New Clustering-based Forecasting Method for Disaggregated End-consumer Electricity Load Using Smart Grid Data

Peter Laurinec, and Mária Lucká

14.11.2017

Slovak University of Technology in Bratislava

Motivation

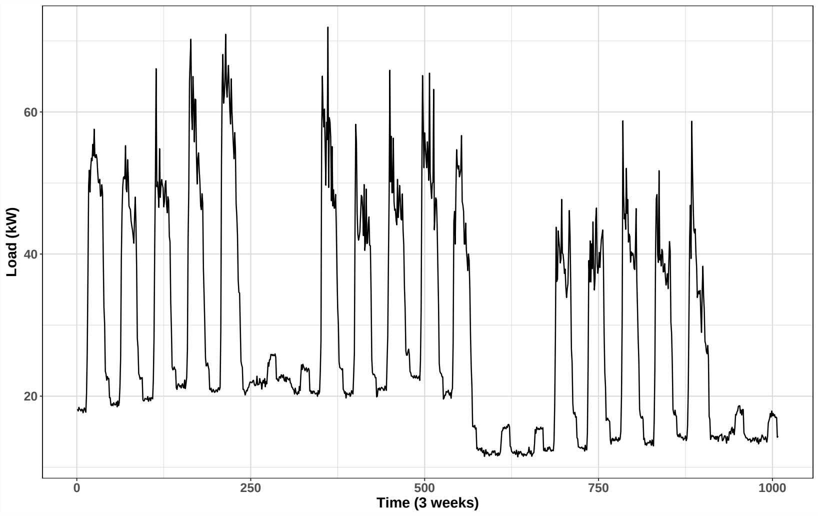
More accurate forecast of electricity consumption is needed due to:

- Optimization of electricity consumption.
- Distribution (utility) companies. Deregulation of the market. Purchase and sale of electricity.
- Ecological factors.

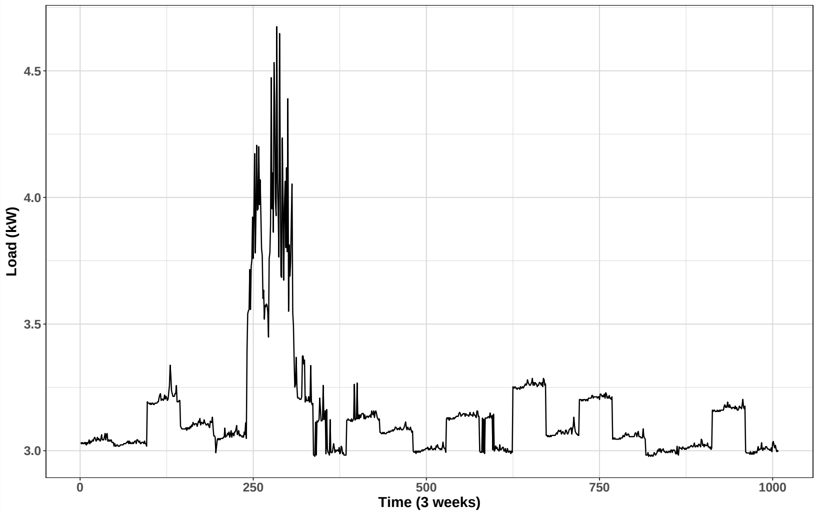
However, it is very difficult task for **individual end-consumers** due to:

- Stochastic behaviour (processes).
- Many factors influencing the consumption:
  - Seasonality
  - Weather
  - Holidays
  - Market

