D4.2.1

Methods for using AMR measurements in load profiling

Revision History

Edition	Date	Status	Editor
v0.1	26.3.2013	Created	A. Mutanen
v1.0	8.4.2013	1st version	A. Mutanen

Abstract

This report is a summary of several international publications [1]–[5] published during the SGEM project. In this report, we will review the developed AMR measurement based load profiling methods and present the main results from the aforementioned publications.

This report is part of SGEM Task 4.2.1 “Novel load modelling methods using new available data” and comprises mostly of work done in Tampere University of Technology.

Contents

1 Preface

This report was done as a part of the Finnish national research project "Smart Grid and Energy Market" SGEM. It was funded by Tekes – the Finnish Funding Agency for Technology and Innovation and the project partners.

This report is a summary of several international publications [1]–[5] published during the SGEM project. In this report, we will review the developed AMR measurement based load profiling methods and present the main results from the aforementioned publications.

2 Introduction

Automatic meter reading (AMR) is becoming common in many European countries. In Finland, distribution network operators (DNOs) are required to install AMR meters to at least 80 % of their consumption sites in their distribution networks by the end of 2013 [6]. Many DNOs plan to install AMR meters to all customers. AMR provides DNOs with accurate and up-to-date electricity consumption data. In addition to other functions, this data can be used to update load profiles and classify customers. The availability of AMR data also enables new and more accurate methods of modelling distribution network loads. Accurate load profiles are needed in daily used distribution network calculation, for example in load flow calculation, state estimation, planning calculation and tariff planning.

Distribution network customers are commonly classified to predefined customer classes, and the load of each customer is then estimated with customer class specific hourly load profiles. Currently, this method involves several error sources.

1) Sampling error. Parameters in the existing customer class load profiles can be based on measurements, which are misclassified or comprise an insufficient number of measurement points.

2) Geographical generalization. Load profiles are typically defined in national load research projects. Some of the accuracy is lost due to geographical generalization and within-country differences in electricity consumption are left unmodelled.

3) Profile drift. Electricity consumption is constantly changing but the load profiles are rarely updated.

4) Customer classification. DNOs have limited information on the type of the customers. The type of the customer is usually determined through a questionnaire when the electricity connection is contracted. However, the customer type may later change for instance because of a change in the heating solution.

5) Outliers. Some customers may have such an exceptional behaviour that they do not fit in any of the predefined customer class load profiles.

The above mentioned problems could be solved with the help of AMR measurements. The customer classification and load profiling could be done according to actual consumption data. Since AMR data is collected continuously, the classification and load profiles would remain up-to-date at all times. The classification and accuracy of the load profiles could be checked automatically for instance once a year. The load profiles could also be calculated separately for each DNO or region, thus avoiding the errors caused by geographical generalization. Outliers could be detected and individual load profiles could be formed for the outliers. Individual load profiles could also be calculated for some of the largest customers to improve the load estimation accuracy.

The load profiling methods presented in this report are based on AMR measurements which are used to update customer class load profiles and reclassify customers. Different classification methods from simple reclassification to existing customer classes to K-means and ISODATA clustering (Iterative Self-Organising Data Analysis Technique), GMM (Gaussian Mixture model) and MFA (Mixture of Factor Analysers) are tested [1]–[4]. The results are compared with the existing customer class load profiles. This report shows that updated DNO specific customer class load profiles are more accurate than the existing customer class load profiles. Good results are achieved with all the clustering methods presented in this report, but there are some differences and limitations in each method. These differences and limitation are discussed in this report.

3 Load profiling

3.1 Current load profiling practices

Distribution system loads are commonly estimated with customer class load profiles. Each customer is linked to one of the predefined customer classes, and the load of each customer is then estimated with the customer class specific hourly load profile [7]. This method assumes that the distribution system operator knows which customer belongs to which customer class. In practice, classification errors are common.

In Finnish distribution network calculation applications the load profiles are presented as topographies where expectation value and standard deviation for the customer's hourly load is presented as a linear function of the customer's annual energy consumption. Typical load profile contains 8760 expectation and standard deviation values, one for each hour of the year, and is normalized to base energy consumption of 10 MWh/year. When used, the load profiles are scaled to match the customer's annual energy consumption.

