract optimal amount of information out of it. The current trend towards the society's registers being constantly more open and accessible emphasizes the usefulness of investigating new methods to utilize the data.

The environmental data often contains several restrictions that can be technical or legal. The issues and ways to overcome them are discussed in this report. Moreover, an example of modelling with environmental data in case of Plug-in Hybrid Electric Vehicle (PHEV) adoption is briefly presented.

2 Load modelling and Distributed Energy Resources

Load modelling is needed in many sectors of electricity market. For example DSOs need to be able to model loads for planning and managing the distribution network. In network planning the specific need is for models that can predict the loads on medium and long term. For long term predictions the time horizon can span as many as 30 years. The better the magnitude, location and time of the load can be predicted the more savings can be made on investments to the distribution network infrastructure (Lakervi & Holmes 1995).

DER are small scale distributed power generation and energy storage technologies such as solar panels, wind turbines, micro CHP and battery storage. Because of their batteries which can be used to store electric power for feeding it later back to the grid, the plug-in electric vehicles (PHEV) can also be considered as DER. Depending of the situation, the DER can reduce, shift or in some cases increase the momentary loads and are therefore an important aspect to consider when planning the distribution network.

When evaluating the distribution infrastructure's withstanding capability, it is possible to consider the amount of electricity customers, the current loads and the possible future loads in certain scenarios. Possible future loads could be simulated for instance by Monte Carlo experiments or some realistic fixed adoption percentage, based on expert knowledge, could be used for all the regions. However for better capability to prioritize the investments, a model should be used alongside which is able to estimate the DER adoption spatially.

3 Environmental data

In this case, the term environmental data refers to external information that is obtained from the surrounding environment and which is independent of the phenomenon being inspected. In the case of power-distribution network, temperature measurements, socio-economical and land use information are some examples of such external environmental data. Often it includes spatial reference, in which case it can also be referred to as geo-information.

In some applications of load modelling, environmental data is already being utilized. Temperature measurements, for instance, are often used for predicting loads as heating and cooling are normally big part of the overall electric load. However, the huge amount of environmental data sources, which are increasingly available, open new interesting possibilities for load modelling. One possibility could be for example to improve the accuracy of the existing load models by using appropriate environmental data as additional model inputs. Spatial load models could be improved by suitable spatially referenced data sets. Furthermore, direct modelling of the load affecting phenomena e.g. the diffusion of innovations such as PHEVs can be done for better planning of the distribution network.

Many limitations are also present when dealing with environmental data. The availability of the data might have legal or technical limitations. It can be very expensive, its usage can require special permissions or be otherwise restricted (e.g. for commercial use) and instead of easily being available in the Internet, it might involve lot of bureaucracy and manual work. The importance of free use of information has been lately better recognized in the public however. The Finnish Transport and Communication Ministry's decision in principle concerning the availability of public information and the topographic data, soon to be opened for free use, by National Land Survey of Finland serve as good examples of the recent advancements in the openness of information.

The accuracy of the data itself can be sometimes limited. For privacy and technical reasons the data is often restricted to some regional scale (e.g. the Socio-economical Grid Database from Statistics Finland). The limited accuracy of the data suggests developing regional modelling and spatial analysis approaches.

In some cases there already exists a problem with the sheer amount of usable data, also known as “information overload”. When dealing with large amount of (multidimensional) information, it is often difficult for humans to notice the patterns and interrelationships contained within the data. In such data rich conditions, for example computational intelligence (CI), data mining, statistical and geocomputing methods become useful (Niska & Saarenpää 2011). Especially the CI methods (e.g. neurocomputing and genetic algorithms) can aid in structuring and analysing large amounts of data. Moreover, when the phenomenon under study is poorly understood or is unknown, such methods help in finding useful patterns for further study.

4 Modelling smart grid related technology diffusion

The diffusion of technology happens both in spatial and temporal scales. It is dependent of practical factors (e.g. taxation, fuel price and utility) but also of consumers’ personal preferences (e.g. environmental consciousness, comfort). Both are to some extent related with socio-economy and geography. Using publicly available data, the individual cases of adoption are extremely hard to predict, but on group level the behaviour becomes more apparent.

