**Preprint-version** submitted to Energy Conversion and Management:X
Date: 19/03/2021
DOI: doi:10.17170/kobra-202103233576

# Annual industrial and commercial heat load profiles: modeling based on k-Means clustering and regression analysis

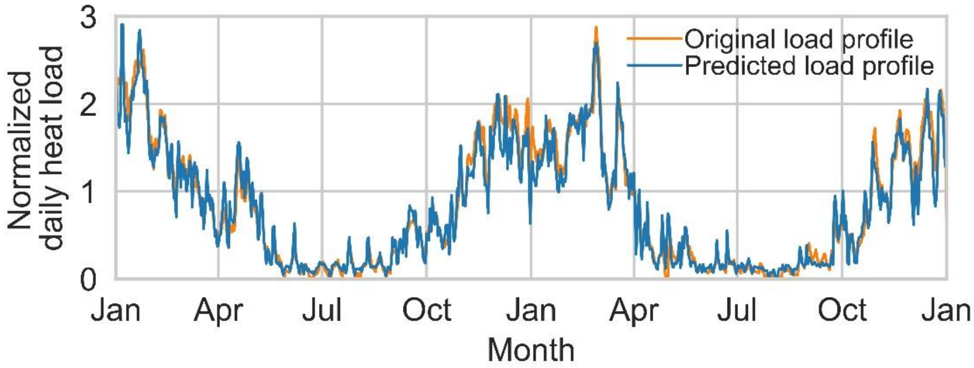
*Mateo Jesper[a], Felix Pag, Klaus Vajen, Ulrike Jordan*

*University of Kassel, Department of Solar and Systems Engineering, Kurt-Wolters-Str. 3, 34125 Kassel, Germany*

## Abstract

An accurate method to predict annual heat load profiles is fundamental to many studies, e.g., preliminary design or potential studies on renewable heating systems. This study presents a method to predict annual heat load profiles with a daily resolution for industry and commerce, based on an analysis of 800 natural gas load profiles (≥ 1.5 GWh/a). To derive heat load profiles, these natural gas load profiles are normalized and those with a potentially non-linear relationship between heat demand and natural gas consumption are excluded. The heat load profiles are clustered using the k-means algorithm according to their respective dependency on mean daily ambient temperature. The results reveal that the heat demand of most consumers is characterized by a clear dependency on mean daily ambient temperature, even in industry. The assignment of the load profiles to the clusters can be explained by the respective composition of each consumers' heat sinks. In a regression analysis, individual regressions for each load profile are only slightly more accurate than the regressions for all load profiles assigned to one of the respective clusters. In terms of accuracy and user-friendliness, the developed cluster regression-based correlations for load profile prediction offer a significant improvement on previous methods.

**Graphical abstract [color]:**



**Highlights:**

- Analyses 797 annual natural gas load profiles (> 1.5 GWh/a).
- K-means clustering according to dependency on mean daily ambient temperature.
- Heat demand of most consumers depends on ambient temperature, even in industry.
- Cluster correlations are almost as accurate as individual correlations.
- Accurate and user-friendly annual heat load profile correlations (resolution: one day).

**Keywords:**

annual heat load profiles, correlations, industry, commerce, k-means clustering, standard load profiles

[a] Corresponding author.
E-mail address: solar@uni-kassel.de

## Nomenclature

| | |
|---|---|
| a | mean distance between a sample and all other points in the same cluster [-] |
| A | fit parameter [-] |
| b | y-axis intercept [1/d], mean distance between a sample and all other points in the next nearest cluster [-] |
| B | fit parameter [-] |
| BDEW | German Federal Association of the Energy and Water Industry |
| c | cluster [-] |
| C | fit parameter [-] |
| CHP | combined heat and power |
| d | distortion (sum of the squared Euclidean distances from the cluster centroids) [-] |
| D | fit parameter [-] |
| DWD | German Meteorological Service |
| h | normalized daily natural gas consumption/ heat load [-] |
| HLNUG | Hessian State Agency for Nature Conservation, Environment, and Geology |
| i | count variable [-] |
| j | count variable [-] |
| k | number of clusters [-] |
| lin | linear |
| m | slope [-] |
| MaStR | Marktstammdatenregister (German register off all units producing electricity) |
| n | number [-] |
| Q | natural gas consumption/ heat demand [kWh] |
| $R^2$ | coefficient of determination [-] |
| s | silhouette coefficient [-] |
| SGB | standard natural gas boiler |
| sig | sigmoid |
| siglin | sigmoid linear |
| SLP | Standard Load Profile |
| T | temperature [°C] |
| wd | working day |
| wknd | weekends and holidays (idle days) |
| x | sample (load profile) |

Greek symbols

| | |
|---|---|
| $\mu$ | cluster centroid [-] |
| $\sigma$ | standard deviation [-] |

Subscripts

| | |
|---|---|
| 0 | present day |
| -1 | one day before |
| -2 | two days before |
| -3 | three days before |
| amb | ambient |
| d | day/ daily |
| gs | geometric series |
| h | space heating |
| hl | heating limit |
| i | cluster number |
| j | sample number |
| sig | sigmoid |
| siglin | sigmoid linear |
| w | domestic hot water |
| wd | working day |
| wknd | weekends and holidays (idle days) |

## 1. Introduction and objective

Renewable heat generators like solar thermal or heat pumps are key technologies to decarbonize the heating sector. Heat generation accounts for the largest share of global final energy consumption but is still dominated by fossil fuels [1]. Renewable heat generators besides biomass, such as solar thermal or heat pumps, accounted for only 3.5 % of total heat consumption in 2019 [1]. One reason for the low market penetration, especially in large-scale applications, is the relative high complexity of renewable heating systems in comparison to conventional heating systems. For example, Lauterbach et al. [2] highlight the variety of possible heat sinks in industry causing a broad range of possible hydraulic set ups and components to be used in solar thermal heating systems. Moreover, Schmitt [3] emphasizes that pre-dimensioning and yield assessment of a solar heating system can be very complex and time consuming. The same applies to large-scale heat pumps. The lack of knowledge of important installers and decision-makers about the requirements of a broad range of possible heat sinks is regarded as an important barrier to market acceptance [4].

The main requirements for heating systems, determined by the respective composition of heat sinks at a consumer's site, are the temperature level of heat supply and the temporal course of the heat load (load profile). While sufficient information on the temperature level of common large-scale heat sinks is available [5,6], little has been published on load profiles of common large-scale applications, especially in manufacturing industry. The literature on annual heat load profiles is limited to residential and small or non-manufacturing commercial consumers. For these types of consumers, the Standard Load Profile (SLP) methodology (section 2.1) allows utilities to predict gas consumption for the next few days based on a weather forecast. Another topic that has gained importance in studies on load profiles in recent years is load profile clustering (section 2.2). Many studies have been published on clustering based daily pattern recognition in residential load profiles. The aim of most of these studies is to support the development of advanced building control, fault detection, or demand side management by improving the understanding of energy demand characteristics of various consumer groups.

The objective of this study is to develop a methodology to predict heat load profiles of large-scale heat consumers in commercial, industrial, public or residential sectors with a focus on manufacturing industry. Since the heat load is usually not measured, acquiring a comprehensive database on heat load profiles is a major challenge. In contrast to heat load, natural gas consumption is often measured by German utilities with an hourly resolution. To reach the objective of this study, almost 800 natural gas load profiles are analyzed, most with a consumption of more than 1.5 GWh/a. Firstly, load profiles are normalized to enable comparability. In the next step, those load profiles from consumers with a potentially non-linear correlation of natural gas consumption and heat demand are excluded, e.g., consumers operating a natural gas-fired combined heat and power plant (CHP). For all other consumers, normalized heat load profiles and normalized natural gas load profiles are assumed to be equivalent. Normalized heat load profiles are clustered according to their specific correlation between daily mean ambient temperature and daily heat demand. The evaluation of the clustering results is based on an analysis of how the cluster assignment can be explained by the respective composition of the heat sinks at the consumer sites. For each of the found clusters, the correlation between daily mean temperature and natural gas consumption is mathematically captured in a regression analysis. Finally, the methodology to create heat load profiles with a daily resolution by applying the results of the previously described analysis is outlined.

## 2. Related work

Existing standards like VDI 4655 [7] or SLP methodology [8] focus on residential and commercial buildings but do not cover load profiles from manufacturing industrial or commercial companies. Therefore, existing studies on renewable heating systems in industry or commerce are based on simplified load profile approximations. For instance, the potential studies by Lauterbach [9] and Wolf [6] employ related heat load profile generators that create synthetic load profiles by manually selecting and combining typical daily, weekly and annual patterns. Since available data about industrial heat load profiles are rare, these load profile generators partly use electricity consumption patterns instead of heat consumption patterns. The German Association of Engineers (VDI) confirm the lack of a methodology to estimate accurate reference load profiles for consumers in industry and commerce and at the same time emphasize the importance of developing such a methodology to enable a transparent and reproducible comparison of heating systems in terms of efficiency-potential and cost-effectiveness [10].

This section reviews standards and studies, primarily on residential load profiles, to identify promising approaches that can be applied to large-scale industrial and commercial consumers. Firstly, the SLP methodology is summarized (section 2.1), which covers a comprehensive analysis of the correlation between mean daily ambient temperature and natural gas consumption. This correlation is a basis to this study. Secondly, Section 2.2 provides a general overview of other standards and studies on load profiles. This includes summaries of an earlier study by the authors and the VDI 4655 standard [7]. Additionally, recent scientific publications are reviewed, especially in the area of residential load profile clustering. Finally, section 2.3 summarizes implications from related work for this study.

## 2.1 *Standard load profiles*

To ensure the security of supply in natural gas networks, operating utilities need information about current and estimated natural gas consumption of all consumers connected. Therefore, consumers with a natural gas consumption of more than 1.5 GWh/a or 500 kWh/h are usually online metered with an hourly resolution [11]. Natural gas consumption of not online metered consumers is estimated based on SLPs which were developed by Hellwig in 2003 [8]. The SLP method is based on a correlation between ambient temperature and daily natural gas consumption. According to this methodology, a sigmoid (sig) function is best suited to mathematically model the correlation between mean daily ambient temperature and daily natural gas consumption (Eq. 2.1). Since the natural gas consumption differs by orders of magnitude, comparability is ensured by normalizing natural gas consumption to the mean natural gas consumption on days with an ambient temperature of 8 °C. The highest accuracy of natural gas demand prediction is achieved if a geometric series of the daily mean temperatures over the last four days is used instead of the simple daily mean temperature (Eq. 2.2) [8].

$$h_{sig}(T_{amb}) = \frac{A}{1 + \left(\frac{B}{T_{amb,gs} - 40}\right)^{C}} + D \qquad \text{Eq. 2.1}$$

$$T_{amb,gs} = 1 \cdot T_{amb,0} + 0.5 \cdot T_{amb,-1} + 0.25 \cdot T_{amb,-2} + 012.5 \cdot T_{amb,-3} \qquad \text{Eq. 2.2}$$

| | |
|---|---|
| $A, B, C, D$ | fit parameter of sigmoid function [-] |
| $h_{sig}(T_{amb})$ | normalized daily natural gas consumption as sigmoid function of $T_{amb}$ [-] |
| $T_{amb,gs}$ | geometric series of daily mean temperatures over last 4 days [°C] (insert unitless)) |
| $T_{amb,0}$ | mean ambient temperature at present day [°C] |
| $T_{amb,1}$ | mean ambient temperature one day ago [°C] |
| $T_{amb,2}$ | mean ambient temperature two days ago [°C] |
| $T_{amb,3}$ | mean ambient temperature three days ago [°C] |

Hellwig [8] defines SLPs for 14 different consumer groups. The focus is on non-manufacturing consumers like households, retail trade, banks, or accommodation businesses. Four groups also contain manufacturing companies (bakeries, laundries, metal & automotive, paper & print). For each of the consumer groups, up to five different shapes are given which represent low to high shares of process heat load on overall heat load. In contrast to space heating, process heat load is assumed to be independent from ambient temperature.

The SLP methodology developed by Hellwig was then adopted by German Federal Association of the Energy and Water Industry (BDEW) in a guideline [12]. A status report on the SLP methodology was compiled by the Forschungsgesellschaft für Energiewirtschaft (FFE), Germany, in 2014 [13]. This status report outlines two approaches to optimize the SLP methodology:

1. Hellwig explains the horizontal flattening of the sigmoid SLPs for temperatures below 0 °C by a changed user behavior (Figure 1). When it is freezing, users tend to reduce manual ventilation which leads to reduced heat losses [8]. In contrast to that, FFE states that sigmoid SLPs often lead to an underprediction of natural gas consumption for these days [13]. Linearized sigmoid (siglin) SLPs that reduce this error were published by FFE in 2015 [14] and also became part of the BDEW guideline [15]. The siglin SLP is the sum of a linear (lin) SLP and a sig SLP (Eq. 2.3 and Eq. 2.4). Lin SLPs consist of two lines. The right line is representing domestic hot water preparation or other ambient temperature independent heat loads which are almost constant throughout the year. The left line additionally includes space heating or other ambient temperature dependent heat loads which just occur when temperature falls below the heating limit temperature ($T_{hl}$). Figure 1 visualizes lin, sig and siglin SLPs using the example of German households.

$$h_{lin}(T_{amb}) = max \begin{Bmatrix} m_h \cdot T_{amb,gs} + b_h \\ m_w \cdot T_{amb,gs} + b_w \end{Bmatrix} \qquad \text{Eq. 2.3}$$

$$h_{siglin}(T_{amb}) = w_{lin} \cdot h_{lin}(T_{amb}) + (1 - w_{lin}) \cdot h_{sig}(T_{amb}) \qquad \text{Eq. 2.4}$$

| | |
|---|---|
| $b_h$ | y-axis intercept of space heating line [-] |
| $b_w$ | y-axis intercept of domestic hot water line [-] |
| $h_{lin}(T_{amb})$ | normalized daily natural gas consumption as linear function of $T_{amb}$ [-] |

| | |
|---|---|
| $h_{sig}(T_{amb})$ | normalized daily natural gas consumption as sigmoid function of $T_{amb}$ [-] |
| $m_h$ | slope of space heating line [-] |
| $m_w$ | slope of domestic hot water line [-] |
| $T_{amb,gs}$ | geometric series of daily mean ambient temperature [°C] (insert unitless) |
| $w_{lin}$ | weight of linear SLP [-] |

2. The consideration of additional weather parameters like humidity, irradiation or wind speed can lead to a higher accuracy of the SLP methodology. A method to consider additional weather parameters was introduced by the latest revision of the BDEW guideline [16] and is applied in collaboration with meteorological services. If historical natural gas consumption is available for a specific part of a natural gas network, meteorological services use this data to calculate the temperature which results in the lowest residuals between real natural gas consumption and predicted natural gas consumption. In the next step, model parameters using various input variables like humidity, irradiation or wind speed are fitted to predict this optimal "natural gas prediction temperature". Consequently, the "natural gas prediction temperature" is not a physical temperature, but a parameter combining various weather parameters. The weights of these parameters are individually fitted to a specific part of a natural gas network and should be checked at least once a year. The natural gas prediction temperature model has not been published yet.

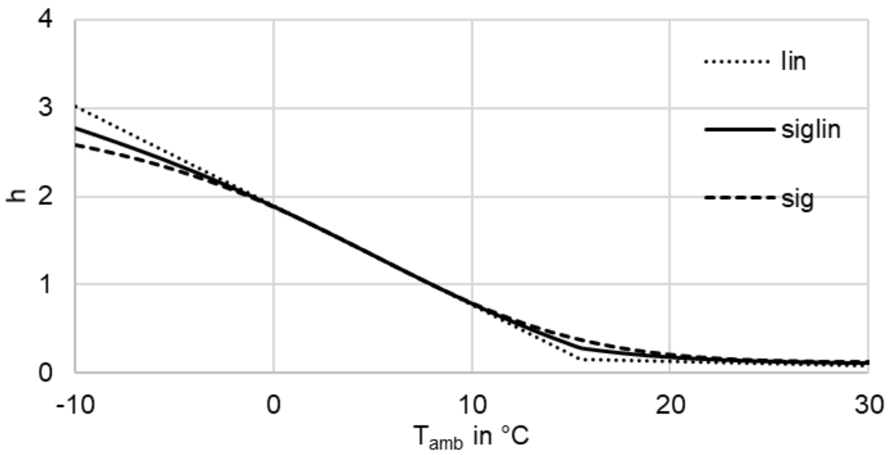


**Figure 1: Linear, sigmoid and linearized sigmoid SLPs for German households with a natural gas consumption of less than 50 MWh/a (siglin = 44.4 % · lin + 55.6 % · sig) [16]. [no color]**

To predict the absolute daily natural gas consumption ($Q_d$) for a specific consumer (Eq. 2.5), the respective sig (Eq. 2.1) or siglin (Eq. 2.4) function and either the geometric series of conventional temperature predictions (Eq. 2.2) or, if available, the natural gas prediction temperature are used [16]. The variations of heat load caused by type of day (working day or weekend and holiday) are considered by a consumer group specific weekday factor ($F_d$) [8]. The load profile is scaled to adapt to the absolute natural gas consumption of the respective consumer using the mean natural gas consumption on days with a mean ambient temperature of 8 °C ($Q_d$(8 °C)). The latter is calculated based on the historical natural gas consumption of a season of at least 300 days ($Q_s$) in relation to the summed normalized daily natural gas consumptions in this season (Eq. 2.6).

$$Q_d = h(T_{amb}) \cdot F_d \cdot Q_d(8°C) \quad \text{Eq. 2.5}$$

$$Q_d(8°C) = \frac{Q_s}{\sum_{i=1}^{j} h(T_{amb})_i} \quad \text{Eq. 2.6}$$

| | |
|---|---|
| $F_d$ | weekday factor [-] |
| $h(T_{amb})$ | normalized daily natural gas consumption as sig or siglin function of $T_{amb}$ [-] |
| $j$ | number of days in examined season [-] |
| $Q_d$ | daily natural gas consumption [kWh] |
| $Q_d(8°C)$ | natural gas consumption on days with 8 °C mean ambient temperature [kWh] |
| $Q_s$ | seasonal (usually annual) natural gas consumption (minimum 300 days) [kWh] |

## 2.2 *Other studies on load profile clustering and prediction*

In a previous study, the authors analyze 77 industrial natural gas load profiles from nine industrial economy divisions regarding their correlations between mean daily ambient temperature and natural gas consumption [17]. Within the economy divisions, most load profiles show similar patterns. At the same time, differences between the divisions are obvious. While no correlation is observed for four of the economy divisions, five of the divisions show a significant correlation between ambient temperature and natural gas consumption. As far as the authors know, this is the only systematic analysis of a large industrial heat load profiles available.

The VDI 4655 standard [6] includes a methodology to create reference load profiles for space heating and domestic hot water preparation in single family houses with up to six inhabitants and multifamily houses with up to 25 dwelling units. Next to heat load profiles, the VDI 4655 also covers profiles for electricity consumption and photovoltaic generation. The presented methodology divides Germany into 15 climate zones [18,19]. For each of these climate zones, the respective number of days from a specific type are given, e.g., summer-Sunday or winter-working day-cloudy. Reference load profiles are created based on the respective number of day types and the respective building type.

Two recent studies examine daily heat load profiles from Scandinavian district heating consumers. Calikus et al. [20] develop a methodology to automatically identify normal and abnormal patterns. The methodology is intended to be used for the optimization of district heating operation and management. In total, 1,222 consumers are assigned to 15 different normal patterns (clusters). An abnormal pattern is identified for 163 consumers. The clustering analysis is based on k-shape algorithm, which is similar to k-means algorithm. The optimal number of clusters is determined by a silhouette coefficient analysis. A similar clustering-based knowledge discovery to optimize design, operation, and demand-side management of district heating networks is conducted by Gianniou et al. [21]. Based on a comprehensive database that includes 8,293 single-family households, they find five clusters in the absolute load profile dataset and nine clusters in the normalized load profile dataset. The optimal cluster number of k-means clustering is determined based on Bayesian Information Criterion. Clustering results are evaluated using silhouette coefficients. Because most consumers have a regular and predictable consumption, Gianniou et al. conclude that a clustering-based short-term load forecasting is feasible and will be investigated in future work.