Finish Electricity Association Sener has defined customer class load profiles for 46 different customer groups. Different kinds of customers are well presented in these 46 customer classes but the load profiles are very old and suffer from errors presented in Chapter 2. The Sener load profiles were published in 1992 and are based on measurements done during late 80's and early 90's [8]. Since then, more research has been done to update some of the load profiles and correct the errors in the load modelling procedure [9]. Some electric utilities have also had their own load research projects where they have for example defined individual load profiles for their largest customers. However, the original Sener load profiles are still the only publicly available load profiles. In this report, the new AMR-based load profiles are compared to the original Sener load profiles.

Finland is a cold country and the outdoor temperature has a big effect on the electricity demand. The Sener load research report [8] contained temperature dependency parameters for each load profile but since only January's parameters were published and the importance of using temperature parameters was not properly emphasized, they are rarely used in distribution network calculation.

3.2 AMR-based load profiling

AMR meters provide huge amounts of customer level electricity consumption information and this information can be used to improve the load profiling accuracy. Intuitionally, the two most obvious ways to use AMR meters to improve the load profiling accuracy is to reclassify the customers according their real measured electricity consumption and to update the load profiles.

With AMR measurements, every customer can be classified to customer class which load profile is closes to the customer's real measure consumption. When customers are reclassified with the help of AMR measurements, a considerable portion of customers (up to 90 %) changes customer class. More accurate customer classification means better load profiling accuracy but since the currently used load profiles are not very accurate, there is still much to improve.

A simple load profile updating would solve many issues. Old load profiles and the constant change in electricity consumption habits have caused significant profile drift to the customer class load profiles. During the last decade the use of entertainment electronics has increased, heat pumps and air conditioners have become more common and lighting efficiency has increased, just to name a few changes. AMR measurements can be used to update customer class load profiles. This will have several benefits. Regularly, for instance once a year, done load profile update would keep the load profiles up-to-date at all times. This would ensure that the load profiles keep up with the changing electricity consumption habits. Also, errors that are associated with sampling and geographical generalization would decrease. The sampling errors decrease when measurements from all or almost all customers are used in the load profile calculation. The geographical generalization could be avoided by calculating the load profiles separately for each distribution network area or region. Load profile updating is a lot more effective way to improve the load profiling accuracy than the customer reclassification. Customer reclassification and load profile updating are compared in Figure 1.

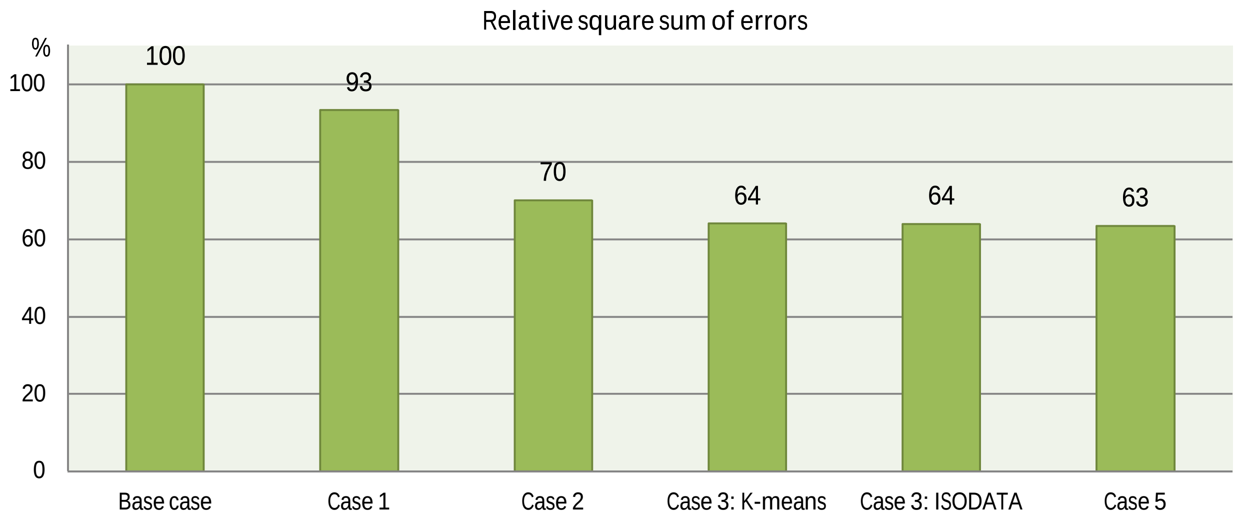

Figure 1. *Comparison between different load profiling methods when modelling residential customers [2].*

