

## THE IMPROVED MODEL FOR THE SPATIAL LOAD FORECASTING OF THE SLOVENIAN DISTRIBUTION NETWORK

Miha GRABNER  
Slovenia  
miha.grabner@eimv.si

Zvonko BREGAR  
Slovenia  
zvonko.bregar@eimv.si

Štefan IVANJKO  
Slovenia  
stefan.ivanjko@eimv.si

Leon VALENCIČ  
Slovenia  
leon.valencic@eimv.si

### ABSTRACT

*In accordance with the EU directive (EU 2009/72/EC) at least 80 % of consumers will have to be equipped with smart meters until 2020, if a cost-benefit analysis is positive for a member state. Therefore, the distribution companies are currently massively replacing old Ferraris meters with the new AMI (Advanced Metering Infrastructure) meters. The analyses of metering data from smart meters allows better understanding of the network conditions in all operating states and help accurately assess the load of existing and new consumers.*

*The paper presents a new analytics application based on big data from smart meters. Using unsupervised machine learning methods of grouping (clustering), the daily load profiles can be determined from a large amount of input data. By examining the load probability distribution in each cluster, consumers' stochastic models are made. The original daily load profiles are reproduced by using the Monte Carlo method, which allows very accurate analysis of LV and MV networks.*

*The results obtained are used for spatial load forecasting. One of the major problems faced by distribution companies in the network planning is to assess the load and location of new consumers. Detailed analyses of existing consumers help solving this problem. The forecasting process was upgraded with newly acquired GIS (geographic information system) data on land plots intended for construction. This gives a detailed view of the area saturation and allows better load forecasting at micro locations.*

*The paper briefly presents how it all fits together to evaluate the future load development for the entire considered area.*

### INTRODUCTION

Milan Vidmar Electric Power Research Institute (EIMV) has long been cooperating with Slovenian distribution companies by providing development plans of MV networks and consumption growth. The long-term load forecasting requires a detailed analysis of the existing situation and future years covering short-, medium- and long-term development. The results obtained in this study are then used for further network development planning.

In accordance with the EU directive (EU 2009/72/EC) at least 80 % of consumers will have to be equipped with smart meters until 2020, if a cost-benefit analysis is positive for a member state. Therefore, the distribution companies are currently massively replacing old Ferraris meters with the AMI meters. This will enable acquiring better data basis for network analyses in terms of load profiles and spatial load distribution.

### BRIEF ANALYSES OF LV CONSUMER BEHAVIOUR

At first it seems that individual LV consumers behave randomly on a daily basis. However, by analysing multiple consumers at the level of transformer station (TS) or at some other higher level in the network the shape of typical daily load profiles can be identified. Figure 1 shows the daily load profile for one household, followed by daily load profiles for 10 and 200 households. We can see that by adding the load profiles we get smoother shapes.



Figure 1: Daily load profiles.

Since we want to examine the habits of individual consumer groups, we need to combine only their load profiles. For example, we may want to combine the daily load profiles of only one certain type of one-phase households. This seems to be an impossible task without complex computer analysis since a single consumer is on an annual basis represented by as many as 35,040 data points. The paper presents an unsupervised machine learning method that allows grouping similar daily load profiles together. By using the approximation of load probability distribution and the Monte Carlo method, the original daily load profiles are reproduced.

### SMART METER TIME SERIES ANALYSES

The article will consider only one part of the distribution company Elektro Ljubljana. Baseline electric load data are taken from 2015. We analyzed the area of Domžale, which is powering a little less than 20,000 consumers with

approx. 300 TS. In the base year TS were supplied by 110/20 kV Domžale substation, which is equipped with two 31.5 MVA transformers.

### Data preparation

The project area of Domžale was chosen because it is entirely covered with smart meters. AMI meters measure the energy consumed in time intervals of 15 minutes. These measurements are then recalculated to an average active power. There are 96 values of the data on a single day theoretically available for each consumer and, consequently, 35,040 values for the whole year.

Due to the seasonal interdependence of consumption and the differences in power consumption between working days and weekends (holidays) we divided the consumer daily load profiles in 6 groups. The first division was made according to the season in three groups, and then each group was additionally subdivided in two groups, (working days and weekends). So we modelled three periods: winter, summer and inter-season. The exact boundaries between different seasons were determined on the basis of average daily temperatures.

Consumers are further divided in 5 groups: 1-phase and 3-phase households, 1-phase and 3-phase other LV consumption and one MV group.