Do Carmo and Christensen [22] present a cluster and regression analysis of 139 load profiles from Danish dwellings supplied by decentral heat pumps. The objective is to optimize demand-side management based on a better understanding of the temporality of energy demand. The k-means clustering results show two main clusters of daily load profiles for both weekdays and weekends and across load segments. Differences between those two clusters are mainly correlated to characteristics of the respective dwelling like floor area, building year and type of space heating distribution system.

Load profile clustering based on k-means algorithm is dominating literature. Nevertheless, some studies also incorporate other algorithms. For example, Lu et al. [23] cluster hourly heat loads of six offices in China based on Gaussian-Mixture-Model to optimize management and operation of district heating systems. They find four typical patterns which are used in combination with other inputs like time of day, type of day and ambient temperature to model the hourly heat load. In the next step, different algorithms for load prediction are compared. The models based on multiple linear regression or artificial neural network yield the highest Pearson coefficients of 0.93 and 0.92, but simple linear regression based only on ambient temperature still results in a Pearson coefficient of 0.88. Ma et al. [24] compare clustering results of daily load profiles of 19 higher education buildings in Sweden based on k-means algorithm and Partitioning Around Medoids (PAM) algorithm. They conclude that both algorithms lead to similar results that can be used to assist in the development of advanced building control, fault detection, or demand side management. Nevertheless, a key advantage of the PAM algorithm is stated to be that regularly repeating load peaks are better reflected by the PAM-clusters compared to the k-means clusters.

## 2.3 *Implications from related work*

Based on correlations between daily heat load and weather, the SLP-methodology predicts the daily natural gas consumption of a group of residential or small commercial consumers. Daily ambient temperature is the weather parameter with the highest influence on daily heat load but other parameters like irradiation, humidity or wind speed can be used to increase the accuracy of prediction. Pag et al. [17] prove in their analysis that the correlation between daily mean temperature and daily natural gas consumption also exists for many consumers in industry. At the same time, the strength of this correlation seems to be specific to the investigated economy divisions. Therefore, this study examines whether the strength of this correlation can be used to cluster load profiles reasonably.

The methodology of VDI 4655 [7] is not considered in this study because it cannot be transferred from the relatively homogeneous group of residential consumers to the much more diverse group of industrial and commercial consumers.

Several studies use machine learning algorithms to cluster heat load profiles. These studies focus on identifying daily patterns discernible in residential load profiles with an hourly resolution. In contrast, this study aims to identify annual patterns in industrial heat load profiles with a resolution of one day. Nevertheless, it is evaluated whether the k-means algorithm, which has been successfully used several times for clustering load profiles, can be adopted for the objective of this study (section 3.3).

SLP correlations are based on sig or siglin regressions [25]. Pag et al. [17] apply simple linear regression to model industrial load profiles. Compared to a simple linear regression, recent research shows that the accuracy of load profile models can be increased by supervised machine learning algorithms like multiple linear regression or artificial neural networks [23]. A major drawback of these supervised machine learning models is the limited and more complex applicability for third parties. To train these models, a comprehensive load profile database as used in this study is fundamental but cannot be published due to data protection reasons. At the same time, Lu et al. [23] find that the accuracy of supervised machine learning models is just slightly increased compared to a simple linear regression. For these reasons, machine learning is used only for clustering in this study, but the final heat load profile model is intended to be based on simple regressions, which ensures transferability, applicability, and user-friendliness.

## 3. Database and methods

This section summarizes the analysis of a comprehensive natural gas load profile database to derive load profile correlations of commercial and industrial heat consumers. Section 3.1 gives an overview on the load profile database and annual natural gas consumption of the examined consumers. Section 3.2 outlines the pre-processing for load profile clustering. The methodology used for k-means clustering is summarized in section 3.3. In section 3.4, the methodology of the regression analysis carried out for each of the previously identified clusters is delineated.

The analysis presented in this study is based on the programming language Python. Appendix A gives an overview on the used software libraries.

### 3.1 *Overview of the database and the annual natural gas consumption*

This study is based on the metered natural gas consumption of 797 large-scale consumers provided by German utilities. The resolution of the data covers hourly averages of natural gas consumption. Although consumer names and addresses are available and used within this study, company names or other information that could be used to identify the consumers are not published for data protection reasons. For most of the consumers (90 %), natural gas load profiles are available for the years 2017 and 2018. Another 10 % of the load profiles are from the year 2016. The natural gas consumers are located in Germany, in the metropolitan area of Stuttgart as well as in northern Hesse and neighboring regions.

Figure 2 shows a boxplot of the annual natural gas consumptions for the most common economy divisions (section 3.2.1.4) within the database. Only economy divisions that are represented by ten consumers minimum are visualized separately. All other economy divisions are grouped in “others”. A vertical line with an annual natural gas consumption of 1.5 GWh is displayed, as this is the generally used minimum value for online metering (section 2.1). Since this threshold can be adjusted to ensure the security of supply, some smaller annual natural gas consumptions can also be observed. The plot shows that the range of natural gas consumptions is rather wide, both across all economy divisions and in each economy division individually. The first eight divisions at the top, which are all consumers from manufacturing industry, tend to show a higher natural gas consumption compared to the other divisions. Additionally, most of the interquartile distances, represented by the boxes, are also larger in these divisions. When the overall high variance of natural gas consumption is considered, it can be concluded that no annual natural gas consumption benchmark can be derived from the analyzed dataset. Further investigations to define specific benchmarks, e.g., natural gas consumption per area production hall, per employee or per turnover, are necessary, but would go beyond the scope of this work.

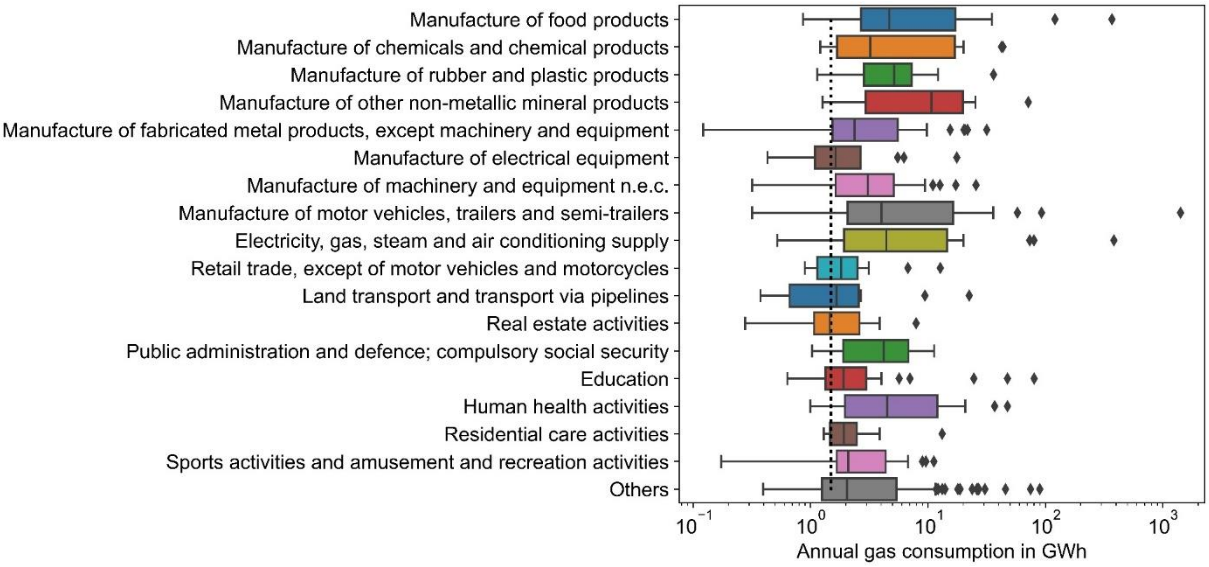


**Figure 2: Boxplots showing the annual natural gas consumption in GWh in logarithmic scaling. Each boxplot represents one economy division according to NACE Rev. 2 [26]. Only divisions with at least ten consumers are shown, all others are grouped under "Others". The vertical line at 1.5 GWh is representing the usual threshold for online metering (section 2.1). [no color]**

## 3.2 *Pre-processing*

In this section, the pre-processing done for clustering is described. Firstly, parameters influencing the heat load and availability of information about these parameters are summarized (section 3.2.1). In the next steps, the methodologies used for normalization (section 3.2.2) and plausibility check (section 3.2.3) are exemplified. Finally, the way the load profiles are transformed to uniform data vectors suitable as input for the clustering algorithm is outlined (section 3.2.4). These uniform data vectors contain all relevant information to cluster the consumers according to their specific relationship between ambient temperature, type of day and heat load.

### *3.2.1 Influences on heat load*

This section summarizes parameters that influence the examined heat load profiles. Table 1 gives an overview on availability of information about these parameters. In the following, each of the influencing parameters is outlined. For those parameters for which no information is available, a description is provided of how the data is processed to reduce the impact of these parameters on the load profile and to reduce potential misleading effects on clustering. For the parameters for which information is available, the source and quality of the information is described.

**Table 1: Summary of available and unavailable information that are linked to the heat load.**

| **Unavailable information** | **Available information** |
|---|---|
| Heat sinks (section 3.2.1.1) | Economy division (section 3.2.1.4) |
| Heating system (section 3.2.1.2) | Type of day (section 3.2.1.5) |
| User behavior (section 3.2.1.3) | Weather (section 3.2.1.6) |

#### *3.2.1.1 Heat sinks*

This study is based on the hypothesis that the respective composition of heat sinks is specific to the consumers and a main reason for differences of the normalized heat load profiles. Space heating and domestic hot water heating are the most common heat sinks in households [8]. In contrast, heat sinks are much more diverse for industrial or commercial consumers. The range of heat consuming processes in industry and commerce is broad and even characteristics of space or water heating can vary strongly from consumer to consumer. Consequently, industrial and commercial heat load profiles are much more diverse, ranging from nearly constant throughout the year to load profiles with significant seasonality. To support this, all load profiles from one example economy division (manufacture of motor vehicles, trailers and semi-trailers) are illustrated in Appendix F. Additionally, plots of all 797 natural gas load profiles are available in a data repository [27]. However, no detailed information about heat sinks is available but some information can be found on consumer webpages or derived from the respective economy divisions (section 3.2.1.4). This available information on heat sinks is not consistent and therefore cannot be used for clustering but is used to validate the hypothesis formulated at the beginning of this section and to evaluate the clustering results (section 5.1).

#### *3.2.1.2 Heating system*

To ensure the transferability of the results of this study, it is necessary to correlate natural gas consumption profiles to heat load profiles. However, the heat load is not affected by the heating system design. Instead, the heating system determines the correlation between heat load and natural gas consumption. For most natural gas-fired heat generators like standard natural gas boiler (SGB), steam boilers, direct burners or ovens, the correlation between heat load and natural gas consumption is almost linear. For instance, SGB are just slightly less efficient in summer compared to winter [8]. Therefore, the normalized heat load profiles and normalized natural gas load profiles are assumed to be equal. Nevertheless, this assumption does not apply for all of the consumers or heating systems and the following restrictions must be considered:

- Combined heat and power (CHP):
  Different operating strategies are common, e.g., to primarily meet heating demand or electricity demand [28]. Additionally, CHP are often designed to supply just the base heat load. Therefore, the relationship between heat load and natural gas consumption is not predictable and consumers operating CHP plants are excluded using the German “Marktstammdatenregister” (MaStR) [29] (section 3.2.3).
- Other natural gas-fired heat generators:
  In the case of some other heat generators (e.g., natural gas absorption heat pump), additional influences on natural gas consumption (e.g., heat source temperature) cause a relationship between heat load and natural gas consumption which is unpredictable without additional information. Consequently, the usage of other heat generators is a potential source of error. Nevertheless, according to the author’s experience, market penetration of other natural gas-fired large-scale heat generators like natural gas absorption heat pumps is low and consequently potential errors caused by other heat generators are negligible. This is supported by a low market availability of, for example, large-scale natural gas absorption heat pumps [30].
- Other natural gas uses:
  Natural gas can also be used to generate mechanical energy or as a material, e.g., in the chemical industry. However, the use of natural gas as an energy carrier accounts for 95 % of total natural gas consumption in manufacturing industry [31]. The use of natural gas as an energy carrier in the German industry is almost completely (97 %) is for heating purposes [32]. Other uses are therefore almost negligible but can still be a minor source of error to the results of this study.
- Other heat sources:
  Other heat sources such as excess heat recovery or heat generators powered by other final energies could be operated in parallel with natural gas-fired heating systems. If the share of natural gas-fired generators and other heat sources is nearly constant over a year, this has no negative impact on the results of this study due to normalization. In all other cases, the relationship between heat demand and natural gas consumption is not linear, which leads to errors that cannot be eliminated or evaluated due to the lack of detailed information about the respective heating systems. Nevertheless, other parallel heat sources are assumed to be rare. This assumption is supported by several observations:
  - Natural gas dominates industrial heat generation (Figure 3) but it is not available at all industrial sites. If natural gas (26.3 €/MWh) is available, it is significantly cheaper than electricity (107.7 €/MWh) or other fossil fuels like oil (56.4 €/MWh) [33]. Therefore, the assumption is viable that parallel heating systems operated by other fossil fuels or electricity are avoided if natural gas is available.
  - Coal accounts for 22.3 % of total heat generation in German industry but it is almost exclusively (92 %) used in manufacture of basic metals, manufacture of other non-metallic mineral products, manufacture of paper and paper products, or manufacture

of chemicals and chemical products [32]. In all other economy divisions, parallel coal and natural gas-fired heating systems are therefore very rare.

- Some processes cannot be supplied by district heating or some renewable heat generators like heat pumps or solar thermal due to limitation of supply temperature. Therefore, if an industrial consumer is supplied by these, it is likely that a parallel natural gas-fired heating system is used. Nevertheless, renewable heating systems and district heating only account for 16 % of total industrial heat generation (Figure 3). Parallel natural gas-fired and renewable or district heating systems are therefore assumed to be rare.
- Reckzügel et al. [34] provide a comprehensive survey to estimate the potential of industrial excess heat utilization in North Rhine Westphalia, a German state. They investigated the availability of excess heat streams of 528 companies. These companies provided information on 588 processes. The survey yields that measured data on excess heat is available for only 10% of these processes. Almost half of the surveyed companies (45 %) could not even estimate their waste heat potentials from energy or process plants. It can be assumed that if there is no data on excess heat, excess heat is not recovered. Consequently, excess heat utilization is still an exception.

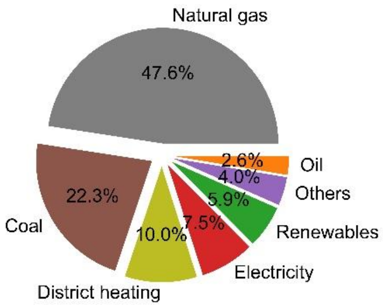


**Figure 3: Final energy usage for heating purposes in German industry in 2018 [32]. [no color]**

### *3.2.1.3 User behavior*

The user behavior covers a wide range of possible influences on natural gas consumption including production times or the degree of production capacity utilization. By reducing the resolution to one day (section 3.2.2), variances of user behavior appearing for periods smaller than one day (e.g., working shifts) can partly be eliminated. Individual differences of user behavior appearing for longer periods (e.g., plant holidays, collapse of the order situation, etc.) cannot be eliminated and are a source of error to the results of this study.

### *3.2.1.4 Classification of economic activities*

NACE Rev. 2 is a systematic classification developed by the Statistical Office of the European Communities (Eurostat) to classify economic activities [26]. The hierarchical structure of this systematic classification covers four levels:

1. Section: alphabetical code (e.g., C - Manufacturing).
2. Division: two-digit numerical code (e.g., 10 - Manufacture of food products).
3. Group: three-digit numerical code (e.g., 10.7 - Manufacture of bakery and farinaceous products).
4. Class: four-digit numerical code (e.g., 10.71 - Manufacture of bread; manufacture of fresh pastry goods and cakes).

To identify the section, division, group, and class of a specific company, a top-down method depending on the share of value added is applied. The section with the highest share of value added is selected first. In the next steps, the divisions, groups, and classes within the respective superordinate levels are selected [26].

All natural gas consumers from the examined database were classified based on their names and addresses in an online research. The available information from names and homepages is sufficient to assign sections and divisions, as opposed to groups and classes, which often can only be estimated. In total, 58 divisions were identified. Since some consumers are large residential buildings or office buildings that house multiple businesses, these consumers cannot be classified according to NACE Rev. 2. For these consumers, the division "0 – Residential building or office" is added in this study. Figure 4 visualizes all divisions containing at least ten consumers. Additionally, the groups within the depicted divisions are visualized by different colors. As detailed information on the value added by specific products is missing, the definition of groups and classes is inaccurate. In addition, many groups contain only a few consumers. Therefore, no reliable conclusions can be drawn about the groups and classes and only classification on division level will be used in the following.

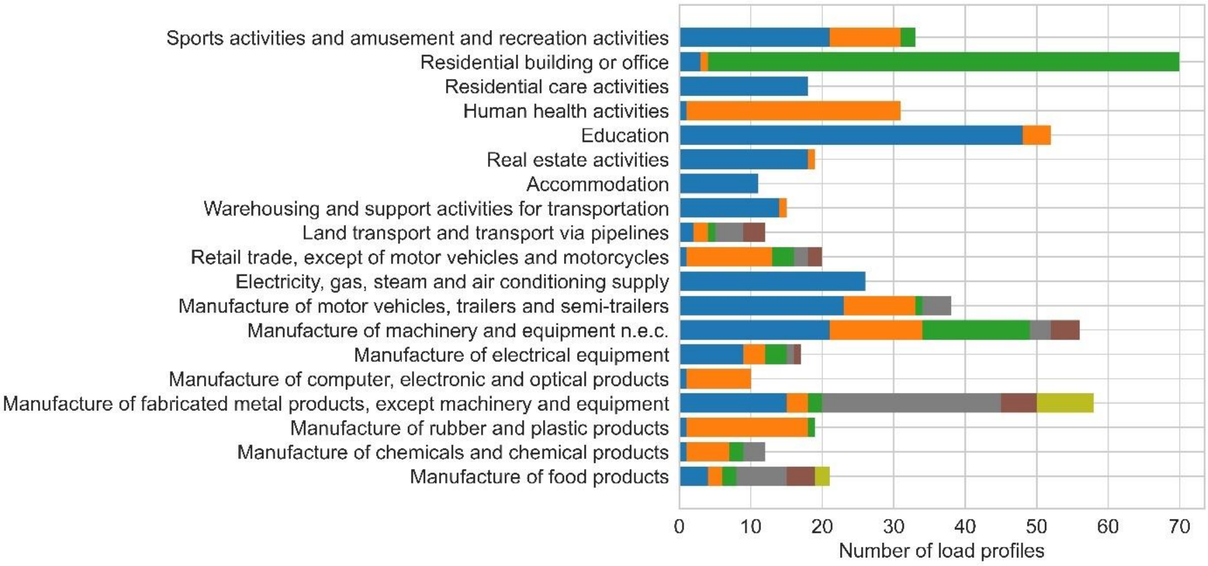


**Figure 4: Classification of load profiles according to Eurostat [26] (only the divisions with at least 10 load profiles are illustrated, colors of bars represent different economy groups within the specific division). "Residential building or office" is not part of NACE Rev. 2 and was added by the authors. [no color]**

The annual course of load profiles within a specific economy division shows major variances, and it becomes evident that the economy division is not an adequate cluster criterion. For example, Appendix F visualizes load profiles of the division "manufacture of motor-vehicles, trailers and semi-trailers". These load profiles are ranging from almost constant over the whole year to load profiles with no load in summer and a clear winter peak. This broad range is also noticeable for most of the other economy divisions (see data repository [27]).

### *3.2.1.5 Type of day*

Since heat load of many heat sinks (processes, staff showers, etc.) differs depending on the type of day, load profiles are separately clustered for working days (wd) and weekends and holidays (wknd). The regression analysis is also carried out separately for wd and wknd. In this analysis, Mondays to Fridays are generally defined as wd. Saturdays, Sundays, and regional holidays are defined as wknd.

### *3.2.1.6 Weather data*

Natural gas consumers examined in this study are located in Germany, in the metropolitan region of Stuttgart as well as in northern Hesse and neighboring regions. Just one weather dataset with an hourly resolution is used for each of these regions. In the case of northern Hesse, data measured at the station "Kassel-center" operated by Hessian State Agency for Nature Conservation, Environment, and Geology (HLNUG) is used [35]. For those consumers located in the metropolitan region of Stuttgart, data measured at the station Stuttgart/Echterdingen operated by the German Meteorological Service (DWD) is used [36].