Base case: original classification and Sener load profiles

Case 1: customer reclassification and Sener load profiles

Case 2: original classification and updated load profiles

Case 3: customer reclassification and load profile updating with K-means clustering

Case 4: customer reclassification and load profile updating with ISODATA clustering

Case 5: individual load profiles

The load profile update has a bigger effect on the load profiling accuracy than the customer reclassification. The load profile update and the customer reclassification should of course be combined to achieve the best result. However, if the load profile update is done after the customer reclassification, the updated customer class load profile is no longer the nearest load profile for all customers. The customer class reassignment and load profile update should be done again and again until none of the customers change customer class during the reclassification process. Basically, this is a clustering problem.

Clustering is an efficient technique for grouping customers with similar behaviour. In literature, several different clustering methods have been applied to electricity customer classification [10], [11]. During SGEM-project, ISODATA, K-means, GMM and MFA clustering algorithms were used to solve the customer classification problem. The clustering methods are compared and discussed in Chapter 5.

Some customers have such an abnormal behaviour that they do not fit well in any of the predefined customer classes or clusters. In these cases, individual load profiles can be calculated. Individual load profiles can also be used to improve the load profiling accuracy of very large customers. The bigger the customer's electricity consumption is, the more important its load profiling accuracy is. Figure 1 shows that individual load profiling does not benefit small residential customers but as Figure 2 shows, large industrial customers benefit from the individual load profiling. However, using individual load profiles adds strain to the network calculation applications. Therefore, only those customers who benefit the most from the individual load profiling should be modelled individually.

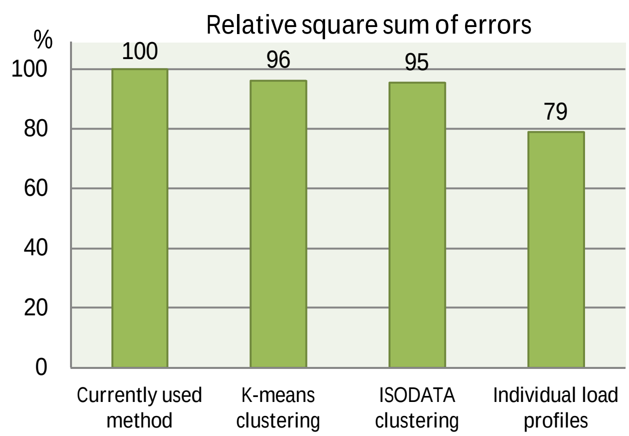

Figure 2. *Comparison between different load profiling methods when modelling large interval metered customers [2].*

In this study, the load profile structure (topography with expectation and standard deviation values for each hour of the year) is kept unchanged and only the effects of reclassification, load profile updating, clustering, individual load profiling and better temperature dependency parameters are studied. The Finnish AMR meters provide measurement information with only one hour resolution so the resolution of load profiles cannot be increased and since most AMR meters measure only active power, separate reactive power profiles cannot be calculated.

4 Preparing for clustering

Prior to clustering, the AMR data has to go through validation and verification, temperature normalization and dimension reduction, regardless of the clustering method used.

4.1 Validation and verification

Depending on the meter data management system (MDMS), the AMR data in the metering database has already gone through various degrees of data validation and verification. Still, since this degree varies from system to system, it is important to validate and verify the data before proceeding to the next steps of load profiling. During this project, we have seen AMR data with polarity errors, unit errors, missing and duplicate values, and changes from cumulative time series to non-cumulative and vice versa. Luckily, errors in AMR data have decreased in recent years when the metering systems have evolved, but especially the missing data continues to pose a problem and must be addressed.