For certain consumers there are occasional outages measurements through the period, so it is necessary to pre-process the raw measurements. Consumers whose measurements are missing are not taken into account when calculating the optimal groups. However, after the optimal groups have been calculated, that type of consumers is recalculated back to one of the groups obtained.

Daily load profiles of low voltage consumers are thus represented by a total of 30 matrices. The most numerous group are the 3-phase households having the matrix size of 600,000 x 96. Furthermore, the daily load profiles of consumers must be normalised since this improves the quality of the clustering algorithm. The problem of grouping time series data is solved with the help of machine learning.

### Machine learning clustering algorithm

Machine learning is the subfield of computer science that “gives computers the ability to learn without being explicitly programmed”. In unsupervised machine learning, no labels are given to the learning algorithm, leaving it on its own to find structure in its input. Unsupervised learning can help us discover hidden patterns in data [1].

Cluster analysis or clustering is the task of grouping a set of objects in such a way that objects in the same group (called a cluster) are more similar to each other than to those in other groups (clusters). Cluster analysis can be achieved by various algorithms that differ significantly in their notion of what constitutes a cluster and how to efficiently find them [2]. A classification algorithm finds relationship between a set of input features and a set of discrete output classes. There are many different

classification algorithms but they mostly assume that the input features are independent. However, this supposition does not hold when dealing with time series data since the values in near time intervals tend to be similar. Suitable classification algorithms for time series rely on the notion of similarity between time series.

To better express the notion of distance between similar time series we used *dynamic time warping* (DTW) distance. As when calculating Euclidean distance here too we take the square root of sum of squares:

$$dtw(Q, C) = \min_{(i,j)} \sqrt{\sum_{k=1}^K (Q(i) - C(j))^2}$$

but here the two time stamps  $i$  and  $j$  involved in the innermost subtraction  $(Q(i) - C(j))$  at both time series do not need to be the same. In effect this means that it is allowed to compare different time steps thus giving the impression of time warping hence the name of the method. Of course, the combinations of indices  $i$  and  $j$  are not arbitrary; the goal is to find such combinations that the overall criterion function in total  $K$  steps would be minimised. The optimal path is found by calculating the rectangular *DTW* matrix. Initially the upper-left corner of this matrix is set to 0 whereas all other elements of the first row and first column are set to  $\infty$ . Then the following recursion is applied:

$DTW(i, j) = (Q(i) - C(j))^2 + \min(DTW(i - 1, j - 1), DTW(i - 1, j), DTW(i, j - 1))$   
yielding the distance  $dtw(Q, C)$  in the lower-right corner  $\sqrt{DTW(n, n)}$ . Calculating all  $n^2$  elements of the *DTW* matrix might be computationally expensive if input time series  $Q$  and  $C$  are large. Since it is not very likely that the resulting path from the upper-left corner to the lower-right corner in the *DTW* matrix would strongly deviate from the main matrix diagonal we may limit the calculation to only a few sub and super diagonals to speed-up the calculation. The other way to enhance the calculating speed is to apply lower bounding that can be calculated in linear time, such as the:

$$LBKeogh(Q, C) = \sum_{i=1}^n (C(i) - U_i)^2 I(C(i) > U_i) + (C(i) - L_i)^2 I(C(i) < L_i)$$

where the upper bound  $U_i$  is calculated as  $\max(Q(i - r; i + r))$ , the lower bound  $L_i$  is calculated as  $\min(Q(i - r; i + r))$ , for a certain reach  $r$  and indicator function  $I$  [4], [11].

To arrange time series of daily 15-minutes electricity measurements into clusters we used the classification algorithm named K-means which is the most widely used algorithm for unsupervised learning. The total number of clusters has to be set manually. Initially clusters are chosen randomly. Then a procedure is repeated as many times as necessary to achieve convergence. In the first step the time series are assigned to the nearest clusters. It is here that the dynamic time warping distance (DTW) (including lower bounding) is applied to determine the similarity (equals

distance) between time series. In the second step the centroids of clusters are recalculated. We programmed the core K-means algorithm and DTW with lower bounding in Fortran for numerical efficiency. On the other side, the data manipulations, visualisations and histogram curve fitting have been programmed in Python.

### Clustering results

The application groups each daily load profiles in a particular cluster. As we present each consumer with multiple daily load profiles the result has to be interpreted as a probability of individual daily consumer load profile being in a given cluster. Due to the overly broad scope, the method will be explained only by choosing an example of one-phase households during winter workdays.

