Treatment and analysis of smart energy meter data from a cluster of buildings connected to district heating

A Danish case

Johra, Hicham; Pereira, Daniel Henrique Leiria E.; Heiselberg, Per; Marszal-Pomianowska, Anna; Tvedebrink, Torben

Published in:
Proceedings of the 12th Nordic Symposium on Building Physics

DOI (link to publication from Publisher):
https://doi.org/10.1051/e3sconf/202017212004

Publication date:
2020

Document Version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Johra, H., Pereira, D. H. L. E., Heiselberg, P., Marszal-Pomianowska, A., & Tvedebrink, T. (2020). Treatment and analysis of smart energy meter data from a cluster of buildings connected to district heating: A Danish case. In *Proceedings of the 12th Nordic Symposium on Building Physics : NSB 2020* (Vol. 172) https://doi.org/10.1051/e3sconf/202017212004

Treatment and analysis of smart energy meter data from a cluster of buildings connected to district heating: A Danish case

Hicham Johra[1,*], *Daniel* Leiria[1], *Per* Heiselberg[1], *Anna* Marszal-Pomianowska[1], and *Torben* Tvedebrink[2]

[1]Aalborg University, Department of Civil Engineering, Thomas Manns Vej 23, 9220 Aalborg Ø, Denmark
[2]Aalborg University, Department of Mathematical Sciences, Skjernvej 4A, 9220 Aalborg Ø, Denmark

Abstract. District heating has been found to be a key component of future and reliable smart energy grids comprising 100% of renewable energy sources for countries with dominant heating season. However, these systems face challenges that require a deeper understanding of the coupling between the distribution networks and the connected buildings, to enable demand-side management and balance the intermittence of renewables. In recent years, many smart energy meters have been installed on the heating systems of Danish dwellings connected to district heating, and the first yearly measurement data sets of large building clusters are now available. This article presents the methodology for the pre-processing and cluster analysis (K-means clustering) of a one-year-long smart energy meter measurement data from 1665 Danish dwellings connected to district heating. The aim is to identify typical household daily profiles of heat energy use, return temperature, and temperature difference between the supply and the return fluid. The study is performed with the free software environment "R", which enables the rapid extraction of information to be shared with professionals of the building and energy sectors. After presenting the preliminary results of the clustering analysis, the article closes with the future work to be conducted on this study case.

1 Introduction

For countries of temperate and cold climates with a dominant heating season, district heating (DH) has been found to be the most efficient, sustainable and cost-effective solution to provide heat to buildings in urban areas. It is also a key component for future and reliable smart grid systems with 100% renewable energy sources [1,2]. However, the current DH systems are facing several production and distribution challenges:

- Detecting faults in the network.
- Identifying critical distribution points, congestion distribution bottlenecks, problematic buildings, and energy-intensive user behaviors.
- Optimizing the system's efficiency by decreasing the overall temperature in the DH network (from typically 80 °C / 45 °C supply/return temperature, down to 50 °C / 25 °C in some extreme cases) to reduce the heat losses in the distribution pipes.
- Decreasing the supply temperature to allow the integration of low-temperature waste heat input from the surrounding industry ecosystem.
- Minimizing the return temperature of the DH network to increase the energy efficiency of the heat production units.
- Diminishing the demand peaks (typically occurring in the morning) to avoid using or installing fossil-fuel peak boilers or temperature boosters with high operation and maintenance costs, and to enable downsizing new distribution networks.
- Matching the heat energy demand with the heat energy supply and bear the increasing share of intermittent renewable energy sources.

Smart-meter monitoring and demand-side management strategies can greatly help to address the aforementioned challenges. However, large efforts are still needed to gain a better understanding of the dynamic interactions between the DH networks and the connected buildings, in order to develop and calibrate efficient numerical urban-scale models of those integrated systems. The latter, together with the feedback information from the Smart Grid meters, will ease planning and optimum control of the DH plants and networks, making use of the aggregated energy flexibility potential of building clusters to perform load shifting and peak shaving at a district level and tackle some of the aforementioned issues.