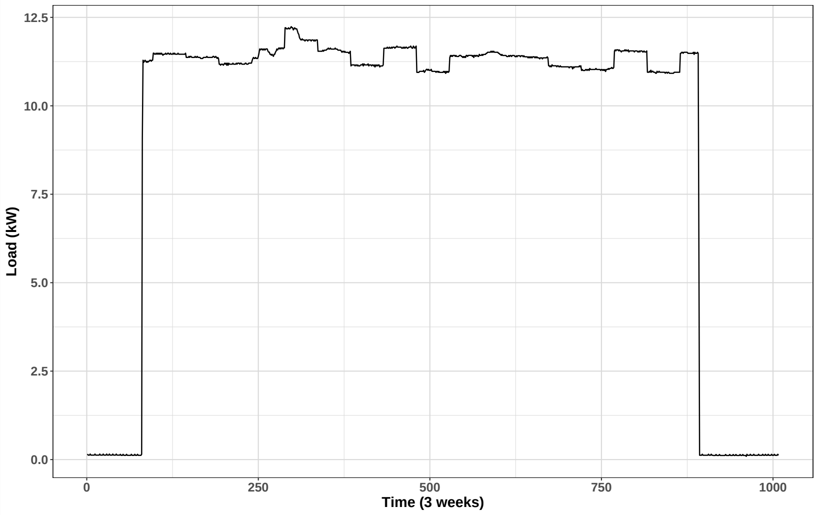
Example of consumers electricity load



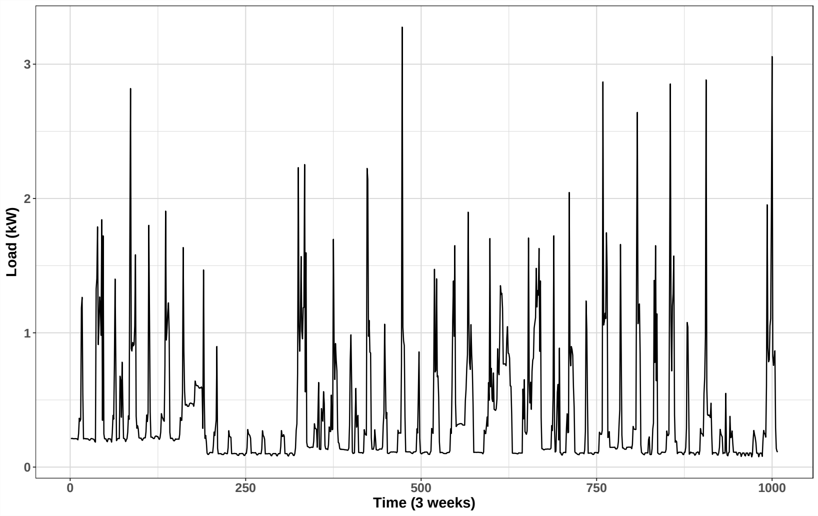
Example of consumers electricity load



Example of consumers electricity load



Example of consumers electricity load - residential



Classical vs. our approach

The classical way is to train a model for every consumer separately (drawbacks).

Our approach uses data from **all consumers** in a smart grid to overcome stochastic changes and noisy character of data (time series).

Solution: clustering of all consumers.

Our method

We will suppose that N is a number of consumers, the length of the training set is 21 days (3 weeks) whereby in every day we will consider $24 \times 2 = 48$ measurements, and we will execute one hour ahead forecasts.

1. Starting with iteration $iter = 0$.
2. Creating of time series for each consumer of the lengths of three weeks.
3. Normalisation of each time series by z-score (keeping a mean and a standard deviation in memory for every time series).
4. Computation of representations of each time series.
5. K-means clustering of representations and an optimal number of clusters is computed.
6. The extraction of K centroids and using them as training set to any forecasting method.
7. The denormalisation of K forecasts using the stored mean and standard deviation to produce N forecasts.
8. $iter = iter + 1$. If $iter$ is divisible by 24 ($iter \bmod 24 = 0 \bmod 24$) then steps 4) and 5) are performed otherwise they are skipped and the stored centroids are used.

Representation of time series

After normalisation -> computation of representations of time series.

