CITIES
Centre for IT Intelligent Energy Systems

Clustering of energy consumption for energy planning

Per Sieverts Nielsen, Xiufeng Liu, DTU Management
Adam Brun, Aarhus School of Marine and Technical Engineering

CITIES AGM
DTU Lyngby

9th November 2020

Content

Topics:

- Why do we want to cluster energy consumption profiles?
- Methods used for clustering
- Cases and challenges of making stable cluster

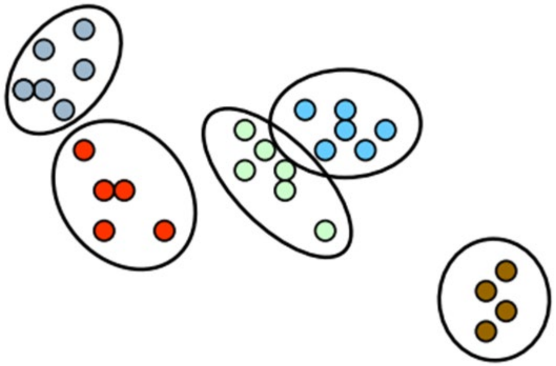
Cases:

- Clustering residential heat consumption profiles
- Clustering heat consumption profiles from heat exchangers
- Clustering industrial electricity consumption profiles
- Clustering residential electricity consumption profiles - How stable are the clusters?

Why do we want to cluster energy consumption profiles?

Transforming dynamic consumption data into useful information for the energy sector

- We get larger and larger amount of very dynamic energy consumption data - the more fine grained the more dynamic
- The energy system, however, is very static
- Clustering and segmentation is a way to identify and extract valuable information in the dynamic data to be used for prediction of energy consumption for short and long term planning in the energy sector

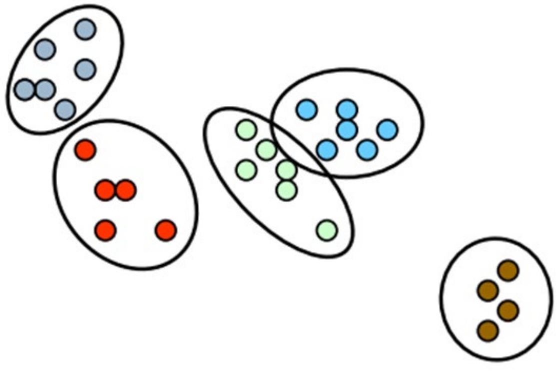


Why do we want to cluster energy consumption profiles?

Large amount of fine-grained smart meter data are available:

- The large amount of data needs to be processed and transformed into a useful format where the important information can be identified and extracted

- Segment customer into groups with similar consumption behaviors
- Identify anomalous energy consumption
- Predicting future energy consumption
- Can for instance be used for demand-response program



Why clustering - for the heating sector?

Cases we will be talking about:

- Clustering residential heat consumption profiles (smart meter data from district heating)
- Clustering heat consumption profiles from heat exchangers (district heating network data)
- Clustering industrial electricity consumption profiles (smart meter data from electricity customers)
- Clustering residential electricity consumption profiles - How stable are the clusters? (smart meter data from electricity customers)

In the heating sector the research can be used to:

- Improve energy efficiency and decarbonisation
- Building energy performance benchmarking (EPCs)
- Analyze the regulations on energy efficiency of buildings
- Provide personalised energy services and design demand-response programs

The story

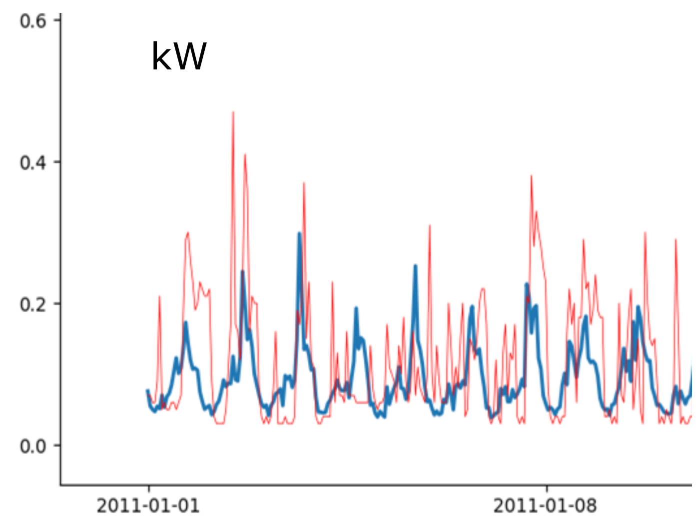
- There are 5.8 mio people in Denmark with different preferences for having a good life
- Some of us live in a single family house, others in apartments
- The energy consumption vary from family to family and in reality we have 3 mio different energy consumption profiles in Denmark - one for each household in Denmark
- Consumption is highly dynamic and - the more fine-grained the data are the more dynamic it becomes
- It is highly dynamic when we look at the individual household profile with 15 minutes data.
- The dynamic nature of the data decreases with aggregating to 1 hourly data, or averaging profiles from more household. We start getting a very familiar pattern with a peak in the morning and a peak in the evening when we average consumption from 10 households.
- The well known consumption pattern therefore is an average of many very different consumption profiles

