

ELMAS: a one-year dataset of hourly electrical load profiles from 424 French industrial and tertiary sectors

Supplementary material: data analysis

Kevin Bellinguer¹, Robin Girard¹, Alexis Bocquet¹, and Antoine Chevalier²

¹MINES Paris, PSL University, Centre PERSEE - Centre for Processes, Renewable Energies and Energy Systems, Sophia-Antipolis, 06904, France

²Technical Direction, Enedis, Courbevoie, 92400, France

Objectives

It is challenging to access the electricity consumption records of industrial and tertiary companies due to confidentiality issues. Here, we propose generic electricity consumption profiles associated with 18 relevant business sectors. These profiles are derived from 55,730 consumption time series initially split into 424 business sectors (or Statistical Classification of Economic Activities in the European Community (NACE) sections) and three levels of subscribed capacity.

To preserve anonymity, a two-level clustering approach is employed. First, the time series of the various companies are aggregated w.r.t. to their business sectors and their subscribed level of power. Then, a clustering approach is performed to group business sectors that share similar temporal patterns, before aggregating them.

An analysis of the clusters reveals that numerous NACE sections are scattered over various clusters, which increases the global heterogeneity of the clustering while spoiling the interpretation of the clustered data. 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, while taking into account the specificity of the section.

The remainder of this document provides a throughout description of the 18 derived clusters and an analysis of their properties.

Clusters description and identification

First of all, Figure 1 highlights that clusters 1, 2, and 5 gather the majority of NACE classes but also the highest shares of annual energy consumption. In that respect, clusters 10 and 12, exhibit a high energy consumption despite possessing a relatively low number of NACE classes.

Figures 2-4 provide a finer insight of the distribution of the NACE sections within each cluster. Typically, we observe that the clusters are characterised by prevailing sections in terms of consumed energy. An in-depth analysis of the NACE class distribution leads to an identification of the predominant business sectors:

Cluster 1 is mainly composed of activities that belong to the manufacturing sector, and especially those specialised in the production of plastic products and pharmaceutical preparations. Therefore, this cluster is named **Manufacturing processes**.

Cluster 2 At first sight, clusters 2 and 3 seem redundant, both contain high shares of activities related to the wholesale and retail trade. However, a closer analysis reveals that the former is mainly characterised by non-food-related businesses. Thereafter, this cluster is defined as **Trades (non food)**.

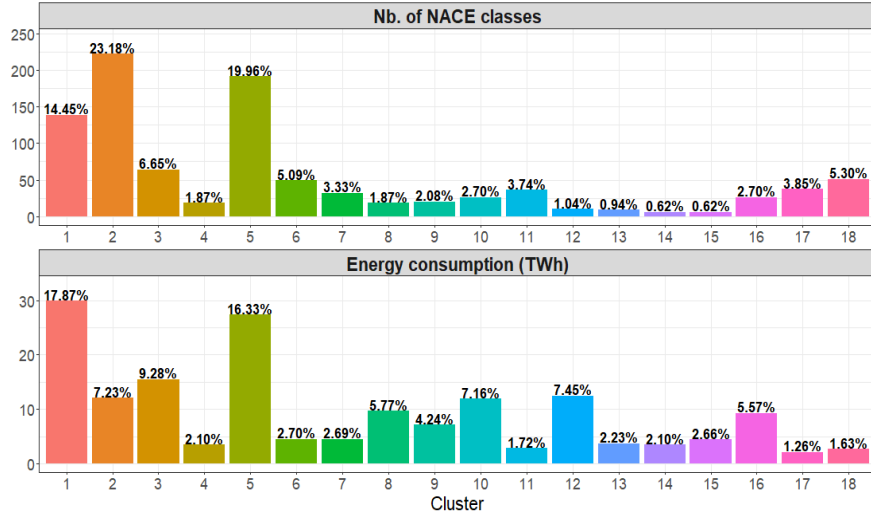


Figure 1: Distribution of the number of NACE classes and annual energy consumption w.r.t. the different clusters.

Cluster 3 On the contrary, cluster 3 is mainly influenced by food-related trades, and more precisely by the class (47.11) *Retail sale in non-specialised stores with food, beverages or tobacco predominating*. Thus, this cluster is designated as **Trades (food)**.

Cluster 4 This cluster is mainly composed of education-related activities. The alias **Education** is given to cluster 4.