Denmark has a long history of district heating use with networks that are widespread over the entire country. Indeed, Denmark has one of the largest shares of heating production covered by DH systems. 63% of all private houses and citizens of Denmark are connected to the district heating network for space heating and domestic hot water (DHW) production. For comparison, in Europe, only Iceland and Latvia have a higher share of DH coverage (90% and 65%, respectively), followed by Finland, Lithuania, Poland and Sweden (more than 50%). DH systems are primarily found in urban and suburban areas where the heating need density is high. However, smaller towns and villages (500 households)

[*] Corresponding author: hj@civil.aau.dk

can be also be equipped with a DH network. In addition to the vast central DH networks implemented in the 6 largest urban areas of Denmark, around 400 smaller ones can be found in other towns of modest size around the country. Those various DH installations present a large diversity of distribution network sizes, configurations, fuels mixes, and production plants, comprising some short term heat storages (typically 12 hours of full load heat production from the plant), supplementary solar heating or electric boilers, and integrating parts of the surplus heat generation from local industries. In 2013, more than 70% of all DH energy in Denmark was produced in cogeneration with electricity by efficient combined heat and power (CHP) plants. Consequently, more than 60% of the national electricity was produced by CHPs [3].

For all of those aforementioned reasons, Denmark is a particularly interesting case to study the many aspects of the district heating technologies. Fortunately, Denmark has recently started a massive and systematic campaign of smart energy meters installation in all buildings connected to DH. In addition, Denmark has a consistent national building and household information database, which enables statistical studies about the correlations between socio-economic context and energy-related practices, for instance.

The first complete yearly data sets of large clusters of buildings connected to DH networks in different Danish cities are now available for treatment and analysis. Although big data mining has now become a common concept and practice in many fields of research and in the industry such as electricity grid operators (because data from electricity smart meters have been readily available for several years now), this is still a new thing for urban-scale district heating systems.

In recent years, some Danish research projects started investigating how to extract valuable information for the buildings and energy sector from those smart energy meters' data sets.

Gianniou et al. [4] performed a clustering analysis on the heating usage data of 8293 Danish dwellings connected to a district heating network. The researchers could thus classify the different households into specific groups with defined consumption intensity and representative patterns. They also looked at the correlations between the energy intensity, the building's characteristics, the type of occupants, the load profiles of the households, their consumption behavior, and the changes of the latter over time.

In another publication, Gianniou et al. [5] have estimated the indoor set-point temperature (a very sensitive assumption for building energy models) and the building heat losses of 14,000 dwellings by applying linear regression and heat balance calculation to a large data set of DH smart energy meters, weather data and information from the national building register.

Kristensen et al. [6] used smart energy meter data (50 training buildings and 100 test buildings) to create dynamic physics-based building energy models of archetypes using the Bayesian calibration framework. Those building models can produce a good prediction of the aggregated energy use of single-family houses connected to DH.

Hedegaard et al. [7] adopted a bottom-up modeling approach to create, calibrate and validate an urban-scale model of the district heating consumption of 159 houses, using data from public building registers, local weather stations, and smart energy meters of this building cluster. This model has then enabled testing the effectiveness of a simple price-based aggregated demand response strategy to reduce demand peaks in the district heating network.

In this paper, the preliminary result analysis of a one-year data collection of 1665 smart energy meters in a Danish town connected to a large DH network is presented. The data is analyzed with the K-means clustering method to help identify typical household daily profiles of heat energy usage, return temperature to the DH network, and the temperature difference between the supply and the return fluid.

2 Study case

2.1. Building cluster

The study case is a building cluster consisting of 1665 residential dwellings (predominantly detached houses.), all located in a small town of Northern Jutland in Denmark.

For space heating and domestic hot water supply, the entire building cluster is connected to the large district heating network of the municipality of Aalborg (see Figure 1). For space heating supply, the local network inside each building comprising the heating terminals (radiators or under-floor heating) is directly connected to the main DH network. On the other hand, the instantaneous domestic hot water production is performed by a heat exchanger (indirect connection to the main DH network).