4.2 Temperature dependency calculation and normalization

The weather influences electricity demand in many ways but the outdoor temperature is clearly the most important weather factor. Temperatures vary from year to year and in order to compare AMR measurements from different years, we must normalize the measurements to the same (monthly) temperatures. For that we need to calculate customer specific temperature dependency parameters. In [1], seasonal temperature dependency parameters were calculated form daily average temperatures and daily energies using linear regression. Once temperature dependencies have been calculated, the AMR measurements can be normalised to the same temperatures. Monthly long time averages are used as target temperatures. All this is done before the clustering.

After the clustering, customer class specific temperature dependency parameters are calculated using the same method. The final load profiles are calculated from the temperature normalized sum loads and are thus representing electricity consumption in the long time monthly average temperature. If there is a need to scale the load profiles to different temperatures, the following equation is used to calculate the electricity consumption difference between the expected long time average temperature and the target temperature [1]:

$$\Delta P(t) = \alpha \cdot (T_{ave} - \mathrm{E}[T(t)]) \cdot \mathrm{E}[P(t)],$$

where

$\Delta P(t)$ is the outdoor temperature dependent part of the load P at time t,

T_{ave} is the average temperature of the previous 24 hours, E[$T(t)$] is the expectation value of the outdoor temperature at time t (long-term daily average temperature),

α is the seasonal temperature dependency parameter [%/°C] and

E[$P(t)$] is the expectation value of the load at time t.

4.3 Dimension reduction

Dimension reduction is applied to AMR measurements before the clustering in order to speed up the calculation. Clustering could also be done based on the raw measurements but then the high dimensionality of hourly AMR data (8760 dimensions for one year measurements) would slow down the clustering. In publications [1]–[3], [5] the dimension reduction was done using pattern vectors describing the average consumption of each customer. In [1], the pattern vectors consisted of four seasonal temperature dependency values and 2016 values (12 months × 7 days × 24 hours = 2016) describing the average hourly consumption. In [2] and [3], the temperature dependency values were omitted and only the 2016 values describing the average hourly consumption were included in the pattern vectors. Similarities in temperature dependencies were captured from the yearly load variations. In [4], the clustering problem was studied from the British point of view. The goal was to define typical load profiles (TLPs) from half hourly measured advances. In this case, the data had only 48 dimensions and dimension reduction was not considered necessary.

The 2016 dimensions in a pattern vector is still a large number. It is granted that a higher degree of dimension reduction could have been achieved by using principal component analysis (PCA) or some other similar method. However, the benefit of pattern vectors is their understandable nature. Furthermore, the pattern vectors are also a great tool for creating individual customer specific load profiles. Since each element in a pattern vector is a mean of several values, the random variations of electricity consumption are smoothed out in the pattern vectors. It is also possible to calculate standard deviation values for pattern vector elements.

5 Clustering

Clustering is an analysis technique aimed to determine how the data is organized. Clustering algorithms divide a set of observations into subsets (clusters) so that the observations in the same cluster are similar in some sense. There are several clustering methods suitable for customer classification, some of which are introduced below.

5.1 K-means

The *k*-means method is one of the most popular clustering methods used in statistical data analysis. The *k*-means algorithm assigns each point to the cluster whose centre (centroid) is the nearest. The centroid is the average of all the points in a cluster. The algorithm steps are:

1. Choose the number of clusters, *k*.
2. Randomly assign *k* points as cluster centres.
3. Assign each point to the nearest cluster centre.
4. Recompute the new cluster centres.
5. Repeat steps 3 and 4 until the assignment does not change.

The main problem with *k*-means method is that it requires the number of clusters as an input and the optimal number of clusters is unknown. The number of clusters is always a compromise between clustering accuracy and how easy it is to manage and understand the clusters. In load profiling, a high number of clusters results in a better numerical accuracy but it is difficult to keep track what kind of customers each cluster contains. In [3], the number of clusters was chosen to be the same as the number of original Sener customer classes. This enabled us to use the original customer classification as a starting point of the clustering. On average, this yields better results than the random initialization of cluster centres and we do not need to calculate multiple replicates to get good results. The optimum number of clusters could also be defined with the help of a knee-point criterion, as was done in [1]. However, before we can use the knee-point detection, we must do the clustering with all possible values of k and calculate accuracy indices to each result. This is not very practical due to the excessive computation required.