Figure 2 shows daily load profiles of one consumer classified in each of the clusters. The plot in the upper-left shows all daily load profiles for one household. In the following figures, we have plotted the daily load profiles of one household, which belong to each group. Numbers in the title brackets show the share of daily load profiles of that household in each of the clusters.

Figure 3 shows the resulting load profiles for all one phase households during the winter workdays. This type of consumption can be described by three average load profiles. The grey colour line in the upper-left plot marks the average load profile. In the following figures, we have plotted all daily load profiles, which belong to each group. Blue colour lines show all the load profiles of a particular group, with the red colour line on a secondary y-axis being the average of the group. Green histograms on the right side of the figure show load probability distribution. Here we can see that because of great complexity it would be impossible to manually classify these load profiles.

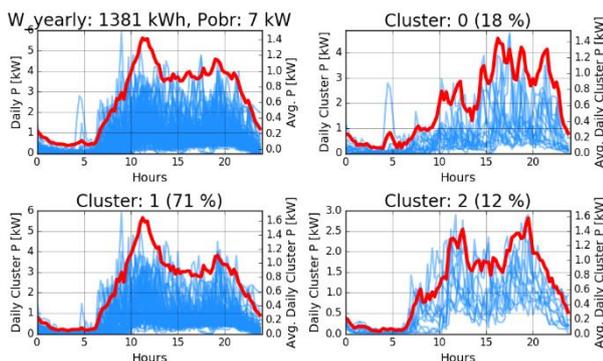


Figure 2: Load profiles for one consumer [5].

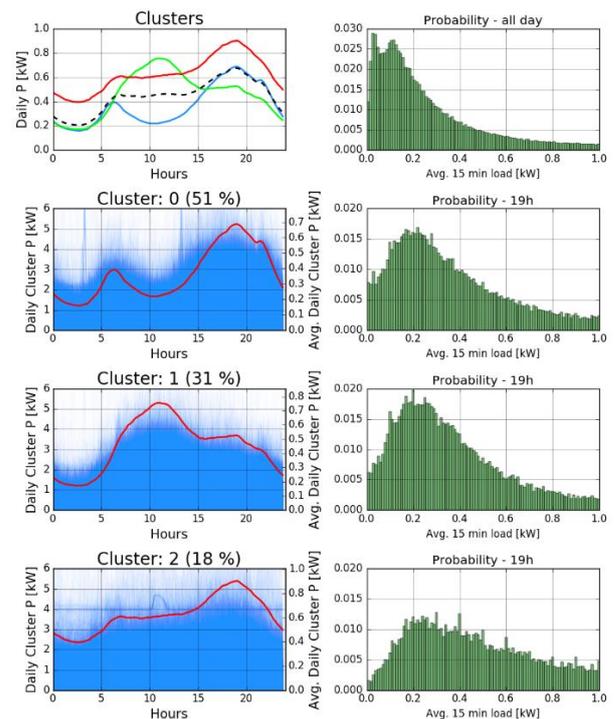


Figure 3: Clustering results [5].

### Building stochastic LV consumer model

Examining the probability density function of each cluster for all time intervals help us make stochastic models of LV consumers. This allows analysing all possible network operating states. Therefore, the distributions can be well approximated by the Weibull probability density functions. Figure 4 shows the load probability distribution for a selected time interval in a cluster and Weibull function fit.

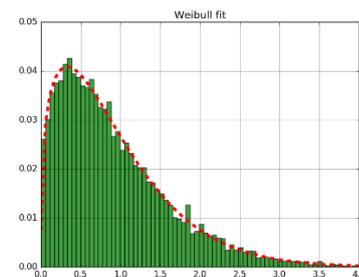


Figure 4: Load histogram and Weibull function fit [5].

Characteristics of each cluster will be described by a probabilistic distribution functions. The equation of the Weibull function is as follows:

$$f(x, \lambda, k) = \frac{k}{\lambda} \cdot \left(\frac{x}{\lambda}\right)^{k-1} \cdot e^{-\left(\frac{x}{\lambda}\right)^k}$$

This means that we have to calculate the value of  $k$  and  $\lambda$  for all clusters for each time interval. Daily load profiles could be described with fewer functions if only certain hours of the day were selected. For example, we could study only peak states of all clusters. This has to be additionally studied.

Figure 5 explains the reproduction of daily load profiles for one household. First we can see that each consumer is presented with its share (probability) of daily load profiles in each cluster. Therefore, the figure presents the average daily load profile for each cluster. Load distribution for each time interval in every cluster is then approximated with Weibull functions.