Geo-demographic segmentation is a technique used for classifying different geographical areas to groups by their demographic/socio-economic features (Spielman & Thill 2008). It is for instance often utilized for marketing purposes because of the idea that the behaviour of people being part of the same (socio-economic) group is similar. The usual procedure in geo-demographic segmentation is to cluster the areas to homogenous groups which are later attached with labels describing the group’s consumer type.

The same kind of approach can be used for mapping geographical areas by their technology adoption potential by forming homogenous consumer groups inside of which the adoption ratio is expected to be similar. The higher adoption ratio among a group can be interpreted as higher adoption potential. To get the exact socio-economic features which determine the similarity of a group and which correlate in some way with the adoption, feature selection must be performed. Moreover, the clustering and its correspondence with the adoption rates being examined must be validated. This can be done by comparing the adoption rates of each clustering group with those of some different area that has not been part of the original clustering.

Many Smart Grid related technologies (such as DER or EV) are still not available or have just entered the market. In any case, there is often very little data available from actual market transactions, which makes predicting their diffusion even more challenging task. Thus, it is problematic to apply data-driven modelling (DDM) methods to the problem directly. That said, by finding another product, reasonably similar but still different enough to not be considered as just a different generation of the same product, the adoption data of the existing product can be used to predict the adoption of the new one (Bass et al. 2001). This method, called prediction by analogy, has been used before in marketing research to predict the sales performance of a new product. The similarities and differences of the two products should be closely analyzed when making analogies between them.

4.1 Case PHEV

PHEVs are thought to become increasingly common over the coming years. Having batteries that can be charged from the mains, they will bring new type of load to the grid but also provide interesting possibilities from the energy storage and ancillary service points of view. The first PHEVs are expected to have smaller batteries which, compared to full EVs, make them less useful as energy storages. Despite the lack of battery capacity in the first PHEV models, the capacities will grow as the prices come down, improving the potential for vehicle-to-grid applications. Moreover, different ancillary services, such as frequency regulation with faster response time, could be worthwhile even with smaller battery capacities.

In comparison to other distributed energy resources, it is in some ways harder to predict the availability of PHEVs for vehicle-to-grid services. Because of their mobile nature it is difficult to tell how much capacity is there in the battery at any given moment as well as where the vehicles are located. On the other hand, possibly more data is available about PHEV ownership because of the mandatory registration process for vehicles used in traffic. In this case however, the PHEVs are still a new or in many markets even nonexistent product, and consequently no data is available from the actual product adoption.

PHEVs and their non-plug-in counterparts can be considered as analogous products for several reasons. Both are automobile products with environmental friendliness or fuel savings as their main selling points. Their use cases are very similar but on the other hand PHEV differs from HEV sufficiently to not to consider it as just newer generation of the same product. Both continue to be sold simultaneously and certain dissimilarities make them desirable for different type of usage. Plug-in hybrids have longer all-electric range and speed than the normal hybrid vehicles. It can be expected that the same type of consumers who buy HEVs have higher probability to buy PHEVs.

HEVs have been in the market for some time already and there is data available from the vehicle registrations. This provides opportunity to identify the socio-economic group that adopts HEVs and by analogy the same group is potential adopter of PHEVs. By knowing where the people part of this group live, the geographical areas having increasing amount of PHEVs in the future can be reduced to fever possibilities. Respectively, this helps in the distribution network planning and could also provide decision support for aggregator and charging station businesses.

4.1.1 Data-driven method for predicting the potential areas of PHEV adoption

To demonstrate the use of environmental data for decision support in distribution network planning, a model for predicting the regional PHEV adoption was developed. The steps of the modelling process are presented by Figure 1. As of the beginning of 2012, there are still no PHEVs in the Finnish market. Also the amount of HEV data is still limited. Nevertheless, HEV adoption data was used to create the model which based on the PHEV-HEV analogy hypothesis is expected to provide supportive information for assessing PHEV adoption.