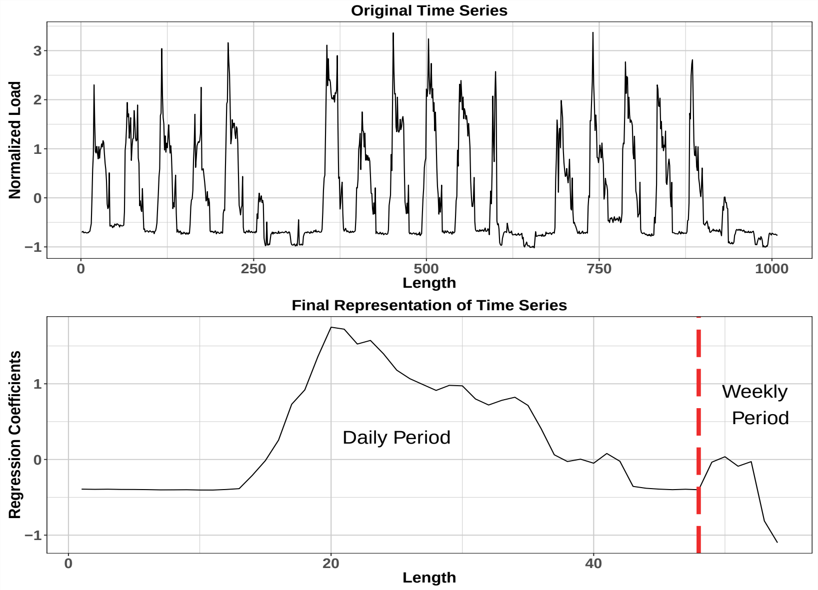
We conducted from our previous works [1] that clustering **model-based representations** significantly improves accuracy of the forecast of the global (aggregate) consumption.

For a representation, **regression coefficients** from the **multiple linear regression** is used. The linear model is composed of daily and weekly seasonal parameters.

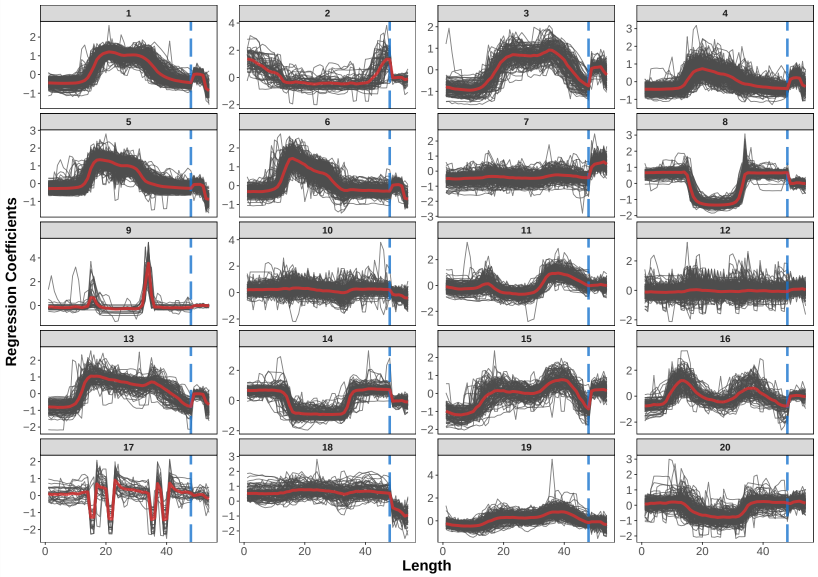
$$x_t = \beta_{d1} u_{td1} + \cdots + \beta_{ds} u_{tds} + \beta_{w1} u_{tw1} + \cdots + \beta_{w6} u_{tw6} + \varepsilon_t$$

[1] Laurinec et al., WCECS (2016) and ICDMW (2016)