### *3.2.2 Resolution and normalization*

Hellwig [8] and Pag et al. [17] detected a correlation between daily natural gas consumption and daily mean temperatures. At a higher resolution (e.g., hourly), this correlation is overlaid by other parameters, especially user behavior. Therefore, the resolution of all data used in this study is reduced to one day. In the case of natural gas consumption, daily sums of natural gas consumption ($Q_d$) are used. In the case of the ambient temperature, daily arithmetic mean values ($T_{amb}$) are applied.

As shown in section 3.1, no overall benchmark of absolute natural gas consumption or benchmark within the various economy divisions can be derived. To eliminate absolute natural gas consumption and to be able to identify relative similarities, the load profiles are normalized to the mean natural gas consumption on working days with a mean ambient temperature of 8 °C. Compared to a normalization on maximum or minimum values, this is less vulnerable to outliers. In contrast to the SLP methodology (section 2.1), which normalizes to the mean natural gas consumption on all days (wd and wknd) with a mean ambient temperature of 8 °C, only natural gas consumption on weekdays with a mean ambient temperature of 8 °C is considered for normalization in this study due to the large differences between wd and wknd natural gas consumption of some consumers. Figure 5 exemplifies normalization for two example companies. Similarities of the absolute load profiles (a) are difficult to detect but similarities of the normalized load profiles (b) are obvious, as the general trend and scale are comparable.

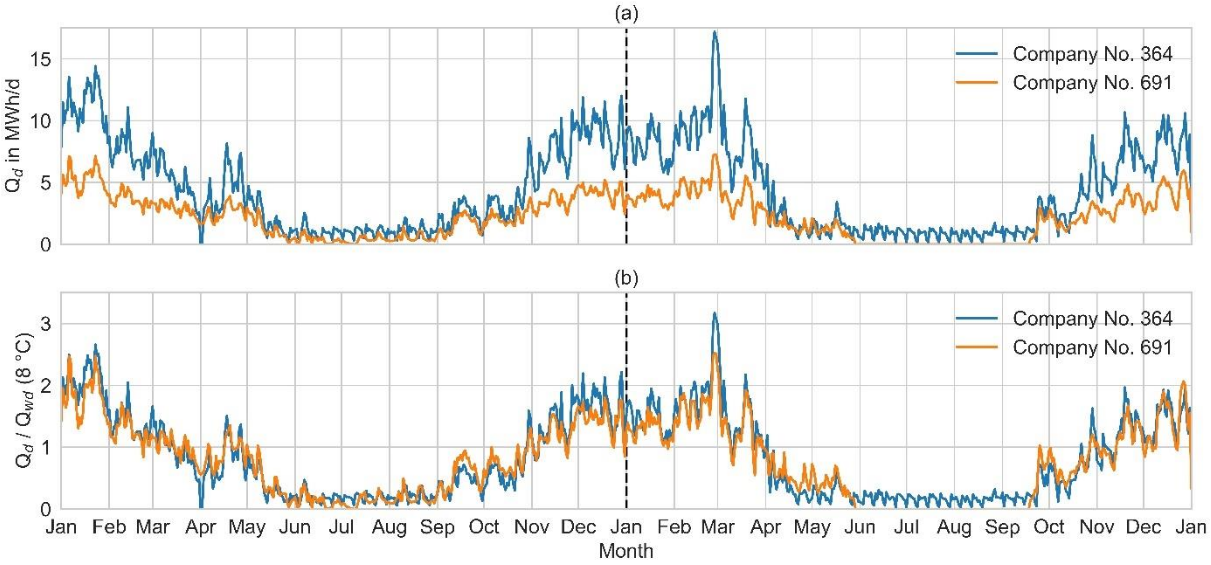


**Figure 5: Absolute (a) and normalized (b) load profiles for the years 2017 and 2018 of two example companies. [color]**

### *3.2.3 Data-filtering and plausibility check*

Some consumers have two separated natural gas consumption profiles and respective measurements, e.g., due to subsidiaries or a heat supply operated by a third party. If two or more load profiles are identified at the same address, load profiles are aggregated. This reduces the number of load profiles by 48 (6.0 % of total dataset).

Only complete and plausible load profiles are supposed to be used in the following analysis. Some load profiles seem to be incomplete, for example because supply just starts in the middle of a year. To sort these out, all load profiles are excluded which contain zero load in more than 83 % of the working day hours. This equals zero load for more than 20 hours on working days or 7,720 hours a year. As a result, 53 load profiles (6,6 % of total dataset) are excluded.

If consumers use a CHP plant for heat generation, the correlation between heat load and natural gas consumption is unpredictable without additional but unavailable information such as the operation mode (section 3.2.1.2). Therefore, consumers operating a CHP plant are excluded using MaStR [29]. MaStR is a register operated by German Federal Network Agency which includes natural gas and power production, storage, and consumption plants. It is possible that some consumers operating CHP plants had not registered at the time this study was prepared and remain undetected. In total, 77 consumers (9.7 % of total dataset) are operating a CHP plant and are excluded from the following cluster analysis.

Some of the remaining load profiles still show sections which seem to be caused by unusual user behavior, e.g., a collapse of the order situation and capacity utilization. To identify these consumers, two regression lines are fitted to the daily natural gas consumption using Eq. 3.1 which is a modified version of the SLP lin function (Eq. 2.3). The SLP lin functions are not to be used on their own but they are only used to linearize the sig functions to derive siglin functions. The parameters of the SLP lin functions are determined based on the respective associated sig functions. In contrast, the lin function (Eq. 3.1) developed in this study is supposed to be fitted and used independently without an associated sig function. The least squares approximation iteratively determines those function parameters including slopes (m), y-axis intercepts (b) and the heating limit temperature ($T_{hl}$, see section 2.1) that result in the minimum overall sum of squared residuals. In total, 53 consumers (6,6 % of total dataset) having a standard deviation (σ) that is higher than 0.75 are excluded.

In this study, the fit of all functions is done individually for wd and wknd (section 3.2.1.5). Therefore, differences of heat load caused by the type of day are already considered. Consequently, weekday factors ($F_d$) as used by the SLP methodology (section 2.1) are obsolete and Eq. 2.5 is simplified to Eq. 3.2.

$$h(T_{amb}) = \begin{cases} m_h \cdot T_{amb} + b_h & if \quad T_{amb} < T_{hl} \\ m_w \cdot T_{amb} + b_w & if \quad T_{amb} \geq T_{hl} \end{cases} \qquad \text{Eq. 3.1}$$

$$Q_d/Q_d(8°C) = h(T_{amb}) \qquad \text{Eq. 3.2}$$

| | |
|---|---|
| $b_h$ | y-axis intercept of heating line [-] |
| $b_w$ | y-axis intercept of domestic hot water line [-] |
| $h(T_{amb})$ | normalized daily natural gas consumption/ heat load as function of $T_{amb}$ [-] |
| $m_h$ | slope of heating line [-] |
| $m_w$ | slope of water line [-] |
| $w_{lin}$ | weight of linear SLP [-] |
| $Q_d/Q_d(8°C)$ | normalized daily natural gas consumption/ heat load [-] |
| $T_{amb}$ | daily mean ambient temperature [°C] (insert unitless) |
| $T_{hl}$ | heating limit temperature [°C] (insert unitless) |

The plausibility check based on the regression lines is exemplified by Figure 6 showing the load profiles of two consumers as time series [(a) and (b)] and depending on ambient temperature [(c) and (d)]. On the left side (company no. 143), an inconspicuous load profile as seen many times in the database is visualized. The correlation between natural gas consumption and ambient temperature is clearly visible in this case (c). This is underscored by the low σ of the two regression lines in the lower left diagram. On the right side (company no. 679), the load from January to March 2017 is conspicuously low. In the following year 2018, the load in January to March is as high as would be expected based on the load in the remaining months. The trend deviating from the rest of the profile in the first quarter of 2017 leads to an increased standard deviation of σ > 0.75. Consequently, company no. 679 is excluded from the analysis presented in the following.

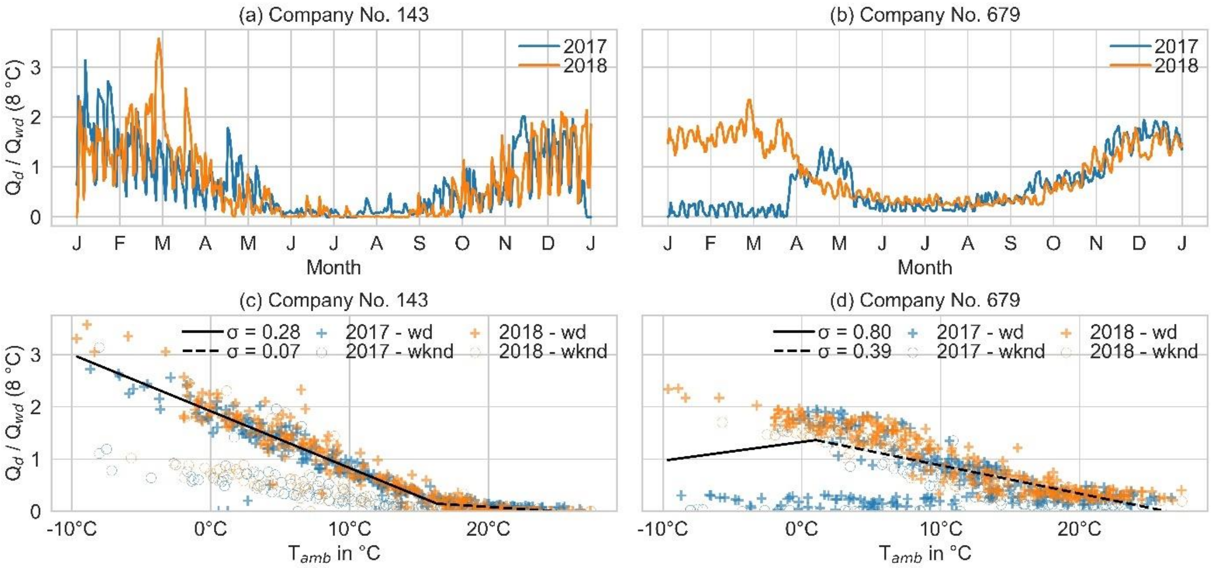


**Figure 6: Normalized load profiles of two exemplarily companies: (a) and (b) plot of daily normalized natural gas consumption as timeseries; (c) and (d): plot of daily natural gas consumption depending on daily ambient temperature). [color]**

After data filtering and plausibility check, 566 load profiles are remaining (Figure 7) and used in cluster analysis. In sum, 233 load profiles (29.0 %) are excluded for the reasons described above. Both thresholds used in completeness and plausibility checks, i.e., concerning the maximum share of zero values (of 83 %) and the maximum standard deviation of the lin function (of 0.75), have a high influence on the number of excluded load profiles. These thresholds are found iteratively. The thresholds used in this analysis appear to be a viable compromise. Diagrams of all remaining and excluded load profiles are shown graphically in Appendix F and in the data repository [27]. These graphs can be used to review the thresholds.

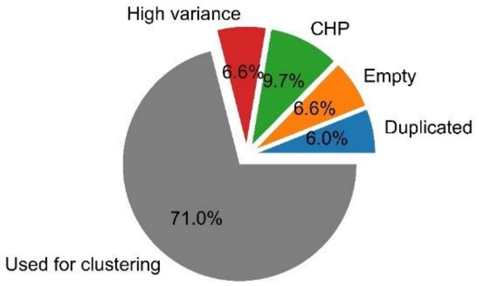


**Figure 7: Summary of plausibility check (number total: 797, number used for clustering: 566). [no color]**

### *3.2.4 Creation of uniform data vectors for clustering*

To identify similarities between the examined load profiles using the k-means algorithm, values in each dimension of the input data vectors must be comparable. The load profiles are from three different years and two different regions. In the context of this study, the main influences on daily heat load are mean daily ambient temperature and type of day (wd or wknd). Since both influencing parameters vary for a specific day of year from location to location and year to year, load profiles are not comparable in the original form of a time series. Consequently, the load profiles must be reshaped in a way that the heat loads of identical types of day with similar mean daily outdoor temperatures are in the same dimensions. This is done by forming daily mean ambient temperature intervals with a size of 0.5 K each. In the next step, the frequencies in which the ambient temperature lies in each interval are determined separately for the two regions. Then, the minimum frequencies from both load profile regions are identified separately for wd and wknd. Since only one weather station is used for each of the load profile regions (section 3.2.1.6), only two ambient temperature frequency distributions must be considered for each type of day. The minimum frequencies are exemplified in Figure 8 (a) for wd and contain 233 values. In the case of wknd, the minimum frequency distribution contains 105 values. For each load profile, two data vectors are formed that contain exactly as many days in each ambient temperature interval as specified by the respective frequency distribution for wd or wknd. If an original load profile contains more data in an interval than specified by the minimum frequency distribution, days in this interval are picked randomly. Figure 8 (b) visualizes total available and picked days for one consumer as an example. Finally, the created data vectors are sorted in ascending order by mean daily ambient temperature.

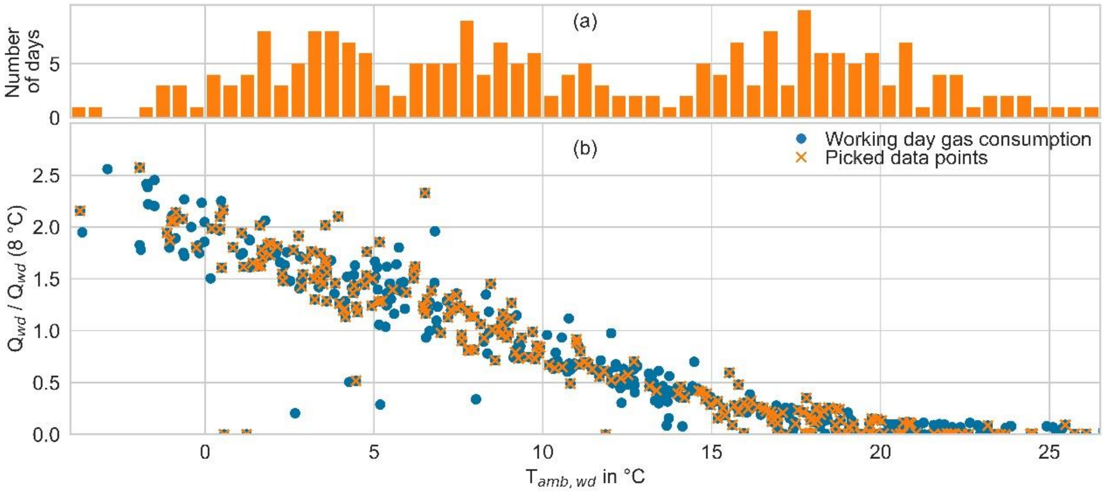


**Figure 8: (a) Histogram (interval size: 0.5 K) showing minimum frequency of working days with a specific ambient temperature in all covered datasets; (b) picked days for one exemplary company. [no color]**

## 3.3 *Clustering*

The presented study is based upon the programming language python [37]. The software library Scikit-learn [38] is used for load profile clustering. Various cluster algorithms are implemented to this library. The k-means algorithm, which is used in this study, is a general purpose algorithm and has already been used in literature for load profile clustering (section 2.2). The aim of the k-means algorithm is to choose cluster centroids (cluster means) that minimize the sum of the squared Euclidean distances of each sample from its assigned cluster centroid (Eq. 3.3) [39]. Figure 9 illustrates k-means clustering for two-dimensional datasets. In this study, each sample (load profile) is represented by a 233-dimensional data vector for wd and a 105-dimensional data vector for wknd (section 3.2.4).

$$d = \sum_{i=1}^{k} \sum_{x_j \in c_i}^{n} \left( \left\| x_j - \mu_i \right\|^2 \right) \qquad \text{Eq. 3.3}$$

| | |
|---|---|
| $c_i$ | cluster i [-] |
| $d$ | distortion (sum of the squared Euclidean distances from the cluster centroids) [-] |
| $k$ | number of clusters [-] |
| $n$ | number of samples (load profiles) in the respective cluster $C_i$ [-] |
| $x_j$ | sample (load profile) [-] |
| $\mu_i$ | centroid of cluster i [-] |

To find the cluster centroids, the k-means algorithm uses three steps. Step 1 is carried out only once. Step 2 and 3 are iterated until the new centroids do not significantly differ from the previous ones [39]:

1. Choose k initial cluster centroids.
2. Assign each sample to its nearest cluster centroid (the centroid where d (Eq. 3.3) increases the least).
3. Compute new cluster centroids using all samples assigned to each of the clusters.

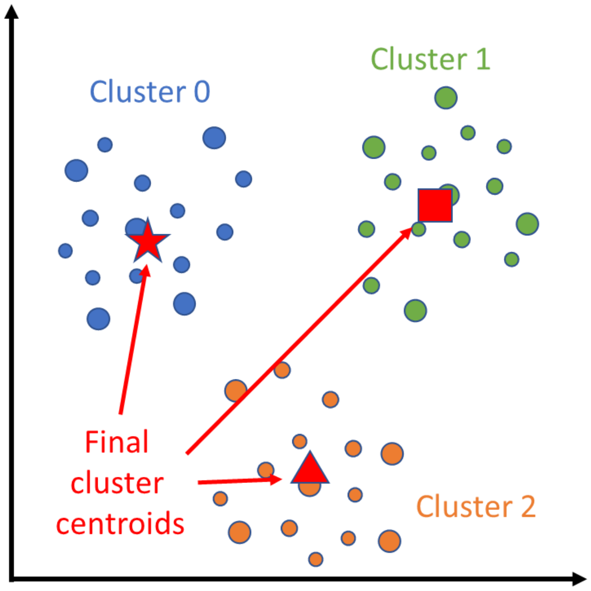


**Figure 9: k-means clustering for a two-dimensional dataset. [no color]**

The initial denomination of all clusters after k-means clustering is random and therefore meaningless. To add a structure to the cluster denomination, the difference between the means of all samples in each cluster at high ($T_{amb} > 20$ °C) and low ($T_{amb} < 0$ °C) temperatures are calculated. In the next step, the clusters are renumbered in ascending order according to these differences in sample means. For example, wd and wknd clusters 0 are the clusters with the smallest difference between the means of all samples at high and low temperatures. Consequently, wd cluster 3 or wknd cluster 4 are the clusters with the highest difference between high and low temperatures.

The k-means algorithm needs the number of clusters to be specified. The higher the number of clusters is, the lower is the overall distortion. To find a good compromise between a low distortion and a manageable low number of clusters, an elbow plot is used in this study (Figure 10). An elbow plot visualizes the overall distortion depending on the number of clusters. The point with the maximum curvature (“elbow” or “knee”) is a good tradeoff between additional input (number of clusters) and resulting output (low distortion). The Yellowbrick [40] software library is used to automatically identify the elbow point.

The clustering performance is evaluated using silhouette coefficients (Eq. 3.4) [41]. Silhouette coefficients can take values between -1 and 1. Values near 1 indicate a correct clustering with clearly separated clusters. Values near -1 indicate an incorrect clustering with overlapping clusters.

$$s = \frac{b - a}{\max{(a, b)}} \qquad \text{Eq. 3.4}$$

$a$ mean distance between a sample and all other samples in the same cluster [-]
$b$ mean distance between a sample and all other samples in the next nearest cluster [-]
$s$ silhouette coefficient [-]

A silhouette plot visualizes the silhouette scores of all clustered samples. Samples are sorted by clusters and within each cluster drawn as ascending horizontal bars. This results in the triangular or sail-shaped appearance that is recognizable in the silhouette plots in Figure 11 and Appendix B. A silhouette plot provides an overview on the silhouette score of all samples, even in a large dataset.

## 3.4 *Regression analysis*

In a least-squares approximation, the sig and siglin functions presented in section 2.1 and the modified lin function presented in section 3.2.3 are fitted to each load profile individually and jointly to all load profiles assigned to one of clusters. The quality of these regressions is evaluated using the coefficient of determination ($R^2$) and the standard deviation of the residuals (σ). Residuals are given by the difference of predicted load using the regression functions and the real load. $R^2$ is a commonly used metric for the quality of a regression but is not applicable to horizontal trends. By definition, $R^2$ is 0 for a horizontal trend, even if the regression fits perfectly. To account for possible horizontal trends that imply only a small ambient temperature dependence of the load profile, σ is used as a second metric. In contrast to $R^2$, σ does not depend on the orientation of a trend and results in same values for same deviations at different orientations.