The model input data was obtained from two sources. Finnish transport safety agency’s Vehicular and driver data register contains information about all the registered vehicles including HEVs. HEVs and their holders' addresses were collected and further geo-coded to geographical coordinate form. Socio-economic grid database from Statistics Finland was used to couple HEV ownership with socio-economic characteristics. Moreover, some spatial variables were calculated, such as the distance to nearest city centre and shopping centre. The grid database constitutes from 250m x 250m squares covering whole Finland. Finally, by combining the two datasets, the amount of HEVs per grid square was obtained.

Analysis based on Support Vector Regression (SVR) was performed. It turned out that the socio-economic characteristics provided by the grid database were not enough to model HEV adoption directly for each square. Many squares, some even geographically remote ones, contain similar socio-economics but dissimilar adoption rates. The available socio-economic data is simply not extensive enough to predict adoption on such a small scale. The current low amount of HEVs makes it even more difficult to find common characteristics among the adopters. The personal preferences and individual behaviour are what largely affect the adoption habits. It is not possible to model such level of individualism using the grid database.

As an experiment, Self-Organizing Map (SOM) was used to cluster similar type of squares to same group by which a better accuracy was obtained for the regression model. By compressing the data to more coarse presentation, the individual behaviour gets diminished while the group behaviour is more observable and the adoption can be predicted in larger groups. This however, also reduces the model's geographical accuracy. Therefore a compromise must be done between the model's prediction capability and accuracy. For each group, new adoption rate was calculated by dividing the amount of HEVs in the group by the amount of households. Additionally, feature selection was performed to the full set of variables in the grid database. The most meaningful features were found by training and cross-validating the model iteratively.