Individual household electricity consumption profile vs mean from 10 households

The actual electricity consumption in the individual households is highly dynamic

Only by averaging at least 10 households are we getting the conventional profile

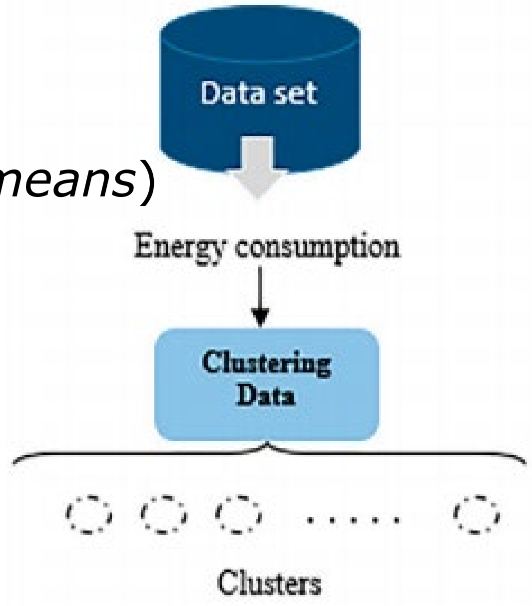
It is therefore also clear that the challenge is to cluster these individual consumption profiles into meaningful clusters



Alexander Tureczek, 2017

What is clustering?

*"**Clustering** is the process of putting data in a group so that they have the greatest similarity and are very dissimilar from data from other clusters."*



Four main categories of clustering methods:

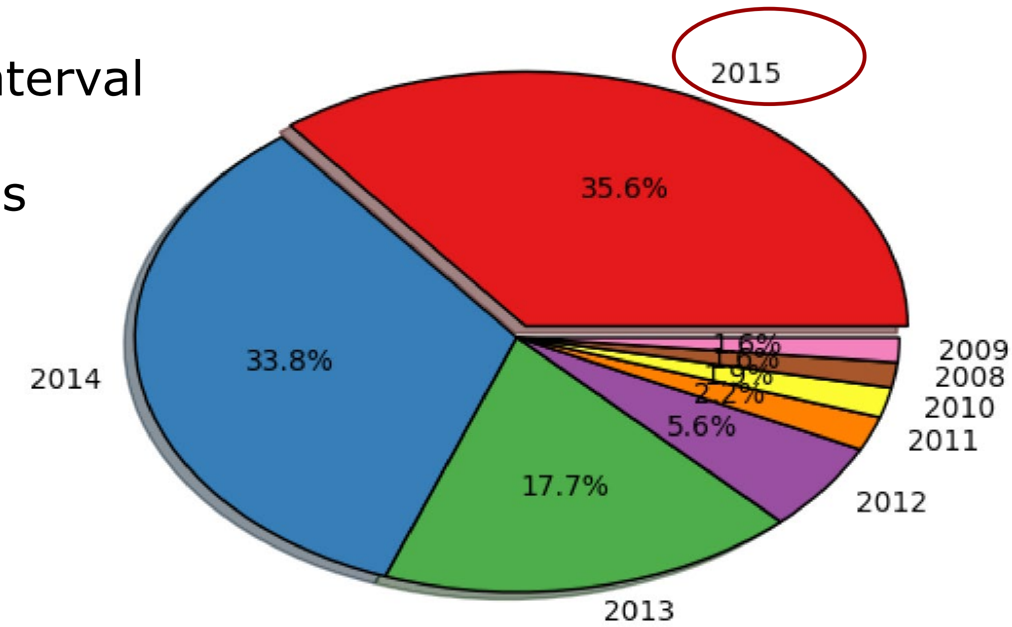
- Partitioning-based (*e.g., k-means, fuzzy C-means*)
- Hierarchical-based (*Agglomerative*)
- Density-based (*e.g., DBSCAN, kNN Density*)
- Model-based (*e.g., applying GMM model*)

The following will make the clustering-based knowledge discovery from residential heating consumption data.