In its basic form, *k*-means clustering assumes that each input vector has equal effect on the location of cluster centres. In reality, we are clustering customers of different sizes and large customers should have a bigger effect on the location of cluster centres than the small customer. In [3], a weighted *k*-means algorithm was used. The k-means algorithm was customized to weight the input vectors (pattern vectors) with the yearly energy consumption of the corresponding customers.

5.2 ISODATA

In [1], an iterative self-organizing data-analysis technique (ISODATA) algorithm was used to cluster interval metered customers. Basically, ISODATA algorithm is a variant of the k-means algorithm with added heuristic provisions for splitting and merging existing clusters. In theory, the method is unsupervised and adjusts the number of clusters automatically to fit data – the operator does not need to know the exact number of clusters before the clustering is completed. In practice, a starting value for the number of clusters and threshold values for splitting and merging are required.

The final number of clusters is between $K/2$ and $2K$, where K is the user given starting value for the number of clusters. Therefore, the operator has to have a rough idea what is a suitable number of clusters. Even more expertise is needed when selecting appropriate values for the splitting and merging thresholds, which also affect the final number of clusters. Choosing the right threshold values requires advance information on the type of the customers or use of the trial-and-error technique. High threshold values are chosen when clustering customers with high stochasticity and low thresholds are chosen when clustering customers with low stochasticity. Also, the number of customers affects the threshold values. Table I shows the consequences of choosing too small or too large threshold values. A detailed description of the ISODATA algorithm can be found in [1].

TABLE I
EFFECT OF THRESHOLD PARAMETERS (S_S splitting threshold, D merging threshold) [1]

parameter		number of actions		consequence
S_s	D	split	merge	
small	small	high	low	large number of clusters
small	large	low	low	bad classification accuracy
large	small	high	high	long computation time
large	large	low	high	small number of clusters

When the input parameters are chosen correctly, the ISODATA algorithm gives betters results than the k-means algorithm. However, this comes with a price of longer computation time and requirement of used expertise. The results achieved with ISODATA and *k*-means methods are compared in Figures 1 and 2. In practise, the benefits of using ISODATA instead of the faster and simpler *k*-means algorithm are marginal. Due to these reasons, the *k*-means algorithm has been used in the most recent publications [3] and [5].

5.3 GMM

When preparing for [4], several machine learning methods were applied to the problem of customer classification. The following methods were tested:

- Dirichlet Process Mixture Model (DPMM)
- Infinite Gaussian Mixture Model (IGMM)
- Gaussian Mixture Model (GMM)
- Mixture of Factor Analysers (MFA).

In [4], our aim was to study half hourly daily load profiles, i.e. the data had 48 dimensions. Although theoretically interesting, the DPMM and IGMM did not work with daily load profile data. 48 dimensions were too much for these methods. The GMM and MFA worked and are discussed next.

GMM has some interesting properties. In k-means clustering, the clusters are spherical, separate and data points can be assigned only to one cluster, where as in GMM the clusters can be elliptical, overlapping and fuzzy. With full covariance matrix, GMM can be used to evaluate dependencies between data points. Unfortunately, GMM with full covariance matrix is very sensitive to the number of dimensions and needs a large amount of observations. In [4], GMM was trained with diagonal covariance structure. Although the dependencies between advance times could not be studied in [4], the GMM worked as a classifier. Furthermore, the profile means and standard deviations could be derived directly from the component means and covariance matrixes.

GMM suffers from the same problem as *k*-means; the number of clusters has to be defined a priori. In [4], Bayesian Information Criterion (BIC) was used to define the optimum number of clusters. However, as the knee-point-detection methods used in conjunction with *k*-means and ISODATA methods, this method is slow and computationally intensive. GMM requires that the data is Gaussian and normalized to zero mean and standard deviation of one. In [4], the smart meter advances were log-normalized, so that the data would be Gaussian. Figure 3 shows how well the log-normal distribution fits to the residential smart meter data. This result is in line with the results presented in [7].