The aim of the regression analysis is to develop and test the suitability of individual and cluster regressions for predicting a load profile just based on the ambient temperature. At the same time, it is examined which losses in accuracy have to be accepted if cluster regressions are used instead of individual regressions. Clustering and regression analysis is performed individually for wd and wknd. This partly replaces the purpose of weekday factors ($F_d$) as used by SLP methodology. In contrast to the SLP methodology, no further distinction is made between weekdays (Monday to Friday) or weekends (Saturday and Sunday). As a result, Eq. 3.2 is valid for all regressions developed in this study.

The SLP methodology (section 2.1) uses a geometric series of the ambient temperature to consider the inertia of the correlation between ambient temperature and heat load. This study investigates if the accuracy of load profile estimation can be increased by the usage of a geometric series (Eq. 2.2) instead of the usage a simple time series. Therefore, the results of two complete runs of the clustering and regression analysis, one with a simple time series and one with a geometric series, are compared.

# 4. Results

This section summarizes the results of the previously described clustering (section 4.1) and regression analysis (section 4.2).

## 4.1 *k-means clustering*

The elbow plots (Figure 10) indicate four wd clusters and five wknd clusters. In the case of wd (a), the elbow is clearly visible. Nevertheless, the angle of the elbow is relatively obtuse. In the case of wknd (b), no elbow is visible at all. The point of the highest curvature, which is equivalent to the point of the elbow, can only be derived mathematically.

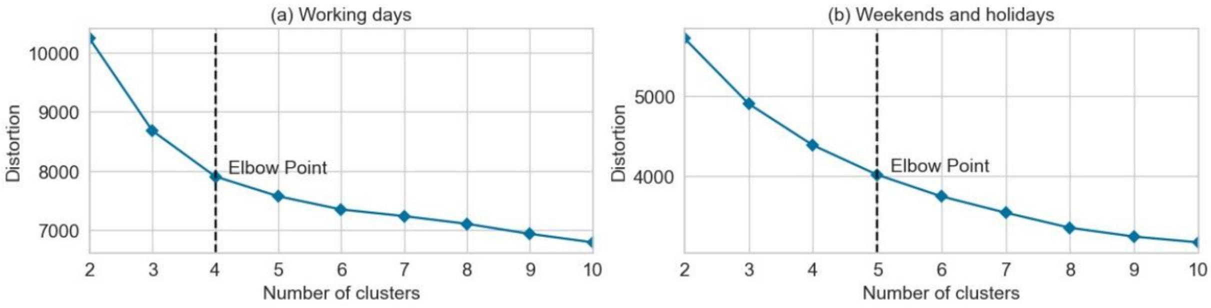


**Figure 10: Elbow plot of k-means clustering for working days (a) and weekends and holidays (b). [no color]**

The silhouette plots with four wd and five wknd clusters (Figure 11) indicate a weak separation of the clusters which is indicated by the overall low silhouette coefficients. Silhouette coefficients are increased when the number of clusters is reduced. In the case of wd, two clusters lead to a significantly increased cluster separation (Appendix B, Figure B.1). In the case of wknd, two and three clusters result in significantly increased silhouette coefficients (Appendix B, Figure B.2).

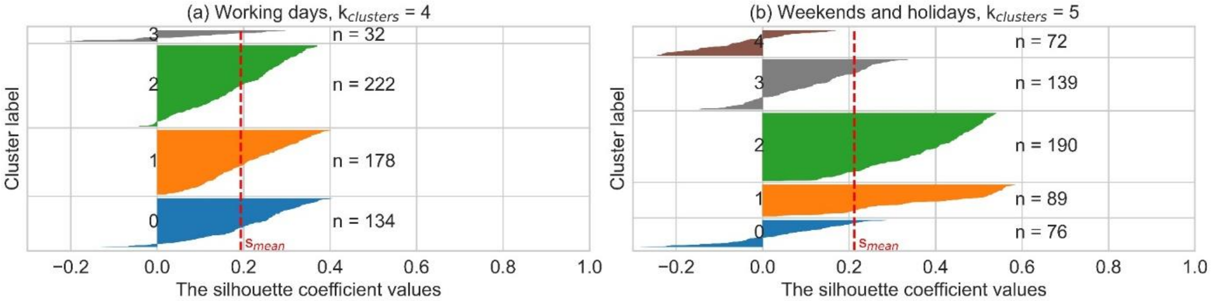


**Figure 11: Silhouette Plot of k-means clustering for four wd clusters (a) and five wknd clusters (b). [no color]**

Figure 12 and Figure 13 show all load profiles assigned to each of the identified clusters. To provide an unhindered overview of the general trend within each cluster, triangular weighted moving averages (± 15 periods) are used to keep the figures clean from outliers. Additionally, two linear regression lines (Eq. 3.1) are visualized for each cluster to facilitate comparability of the general cluster trends. In the case of wd, dependency of natural gas consumption on mean daily ambient temperature increases from cluster 0 to cluster 3. The slope of the left regression line ($m_h$) is an indicator for ambient temperature dependency (Table 2). The higher the absolute value of $m_h$ is, the more the heat demand increases when the ambient temperature decreases.

Aside from the general variance, natural gas consumption is nearly constant for all consumers assigned to cluster 0. From cluster 1 to 3, the share of natural gas consumption on days with a high ambient temperature (summer days) is decreasing. Since the right regression line is almost horizontal for all clusters, the y-axis intercept of this line ($b_w$) is an indicator for the summerly base load in each cluster. Table 2 classifies all wd and wknd clusters according to ambient temperature dependency ($m_h$) and summerly base load ($b_w$).

**Table 2: Cluster classification according to $m_h$ (slope of left regression line) and $b_w$ (y-intercept of right regression line).**

| | **$b_w > 1.0$** | **$1.0 \geq b_w > 0.6$** | **$0.6 \geq b_w > 0.2$** | **$b_w \leq 0.2$** |
|---|---|---|---|---|
| **$m_h \geq -0.02$** | wd 0 | wknd 0 | | wknd 1 |
| **$-0.06 \leq m_h < -0.02$** | | wknd 2 | wd 2, wknd 3 | |
| **$-0.10 \leq m_h < -0.06$** | | wd 1 | | |
| **$-0.14 \leq m_h < -0.10$** | | | wknd 4 | |
| **$m_h < -0.14$** | | | wd 3 | |

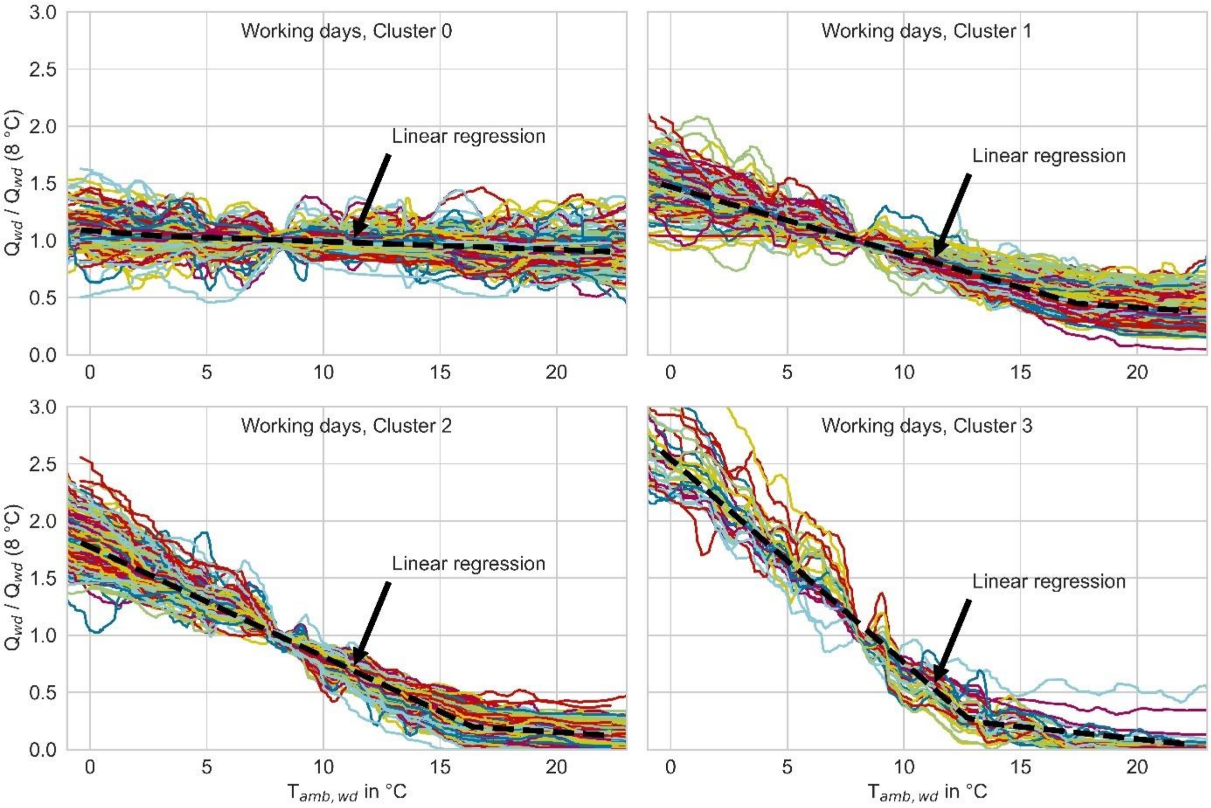


**Figure 12: Results of k-means clustering of natural gas consumption on wd (to keep the figure clean of outliers, the triangular weighted moving average (± 15 periods) of each load profile is illustrated). [color]**

The clustering natural gas consumption on wknd leads to similar results compared to wd. Nevertheless, four main differences are evident from Figure 11, Figure 12 and Table 2:

1. Load profiles are normalized to the heat load on wd with a mean daily ambient temperature of 8 °C. Consequently, all wd clusters have a mean normalized load of 1 on days with a mean daily ambient temperature of 8 °C. Although the heat demand on wknd is not relevant for normalization, wknd clusters 0, 2 and 4 also show a normalized load of approximately 1 on days with a mean ambient temperature of 8 °C.
2. Wknd clusters 1 and 3 show a heat load that is significantly below 1 on days with a mean ambient temperature of 8 °C. Wknd cluster 1 includes consumers with a significantly reduced natural gas consumption on wknd compared to wd. Some of these consumers have an almost constant consumption. Some other consumers in wknd cluster 1 show a slightly decreasing consumption when ambient temperature rises. All consumers in wknd cluster 3 show a clear dependency on ambient temperature with a small summer load and a medium winter load.
3. There is no wknd equivalent to wd cluster 3. The mean normalized load curve of wknd cluster 4, the wknd cluster with the highest normalized winter and lowest summer natural gas consumption, lies underneath the one of wd cluster 3.
4. Wknd lines are smoother.

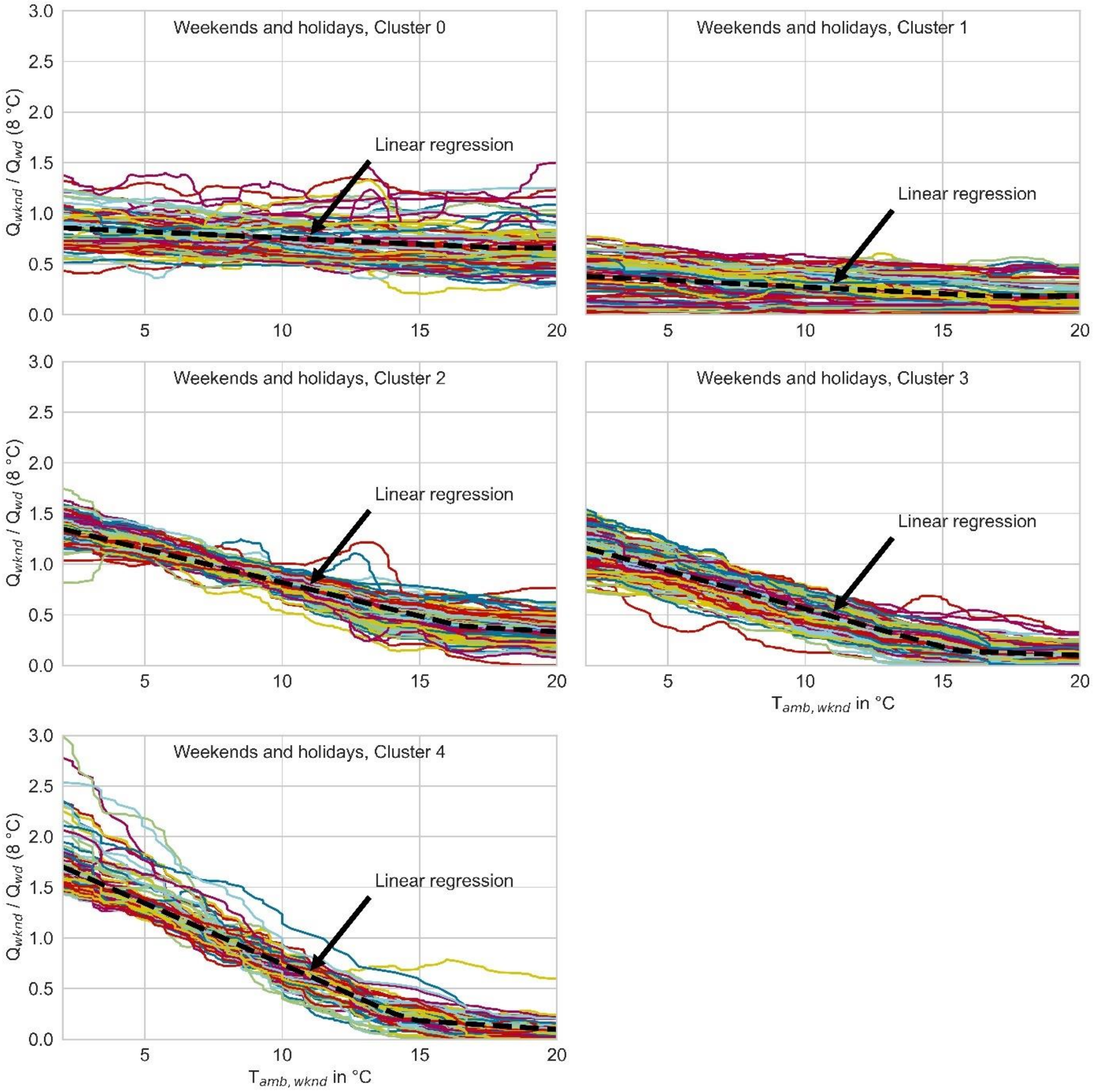


**Figure 13: Results of k-means clustering of natural gas consumption on wknd (to keep the figure clean of outliers, the triangular weighted moving average (± 15 periods) of each load profile is illustrated). [color]**

Figure 14 visualizes the proportion of wknd clusters within the individual wd clusters. A total of 20 different cluster combinations are possible but only 14 cluster combinations occur in the results. This is because combinations between low wknd clusters and high wd clusters or the other way around do not appear in any of the investigated load profiles. Consequently, the general characteristics of the dependence of natural gas consumption on ambient temperature at weekdays on the one hand and weekends on the other hand are similar for all load profiles but may show slight variations in absolute values.

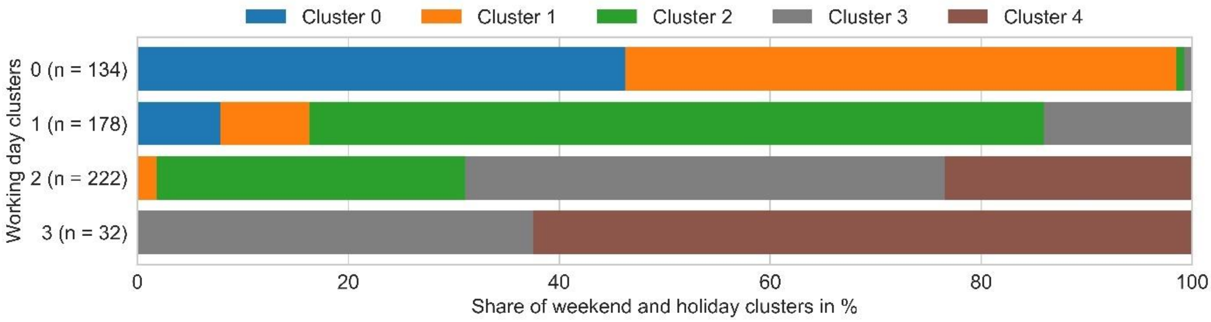


**Figure 14: Relationships of wd- and wknd-clusters. [color]**

Figure 15 shows the clustering results within the economy sectors. The same do Figure 16 and Appendix C for economy divisions. The share of wd clusters 2 and 3 is similar for all economy sectors. Significant differences in cluster share are only evident for wd cluster 0 and 1. While wd cluster 0 is relatively rare in the tertiary sector (commercial services), wd cluster 0 dominates secondary sector (manufactures and assembly goods).

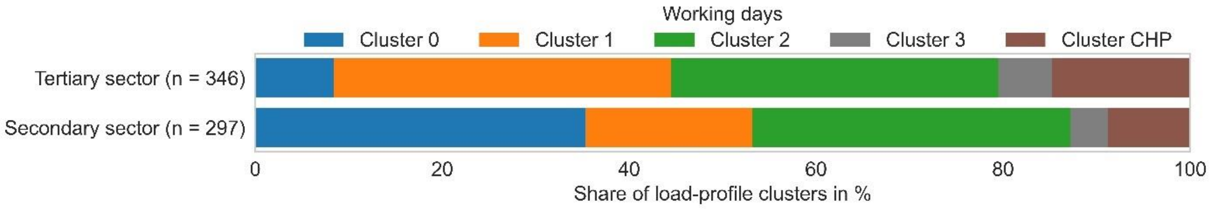


**Figure 15: Shares of wd-Clusters in the secondary and tertiary sectors. [color]**

For most of the economy divisions in secondary sector (Figure 16), more than one wd cluster is frequent. For instance, the very different wd clusters 0 and 2 have almost the same share in manufacture of furniture. In contrast to that, only a few economy divisions are clearly dominated by one wd cluster, e.g., manufacture of food products. Nevertheless, a trend towards one of the wd clusters can be observed for most economic divisions. The same characteristic appears in tertiary sector and for wknd clustering (Appendix C).

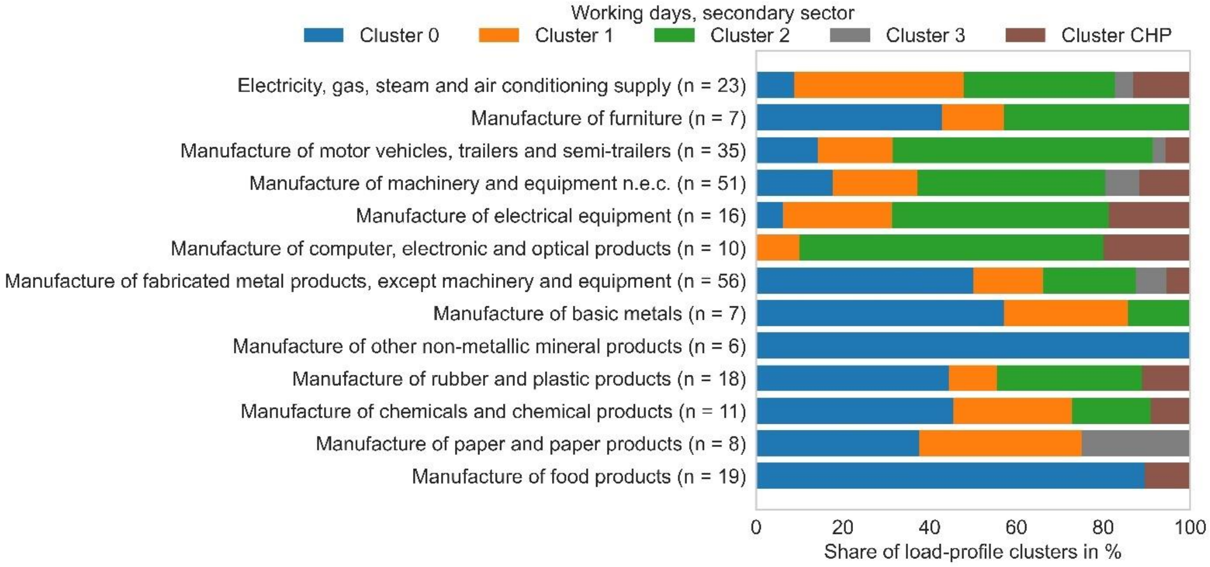


**Figure 16: Shares of wd-Clusters in secondary sector (manufactures and assembly goods); economy divisions with at least five load profiles only. [color]**

## 4.2 *Regression analysis*

Figure 17 visualizes the global accuracy of the cluster regression-based load profile prediction by illustrating residuals between daily predicted heat loads and real heat loads for all days of all clustered load profiles in form of a histogram. The mean value of the residual is 0.00 and the one of σ is 0.27. There are no significant differences between the three regression function types (lin, sig, or siglin).