Clustering analysis on district heating data

The data set: *AVA district heating consumption data*

- With a resolution of 1 hour interval
- 15,065 installations
- 9,051,636 hourly load profiles
- Range from December 2008 to November 2015

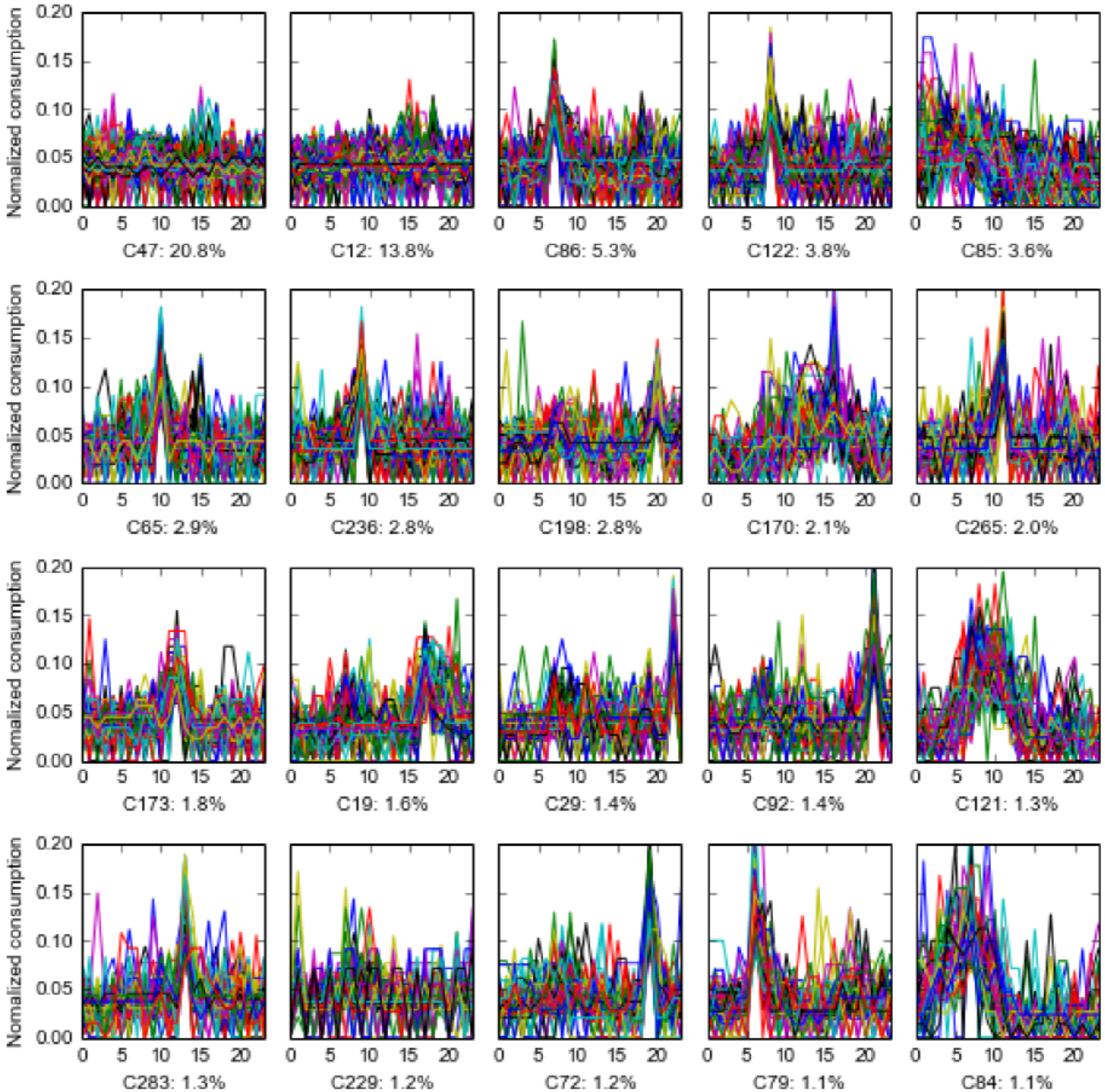


This study:

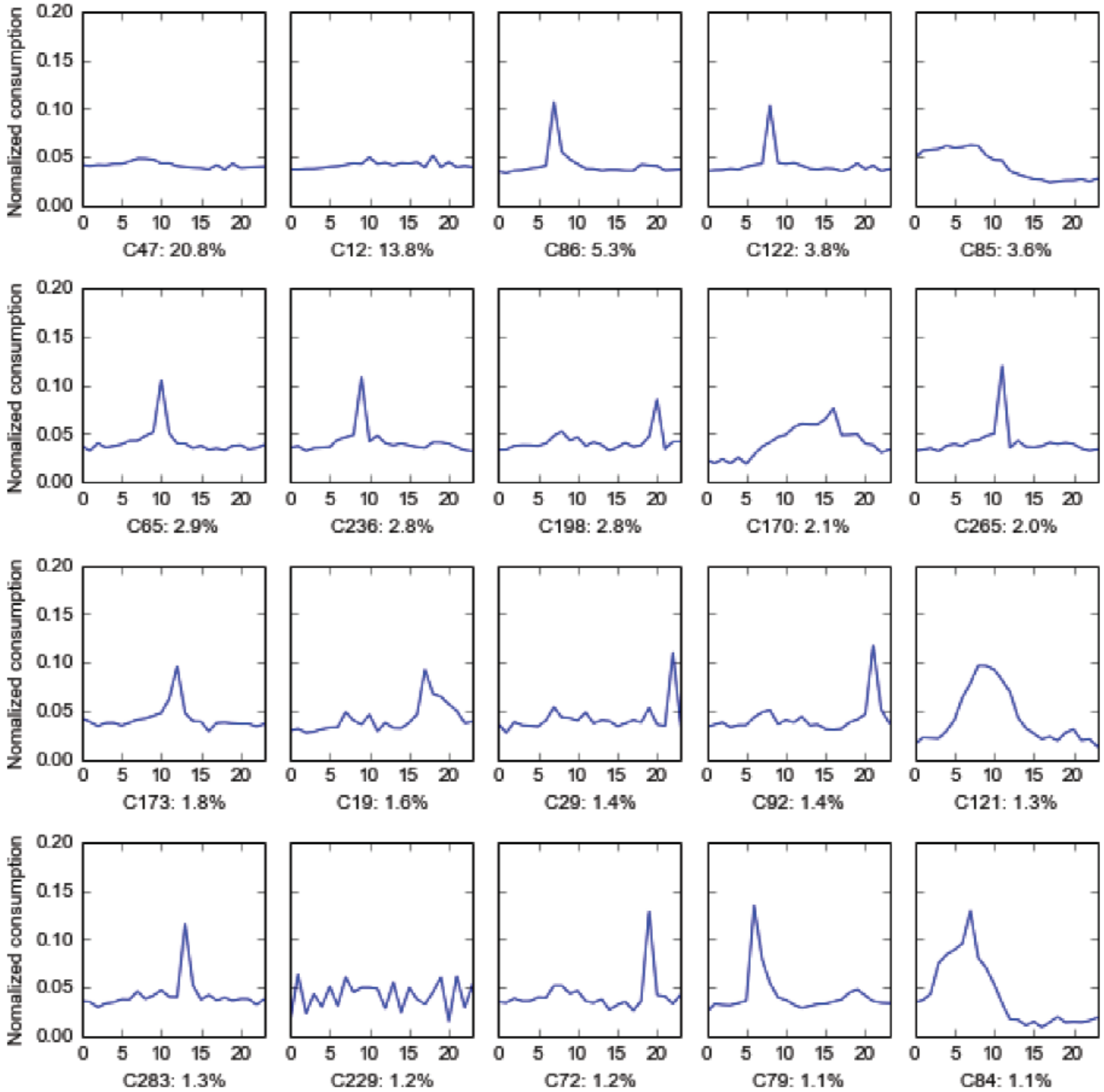
- Uses the data of 2015
- Select measurements only from single-family houses

P Gianniou, X Liu∗, A Heller, P S Nielsen, C Rode. Clustering-based analysis for residential district heating data. Energy Conversion and Management 165 (2018) 840–850

Typical daily consumption of district heating consumption



Typical daily consumption of district heating consumption



Clustering heat consumption profiles from heat exchangers

- A dataset covering 49 heat exchangers in the AVA distributions systems was analysed for analysing heat consumption in 49 districts. The study focused on clustering hourly consumption data for the entire month of January 2017. The optimal number of clusters was identified to 7.
- A surprising result from the ACF clustering is that 5 out of 7 clusters had identical structure only offset by differences in co-efficient's values.
- This suggested that most of the heat exchange stations consumption pattern can be described by the same underlying time series model, regardless of composition of the consumers supplied by the individual heat exchange stations.
- This was surprising as the demographic and consumer composition of an area intuitively was expected to influence the consumption pattern.
- This finding is important to reduce capital expenditures when expanding the district heating grid.