Representation of time series

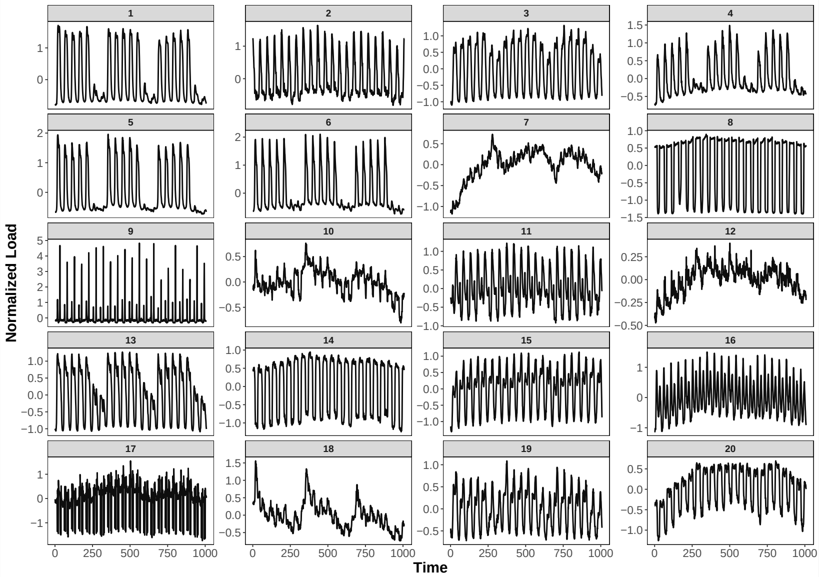


Clustering



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
Regression Coefficients
Length

Final centroids



Forecasting methods

Four methods were implemented

- Seasonal naive method (SNAIVE)
- Multiple Linear Regression (MLR)
- Random Forest (RF)
- Triple exponential smoothing (ES)

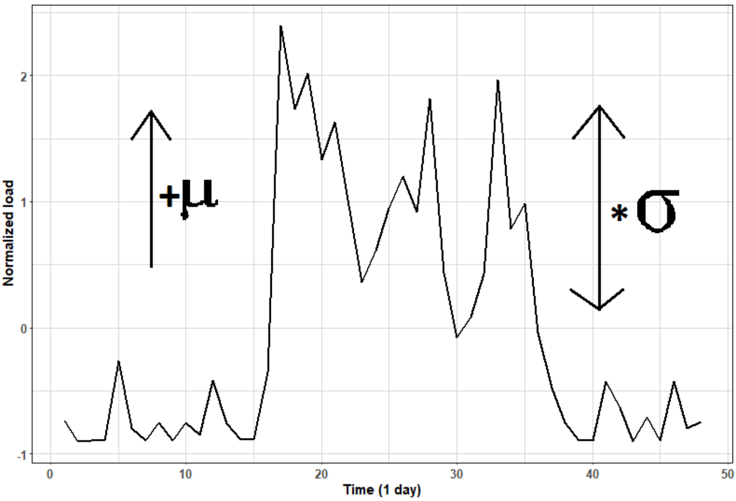
MAE (Mean Absolute Error):

$$\frac{1}{n}\sum_{t=1}^{n}|x_t - \overline{x}_t|,$$

where x_t is a real consumption, $\overline{x}_t$ is the forecasted load and n is a length of data.

Scaling forecasts

Denormalising *K* centroid-based forecasts by stored mean and standard deviation from every consumer (*N*).



Data for experiments

We used two different datasets consisting of a large number of variable patterns that were gathered from smart meters. This measurement data includes Irish and Slovak electricity load data.

For the Irish residential testing dataset (3639 consumers) the data measurements from 1.2.2010 to 21.2.2010.