GMM worked well in the British case where the goal was to cluster daily load profiles with 48 dimensions. However, the 2016 dimensions in the patterns vectors used in the Finnish cases are too much for the GMM. Furthermore, since GMM clusters normalized vectors, the clustering accuracy with original data is not as good as with k-means method. Thus, it is not worthwhile to use GMM solely as a clustering tool.

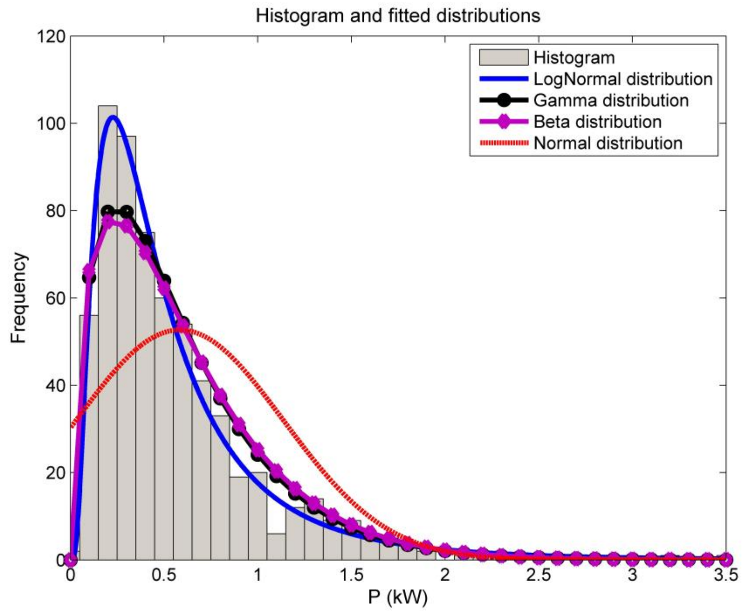

Figure 3. *Histogram and fitted distributions for half hour period 15:00-15:30 in January (weekdays only).*

5.4 MFA

MFA succeeds where the GMM failed, in describing the dependencies between the times of use. As smart meter data contains multiple dimensions, it is difficult to assess which times of use influence each other and how. Multivariate data can sometimes contain correlations between variables and these can be amalgamated allowing only the most informative or uncorrelated variables to be represented in a space of reduced dimensionality. Two examples of models which can reduce the dimension of an observation space and thus discard uninformative variables and reveal dependency structures are Principal Component Analysis [12] and Factor Analysis [13]. Extending the mixture model to factor analysis, allows multiple sub-populations in a sub-space to be captured.

In MFA, full covariance structure can be obtained for all mixture components regardless of the dimensionality of the data or the sparseness of the subpopulation that forms a mixture component. With the covariance matrix, we can study how the times of use influence on each other in a corresponding cluster. In Figure 4, a covariance matrix for 48-dimensional load profile is shown. Dark red areas are strong positive correlations i.e. when a given (row) advance increases, the corresponding (column) advance increases. Blue areas show negative correlation – increases in (row) advance size result in decreases in corresponding (column) advance.