The last step is a reproduction of daily time series. Analyses are carried out with the help of the Monte Carlo method. As every consumer is presented with the probability of being in each cluster we have to use unequal probability sampling for each of the Weibull functions in every cluster.

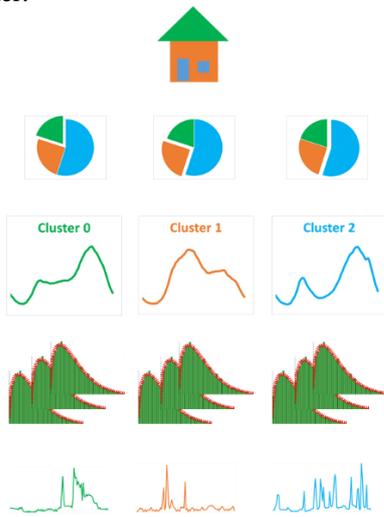


Figure 5: Reproduction of daily load profiles for one household.

## EVALUATING THE LOCATION OF NEW CONSUMERS WITH THE HELP OF GIS DATA

The location of consumers itself has an important role in spatial load forecasting. The process of predicting the location of new consumers and predicting the development of particular micro-locations was upgraded with the help of the newly acquired GIS data. In Slovenia, the land records are maintained in the land register, where the type of land is defined by the plot [6]. These data are then linked to the data on land use from the public register of real estates. We connect each plot with real estate data obtained by municipal plans for the construction of certain types of buildings (e. g. residential areas, commercial zones, etc.). For each plot, we have an indication of what proportion of the surface area of the plot is intended for construction. The consumption of new consumers is then estimated with spatial analysis of the existing consumers.

This allows us to accurately distinguish between existing and future load. Until now we have considered each TS in the future as a kind of micro zone where the growth of existing consumption is also attributable to a hypothetical new consumption. The new method allows better analyses of area saturation. We predict exactly where the possible

future expansion of the settlements, economic zones, etc. might take place.

Figure 6 shows a part of a suburban area, which is powered by two TS. Consumers are indicated by circles that are coloured by the feeding TS. The red polygons highlight existing buildings. Blue polygons mark the land plots for which we have data on what share of them is intended for the construction of buildings. All land plots intended for the construction are in residential areas.

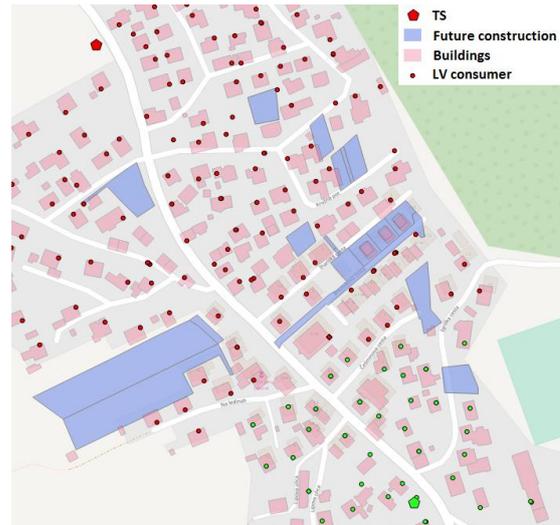


Figure 6: Suburban area – spatial analyses [3].

Next, the load obtained by spatial analysis is added to the closest existing TS. If this TS is anyway intended for replacement in future periods, then, in the context of our development studies, we examine also the problem of searching the new optimal location for this substation.

The method is currently still in the development stage. A further objective is to improve the spatial load analysis of new consumers, using machine learning methods.

## SPATIAL LOAD FORECASTING

In our model, consumers are divided into two initial groups. In the first group, the biggest consumers are considered. A survey about their past, present and future peak load and energy demand is carried out by distribution companies. Data obtained for this consumer group are directly used in further calculations. Until recently, electricity demand forecasting has been realised within the so called zonal model, mentioned in the previous chapter. A surrogate modelled system has been used in the future years instead of the actual real geographic representation. The forecasting process is based on combination of a “top down” and “bottom up” principle. The new methodology originates from the existing consumers and estimated new consumers. Load projections are prepared taking into account the results of analyses presented in the previous chapters. This way the transparency of data and the quality of planning process is improved. Figure 7 shows the peak load substation forecast. The diagram shows the sum of

load demand for each consumer group.

In the planning process of existing non-surveyed consumers we make energy demand projections in accordance with the expected growth rates in the previous period, the expected rate of development at the national level and the expected trend in consumption in comparable countries. This energy growth is the result of changes in the load diagram profiles. In this way the consumers are shifted between different clusters due to electricity appliances.