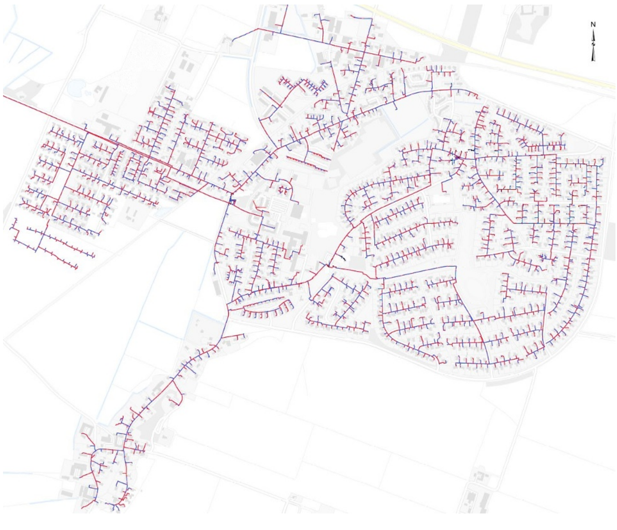


Fig. 1. Overview of the district heating distribution network of the town study case.

Aalborg Forsyning [8] is the utility company in charge of the heat production and distribution, development, operation and maintenance of the DH distribution network of the Aalborg municipality and its surrounding smaller cities. In 2018, Aalborg Forsyning has installed smart energy meters in every dwelling of the town study case: one smart energy meter per building, measuring the aggregated energy usage for space heating and domestic hot water production.

2.2. Smart energy metering system

The smart energy meters installed in the buildings of the study case are state-of-the-art metering devices for hydronic heating systems: Multical® 402, Multical® 403, and Multical® 602 from the specialized manufacturer Kamstrup A/S [9]. The typical uncertainties on the temperature measurements, the fluid flow rate measurements, and the estimates of the energy usage are ± 0.53%, ± 2% and ± 0.22%, respectively. The different measurement variables recorded by the smart energy meters are as follows: timestamp, fluid flow rate, supply temperature, return temperature, cumulative volume usage, and cumulative heat energy usage. In addition, the smart energy meter indicates and records any specific error message if it detects the following problems: wrong flow direction in the fluid flow sensor, air detected in the fluid flow sensor, weak signal from the fluid flow sensor, return temperature is higher than supply temperature, return temperature is too low. The measurements are integrated (averaged) over a time period of around one hour. Consequently, there are around 24 measurements per day for each building.

2.3. Data set

The study of this paper analyses the data collected from the building cluster case during an entire year: from the 1st of October 2018 until the 7th of October 2019. Although the raw data contains more than 14 million lines, the entire data loading, pre-processing, processing, and analysis algorithm developed for this study can be run on a standard computer station in around 2 hours.

3 Methodology

In this study, the entire data processing and analysis are performed with the free software environment and language for statistical computing and graphics “R” [10]. The pre-processing of the raw data files consists in the concatenation of the different log files corresponding to different time periods, followed by resampling to obtain coherent and synchronized hourly times series of the different measurement variables for all the buildings. The quality of the data set is then assessed to determine what criteria is used to discard the sub-sets corresponding to buildings with too many or too large information gaps in the time series. In order to prepare the remaining data subset for clustering analysis, it is necessary to operate an imputation (interpolation or estimation) of the measurement gaps in the time series recorded by the smart energy meters.

The clustering analysis method used in this study to identify subgroups of similar households within the observations of the data set is named “K-means clustering” [11]. This method is simple to employ, widely used, fast to compute, and can be applied to time series, which makes it a perfect tool for the analysis of data from large building clusters.

Antecedently to running the K-means clustering algorithm, the input data set must be standardized, i.e., scaled. The optimum number of clusters (subgroups) is then determined with the “Average Silhouette” method. Finally, the K-means clustering algorithm is executed with a Euclidean distance measure [11].

4 Data analysis, preliminary results, and discussions

4.1. Data set quality and data cleansing

A preliminary data quality assessment is conducted to estimate the amount of missing information in the raw data collected from the 1665 building cases. The missing information is either erroneous measurement points (flagged as an error by the smart energy meter, or containing *NaN* values, or out of realistic range for this DH system) or missing measurements (no measurement has been logged for more than an hour, which creates a gap in the time series). It is found that the raw data set has an overall 3.7% of missing values. One could consider that this percentage of missing information is negligible and use the entire raw data as is. However, consecutive missing data can form large continuous gaps in the time series, which can be problematic when imputing/estimating the latter by means of interpolation.