Cluster 5 gathers office-based activities such as public administration, head offices, holding companies, to name but a few. Thus, the word **Office** is used to describe this cluster.

Cluster 6 At first sight, clusters 6 and 7 seem very alike. Two main sectors compose the former, namely *Agriculture, forestry, and fishing*, and *Transportation and storage*. Crop growing occupies an important place, followed by land transport of passengers. This cluster is defined as **Crop farming and transportation**.

Cluster 7 This cluster is dominated by livestock breeding (e.g. dairy cattle, swine, poultry) and is designated hereafter by **Livestock farming**.

Cluster 8 Water supply (e.g. water collection and treatment) and information-related activities (e.g. telecommunications) represent the biggest shares of annual consumed energy. As a result, this cluster is named **Water supply and telecommunications**.

Cluster 9 Clusters 9 and 13 exhibit great similarity in terms of distributed NACE sections; the (I) *Accommodation and food service activities* sector possesses the highest share of consumed energy for both clusters. However, we note that the former gathers food-related activities, while its counterpart collects accommodation-oriented companies. Cluster 9 is thus called **Restaurants**.

Cluster 10 This cluster is similar to cluster 1 w.r.t. the NACE section distribution. Here, the primary distinction lies in the fact that this cluster is mainly dedicated to the **Food industry**.

Cluster 11 Activities related to the **Wine industry** are predominant in this cluster. The latter gathers most of the components of the wine production chain (e.g. growing of grapes, manufacture of wine).

Cluster 12 This cluster is dedicated to **Energy supply and rental activities**.

Cluster 13 This cluster complements cluster 9; both gather activities belonging to the (I) *Accommodation and food service activities* NACE section. However, this cluster is focused on accommodation-based activities, and is therefore defined as the **Hotels** cluster.

Cluster 14 This cluster, which is dedicated to the **Bakery** field, gathers activities that range from the manufacture of bread to its retailing.

Cluster 15 **Property management companies** represent the main business sector of this cluster.

Cluster 16 can be identified by **Hospital activities** and nursing.

Cluster 17 This group gathers mainly **Recreational and social activities** such as sport-related facilities.

Cluster 18 is exclusively composed of **Construction**-related activities.