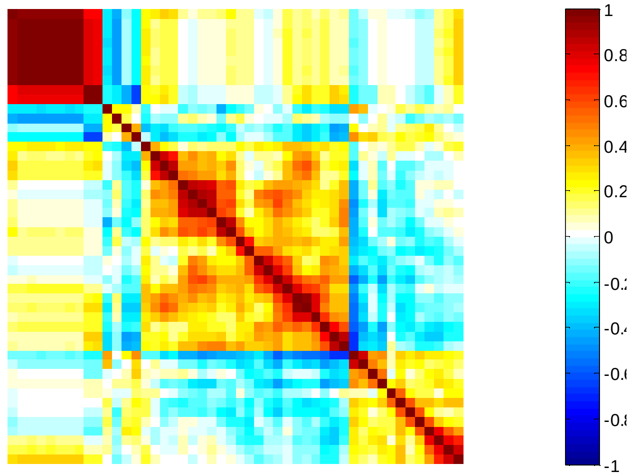

Figure 4. *Example covariance matrix for a 48 dimensional mixture component.*

In a high dimensional representation, we will have difficulties in articulating in the relationships between times of use. The additional advantage of the MFA model is that the factor loading matrix yields a representation of dependence between dimensions as a vector plot in the low dimension subspace. Figure 5 shows one example of this from a single component. The vectors that correspond to each advance can be interpreted as follows: The arrows are the eigenvectors of a covariance matrix with relative directions representing their implied linear dependence: alignment is high correlation while opposition is high negative correlation. Right angles imply linear independence. In the example in Figure 5, advances at time periods 45-47 (10pm to 11:30pm) show a strong correlation reflecting late evening habits with little temporal variation and duration in the order of hours.

In [4], MFA was successfully applied to clustering and modelling of daily load profiles with 48 dimensions. MFA also succeeds in clustering 2016 dimensional pattern vectors but when the number of dimensions is this high, it is difficult to analyze time of use relationships from the covariance matrix. Vector plots could be used for this but then the user should have some advance information which times of use to compare.

MFA also brings more complexity to the model selection process. In MFA, user has to select the number of mixtures (which accommodate various expected load profiles) and the number of subspace dimensions (which capture the drivers of the correlation and variance structure). Finding the optimum number of mixtures and subspace dimensions is even more laborious than the optimization of the number of cluster in *k*-means and GMM methods.

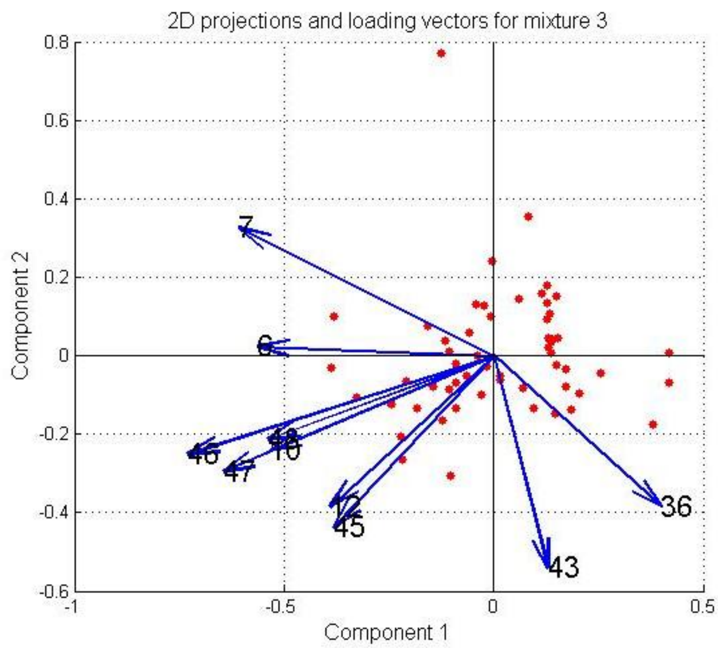

Figure 5. *Vector plot representation of a Factor Analyzer loading matrix.*

6 Other observations

6.1 Individual load profiling

Sometimes the customer's electricity consumption profile is so abnormal that none of the cluster load profiles describes it well enough. In these situations it is worthwhile to consider individual load profiling, especially for large customers whose load models have a large impact on the network calculation accuracy. In [2], it was shown that individual load profiling has only marginal effect on the profiling accuracy of small residential customers but with large industrial customers the individual load profiles improve the load profiling accuracy. These effects have been illustrated in earlier Figures 1 and 2.

In [1]-[3] and [5], the individual load profiles were formed based on the pattern vectors calculated earlier. The use of pattern vectors helps to smooth out the effect of stochastic variation in the load expectation values. Also, the standard deviations can be calculated when each value is a mean of several values. Expansion of pattern vectors into topographies is easy; the average load profile describing one week's consumption is simply duplicated to cover the whole month. In [1], it was shown that individual load profiles calculated this way are better in next day load forecasting than the previous year's measurements used as individual load profiles.

