

The ELMAS dataset

Authors:

- Kevin Bellinguer (kevin.bellinguer@minesparis.psl.eu)
- Robin Girard (robin.girard@minesparis.psl.eu)
- Alexis Bocquet (alexis.bocquet@minesparis.psl.eu)
- Antoine Chevalier

License: This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/> or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

Cite this work as K. Bellinguer, R. Girard, A. Bocquet, and A. Chevalier, ELMAS: a one-year dataset of hourly electrical load profiles from 424 French industrial and tertiary sectors (submitted in 2023)

Description

This dataset provides a set of 18 load profiles that represent main industrial and tertiary sectors in France for the year 2018.

The ELMAS dataset is derived from a total of 55,730 consumption time series initially split into 424 business sectors and three levels of subscribed capacity. The customer's field of activity follows the Statistical Classification of Economic Activities in the European Community (NACE)[1], which is a four-digit industry standard classification used in the European Union composed of 21 sections, 88 divisions, 272 groups, and 615 classes. For anonymity concerns, the initial times series are averaged according to their NACE coding and level of subscribed capacity.

Discrepancies between the temporal patterns of customers that belong to the same NACE section highlight the need to resort to another clustering approach. Thus, a K-means algorithm is used to gather the business groups sharing similar temporal patterns into 18 clusters. The resulting clustering shows that numerous NACE sections are scattered over various clusters, which increases the global heterogeneity of the clustering while spoiling the interpretation. The proportion of these dispersed NACE classes in terms of annual energy consumption remains low, which suggests that a manual reorganisation has little impact on the global consistency of the clusters. This manual reclassification is conducted in such a way that scattered NACE classes are gathered in the cluster that possesses the highest share of the considered NACE section. The energy consumption time series dataset represents a limited panel composed of 55,730 customers, which may bias the output load profiles in comparison with the whole French panel of industrial and tertiary customers. To fill this gap, Enedis provides the annual energy consumption of a wider range of customers for the year 2019. This annual energy consumption dataset is used to generate weights implemented in the clustering approach and to derive weighted average time series for the clusters.

Dataset structure

File	Col. name	Format	Units	Description
<i>NACE_classification.csv</i>	Section / Class	character		Coding of the NACE sections / classes
	Section_description / Class_description	string		Description of the NACE sections / classes
<i>Nb_customers_time_series.csv</i>	Section / Class / Power_level	character		Coding of the NACE sections and classes
	Nb_customer	float		Number of customers
<i>Annual_energy_time_series.csv</i>	Power_level / Section / Class	character		Level of subscribed power / Coding of the NACE sections and classes
	Energy	float	kWh	Annual energy consumption of the considered time series
<i>Annual_energy_weights.csv</i>	Power_level / Section / Class	character		Level of subscribed power / Coding of the NACE sections and classes
	Energy	float	kWh	Annual energy consumption of a wider panel of customers
<i>Mean_sd_nace_classes_time_series.csv</i>	Power_level / Section / Class	character		Level of subscribed power / Coding of the NACE sections and classes
	Mean / Sd	float	kWh	Average and standard deviation of the time serie
<i>Temperature.csv</i>	Time	character		Temporal sequence following the French standard: "DD/MM/YYYY hh:mm"
	Temperature	float	°C	Temperature at the national level
<i>Clusters_before_manual_reclassification.csv</i>	Power_level / Class	character		Level of subscribed power / NACE classes coding
	Cluster	int		Code of the assigned cluster
<i>Clusters_after_manual_reclassification.csv</i>	Power_level / Class	character		Level of subscribed power / NACE classes coding
	Cluster	int		Code of the assigned cluster
<i>Time_series_18_clusters.csv</i>	Time	character		Temporal sequence following the French standard: "DD/MM/YYYY hh:mm"
	1 → 18	float	kWh	Weighted average of the consumption of the clusters
<i>Energy_consumption_per_unit_Surface_area.csv</i>	Class / Description	string		NACE class coding / Description of the NACE class
	Energy	float	kWh	Annual energy consumption
	Surface	float	m ²	Surface area of buildings that belong to the NACE class
	Energy_m2	float	kWh/m ²	Annual energy consumption per unit area

References

[1] EUROSTAT, *NACE rev. 2*. Luxembourg: Office for Official Publications of the European Communities, 2008.

Version history

- V1.0: Initial release (16/03/2023)