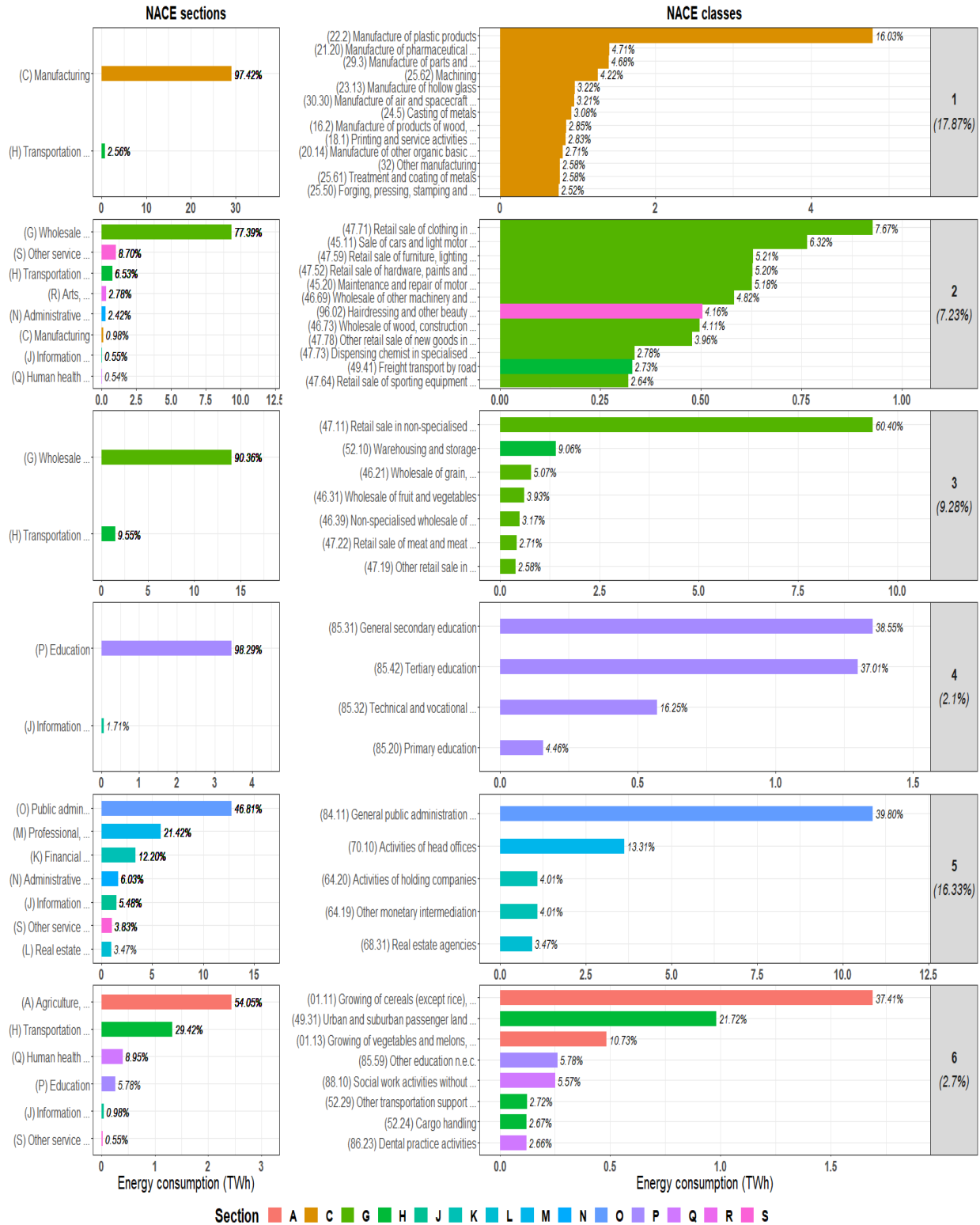


Figure 2: Distribution of the NACE sections and their corresponding classes for the first six clusters. The clusters' identification and their respective shares of annual consumed energy are shown in the grey boxes. The percentage associated with each bar represents the amount of consumed electricity w.r.t. the total energy consumed by the cluster. For the sake of visibility, only NACE sections / classes which account respectively for more than 0.5% / 1.5% of the total amount of consumed energy by clusters are displayed.



Figure 3: Distribution of the NACE sections and their corresponding classes for the 7th to the 12th clusters. The clusters' identification and their respective shares of annual consumed energy are shown in the grey boxes. The percentage associated with each bar represents the amount of consumed electricity w.r.t. the total energy consumed by the cluster. For the sake of visibility, only NACE sections / classes which account respectively for more than 0.5% / 1.5% of the total amount of consumed energy by clusters are displayed.



Figure 4: Distribution of the NACE sections and their corresponding classes for the last six clusters. The clusters' identification and their respective shares of annual consumed energy are shown in the grey boxes. The percentage associated with each bar represents the amount of consumed electricity w.r.t. the total energy consumed by the cluster. For the sake of visibility, only NACE sections / classes which account respectively for more than 0.5% / 1.5% of the total amount of consumed energy by clusters are displayed.

Load profiles description

The daily, weekly, and yearly load profiles of the considered clusters are displayed in Figures 5-6. On the whole, the different clusters exhibit lower electricity consumption during the night and at weekends, and a constant load during working days; however, variations are identified throughout the year, and especially during holidays. In addition, the spaghetti plots of the profiles of the associated NACE classes are presented to judge the sharpness of the clustering process. Such graphs make it possible to deal with the cluster analysis in depth:

- (1) **Manufacturing processes** The profiles of the NACE classes that belong to this cluster are rather homogeneous. The hourly load curve is characterised by fairly constant consumption during the day; however, lower consumption is noted during the early morning (we can presume that day shifts outnumber night shifts). A reduction of the load is observed during the weekend (a slightly higher load is associated with Saturday), and during Christmas and summer vacation periods. We may hypothesise that weekend shifts are not common for this field of activities.
- (2) **Trades (non food)** The profile distribution of the NACE classes is more scattered. We note a larger variation in energy consumption during the day in comparison with the previous cluster due to the absence of workers during the night. The lunchtime break can be observed as well as work activities on Saturdays. The load is higher during the winter period, which may be the result of an increase in the use of thermosensitive-devices in shops. Energy consumption drops associated with end-of-year holidays are not significant.
- (3) **Trades (food)** This cluster is characterised by retailing activities that take place during the day. A slight decrease in energy consumption is associated with the lunchtime break. Energy use is constant during working days, and decreases progressively during the weekend. These activities are rather constant throughout the year; nevertheless, a regime variation is observed around mid-year; the second semester is more energy-intensive.
- (4) **Education** The hourly profile is rather flat during the early morning, consumption increases sharply up to noon, and then decreases progressively. Moreover, the load is stable on weekdays and during the two weekend days due to the closing of schools and universities. The yearly load time series is characterised by consumption decreases that correspond to the school holiday calendar.
- (5) **Office** This cluster reveals day/night, and weekday/weekend patterns, which is in accordance with office-related activities. The yearly profile is characterised by higher consumption during wintertime, however, a slight increase is observed around mid-year, this phenomenon could be explained by the use of air conditioning to cool down both offices and server rooms.
- (6) **Crop farming and transportation** This cluster exhibits day and night consumption patterns. The weekly load profile is uniform, except for a slight decrease during the weekend. Consumption is higher during the winter season, but a pronounced peak is observed during summer, which is associated with a higher work load due to the harvest season.
- (7) **Livestock farming** Contrary to the previous cluster, this one is associated with consumption peaks at 7 a.m. and 6 p.m.. The weekly load profile is uniform. These observations can be explained by the milking process which occurs several times a day, and during the weekend. For this cluster, the distributions of the hourly and yearly NACE class profiles are more "chaotic", which highlights heterogeneous data. The yearly load profile is characterised by an increase in consumption during summertime followed by a progressive decrease. This is associated with camping ground activities (the peak season is during summer break), and the reduction of milking (dairy cattle produce milk for 10 months after calving, the calving season usually takes place during winter).
- (8) **Water supply and telecommunications** The patterns of the various NACE class profiles are quite heterogeneous. This cluster is characterised by a peak consumption during the day, and a low consumption around midnight. The main characteristic of this cluster is its rather steady energy consumption throughout the year, except a peak observed during summer. This increase may be correlated with a greater consumption of water.
- (9) **Restaurants** Restaurant-related activities are defined by a specific bi-modal hourly load profile, with peaks at noon and 7 p.m.. These periods correspond to restaurants' opening times. The weekly and yearly profiles are rather constant.

- (10) **Food industry** This cluster shows steady consumption during the year except for the Christmas break. The variation between the day and night consumption is rather low, which suggests the presence of night shift teams. We observe, at the weekly level, a high heterogeneity in consumption between the different NACE classes. Some of them are characterised by steady consumption which implies weekend shift teams. However, companies with weekday work schedules dominate in terms of energy consumption.
- (11) **Wine industry** This cluster possesses day/night, and weekday/weekend patterns. The distribution of the NACE class profiles are quite "chaotic" for the weekly time series. The weighted average profile is higher at the beginning of the year and during the period spread over August and September, which corresponds to the wine harvest period.
- (12) **Energy supply and rental activities** The main feature of this cluster is its remarkable drop in consumption during summertime. This may be accounted for by the reduction of energy production due to milder temperatures that do not require electric heating.
- (13) **Hotels** Hotel-related activities exhibit specific hourly patterns. A peak in consumption is noted around noon, which can be explained by the cleaning of the rooms, then a second peak appears around 8 p.m., which may be associated with restaurant activities or with customers returning to their rooms. The daily load profile is steady, while the weekly profile is more variable. Consumption peaks are observed at the beginning, middle, and end of the year.
- (14) **Bakery** Bakery activities are preponderant during the morning and then progressively decrease throughout the day. This morning peak is associated with the baking of the daily bread batch. These activities are reduced during the weekend. The yearly profile is steady.
- (15) **Property management companies** This field of activities consumes a relatively constant amount of energy throughout the day and the week. The lighting of advertising signs in real-estate agency windows could account for electricity use during the night and weekend. Consumption is higher during the winter due to thermosensitive uses, while the drop in consumption may result from summer vacations.
- (16) **Hospital activities** Hospital-related activities possess day/night, and weekday/weekend patterns. The yearly consumption is mainly influenced by peaks at the beginning, middle, and end of the year.
- (17) **Recreational and social activities** This group possesses day/night patterns. The decrease in consumption due to the weekend is somewhat low. The main load is located during wintertime, which could be the result of heating sport facilities.
- (18) **Construction** The construction sector is characterised by day/night (with a significant lunchtime break), and weekday/weekend patterns. The yearly load curve is lower during summertime, possibly due to worker vacations.