A.M. Tureczek, P.S. Nielsen and H. Madsen et al. / Energy & Buildings 182 (2019) 144–158

Clustering industrial electricity consumption profiles

- In a preliminary study Peter Nystrup and Henrik Madsen has studied unsupervised learning of patterns in hourly load measurements from smart meters belonging to Danish companies – with a data set over 2 weeks.
- Through cluster validation, they were able to find 8 clearly distinct load profiles, which they compared with the industry classification of the cluster constituents.
- They found a significant overlap between industry classification and load profile. In many cases, most companies within an industry adhere to the same load profile. However, the results also showed that industry classification is not sufficient for load profile segmentation, since companies within some industries have very diverse load profiles.
- The coupling between load profile and industry sector improves the possibility for forecasting the future load profile. With the proposed methods it was possible to quantify the uncertainty of the profiles, which is necessary in order to conduct sensitivity analysis in relation to long-term energy system planning.

How stable are the clusters on residential electricity consumption?

- We typically assume that household energy profiles are similar from week to week.
- In a study where data were analysed from a dataset of 26,562 anonymised customers from Syd Energy (smart meter data), we identified the optimal number of clusters to 6 clusters in the dataset. That means that we expect 15-20% of the profiles in each cluster.
- When we assume the individual profile is similar from week to week, we therefore also expect that if we have identified 15-20% of the consumption profiles have similar profiles from week to week over for instance 3 months.

Tureczek, A. M. (2019). Analysis of High Frequency Smart Meter Energy Consumption Data. Technical University of Denmark (PhD thesis).

How stable are the clusters on residential electricity consumption?

Week 40 overlap with rest of weeks in Q4

Week	40 to 39	40 to 40	40 to 41	40 to 42	40 to 43	40 to 44	40 to 45	40 to 46	40 to 47	40 to 48	40 to 49	40 to 50	40 to 51
Cluster 0	23%	100%	21%	16%	19%	15%	14%	14%	13%	13%	12%	12%	14%
Cluster 1	11%	100%	9%	7%	8%	7%	7%	6%	7%	7%	7%	7%	10%
Cluster 2	10%	100%	10%	7%	9%	10%	9%	10%	9%	9%	9%	9%	9%
Cluster 3	6%	100%	5%	4%	4%	2%	4%	3%	3%	4%	5%	2%	5%
Cluster 4	10%	100%	12%	11%	13%	17%	14%	16%	15%	15%	16%	14%	10%
Cluster 5	7%	100%	6%	6%	6%	6%	6%	6%	6%	7%	7%	6%	9%

- Illustrates Varatio for each cluster combination with week 40. Dark green indicates a 50%+ Varatio coefficient. Light green indicates Varatio estimated at between 20-50%, yellow at 10-20%, light red at 5-10% and dark red at <5% of maximum variance as defined by Varatio.

How stable are the clusters on residential electricity consumption?

- The clusters are at best 20% identical across the weeks in any quarter, meaning that 80% of meters within a cluster rapidly disperse to other clusters in subsequent weeks.
- In other words - a certain family profile didnt have a sufficiently stable consumption profile from week to week - to be able to stay in the same cluster
- Maybe this is due to changes in our consumption profile - as soon as a family member goes on a business trip, a child goes to stay over with a friend, another child has a stayover, everyone goes holiday etc etc ect
- The practical meaning of this is that a cluster is not “clean” enough for ie long term planning (long term investment) – the cluster did not contain the same customers continuously each week over 3 months.

Conclusions

- Energy consumption data are highly dynamic and energy supply is static
- We need to cluster fine grained consumption profiles to identify useful customer information in large data sets – the information is important for short term and long term energy planning
- Clustering customers with similar energy consumption is very useful information for utilities for optimising operation and running campaigns. But we need to make sure the customers are clustered in stable clusters
- We also see that clustering purely based on the data is challenges for instance to keep stable clusters. Including other type of information on the customers appears to help developing stable clusters.

Conclusions

- Future work – the quality of the tools we use for predicting/clustering consumption will determine the need for "buffering" between production and consumption. If we can cluster down to 10 houses that could indicate "local buffering" but then if large consumers is more unpredictable then we need "individual buffering". Or perhaps both.
- Buffering in this context can be technology or tools such as flexibility tools, batteries, heat reservoirs or incentives for adjusting consumption via for instance tariffs.
- Working with prediction (investments in R&D) must be in balance with investments in buffers.