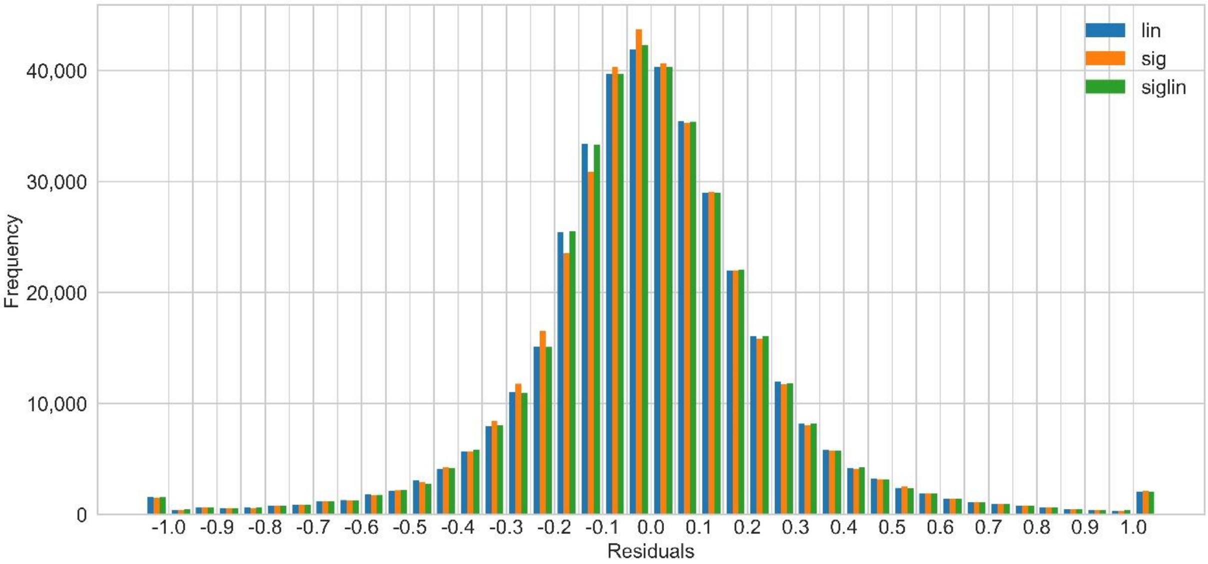


**Figure 17: Histogram of residuals between predicted normalized daily heat load based on cluster regression-based correlations and real normalized daily heat load (388,561 days in total). [color]**

Appendix F illustrates original load profiles and synthetically created load profiles based on lin cluster regressions for all load profiles from manufacture of motor vehicles, trailers, and semi-trailers. The data repository [27] contains the same illustrations for all other investigated economy divisions. To evaluate the accuracy of the cluster regression-based correlations for each load profile as shown in Appendix F or the data repository [27], Figure 18 illustrates σ and R² for each of the clustered load profiles in a histogram and boxplots. Regardless of the type of regression function (sig, lin, or siglin), the mean value of σ is 0.24 and the mean value of R² is 0.71. σ and R² for each of the cluster regressions can be found in Table 3 and Appendix D. σ ranges from 0.23 to 0.43. R² ranges from 0.04 to 0.84 but takes values below 0.76 only for clusters with horizontal trends (wd cluster 0, wknd clusters 0 and 1).

In contrast to the findings of Hellwig [8], substituting mean daily temperatures ($T_{amb}$) by a geometric series (Eq. 2.2) does not increase the accuracy significantly. For individual and cluster regressions, both σ and R² are not improved by more than 1 %-point (not illustrated).

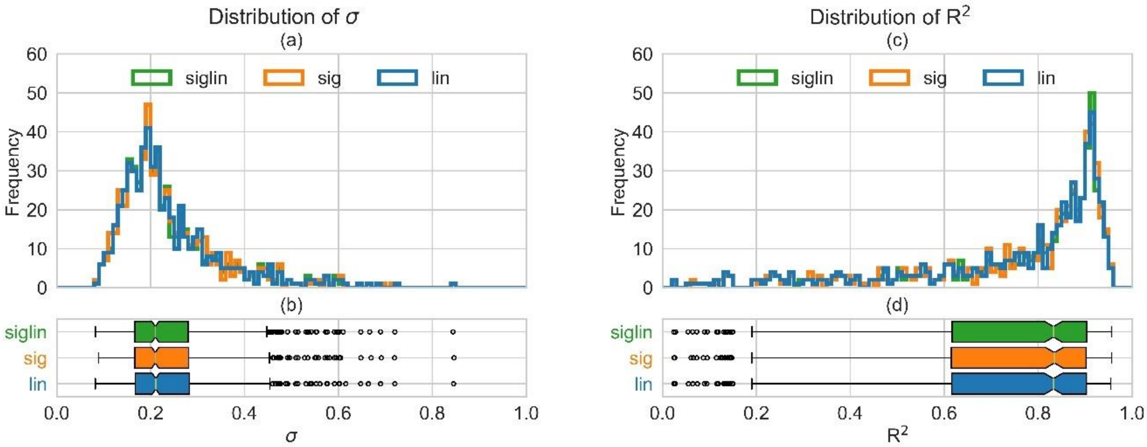


**Figure 18: Accuracy of heat load profile correlations based on lin, sig or siglin cluster regressions; (a) histogram of σ; (b) boxplot of σ; (c) histogram of R²; (d) boxplot of R². [color]**

If regression functions are not fitted jointly for all consumers within each cluster but individually for each consumer, the accuracy of the synthetically created load profiles can be increased slightly. The mean value of σ is 0.21 and mean value of R² is 0.79 for all functions. Figure 19 illustrates the frequency distributions of σ ((a) & (b)) and R² ((c) & (d)) for synthetically created load profiles based on individually fitted lin, sig or siglin functions. No significant differences of σ and R² are detectable between the three examined functions.

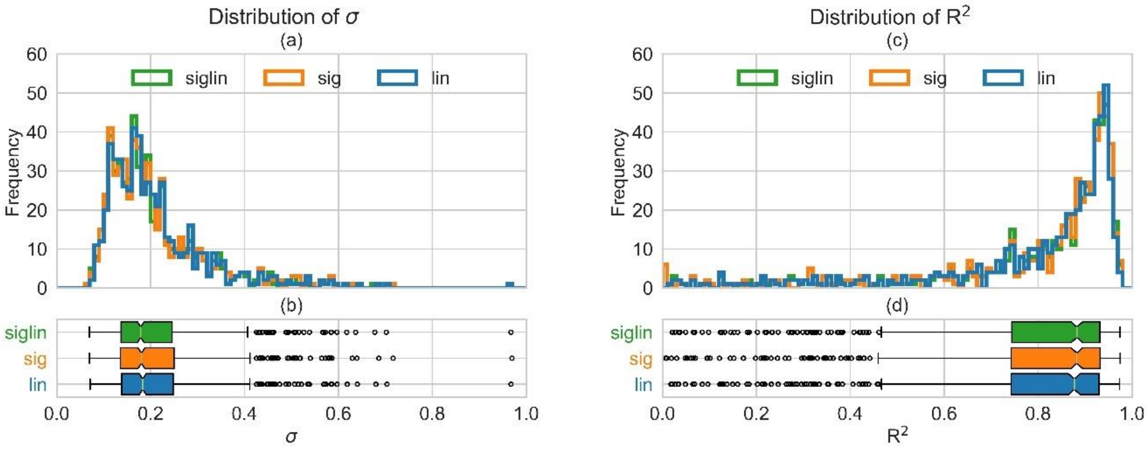


**Figure 19: Accuracy of individually fitted lin, sig or siglin regressions; (a) histogram of σ; (b) boxplot of σ; (c) histogram of R²; (d) boxplot of R². [color]**

Figure 20 visualizes the cluster regression-based correlations. For almost the complete examined temperature range and all clusters, no significant differences between the lin, sig and siglin functions are discernible. Differences between the three functions are limited to ambient temperatures below -5 °C which are rare (compare to Figure 8 (a)) and thus have only a limited influence.

For wd cluster 3 and wknd clusters 3 and 4, negative heat load predictions result when daily mean ambient temperature exceeds about 25 °C (Figure 20). The exact intersection with the x-axis depends on the function type. However, for these temperatures, all functions show nearly horizontal trends. Therefore, the normalized daily heat load predictions do not fall below -0.065 for the examined weather datasets. Negative values occur in 0.14 % of lin daily load predictions and 0.08 % of siglin daily load predictions. For all sig cluster regressions as well as all lin and siglin cluster regressions of all clusters, except wd cluster 3 and wknd clusters 3 and 4, no negative daily load predictions occur at all.

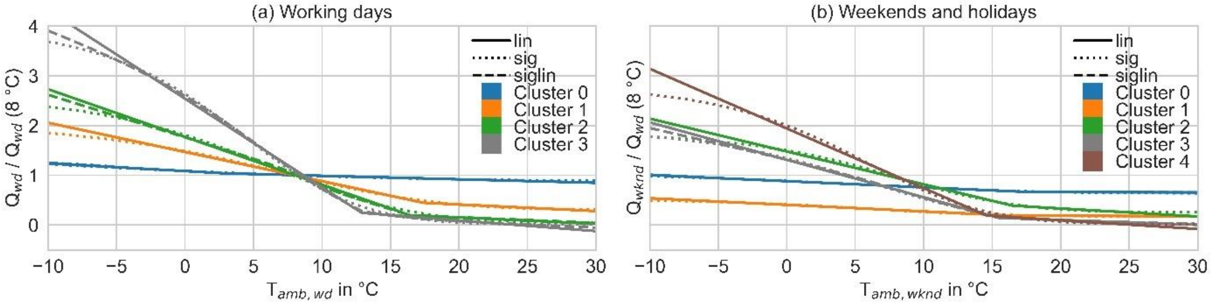


**Figure 20: Cluster regression-based correlations for working days (a) and weekends and holidays (b). [color]**

## 5. Discussion

In this section, the results of clustering (section 5.1) and regression analysis (section 5.2) are discussed. Possible error sources that are relevant for both clustering and regression analysis are examined in section 5.3.

### 5.1 *k-means clustering*

Wd and wknd clustering leads to similar results. For both, the dependency of heat load on ambient temperature increases from cluster 0 to cluster 3 (wd) or cluster 4 (wknd). For wknd clusters 1 and 3, normalized heat load at 8 °C mean daily ambient temperature is below 1. This indicates a reduced heat load at the wknd due to switched-off processes or a reduced room temperature when no production takes place. Consequently, it is reasonable to assume that consumers assigned to wknd cluster 1 and 3 are only producing on five days a week. In contrast, the rest of the clusters (wd clusters 0 to 3 and wknd clusters 0, 2 and 4) show a normalized heat load of approximately 1 for days with a mean ambient temperature of 8 °C. Therefore, it can be concluded that the consumers in these clusters produce seven days a week. For clusters 2 and 4, a five-day production is also possible if space heating demand dominates overall heat load, and the room temperature is not reduced on wknd.

Switched-off processes on wknd may also explain why the load profile lines are smoother for wknd compared to wd. It is plausible that the remaining space heating demand is less influenced by user behavior, resulting in smoother lines.

Elbow and silhouette plots both indicate weakly separated, overlapping clusters. This can be explained by the shape of the identified and evaluated clusters. All the clusters partly overlap or connect to the neighboring ones seamlessly (Figure 12 and Figure 13). Consequently, load profiles in the transition area between two clusters show a poor silhouette coefficient.

To check the clustering results for plausibility, information about heat sinks from consumer’s homepages is linked to their respective cluster assignment (Appendix E). It can be concluded that if industrial production dominates at a consumer's site, clusters 0 or 1 are most frequent. There are several reasons why many processes in industrial production are independent from ambient temperature. For example, more than 90 % of the overall heat demand in manufacture of basic metals or manufacture of non-metallic mineral products is at temperatures of more than 500 °C [42]. Consequently, a change of ambient temperature does not influence heat demand in these economy divisions significantly, resulting in clearly dominant wd clusters 0 and 1. In contrast, almost 60 % of the overall heat demand in manufacture of food is below 100 °C. However, this economy division is still dominated by wd cluster 0 because important processes like cooking, cleaning, steaming or sterilization are based on heated water and are therefore independent of ambient temperature.

However, not all manufacturing processes are characterized by a heat demand that is independent of ambient temperature. Consequently, there are also some manufacturing economy divisions that are dominated by wd cluster 2, e.g., manufacturing of computer, electric and optical products or manufacturing of electrical equipment. Additionally, these economy divisions are characterized by a high share of space heating (> 40 %) on total heat demand which is depending on ambient temperature [42]. Regardless of the economy division to which a consumer belongs, clusters 2 or 3 are more common when a consumer's site is dominated by processes or activities that have usually no heat demand other than space heating. Examples include logistics, general administration, R&D, distribution, or services.

In most economy divisions, cluster assignment clearly tends towards one cluster but, for example, the shares of clusters 0 and 2 are almost equal in manufacture of furniture which appears to be implausible at first. An analysis of the consumers in this economy division reveals that those assigned to cluster 0 produce furniture from metal. They operate a range of processes which are less dependent or even independent of ambient temperature, e.g., surface treatment baths or powder coating. In contrast, consumers from manufacture of furniture assigned to cluster 2 produce wooden furniture and upholstery which only leads to a space heating demand without a significant share of process heat demand.

In summary, the cluster assignment can be explained in most cases based on the respective time schedules or heat sinks at a consumer’s site and is therefore plausible. This confirms the hypothesis that the respective composition of heat sinks is specific to the consumers and a main reason for differences between their normalized load profiles.

## 5.2 *Regression analysis*

The developed heat load profile model based on cluster regressions enables to predict daily normalized heat load with a standard deviation of residuals of 0.27. This means that the standard deviation of the estimated daily heat loads equals 27 % of the mean daily heat load on wd with a daily mean ambient temperature of 8 °C. Nevertheless, since residuals are not biased and the general seasonal trend is captured by the load profile model, the accuracy is sufficient for the intended applications such as preliminary design or potential studies.

The accuracy of cluster regressions is only slightly poorer compared to individual regressions. The heat load profile model accuracy based on cluster regressions could further be improved towards individual regression accuracy if the number of clusters is increased. At the same time, increasing the number of clusters would make the choice of the correct cluster to predict a load profile even more complex and is therefore avoided.

Reducing the number of clusters would result in a better separation of clusters with increased silhouette coefficients but would reduce the overall quality of the regression. Considering the aim of this study, which is to develop an accurate model for heat load profile prediction, reducing the number of clusters would not be constructive and is avoided.

Since the usage of geometric series of daily mean ambient temperatures over the last four days does not increase the accuracy significantly, simple time series instead of geometric series are used in this work. A reason for this could be that the inertia of the correlation between ambient temperature and heat load is less relevant for those large consumers examined in this study compared to those examined by Hellwig [8] (section 2.1). Space heating demand of residential or office buildings is usually inert because of a relatively high thermal mass which can store a significant amount of heat. Consequently, changes of ambient temperature do not directly affect space heating demand. This is different to many consumers in this study. A relatively low thermal mass of industrial buildings, e.g., production halls, causes low inertia of heat load. In addition, heated ambient air is required for the ventilation of production halls or for many production processes. Overall, the delay between changes in ambient temperature and changes in heat demand is therefore shorter in industry compared to the residential sector.

As expected, $R^2$ is almost zero for clusters with a horizontal trend. As this is due to the definition of $R^2$, it does not necessarily indicate a poor accuracy which is confirmed by low values of $\sigma$. In general, the accuracy of the synthetically created normalized load profiles based on cluster regressions seems to be sufficient to be applied, for example, in preliminary design of renewable heating systems. This is confirmed by the visual comparison of the original and synthetic load profiles in Appendix F and the data repository [27].

Significant differences between the examined regression functions can only be observed for mean daily ambient temperatures below -5 °C. Temperatures in this range are rare in the examined dataset (occurring in only 1.9 % and 0.8 % of days in Stuttgart and northern Hesse, respectively). Consequently, none of the examined regression functions shows significant advantages in accuracy. Another explanation could be that the change in user behavior at cold ambient temperatures suggested by Hellwig [8] (section 2.1), which can explain the flattening of natural gas consumption curves, does not apply to the same extend to large-scale consumers such as those examined in this study. Nevertheless, it is possible that sig or siglin functions show advantages compared to the lin function for datasets from colder regions.

For the examined weather datasets, negative heat load predictions are rare and do not fall significantly below zero. Nevertheless, negative heat loads are not reasonable and will significantly be more frequent if the results of this study are applied to warmer locations. Therefore, the lin, sig, and siglin functions are adjusted to avoid negative values when the results of this study are applied in practice (section 6).

### 5.3 *Possible errors*

To ensure the transferability of the developed heat load profile model, normalized natural gas consumption profiles are assumed to be equal to normalized heat load profiles due to a linear relationship between heat load and natural gas consumption. Those consumers for which this assumption does not apply are not considered in this study, e.g., consumers operating a CHP plant. However, there is a high possibility that some consumers with a non-linear correlation of heat load and natural gas consumption are not identified, which is a possible source of error.

There are several other possible sources of errors like unusual user behavior, errors of ambient temperature and natural gas consumption measurements or errors caused by different weather at the consumers' sites and the selected weather stations. All possible errors listed above have in common that they cannot be evaluated due to a lack of more detailed information. Nevertheless, apart from unidentified consumers with a non-linear relationship between heating load and natural gas consumption, most of the possible errors are assumed to be random and therefore do not cause bias. This assumption is supported by Figure 17. Residuals between real and predicted daily heat loads are no biased.

## 6. Application of clustering and regression results

A representative ambient temperature profile with a daily resolution and some basic information about a particular consumer are the only requirements to derive a heat load profile for a given consumer based on the developed method. The prediction of the heat load profile is carried out in three steps:

1. Selection of the load profile clusters according to the company's economy division and production times: Assign wd- and wknd-clusters based on the consumer's economy division (Appendix C). Check the plausibility of assignment based on general information about consumer's activities and processes (section 5.1 and Appendix E). Choose wknd cluster 0 only if the consumer is producing seven days a week. Wknd clusters 1 and 3 are only eligible for five-day operation. Wknd clusters 2 and 4 are eligible for five- and seven-day operation.