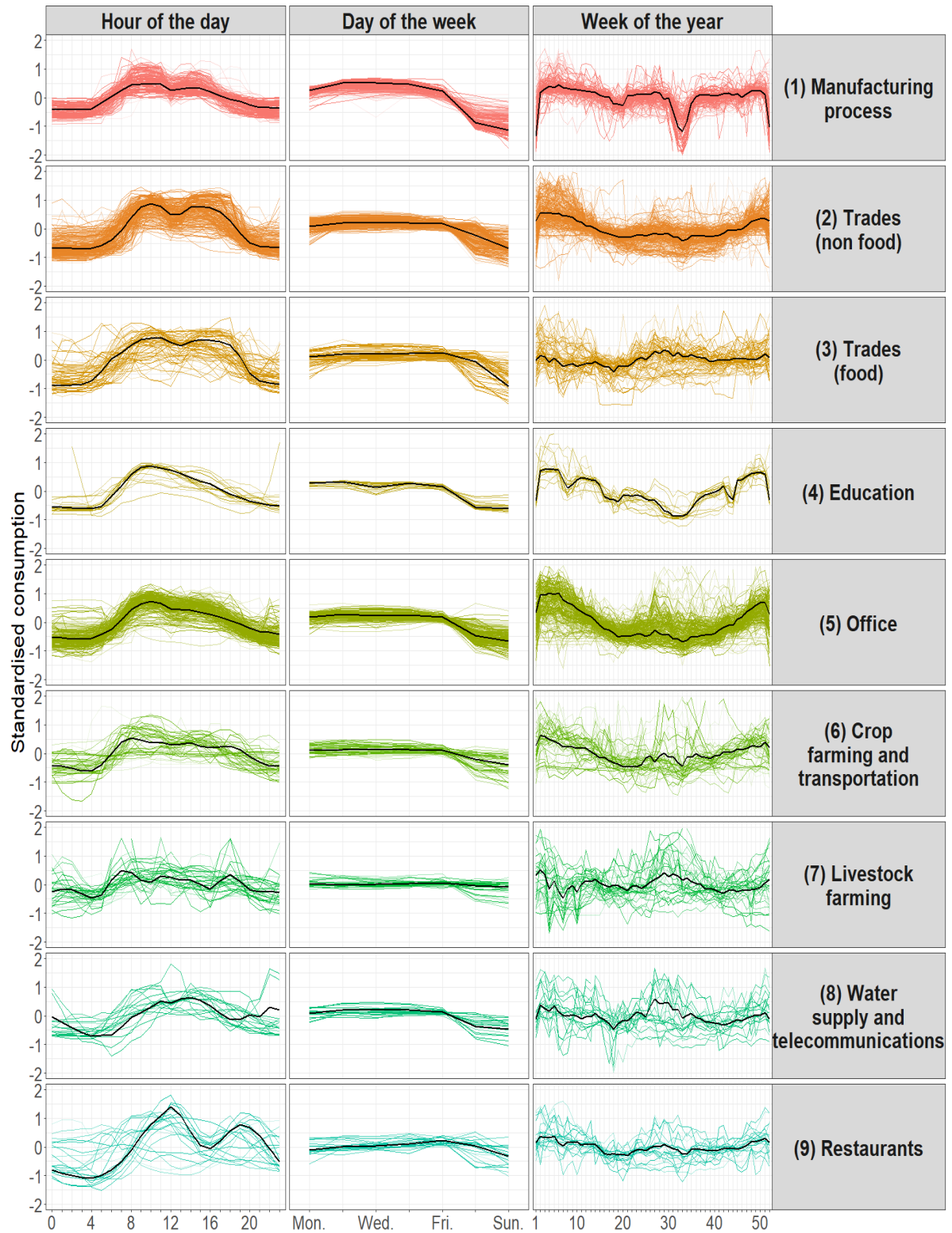


Figure 5: Hourly, daily, and weekly load profiles generated from the first nine clusters. The thick dark lines represent the weighted average of the consumption (weights are derived from the annual energy consumption of the NACE sections that belong to the cluster), while the light coloured lines are the load profiles of the NACE classes that belong to the cluster.