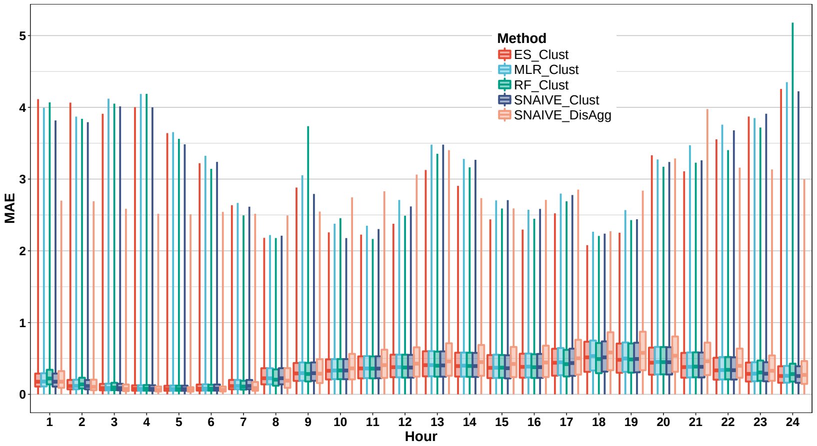
For the Slovak factories testing dataset (3607 consumers) the data measurements from 10.2.2014 to 2.3.2014.

Evaluation

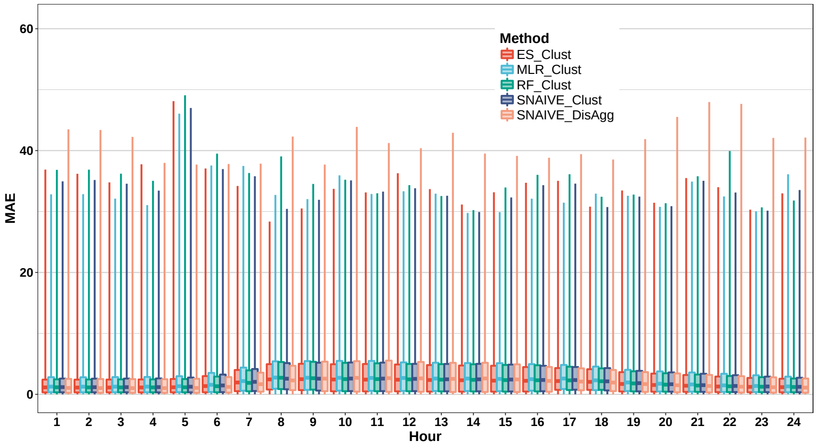
MAE	Ireland dataset		
	Mean	Median	Max
SNAIVE_DisAgg	0.3807 ± 0.203	**0.1928 ± 0.147**	3.014 ± 1.3
SNAIVE_Clust	0.3373 ± 0.178	0.235 ± 0.143	2.6605 ± 1.192
MLR_Clust	0.3403 ± 0.18	0.2394 ± 0.146	**2.6453 ± 1.187**
RF_Clust	0.3394 ± 0.18	0.2425 ± 0.147	2.675 ± 1.192
ES_Clust	**0.3359 ± 0.177**	0.2387 ± 0.144	2.6629 ± 1.189

MAE	Slovak dataset		
	Mean	Median	Max
SNAIVE_DisAgg	2.6903 ± 2.854	**1.769 ± 2.27**	16.1599 ± 14.621
SNAIVE_Clust	2.7873 ± 2.858	2.1479 ± 2.452	14.0958 ± 12.711
MLR_Clust	2.9326 ± 2.984	2.3109 ± 2.612	**14.0306 ± 12.673**
RF_Clust	2.7639 ± 2.836	2.0765 ± 2.388	14.4476 ± 13.081
ES_Clust	**2.6752 ± 2.771**	2.0283 ± 2.357	14.1695 ± 12.816

Ireland dataset results



Slovak dataset results



Conclusion

- Newly proposed **clustering-based** forecasting method for end-consumer load using **all data from a smart grid**.
- We proved that our clustering-based method **decreases the forecasting error** in the meaning of an **average** and the **maximum** (high rates of error).
- However, the error rates did not decrease with respect to the median because of the nature of smart meter data.
- Our method needs to train only K models (in our case about **28**) instead of N models (**thousands**) that is leading to a huge **decrease of the computational load**.

Future work:

- More experiments to find the number of optimal clusters.
- Other centroid-based clustering methods like K-medians, K-medoids and Fuzzy C-means can be also used.