2. Calculation of the normalized load profile:
   Calculate the normalized heat load profile $h(T_{amb})$ using Eq. 6.1, the regression parameters found in Table 3 and a representative seasonal (usually annual) profile of daily mean ambient temperature. Alternatively, choose sig or siglin regressions found in Appendix D. Use the respective wd regression for all working days and the respective wknd regression for weekends and holidays. As motivated in section 5.2, the lin (Eq. 6.1), sig (Eq. D.1) and siglin (Eq. D.2) functions are adjusted to avoid negative daily heat loads.
3. Calculation of the natural gas consumption or heat load profile:
   Normalized natural gas and heat load profiles of those consumers examined in this study are assumed to be equal. Therefore, the absolute natural gas consumption profile or heat load profile can be calculated based on the normalized load profile (Eq. 6.2). An estimation of annual natural gas consumption or heat load is not part of this study but can, for example, be taken from the last natural gas bill.

$$Q_d/Q_d(8°C) = h_{lin}(T_{amb}) = \begin{cases} max(0,\ m_h \cdot T_{amb} + b_h) & if\ \ T_{amb} < T_{hl} \\ max(0,\ m_w \cdot T_{amb} + b_w) & if\ \ T_{amb} \geq T_{hl} \end{cases} \qquad \text{Eq. 6.1}$$

$$Q_d = Q_d/Q_d(8°C) \cdot \frac{Q_s}{\sum_{i=1}^{j} h(T_{amb})_i} \qquad \text{Eq. 6.2}$$

| | |
|---|---|
| $b_h$ | y-axis intercept of space heating line [-] |
| $b_w$ | y-axis intercept of domestic hot water (process heat) line [-] |
| $h_{lin}(T_{amb})$ | normalized daily natural gas consumption/ heat load as function of $T_{amb}$ [-] |
| $j$ | number of days in examined season [-] |
| $m_h$ | slope of space heating line [-] |
| $m_w$ | slope of domestic hot water (process heat) line [-] |
| $Q_d$ | daily natural gas consumption or daily heat load [kWh] |
| $Q_d/Q_d(8°C)$ | normalized daily natural gas consumption/ heat load [-] |
| $Q_s$ | seasonal (usually annual) natural gas consumption (minimum 300 days) [kWh] |
| $T_{amb}$ | daily mean ambient temperature [°C] (insert unitless) |
| $T_{hl}$ | heating limit temperature [°C] (insert unitless) |

**Table 3: Lin cluster regressions.**

| lin | Cluster | $b_h$ [-] | $m_h$ [-] | $b_w$ [-] | $m_w$ [-] | $t_{hl}$ [°C] | $R^2$ [-] | σ [-] |
|---|---|---|---|---|---|---|---|---|
| wd | 0 | 1.0852 | -0.0154 | 1.0610 | -0.0071 | 2.9 | 0.04 | 0.31 |
| | 1 | 1.4695 | -0.0588 | 0.6779 | -0.0133 | 17.4 | 0.76 | 0.23 |
| | 2 | 1.7719 | -0.0960 | 0.4070 | -0.0128 | 16.4 | 0.87 | 0.25 |
| | 3 | 2.5404 | -0.1780 | 0.5210 | -0.0215 | 12.9 | 0.84 | 0.43 |
| wknd | 0 | 0.8814 | -0.0124 | 0.6661 | -0.0003 | 17.9 | 0.04 | 0.39 |
| | 1 | 0.4053 | -0.0132 | 0.1961 | -0.0006 | 16.6 | 0.10 | 0.26 |
| | 2 | 1.4792 | -0.0661 | 0.6527 | -0.0160 | 16.5 | 0.81 | 0.21 |
| | 3 | 1.3112 | -0.0753 | 0.2952 | -0.0098 | 15.5 | 0.76 | 0.26 |
| | 4 | 1.9425 | -0.1201 | 0.4449 | -0.0175 | 14.6 | 0.79 | 0.38 |

## 7. Conclusion

In this study, a method to create normalized annual heat load profiles with a daily resolution for consumers from commercial, industrial, public or residential sectors is developed. To apply this method, only the presented results, an annual ambient temperature profile with daily resolution and some basic information about the products, schedule and heat sinks of a given consumer are required.

The developed method is based on a dataset of metered natural gas load profiles of almost 800 consumers in Germany, most with an annual natural gas consumption of at least 1.5 GWh/a. To derive normalized heat load profiles from natural gas load profiles, consumers with non-linear correlations of natural gas consumption and heat load are eliminated from the database. Based on this, normalized natural gas load profiles are assumed to be equal to normalized heat load profiles.

The simplicity of the developed correlations, based on two regressions, one for working days and one for weekends and holidays, ensures transferability and user-friendliness. At the same time, the developed correlations achieve a sufficient accuracy for the intended applications like preliminary design or potential studies of renewable heating systems. Previous studies of residential and small commercial gas consumers detected that sigmoid or linearized sigmoid functions combined with a geometric series of ambient temperature for the last four days achieve the highest accuracy. In contrast, this study yields that a linear regression without a geometric series of ambient temperature can achieve the same accuracy as these more complex approaches.

Despite the high accuracy of the regression-based correlations, it must be considered that possibly some consumers characterized by a non-linear correlation between heat load and natural gas consumption were not identified and remained in the examined database. This is a possible source of error which cannot be evaluated due to a lack of detailed information about the consumers. For example, consumers operating special natural gas-fired heat generators like natural gas absorption heat pumps, consumers using natural gas as a material, or consumers operating other heat sources in parallel to a natural gas-fired heating system cannot be identified and excluded from this study. However, this possible source of error is considered small based on various statistics. Additionally, some other random error sources are possible, e.g., measurement errors. Despite these potential errors, the developed load profile model is a significant improvement over previous industrial load profile models that create, for example, synthetic load profiles by manually selecting and combining typical daily, weekly, and annual patterns.

The developed method for heat load profile prediction is assumed to be applicable in temperate climate zones where the overall heat demand is influenced by ambient temperature significantly. In sub-tropical or tropical climate zones, it is assumed that ambient temperature has a distinctly reduced influence on overall heating demand and the developed load profile model is therefore not applicable. Unfortunately, the applicability and transferability within the temperate climate zones and to other climate zones could not be validated since no load profiles from other locations are available.

The k-means load profile clustering according to the respective dependency on daily mean ambient temperature is a crucial step in the development of the heat load profile model. Across all economy sectors, consumers that are depending on ambient temperature are most frequent, even in secondary sector (manufactures and assembly of goods). Only for some economy divisions like manufacturing of food or manufacturing of basic metals, heat load of most consumers is not influenced significantly by ambient temperature.

In general, clustering results can be explained by the respective heat sink composition and are therefore plausible. In contrast, cluster separation is relatively poor, as all clusters are partly overlapping and directly connected to the next lower or higher ones. However, this does not have negative consequences for the developed method for load profile prediction.

To apply the results of this study to load profile prediction, a given consumer must be assigned to a cluster, manually. For this purpose, the detected frequencies of each cluster within the different economy divisions (Appendix C) can be used as a key indicator. For some economy divisions, this indicator is unclear because of similar frequencies of two or more clusters. In any case, to assign the cluster is a possible source of error and must be checked for plausibility. This can be done based on an analysis of the respective composition of heat sinks at a consumer's site similar to the review of clustering results presented in this study. If industrial production dominates at a consumer's site, clusters 0 or 1 are most likely. If, on the other hand, assembly, logistics, general administration, R&D, sales or service dominate, clusters 2 or 3 are most likely.

## 8. Directions of future work

Load profiles examined in this study are from two German regions. The transferability of the developed heat load profile model to other locations should be validated. For this purpose, a load profile database covering worldwide locations should be established.

The developed heat load profile model is based on the correlation between ambient temperature and heat demand, only for a resolution of one day. At hourly resolution, this correlation is overlaid by other influences such as consumer behavior. Nevertheless, to increase the resolution of the load profile from one day to one hour, a new methodology should be developed, e.g., by identifying consumer group specific patterns in daily load profiles.

Normalized annual heat load profiles generated based on the developed model can be scaled to any absolute seasonal heat demand. Absolute annual heat demand can be derived, for example, from the last natural gas bill. For cases where information about total seasonal heat demand is not available, e.g., when a new plant is planned, a methodology should be developed to estimate annual total heat demand. This study yields that it is not possible to derive simple benchmarks on total natural gas consumption for the examined economy divisions. Future work should investigate if it is possible to derive economy division specific benchmarks like heat demand per number of employees, per turnover or per area of production hall.

The developed load profile model is based only on the correlation between mean daily ambient temperature and heat demand. Additional influencing parameters like consumer behavior must be considered to further increase the model accuracy. To generate information on consumer behavior, additional sources of information must be evaluated. One possible source of information that is usually available to most industrial consumers are electricity load profiles. Electricity load profiles contain additional information, e.g., about start and end of production, degree of capacity utilization or the type of operated processes at a specific time. Therefore, a methodology based on a supervised machine learning model that evaluates electricity load profiles to increase the accuracy of the developed heat load profile model should be developed.

## Acknowledgements

This work was supported by the German Federal Ministry for Economic Affairs and Energy within the framework of the 7th Energy Research Program [project: "AnanaS", grant number 03ETW014A]. The authors would like to express their sincere thanks to the ministry for the financial support and to the involved utilities for providing natural gas load profiles.

## Data Availability

The original and normalized load profile database examined in this study cannot be published for data protection reasons. Plots of all 797 natural gas load profiles sorted by economy divisions and wd clusters are available in a data repository [27]. The plots show the real load profile and the predicted load profile based on cluster regressions (as illustrated exemplarily in Appendix F for the economy divisions "manufacture of vehicles, trailers and semi-trailers").

Measured weather data is available from Hessian State Agency for Nature Conservation, Environment, and Geology [35] and from German Meteorological Service [36].

## Appendix

**Appendix A: Python-software libraries used in this study.**

| *Library* | *Version* | *Reference* |
|---|---|---|
| Matplotlib | 3.2.2 | [43] |
| Numpy | 1.19.1 | [44] |
| Pandas | 1.1.3 | [45] |
| Python | 3.8.5 | [37] |
| Scikit-learn | 0.23.2 | [38] |
| Scipy | 1.5.0 | [46] |
| Yellowbrick | 1.2 | [40] |

## Appendix B: Silhouette Plot for k-means clustering with k = 2..10 clusters.

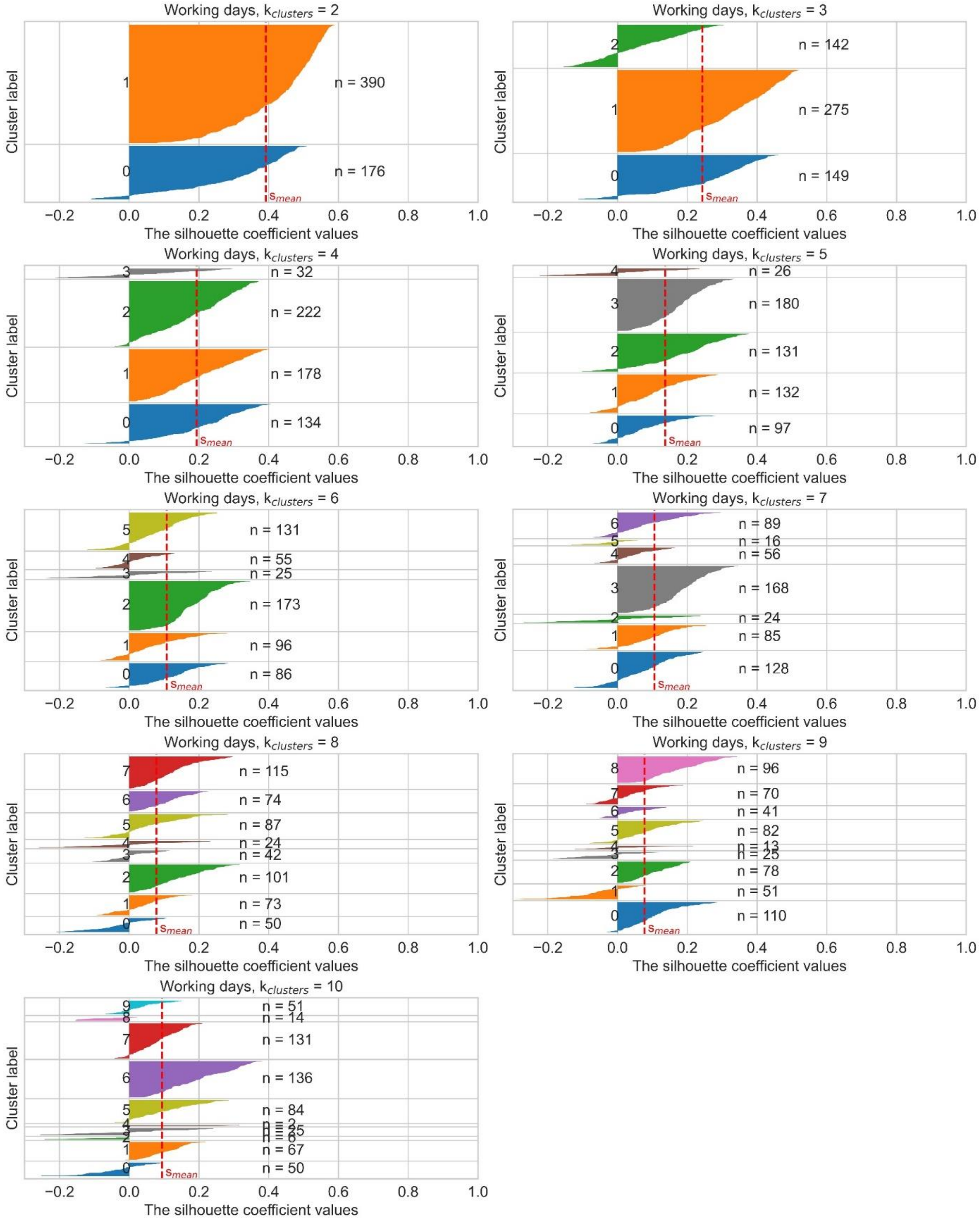


**Figure B.1: Working days. [no color]**

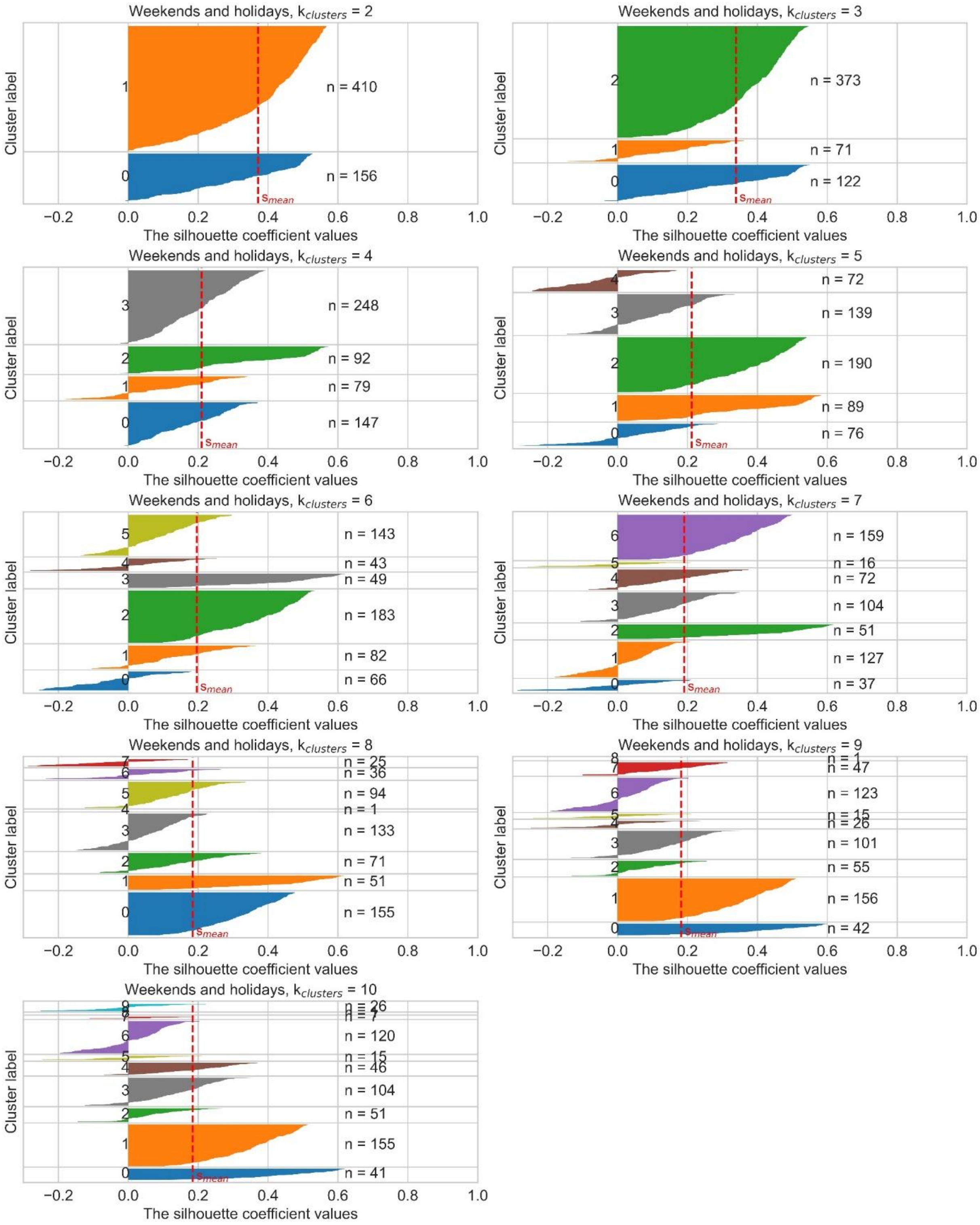


**Figure B.2: Weekends and Holidays. [no color]**

**Appendix C: Shares of clusters in all economy divisions with at least five load profiles ("Residential building or office" is not part of the NACE Rev. 2 systematic and was added by the authors)**

**Table C.1: Working days.**

| Economy division according to NACE Rev. 2 | | Number | Wd cluster share in % | | | | |
|---|---|---|---|---|---|---|---|
| Code | Name | | 0 | 1 | 2 | 3 | CHP |
| 1 | Crop and animal production, hunting and related service activities | 6 | 33.3 | 0.0 | 50.0 | 0.0 | 16.7 |
| 10 | Manufacture of food products | 19 | 89.5 | 0.0 | 0.0 | 0.0 | 10.5 |
| 17 | Manufacture of paper and paper products | 8 | 37.5 | 37.5 | 0.0 | 25.0 | 0.0 |
| 20 | Manufacture of chemicals and chemical products | 11 | 45.5 | 27.3 | 18.2 | 0.0 | 9.1 |
| 22 | Manufacture of rubber and plastic products | 18 | 44.4 | 11.1 | 33.3 | 0.0 | 11.1 |
| 23 | Manufacture of other non-metallic mineral products | 6 | 100 | 0.0 | 0.0 | 0.0 | 0.0 |
| 24 | Manufacture of basic metals | 7 | 57.1 | 28.6 | 14.3 | 0.0 | 0.0 |
| 25 | Manufacture of fabricated metal products. except machinery and equipment | 56 | 50.0 | 16.1 | 21.4 | 7.1 | 5.4 |
| 26 | Manufacture of computer. electronic and optical products | 10 | 0.0 | 10.0 | 70.0 | 0.0 | 20.0 |
| 27 | Manufacture of electrical equipment | 16 | 6.3 | 25.0 | 50.0 | 0.0 | 18.8 |
| 28 | Manufacture of machinery and equipment n.e.c. | 51 | 17.6 | 19.6 | 43.1 | 7.8 | 11.8 |
| 29 | Manufacture of motor vehicles. trailers and semi-trailers | 35 | 14.3 | 17.1 | 60.0 | 2.9 | 5.7 |
| 31 | Manufacture of furniture | 7 | 42.9 | 14.3 | 42.9 | 0.0 | 0.0 |
| 35 | Electricity, natural gas, steam and air conditioning supply | 23 | 8.7 | 39.1 | 34.8 | 4.3 | 13.0 |
| 45 | Wholesale and retail trade and repair of motor vehicles and motorcycles | 5 | 0.0 | 20.0 | 40.0 | 0.0 | 40.0 |
| 46 | Wholesale trade, except of motor vehicles and motorcycles | 5 | 20.0 | 0.0 | 80.0 | 0.0 | 0.0 |
| 47 | Retail trade, except of motor vehicles and motorcycles | 19 | 21.1 | 10.5 | 47.4 | 15.8 | 5.3 |
| 49 | Land transport and transport via pipelines | 10 | 10.0 | 30.0 | 40.0 | 10.0 | 10.0 |
| 52 | Warehousing and support activities for transportation | 10 | 10.0 | 0.0 | 40.0 | 50.0 | 0.0 |
| 55 | Accommodation | 11 | 0.0 | 72.7 | 18.2 | 0.0 | 9.1 |
| 58 | Publishing activities | 6 | 16.7 | 0.0 | 66.7 | 16.7 | 0.0 |
| 65 | Insurance, reinsurance and pension funding, except compulsory social security | 6 | 0.0 | 33.3 | 66.7 | 0.0 | 0.0 |
| 66 | Activities auxiliary to financial services and insurance activities | 7 | 0.0 | 0.0 | 85.7 | 14.3 | 0.0 |
| 68 | Real estate activities | 18 | 0.0 | 50.0 | 38.9 | 5.6 | 5.6 |
| 71 | Architectural and engineering activities; technical testing and analysis | 9 | 44.4 | 11.1 | 22.2 | 11.1 | 11.1 |
| 84 | Public administration and defence; compulsory social security | 9 | 0.0 | 33.3 | 33.3 | 0.0 | 33.3 |
| 85 | Education | 50 | 2.0 | 26.0 | 46.0 | 4.0 | 22.0 |
| 86 | Human health activities | 30 | 10.0 | 50.0 | 13.3 | 0.0 | 26.7 |
| 87 | Residential care activities | 17 | 0.0 | 76.5 | 0.0 | 0.0 | 23.5 |
| 88 | Social work activities without accommodation | 9 | 0.0 | 66.7 | 22.2 | 0.0 | 11.1 |
| 93 | Sports activities and amusement and recreation activities | 27 | 3.7 | 29.6 | 29.6 | 0.0 | 37.0 |
| 96 | Other personal service activities | 8 | 87.5 | 12.5 | 0.0 | 0.0 | 0.0 |
| 0 | Residential building or office | 64 | 1.6 | 56.3 | 32.8 | 1.6 | 7.8 |