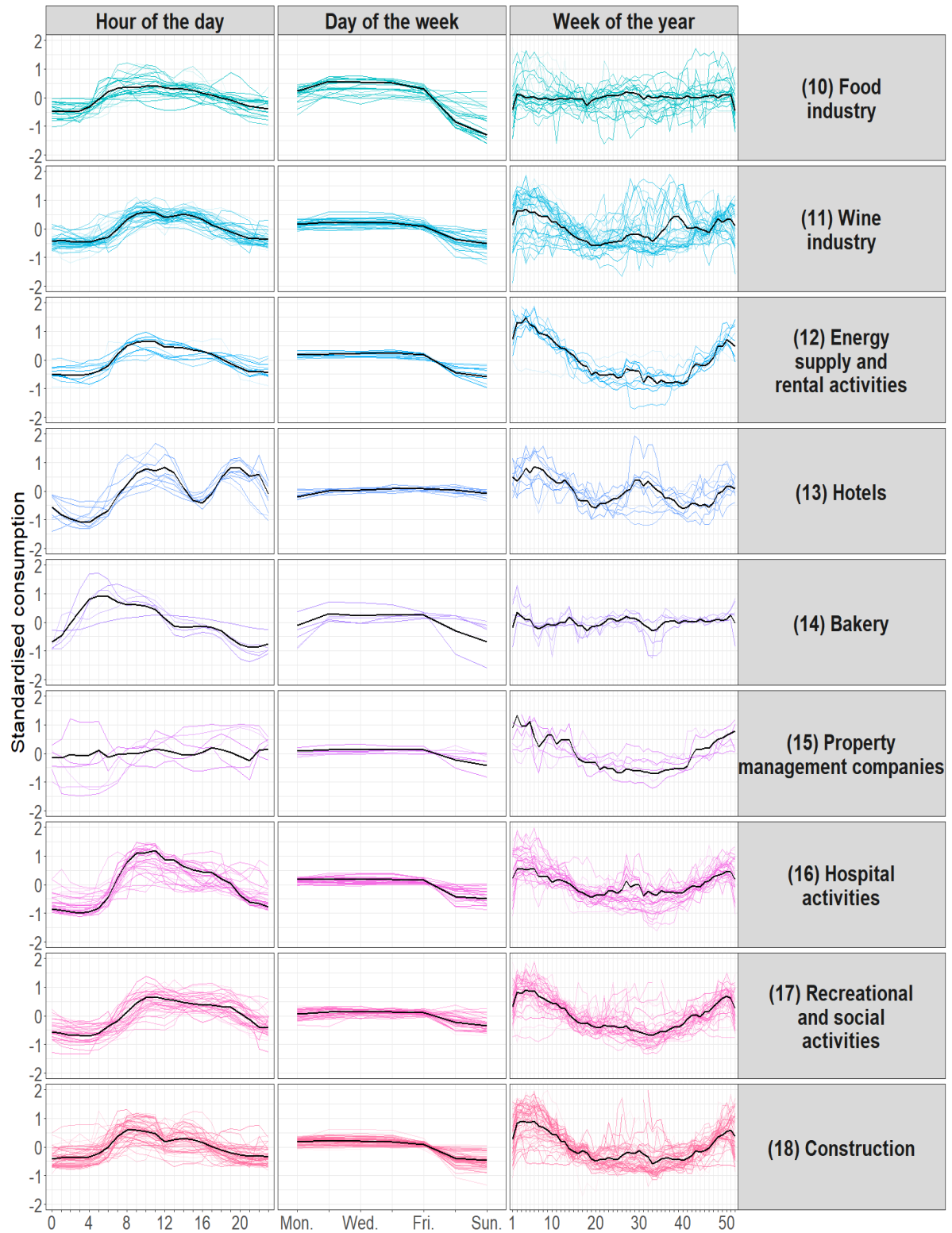


Figure 6: Hourly, daily, and weekly load profiles generated from the last nine clusters. The thick dark lines represent the weighted average of the consumption (weights are derived from the annual energy consumption of the NACE sections that belong to the cluster), while the light coloured lines are the load profiles of the NACE classes that belong to the cluster.