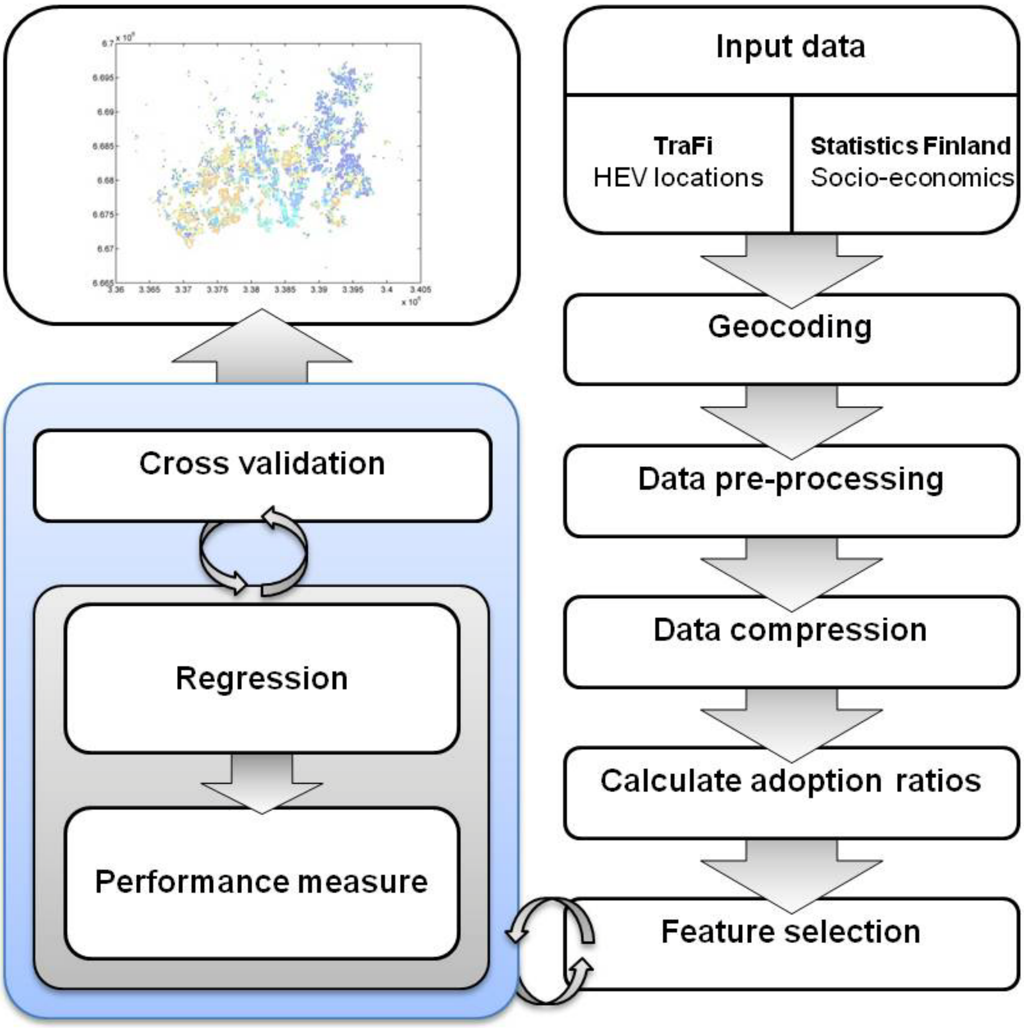


Figure 1. The stages of PHEV adoption modelling process.

To further evaluate the model, its functionality was tested on separate data set from the one used in model training. The literature also provides established theory on diffusion of innovations and the characteristics of the different adopter groups. Good evaluation of model dynamics and its conformity to the theoretical foundation increases the confidence on the model.

4.1.2 The conclusion and additional discussion

Prediction of PHEV adoption involves the problem of not having adequate historical data available. The problem of predicting the adoption of new products is often managed in marketing studies by using older, established analogous product to predict the performance of the new one. In this case the analogous product HEV itself is still such a new product that not much real market transaction data has accumulated. However, as the data represents early adopters, it can be assumed that the model would work to predict the early adopters of PHEV. Moreover, the adopters of HEV have demonstrated the willingness to pay premium price for environmentally friendly and fuel saving car. The same willingness is going to be behind the first purchasers of PHEVs unless a sudden change happens, for instance, in a form of significant technological advancement or tax decision and makes PHEVs more desirable for different type of consumers.

Environmental data is often restricted in accuracy for privacy or measurement technical reasons. The data might have to be generalized so that no individual can be pointed out from the data set. In some cases the information must be left out completely. In this case the grid database represents a data set that has been restricted to groups of geographically outlined squares. The socio-economic data is also restricted in the amount of information. Many variables possibly affecting the adoption are not included because of the difficulty of obtaining such information. Personal details such as lifestyles, hobbies and personal values are some of the things that influence our decision to purchase a product. In the absence of exact personal information the modelling must be done in sufficiently general scale. For the abovementioned reasons it is a necessary action to cluster the statistical squares to adequate number of representative socio-economic groups.

5 Conclusions and recommendations

The environmental data can provide new opportunities in operating and planning the (smart) grid. With appropriate methods the constantly growing amount of data can be harnessed for producing models that have potential to be more accurate than the old ones. However, the environmental data contains many restrictions which have implications on the models based on it. The restrictions, that make the data less accurate or more difficult to obtain, stem from the legal and technical issues. As a consequence, the data might be limited in its spatial or temporal resolution, meaning that it might have been generalized by averaging to certain area or time interval. Additionally, depending of the collection methods and up-to-datedness of the data set, it might contain missing or erroneous values. The data can also be very costly in some cases.

The issues with the spatial accuracy of the data suggest modelling on regional scale. Moreover, the managing of huge amount of information often combined from various different sources can be facilitated by using data-driven or statistical methods. The use of such approach in modelling the PHEV adoption regionally is briefly described by this report. The presented method enables ranking of geographical areas by their tendency to adopt HEVs/PHEVs. The model could be used for instance in estimating regional PHEV/DER potential by an aggregator. DSO could estimate the grid's capability to withstand the future loads when used together with information about the current stress levels of different grid components. For better estimates the model should be further developed to produce actual load curves for different regions, determined by the PHEV usage.

6 References

Bass, F., Gordon, K., Ferguson, T., Githens, M., Directv: Forecasting diffusion of a new technology prior to product launch, Interfaces 31 (2001), no. 3, S82-S93.

Lakervi E., Holmes E. J.: Electricity distribution network design. Institution of Engineering and Technology, London, 1995.

Niska, H. and Saarenpää, J., 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.

Spielman, S., Thill, J., Social area analysis, data mining, and GIS, Computers, Environment and Urban Systems, Volume 32, Issue 2, March 2008, Pages 110-122.