In [3], a two-stage *k*-means clustering was used to separate the large and abnormally behaving customers from others. After the first clustering round, the customers with the largest weighted distance from the cluster centres were selected for individual profiling and the customers with the largest unweighted distances were labelled as outliers and set aside (5 % of the total population). Then the clustering was recalculated. Again, yearly energies were used as weighting factors.

6.2 Accuracy comparisons

In [3], it was shown how AMR based load profiling can improve the forecasting of distribution network peak loads. It was shown that the load temperature dependency information is important when making short term load forecasts but it can be used to improve next year peak load forecasts as well. In cold countries, the peak loads occur during the coldest days of the year and it is quite easy to determine a suitable peak load calculation temperature from the historical temperature information.

In [5], the short-term load forecasting ability of the AMR based load profiles was compared with other short-term load forecasting methods. While the cluster load profiles were not quite as accurate in short-term load forecasting as the best purpose specific load forecasting methods, it was shown that the cluster profiles are significantly better in short-term load forecasting than the currently used customer class load profiles.

7 Conclusions

The large scale introduction of AMR meters is the biggest change load research has encountered in decades. During the SGEM-project we have studied how to utilize these AMR meters in load profiling. The benefits of using AMR measurements in load profile updating and customer classification are indisputable. However, in order to combine load profile updating and customer classification, clustering methods are needed. In this study, we have studied several different clustering methods and found out that many of them are applicable to load profile updating and customer classification. The differences in results achieved with different clustering methods are small; therefore we believe that understandability and computation speed should be more important factors than the absolute accuracy. In the latest development version of our AMR-based clustering method, we have returned to use *k*-means clustering as the basis of our load profiling method.

The next steps in load profiling research done in SGEM-project will focus on quantifying the benefits the AMR-based load profiling brings to the distribution network analysis.

References

[1] A. Mutanen, M. Ruska, S. Repo and P. Järventausta, ”Customer Classification and Load Profiling Method for Distribution Systems,” *IEEE Transactions on Power Delivery*, Vol. 26, No. 3. Jul. 2011.

[2] A. Mutanen, S. Repo and P. Järventausta, ”Customer Classification and Load Profiling Based on AMR measurements,” presented at the 21st International Conference and Exhibition on Electricity Distribution, Frankfurt, Germany, 2011.

[3] A. Mutanen, P. Järventausta, M. Kärenlampi and P. Juuti, ”Improving Distribution Network Analysis with New AMR-Based Load Profiles,” to be presented at the 22nd International Conference and Exhibition on Electricity Distribution, Stockholm, Sweden, 2013.

[4] B. Stephen, A. Mutanen, S. Galloway, G. Burt and P. Järventausta, ”Enhanced Load Profiling for Residential Network Customers,” accepted to *IEEE Transactions on Power Delivery*.

[5] P. Koponen, A. Mutanen and H. Niska, ”Comparison of Short Term Load Forecasting Methods,” submitted to IEEE Innovative Smart Grid Technologies Europe 2013 conference.

[6] Finnish legislation, 66/2009, “Valtioneuvoston asetus sähköntoimituksen selvityksestä ja mittauksesta”. (In Finnish)

[7] A. Seppälä, 1996, “Load research and load estimation in electricity distribution”, Ph.D. dissertation, Helsinki University of Technology.

[8] Sener, ”Sähkön käytön kuormitustutkimus 1992”. Helsinki, Finnish Electricity Association Sener. ISSN 0786-7905.

[9] M. Jalonen, M. Ruska & M. Lehtonen, ”Kuormitustutkimus 2003”, VTT research report, PRO1/P7028/03

[10] G. Chicco, R. Napoli, & F. Piglione, 2006, "Comparison Among Clustering Techniques for Electricity Customer Classification", *IEEE Trans. Power Systems,* vol. 21, 933–940.

[11] G.J. Tsekouras, N.D. Hatziargyriou, & E.N. Dialynas, 2007, “Two-Stage Pattern Recognition of load Curves for Classification of Electricity Customers”, *IEEE Trans. Power Systems*, vol. 22, 1120–1128.

[12] Jolliffe, I.T. Principal Components Analysis, Second Edition. Springer Series in Statistics, 2002.

[13] Everitt, B.S. Introduction to Latent Class Models. Chapman and Hill, London and New York. 1984.