In a next step, it is insightful to consider indicators dedicated to specific electric consumption behaviours. Thus, we compute (1) the thermosensitivity, which measures the variation of the electric consumption w.r.t. the variation of outdoor temperature, (2) the seasonal consumption variation, (3) the variation between working and weekend days, and (4) the load variation between day and night (i.e. for hours comprised between 7 p.m. and 7 a.m.). The thermosensitivity is retrieved from the linear regression between the standardised daily averaged consumption and the daily average temperature, for daily outside temperatures lower than 16°C (it is assumed that below this temperature people resort to electric heating). The seasonal variation of the load is defined as the difference between the average summer and winter consumptions. All of the results are gathered in Figure 7. The clusters (5) *Office*, (11) *Wine industry*, (12) *Energy supply and rental activities*, (13) *Hotels*, (14) *Recreational and social activities*, (16) *Property management companies*, (17) *Hospital activities*, and (18) *Construction* exhibit a significant sensitivity to temperature. This phenomenon is easily explained by the need to heat facilities during winter for clusters 5, 12, 13, 14, 16, and 17. The thermosensitivity of cluster 11 is assumed to result from the different heat treatments that are applied to wine before, during and after the fermentation process (e.g. pasteurisation), while we may expect a reduction of the construction-related workload during summer. The following graphs enable us to judge the similarity of the temporal patterns between the different clusters. The cluster (7) *Livestock farming* is more energy-intensive during the summer period, while other clusters such as (4) *Education*, (12) *Energy supply and rental activities*, (14) *Recreational and social activities*, (17) *Hospital activities*, and (18) *Construction* consume more energy during winter. All of the considered clusters consume more energy during the weekdays, and during the daytime. However, for some clusters the distinction is not that significant, for instance activities gathered in cluster (7) *Livestock farming* are characterised by a significant use of energy during weekends and at night.

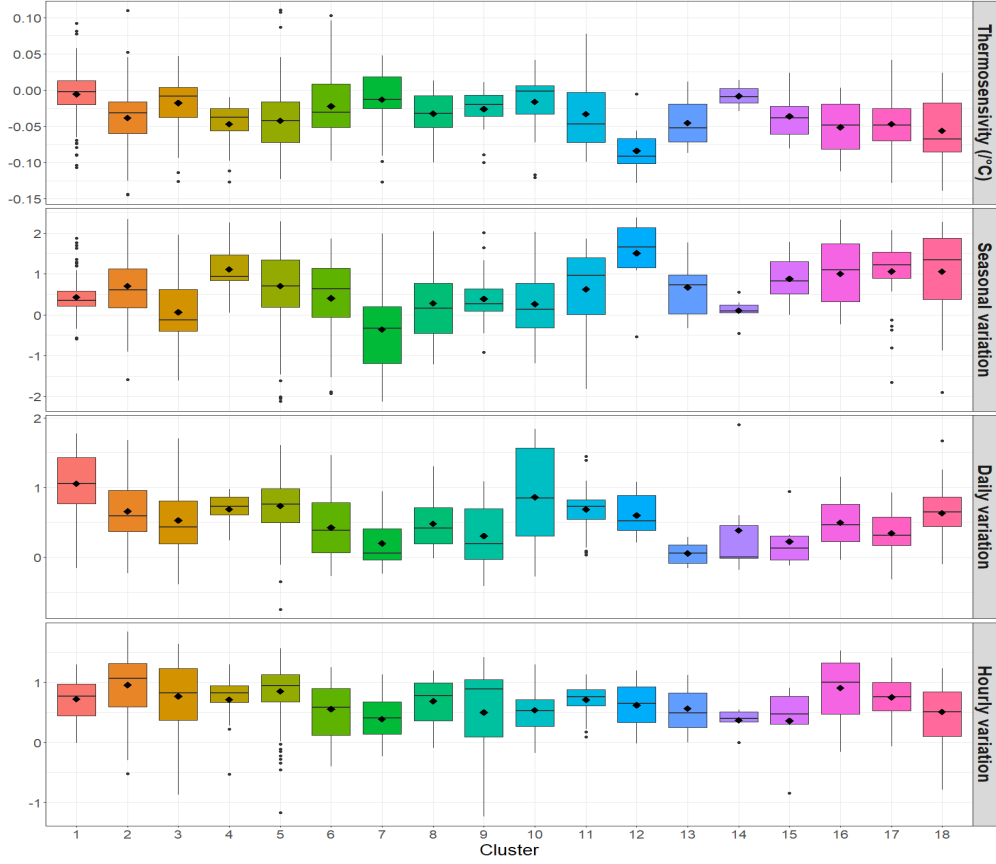


Figure 7: Distributions of indicators used to characterise the load of NACE classes for the 18 clusters considered. The diamond-shaped point represents the mean. These indicators are derived from standardised consumption time series. The seasonal, daily, and hourly variations are computed as the difference of the average consumption between winter and summer, week days and weekends, and day and night.