**Table C.2: Weekends and holidays.**

| Economy division according to NACE Rev. 2 | | Number | Wknd cluster share in % | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Code** | **Name** | | **0** | **1** | **2** | **3** | **4** | **CHP** |
| 1 | Crop and animal production, hunting and related service activities | 6 | 33.3 | 0.0 | 33.3 | 16.7 | 0.0 | 16.7 |
| 10 | Manufacture of food products | 19 | 52.6 | 36.8 | 0.0 | 0.0 | 0.0 | 10.5 |
| 17 | Manufacture of paper and paper products | 8 | 0.0 | 50.0 | 12.5 | 25.0 | 12.5 | 0.0 |
| 20 | Manufacture of chemicals and chemical products | 11 | 27.3 | 27.3 | 9.1 | 18.2 | 9.1 | 9.1 |
| 22 | Manufacture of rubber and plastic products | 18 | 16.7 | 33.3 | 11.1 | 22.2 | 5.6 | 11.1 |
| 23 | Manufacture of other non-metallic mineral products | 6 | 66.7 | 33.3 | 0.0 | 0.0 | 0.0 | 0.0 |
| 24 | Manufacture of basic metals | 7 | 0.0 | 42.9 | 28.6 | 28.6 | 0.0 | 0.0 |
| 25 | Manufacture of fabricated metal products, except machinery and equipment | 56 | 12.5 | 46.4 | 3.6 | 19.6 | 12.5 | 5.4 |
| 26 | Manufacture of computer, electronic and optical products | 10 | 0.0 | 0.0 | 10.0 | 40.0 | 30.0 | 20.0 |
| 27 | Manufacture of electrical equipment | 16 | 12.5 | 0.0 | 12.5 | 43.8 | 12.5 | 18.8 |
| 28 | Manufacture of machinery and equipment n.e.c. | 51 | 11.8 | 15.7 | 3.9 | 43.1 | 13.7 | 11.8 |
| 29 | Manufacture of motor vehicles, trailers and semi-trailers | 35 | 8.6 | 17.1 | 14.3 | 28.6 | 25.7 | 5.7 |
| 31 | Manufacture of furniture | 7 | 0.0 | 71.4 | 14.3 | 14.3 | 0.0 | 0.0 |
| 35 | Electricity, natural gas, steam and air conditioning supply | 23 | 13.0 | 0.0 | 52.2 | 8.7 | 13.0 | 13.0 |
| 45 | Wholesale and retail trade and repair of motor vehicles and motorcycles | 5 | 20.0 | 0.0 | 0.0 | 20.0 | 20.0 | 40.0 |
| 46 | Wholesale trade, except of motor vehicles and motorcycles | 5 | 20.0 | 0.0 | 0.0 | 60.0 | 20.0 | 0.0 |
| 47 | Retail trade, except of motor vehicles and motorcycles | 19 | 21.1 | 0.0 | 31.6 | 26.3 | 15.8 | 5.3 |
| 49 | Land transport and transport via pipelines | 10 | 10.0 | 0.0 | 30.0 | 20.0 | 30.0 | 10.0 |
| 52 | Warehousing and support activities for transportation | 10 | 10.0 | 0.0 | 0.0 | 50.0 | 40.0 | 0.0 |
| 55 | Accommodation | 11 | 0.0 | 0.0 | 90.9 | 0.0 | 0.0 | 9.1 |
| 58 | Publishing activities | 6 | 16.7 | 16.7 | 16.7 | 33.3 | 16.7 | 0.0 |
| 65 | Insurance, reinsurance and pension funding, except compulsory social security | 6 | 0.0 | 0.0 | 33.3 | 66.7 | 0.0 | 0.0 |
| 66 | Activities auxiliary to financial services and insurance activities | 7 | 0.0 | 0.0 | 28.6 | 57.1 | 14.3 | 0.0 |
| 68 | Real estate activities | 18 | 0.0 | 5.6 | 66.7 | 5.6 | 16.7 | 5.6 |
| 71 | Architectural and engineering activities; technical testing and analysis | 9 | 44.4 | 0.0 | 0.0 | 22.2 | 22.2 | 11.1 |
| 84 | Public administration and defence; compulsory social security | 9 | 0.0 | 0.0 | 33.3 | 22.2 | 11.1 | 33.3 |
| 85 | Education | 50 | 8.0 | 0.0 | 30.0 | 36.0 | 4.0 | 22.0 |
| 86 | Human health activities | 30 | 10.0 | 0.0 | 63.3 | 0.0 | 0.0 | 26.7 |
| 87 | Residential care activities | 17 | 0.0 | 0.0 | 76.5 | 0.0 | 0.0 | 23.5 |
| 88 | Social work activities without accommodation | 9 | 11.1 | 0.0 | 77.8 | 0.0 | 0.0 | 11.1 |
| 93 | Sports activities and amusement and recreation activities | 27 | 11.1 | 0.0 | 22.2 | 22.2 | 7.4 | 37.0 |
| 96 | Other personal service activities | 8 | 0.0 | 87.5 | 0.0 | 12.5 | 0.0 | 0.0 |
| 0 | Residential building or office | 64 | 1.6 | 0.0 | 78.1 | 4.7 | 7.8 | 7.8 |

**Appendix D: Sig and siglin cluster regressions.**

**Table D.1: Sig cluster regressions.**

| sig | Cluster | A [-] | B [-] | C [-] | D [-] | R² [-] | σ [-] |
|---|---|---|---|---|---|---|---|
| wd | 0 | 2.4143 | -99.9999 | 2.6003 | 0.8830 | 0.04 | 0.31 |
| | 1 | 1.8240 | -35.4141 | 4.9456 | 0.3079 | 0.76 | 0.23 |
| | 2 | 2.6768 | -35.6469 | 5.7102 | 0.0457 | 0.87 | 0.25 |
| | 3 | 4.0532 | -36.8864 | 7.5258 | 0.0098 | 0.84 | 0.43 |
| wknd | 0 | 0.4415 | -37.8292 | 4.4753 | 0.6375 | 0.04 | 0.39 |
| | 1 | 0.3636 | -35.9533 | 6.0899 | 0.1741 | 0.10 | 0.26 |
| | 2 | 1.6346 | -33.3301 | 6.3828 | 0.2542 | 0.81 | 0.21 |
| | 3 | 1.9819 | -35.7829 | 6.1381 | 0.0275 | 0.76 | 0.27 |
| | 4 | 2.8563 | -35.4605 | 6.9369 | 0.0122 | 0.79 | 0.38 |

$$\frac{Q_d}{Q_d(8°C)} = h_{sig}(T_{amb}) = max\left( 0, \quad \frac{A}{1+\left(\frac{B}{T_{amb}-40}\right)^{C}} + D \right) \tag{Eq. D.1}$$

$A, B, C, D$ fit parameter of sigmoid function [-]

$h_{sig}(T_{amb})$ normalized daily natural gas consumption as sigmoid function of $T_{amb}$ [-]

$Q_d/Q_d(8°C)$ normalized daily natural gas consumption/ heat load [-]

$T_{amb}$ daily mean ambient temperature [°C] (insert unitless))

**Table D.2: Siglin cluster regressions.**

| siglin | Cluster | A [-] | B [-] | C [-] | D [-] | $w_{lin}$ [-] | $R^2$ [-] | σ [-] |
|---|---|---|---|---|---|---|---|---|
| wd | **0** | 1.0368 | -44.6482 | 49.9428 | 0.9624 | 0.9811 | 0.04 | 0.31 |
| | **1** | 0.0000 | -4.7816 | 49.9999 | 0.1578 | 1.0000 | 0.76 | 0.23 |
| | **2** | 1.5058 | -31.2511 | 31.2280 | 0.1962 | 0.8911 | 0.87 | 0.24 |
| | **3** | 3.5011 | -35.4059 | 9.0109 | 0.0367 | 0.5590 | 0.84 | 0.43 |
| wknd | **0** | 3.5455 | -47.4457 | 42.1088 | 0.7042 | 0.9980 | 0.04 | 0.39 |
| | **1** | 0.0000 | -99.5000 | 48.6819 | 0.0554 | 0.9999 | 0.10 | 0.26 |
| | **2** | 0.0001 | -38.1586 | 18.9013 | 1.2633 | 0.9998 | 0.81 | 0.21 |
| | **3** | 1.2936 | -31.8414 | 13.1491 | 0.0918 | 0.8330 | 0.76 | 0.26 |
| | **4** | 0.0001 | -71.7926 | 48.7401 | 1.6555 | 0.9999 | 0.79 | 0.38 |

$$\frac{Q_d}{Q_d(8°C)} = h_{siglin}(T_{amb})$$

$$= \begin{cases} max\left(0,\ w_{lin} \cdot (m_h \cdot T_{amb} + b_h\,) + (1 - w_{lin}) \cdot \left(\frac{A}{1 + \left(\frac{B}{T_{amb} - 40}\right)^C} + D\right)\ \ if\ T_{amb} < T_{hl}\right) \\ max\left(0,\ w_{lin} \cdot (m_w \cdot T_{amb} + b_w)\ + (1 - w_{lin}) \cdot \left(\frac{A}{1 + \left(\frac{B}{T_{amb} - 40}\right)^C} + D\right)\ \ if\ T_{amb} \geq T_{hl}\right) \end{cases} \quad \text{Eq. D.2}$$

$A, B, C, D$ fit parameter of sigmoid function [-]
$b_h$ y-axis intercept of space heating line [-]
$b_w$ y-axis intercept of domestic hot water (process heat) line [-]
$h_{sig}(T_{amb})$ normalized daily natural gas consumption as sigmoid function of $T_{amb}$ [-]
$m_h$ slope of space heating line [-]
$m_w$ slope of domestic hot water (process heat) line [-]
$Q_d/Q_d(8°C)$ normalized daily natural gas consumption/ heat load [-]
$T_{amb}$ daily mean ambient temperature [°C] (insert unitless))
$T_{hl}$ heating limit temperature [°C] (insert unitless)
$w_{lin}$ weight of linear SLP [-]

For the lin regression parameters ($m_h, b_h, m_w, b_w, and\ T_{hl}$) see Table 3.

**Appendix E: Common activities/heat sinks of examined consumers in primary and secondary sector (just economy divisions with at least five load profiles).**

<table>
<tr><th colspan="2">Economy division according to NACE Rev. 2</th><th colspan="4">Number in wd cluster</th><th colspan="3">Heat sinks/ processes in</th></tr>
<tr><th>Code</th><th>Name</th><th>0</th><th>1</th><th>2</th><th>3</th><th>Wd Cluster 0</th><th>Wd Cluster 1</th><th>Wd Clusters 2 + 3</th></tr>
<tr><td>1</td><td>Crop and animal production, hunting and related service activities</td><td>2</td><td>0</td><td>3</td><td>0</td><td>animals (e.g., floor heating systems)</td><td></td><td>plants (e.g. greenhouse)</td></tr>
<tr><td>10</td><td>Manufacture of food products</td><td>17</td><td>0</td><td>0</td><td>0</td><td>slaughterhouse (cleaning), bakery, chocolate (melting)</td><td></td><td></td></tr>
<tr><td rowspan="2">17</td><td rowspan="2">Manufacture of paper and paper products</td><td rowspan="2">3</td><td rowspan="2">3</td><td rowspan="2">0</td><td rowspan="2">2</td><td rowspan="2">production</td><td></td><td>management</td></tr>
<tr><td colspan="2">cutting, slicing</td></tr>
<tr><td>20</td><td>Manufacture of chemicals and chemical products</td><td>5</td><td>3</td><td>2</td><td>0</td><td>synthesis, continuous processes</td><td></td><td>mixing, stirring, discontinuous processes</td></tr>
<tr><td>22</td><td>Manufacture of rubber and plastic products</td><td>8</td><td>2</td><td>6</td><td>0</td><td>production (melting, vulcanizing)</td><td></td><td>management</td></tr>
<tr><td>23</td><td>Manufacture of other non-metallic mineral products</td><td>6</td><td>0</td><td>0</td><td>0</td><td>ceramics, plaster, lime</td><td></td><td></td></tr>
<tr><td>24</td><td>Manufacture of basic metals</td><td>4</td><td>2</td><td>1</td><td>0</td><td>foundries</td><td></td><td>logistics</td></tr>
<tr><td>25</td><td>Manufacture of fabricated metal products, except machinery and equipment</td><td>28</td><td>9</td><td>12</td><td>4</td><td>surface treatment</td><td colspan="2">management, logistics</td></tr>
<tr><td>26</td><td>Manufacture of computer, electronic and optical products</td><td>0</td><td>1</td><td>7</td><td>0</td><td></td><td colspan="2">mixed manufacturing sectors (plastics, metal, electronics, mechanics)</td></tr>
<tr><td>27</td><td>Manufacture of electrical equipment</td><td>1</td><td>4</td><td>8</td><td>0</td><td colspan="2">surface treatment (metal)</td><td>management, assembly</td></tr>
<tr><td rowspan="2">28</td><td rowspan="2">Manufacture of machinery and equipment n.e.c.</td><td rowspan="2">9</td><td rowspan="2">10</td><td rowspan="2">22</td><td rowspan="2">4</td><td rowspan="2">production (own foundry, molding, casting, manufacturing)</td><td></td><td>management, logistics, r & d</td></tr>
<tr><td colspan="2">assembly</td></tr>
<tr><td rowspan="3">29</td><td rowspan="3">Manufacture of motor vehicles, trailers and semi-trailers</td><td rowspan="3">5</td><td rowspan="3">6</td><td rowspan="3">21</td><td rowspan="3">1</td><td rowspan="2">production</td><td>r & d</td><td>sales, management</td></tr>
<tr><td colspan="2">logistics</td></tr>
<tr><td colspan="2">assembly</td><td></td></tr>
<tr><td>31</td><td>Manufacture of furniture</td><td>3</td><td>1</td><td>3</td><td>0</td><td>metal furniture, drying, coating</td><td></td><td>wooden furniture, upholstery, design</td></tr>
<tr><td>35</td><td>Electricity, natural gas, steam and air conditioning supply</td><td>2</td><td>9</td><td>8</td><td>1</td><td>natural gas pressure control systems, natural gas storage</td><td></td><td>energy trading, bionatural gas plants</td></tr>
<tr><td colspan="2"><b>General</b></td><td>93</td><td>50</td><td>93</td><td>12</td><td>production processes</td><td></td><td>assembly, logistics, management, R&D, sales</td></tr>
</table>

### Appendix F: Load profiles for the divisions ”manufacture of motor vehicles, trailers and semi-trailers”.

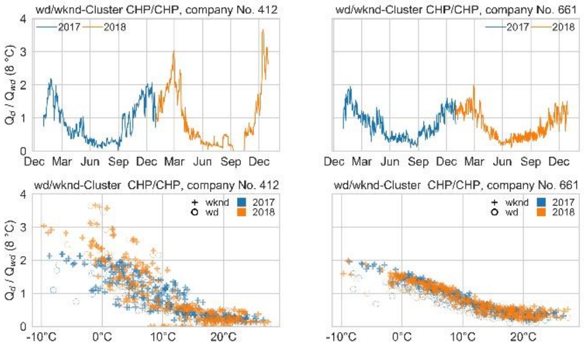


**Figure F.1: Load profiles from consumers operating a CHP. [color]**

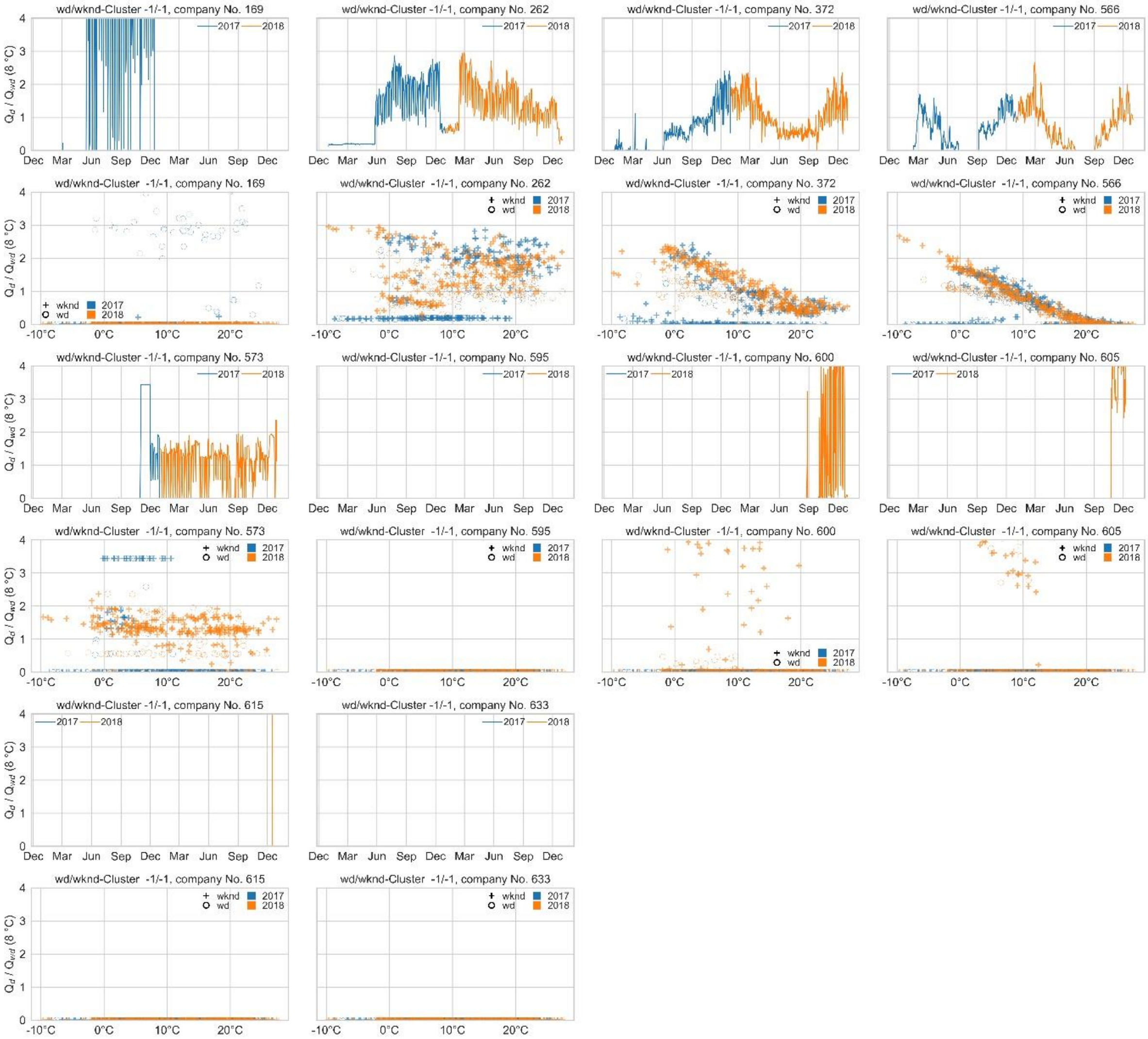
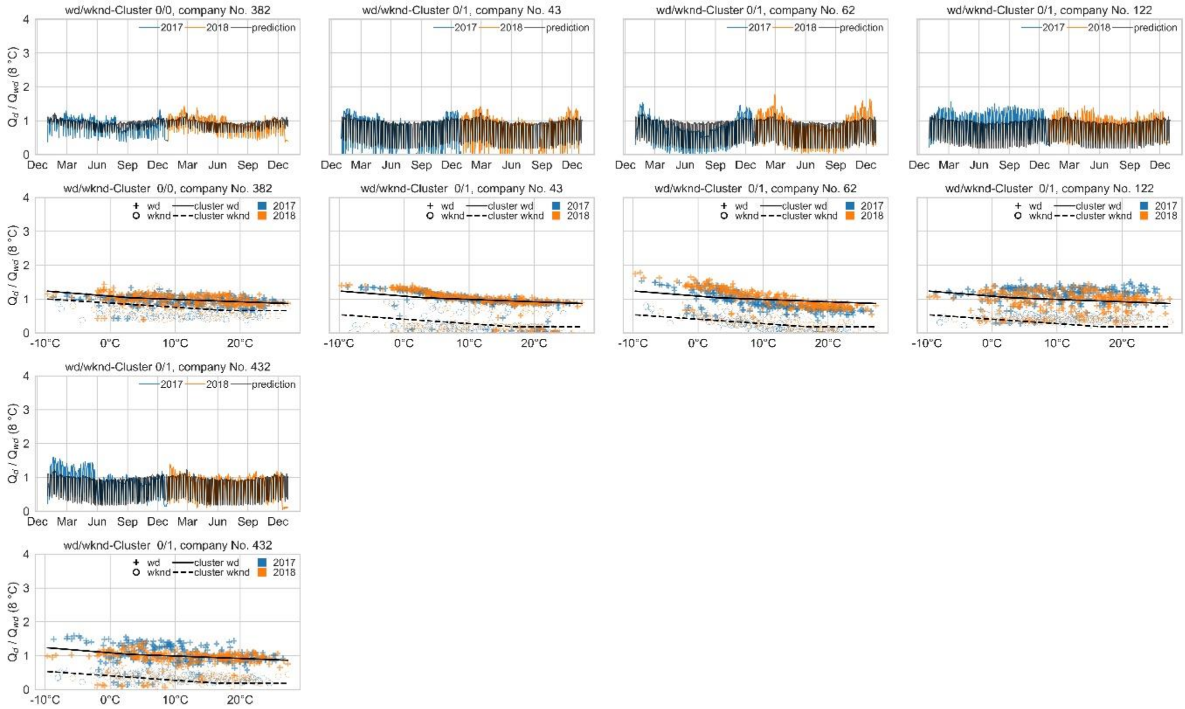