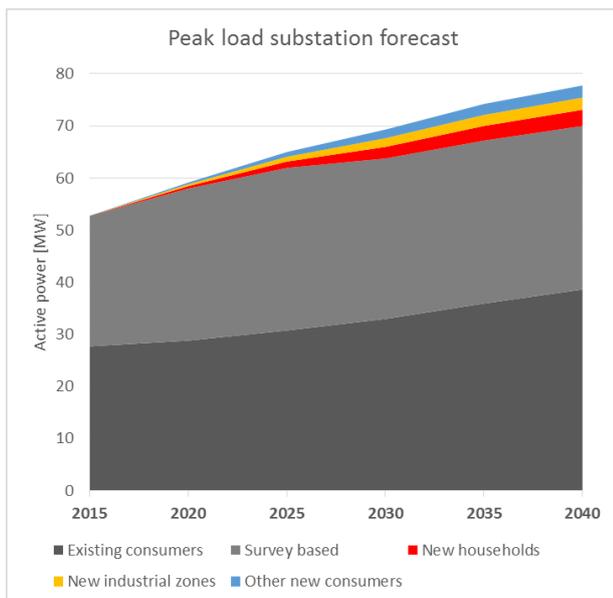


Figure 7: Peak load substation forecast.

New consumers can be seen in the future years. In addition, the available development plans and documents (economic development strategy, development strategy by statistical regions, urban plans, demographic projections, energy policy etc.) must be considered. The load profile of new consumers is also calculated on a basis of existing daily load profiles in accordance with expected changes. This method is still in development, the first provisional results are presented on figure 7.

## CONCLUSION

The paper presented a new analytics application based on big data from smart meters. Using machine learning methods of grouping (clustering), the daily load profiles can be determined from a large amount of input data. Additionally, this allows distinguishing between different types of households and other consumers. By examining the load probability distribution in each cluster, consumers' stochastic models were made. Since each consumer is presented with several load profiles the results have to be understood as the probability of each consumer being in each of the clusters. By using the Monte Carlo method, original daily load profiles can be reproduced, which allows performing accurate analysis of LV and MV networks.

The analyses of metering data from smart meters allows better understanding of the network conditions in all operating states and help to accurately assess the load of existing and new consumers. The results were then used in the spatial load forecasting.

In the future we want to upgrade the modelling of consumers' daily load profiles that will include weather normalisation, real estate type and the number of people living at a certain address etc. This would allow making even better models for existing and consequently for new consumers.

Our goal is to foster better utilization of the smart metering data enabling better network planning and related processes.

## REFERENCES

- [1] [https://en.wikipedia.org/wiki/Machine\\_learning](https://en.wikipedia.org/wiki/Machine_learning).
- [2] [https://en.wikipedia.org/wiki/Cluster\\_analysis](https://en.wikipedia.org/wiki/Cluster_analysis).
- [3] "Map data copyrighted OpenStreetMap contributors and available from <http://www.openstreetmap.org>".
- [4] E. Keogh, C. A. Ratanamahatana, 2005, "Exact indexing of dynamic time warping", *Knowledge and Information Systems*, 7, 358-386.
- [5] Hunter, J. D., 2007, "Matplotlib: A 2D graphics environment", *Computing In Science & Engineering*, vol. 9, 90 -95.
- [6] <http://www.e-prostor.gov.si/>.
- [7] H. Lee. Willis, 2002, *Spatial electric load forecasting*, ABB Inc., Raleigh, North Carolina.
- [8] F. McLoughlin, A. Duffy, M. Conlon, 2015, "A clustering approach to domestic electricity load profile characterisation using smart metering data", *Elsevier Applied Energy*, vol. 141, 190–199.
- [9] Š. Ivanjko, 2005, "Spatial model for the long term electricity demand forecasting of distribution companies", 7. konferenca slovenskih elektroenergetikov.
- [10] Š. Ivanjko, 2013, "The impact of the economic crisis on the development of the electricity consumption", 11. konferenca slovenskih elektroenergetikov.
- [11] [http://www.cs.ucr.edu/~eamonn/LB\\_Keogh.htm](http://www.cs.ucr.edu/~eamonn/LB_Keogh.htm).
- [12] El-Khattam, W., Hegazy, Y.G., Salama, M.M.A., 2006, "Investigating distributed generation systems performance using Monte Carlo simulation", *Power Systems, IEEE Transactions*, vol. 21, no. 2, pp. 524-532.