**Figure F.2: Excluded load profiles due to a high share of zero-values (> 83 % of working day hours) or high standard deviation (σ > 0.75). [color]**


wd/wknd-Cluster 0/0, company No. 382

wd/wknd-Cluster 0/1, company No. 43

wd/wknd-Cluster 0/1, company No. 62

wd/wknd-Cluster 0/1, company No. 122

wd/wknd-Cluster 0/0, company No. 382

wd/wknd-Cluster 0/1, company No. 43

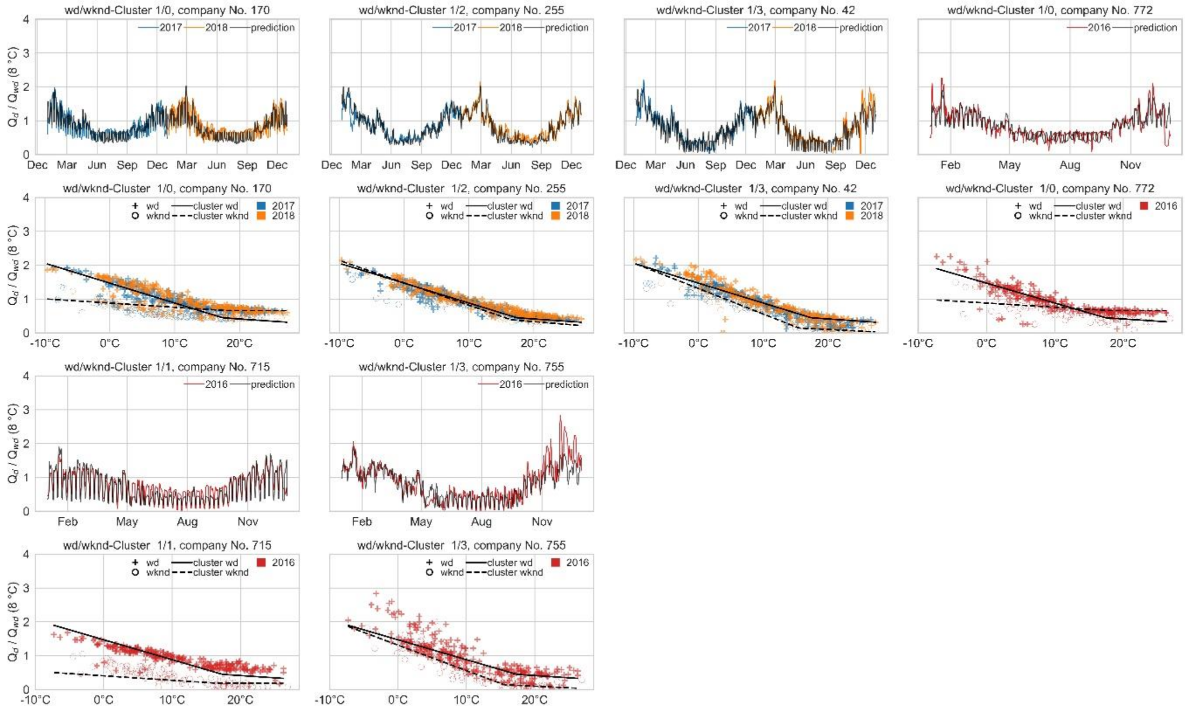
wd/wknd-Cluster 0/1, company No. 62

wd/wknd-Cluster 0/1, company No. 122

wd/wknd-Cluster 0/1, company No. 432

wd/wknd-Cluster 0/1, company No. 432

**Figure F.3: Wd cluster 0. [color]**


wd/wknd-Cluster 1/0, company No. 170

wd/wknd-Cluster 1/2, company No. 255

wd/wknd-Cluster 1/3, company No. 42

wd/wknd-Cluster 1/0, company No. 772

wd/wknd-Cluster 1/0, company No. 170

wd/wknd-Cluster 1/2, company No. 255

wd/wknd-Cluster 1/3, company No. 42

wd/wknd-Cluster 1/0, company No. 772

wd/wknd-Cluster 1/1, company No. 715

wd/wknd-Cluster 1/3, company No. 755

wd/wknd-Cluster 1/1, company No. 715

wd/wknd-Cluster 1/3, company No. 755

**Figure F.4: Wd cluster 1. [color]**

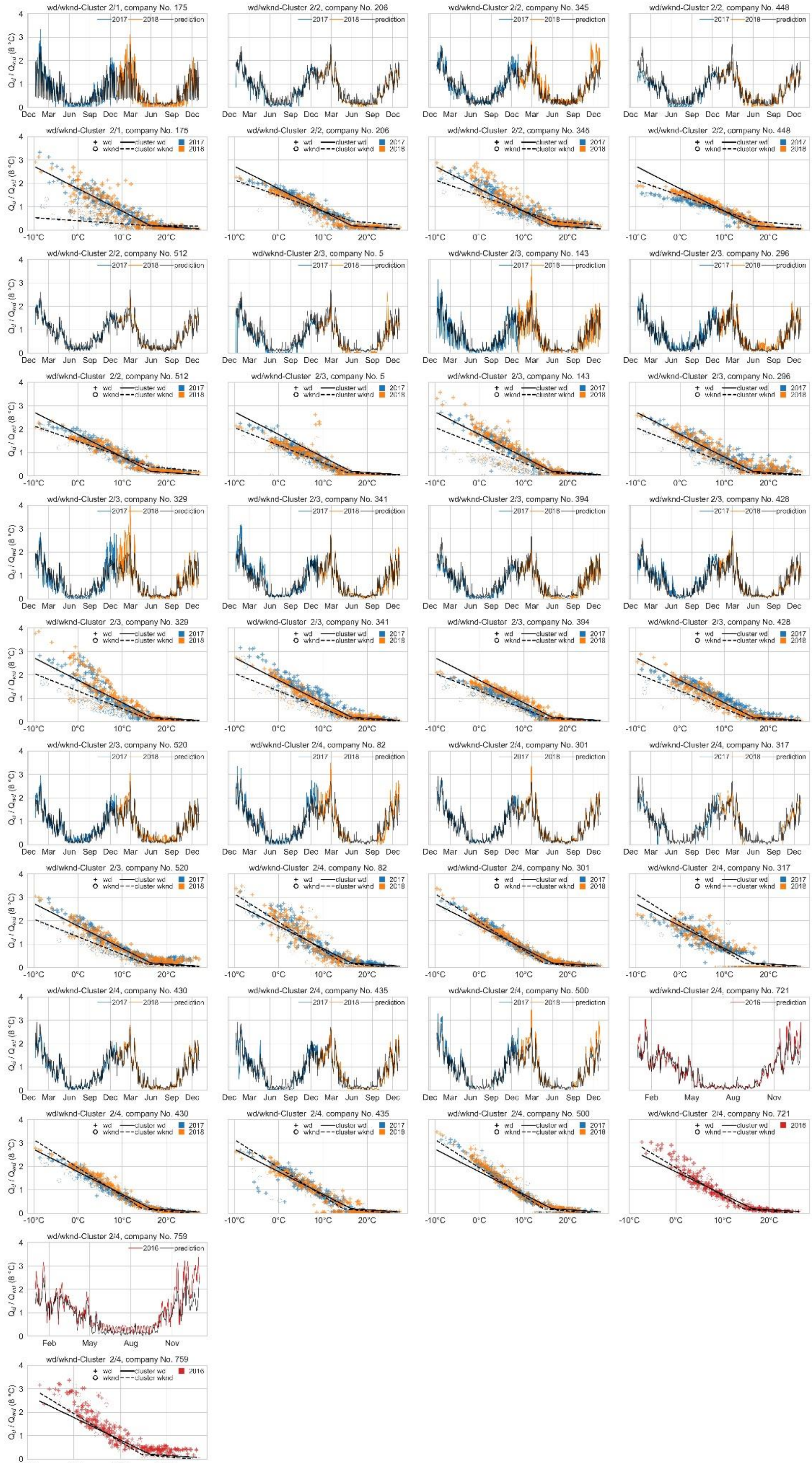


**Figure F.5: Wd cluster 2. [color]**

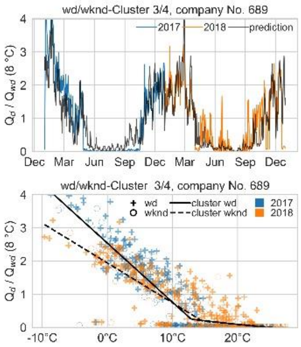


**Figure F.6: Wd cluster 3. [color]**

# References

[1] IEA. Renewables 2020: Analysis and forecast to 2025; 2020; Available from: https://webstore.iea.org/download/direct/4234.

[2] Lauterbach C, Schmitt B, Vajen K. System analysis of a low-temperature solar process heat system. Solar Energy 2014;101:117–30. https://doi.org/10.1016/j.solener.2013.12.014.

[3] Schmitt B. Classification of Industrial Heat Consumers for Integration of Solar Heat. Energy Procedia 2016;91:650–60. https://doi.org/10.1016/j.egypro.2016.06.225.

[4] Annex 35/13. Application of Industrial Heat Pumps: Final Report - Part 1; 2014; Available from: https://iea-industry.org/app/uploads/annex-xiii-part-a.pdf.

[5] Schmitt B. Integration of solar thermal systems for the supply of process heat in industry. Zugl.: Kassel, Univ., Department of Mechanical Engineering, Dissertation, 2014. Aachen: Shaker; 2014 (in German).

[6] Wolf S. Integration of heat pumps into industrial production systems: Potentials and instruments for tapping potential [Dissertation]. Stuttgart: Universität Stuttgart; 2017 (in German).

[7] VDI 4655. Reference load profiles of residential buildings for power, heat and domestic hot water as well as reference generation profiles for photovoltaic plants: draft;ICS 27.010, 27.100, 91.120.10. Berlin: Beuth Verlag; 2019.

[8] Hellwig M. Development and Application of Parameterized Standard Load Profiles [Dissertation]. Munich: TU Munich; 2003 (in German).

[9] Lauterbach C. Potential, system analysis and preliminary design of low-temperature solar process heat systems. Zugl.: Kassel, Univ., Diss. 2014. Kassel: Kassel Univ. Press; 2014.

[10] Association of German Engineers. Load profiles for residential buildings and commerce: For electricity, heating, cooling and domestic hot water; 2019 (In German); Available from: https://www.vdi.de/ueber-uns/presse/publikationen/details/vdi-agenda-lastprofile.

[11] German Federal ministry for Economic Affairs and Energy. Gas Network Access Ordinance: GasNZV; 2019.

[12] BDEW. Processing of Standard Load Profiles Gas. Berlin; 2006 (in German).

[13] FFE. Status report on the Standard load profile method gas. Munich; 2014 (in German); Available from: https://www.ffegmbh.de/download/informationen/508_bdew_slp_gas/Statusbericht_SLP_Gas_FfE_201411.pdf. [November 16, 2020].

[14] FFE. Further Development of the Standard Load Profile Method Gas. Munich; 2015 (in German); Available from: https://www.ffegmbh.de/download/informationen/583_slp_gas_weiterentwicklung/Studie_Weiterentwicklung-SLP-Gas_FfE.pdf. [November 16, 2020].

[15] BDEW. Processing of Standard Load Profiles Gas. Berlin; 2016 (in German).

[16] BDEW. Processing of Standard Load Profiles Gas. Berlin; 2020 (in German); Available from: https://www.bdew.de/media/documents/20200331_KoV_XI_LF_SLP_Gas_clean_final.pdf. [November 13, 2020].

[17] Pag F, Gebele M, Vajen K, Schmitt B. On the importance of ambient temperature dependent process heat and the possibilities of covering it with solar thermal energy. In: Conexio, editor. Proceedings of Solarthermal Symposium; 2018 (in German).

[18] DIN EN 12831. Heating systems in buildings – Method for calculation of the design heat load – National Annex NA;ICS 91.140.10. Berlin: Beuth Verlag; 2008.

[19] DIN 4710. Statistics on German meteorological data for calculating the energy requirement for heating and air conditioning equipment;ICS 07.060; 91.140.30. Berlin: Beuth Verlag; 2003.

[20] Calikus E, Nowaczyk S, Sant'Anna A, Gadd H, Werner S. A data-driven approach for discovering heat load patterns in district heating. Applied Energy 2019;252:113409. https://doi.org/10.1016/j.apenergy.2019.113409.

[21] Gianniou P, Liu X, Heller A, Nielsen PS, Rode C. Clustering-based analysis for residential district heating data. Energy Conversion and Management 2018;165:840–50. https://doi.org/10.1016/j.enconman.2018.03.015.

[22] do Carmo CMR, Christensen TH. Cluster analysis of residential heat load profiles and the role of technical and household characteristics. Energy and Buildings 2016;125:171–80. https://doi.org/10.1016/j.enbuild.2016.04.079.

[23] Lu Y, Tian Z, Peng P, Niu J, Li W, Zhang H. GMM clustering for heating load patterns in-depth identification and prediction model accuracy improvement of district heating system. Energy and Buildings 2019;190:49–60. https://doi.org/10.1016/j.enbuild.2019.02.014.

[24] Ma Z, Yan R, Nord N. A variation focused cluster analysis strategy to identify typical daily heating load profiles of higher education buildings. Energy 2017;134:90–102. https://doi.org/10.1016/j.energy.2017.05.191.

[25] Ravnik J, Hriberšek M. A method for natural gas forecasting and preliminary allocation based on unique standard natural gas consumption profiles. Energy 2019;180:149–62. https://doi.org/10.1016/j.energy.2019.05.084.

[26] Eurostat. NACE Rev.2: Statistical classification of economic activities in the European Community. Luxembourg: Office for Official Publications of the European Communities; 2008.

[27] Jesper M, Pag F, Vajen K, Jordan U. Data repository: Annual industrial and commercial heat load profiles: modeling based on k-Means clustering and regression analysis: Mendeley Data, v1; 2021. https://doi.org/10.17632/nwxv38dxsr.1.

[28] Vögelin P, Koch B, Georges G, Boulouchos K. Heuristic approach for the economic optimisation of combined heat and power (CHP) plants: Operating strategy, heat storage and power. Energy 2017;121:66–77. https://doi.org/10.1016/j.energy.2016.12.133.

[29] BNetzA. Market master data register (MaStR); Available from: https://www.marktstammdatenregister.de/MaStR/.

[30] Jesper M, Schlosser F, Pag F, Walmsley TG, Schmitt B, Vajen K. Large-scale heat pumps: Uptake and performance modelling of market-available devices. Renewable and Sustainable Energy Reviews 2021;137:110646. https://doi.org/10.1016/j.rser.2020.110646.

[31] VCI. Energy statistics in facts and figures; 2020 (in German); Available from: https://www.vci.de/ergaenzende-downloads/energiestatistik-daten-fakten.pdf.

[32] Fraunhofer ISI. Preparation of application balances for the years 2018 to 2020 for the industry and commerce; 2020; Available from: https://ag-energiebilanzen.de/index.php?article_id=29&fileName=isi_anwendungsbilanz_industrie_2019_20200727.pdf.

[33] BMWi. Energy Data. Berlin; 2020 (in German); Available from: https://www.bmwi.de/Redaktion/DE/Binaer/Energiedaten/energiedaten-gesamt-xls.xlsx?__blob=publicationFile&v=131.

[34] Reckzügel, M., Meyer, M., Waldhoff, C., Ludwig, D., Tegeler, A., Schröder, I., Kebschull, O., Magnus, P., Niermann, U., Dering, N., Kruse, A., Vogel, K. Potential study industrial waste heat: LANUV-Fachbericht 96. Recklinghausen, Germany; 2019 (in German).

[35] HLNUG. Weather data (air temperature at 2 m high) - Station 1401 "Kassel-Centre". [November 17, 2020]; Available from: https://www.hlnug.de/?id=9231&station=1401.

[36] DWD. Weather data (air temperature at 2 m high) - Station 04931 "Stuttgart/Echterdingen". [November 17, 2020]; Available from: https://opendata.dwd.de/climate_environment/CDC/observations_germany/climate/hourly/air_temperature/historical/stundenwerte_TU_04931_19880101_20191231_hist.zip.

[37] Oliphant TE. Python for Scientific Computing. Comput. Sci. Eng. 2007;9(3):10–20. https://doi.org/10.1109/MCSE.2007.58.

[38] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O et al. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 2011;12:2825–30.

[39] Arthur D, Vassilvitskii S. K-Means++: The Advantages of Careful Seeding. In: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms. USA: Society for Industrial and Applied Mathematics; 2007, p. 1027–1035.

[40] Bengfort B, Danielsen N, Bilbro R, Gray L, McIntyre K, Richardson G et al. Yellowbrick V0.6. Zenodo; 2018.

[41] Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics 1987;20:53–65. https://doi.org/10.1016/0377-0427(87)90125-7.

[42] Martin Pehnt, Jan Bödeker, Marlene Arens, Eberhard Jochem, Farikha Idrissova. The Utilization of Industrial Waste Heat - Technical and Economic Potentials and Energy Policy Implementation: Report within the framework of the project "Accompanying scientific research on overarching technical, ecological, economic and strategic aspects of the national part of the climate protection initiative".

Heidelberg; 2010 (in German); Available from: http://publica.fraunhofer.de/eprints/urn_nbn_de_0011-n-5690335.pdf.

[43] Hunter JD. Matplotlib: A 2D Graphics Environment. Comput. Sci. Eng. 2007;9(3):90–5. https://doi.org/10.1109/MCSE.2007.55.

[44] Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D et al. Array programming with NumPy. Nature 2020;585(7825):357–62. https://doi.org/10.1038/s41586-020-2649-2.

[45] McKinney W. Data Structures for Statistical Computing in Python. In: Proceedings of the 9th Python in Science Conference. SciPy; 2010, p. 56–61.

[46] Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 2020;17(3):261–72. https://doi.org/10.1038/s41592-019-0686-2.