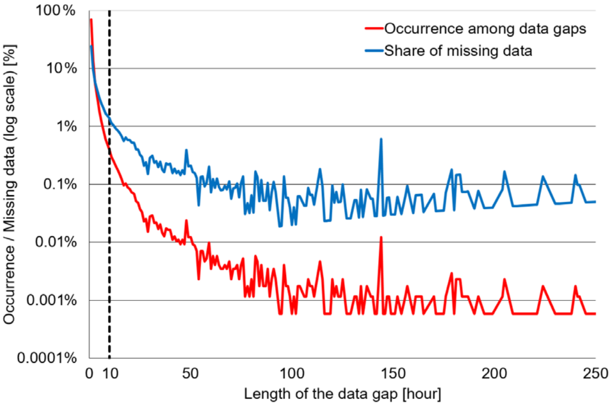


Fig. 2. Occurrence distribution and contribution to the total missing information of the time series gaps as a function of their length in the raw data set.

One can see in Figure 2 the distribution of the gaps in the entire data set as a function of their length. The vast majority of the gaps are short, with the one-hour gaps, two-hour gaps and three-hour gaps representing 70%, 13.7% and 5.5% of the occurrences in the data set, respectively. However, some rare but very large gaps are present in the time series of many buildings. In order to avoid the interpolation of these large gaps, all buildings having information gaps larger than 9 consecutive hours have been discarded. The data subset of the 1028 remaining buildings (out of 1665) has an overall 2.49% of missing information, with a maximum of 10.7% missing values for the time series with the most data gaps. The quality of this subset is considered satisfactory and it is thus used for the clustering analysis.

4.2. Imputation of the missing data

The next step in the data pre-processing is the imputation of the missing values (gaps) in the time series. Because there are numerous methods suitable for the replacement of missing values in univariate time series [12, 13], the latter are benchmarked against each other. One can see in Table 1 the results of this benchmarking. Each method has been tested to estimate information gaps artificially introduced into 10 building time series that originally did not have any missing data. These information gaps have a length distribution that is equal to that of the entire data set (see the previous subsection) and are randomly placed within the 10 building time series. The Root Mean Square Error (RMSE) is then calculated for each measurement variable and used to determine the most suited imputation methods for the current data set, i.e., the ones with the lowest RMSE.

According to the benchmarking results, it is chosen to use linear interpolation for the cumulative measurements (energy, volume, volume x supply temperature, volume x return temperature), and exponential weighted moving average (k=8) for instantaneous measurements (supply temperature, return temperature, fluid flow rate).

Table 1. Benchmarking results for the different missing data imputation methods.

Imputation method	Root Mean Square Error (RMSE)						
	Cumulative measurements				instantaneous measurements		
	Energy	Volume	Volume x supply temperature	Volume x return temperature	Supply temperature	Return temperature	Fluid flow
Linear interpolation	4.0E-01	7.5E-03	2.4E+00	1.8E+00	9.5E-01	7.1E-01	9.4E-03
Spline interpolation	7.0E-01	1.4E-02	4.1E+00	2.6E+00	1.1E+00	8.4E-01	1.1E-02
Stineman interpolation	4.2E-01	7.6E-03	2.5E+00	1.9E+00	9.6E-01	7.1E-01	9.5E-03
Last observ. forward	1.5E+00	3.3E-02	1.2E+01	6.1E+00	1.3E+00	8.8E-01	1.2E-02
Next observ. backward	1.6E+00	3.4E-02	1.2E+01	6.2E+00	1.2E+00	8.7E-01	1.1E-02
Moving average k=2	6.4E-01	1.3E-02	4.9E+00	2.7E+00	9.8E-01	7.1E-01	9.4E-03
Moving average k=4	8.3E-01	1.8E-02	6.6E+00	3.4E+00	1.1E+00	7.2E-01	9.1E-03
Moving average k=6	1.0E+00	2.2E-02	7.9E+00	4.3E+00	1.2E+00	7.5E-01	9.2E-03
Moving average k=8	1.1E+00	2.5E-02	8.8E+00	4.8E+00	1.3E+00	7.8E-01	9.2E-03
Lin. weighted avrg. k=2	5.9E-01	1.2E-02	4.5E+00	2.5E+00	9.6E-01	7.0E-01	9.3E-03
Lin. weighted avrg. k=4	6.8E-01	1.4E-02	5.2E+00	2.8E+00	1.0E+00	6.9E-01	9.0E-03
Lin. weighted avrg. k=6	7.7E-01	1.7E-02	6.0E+00	3.3E+00	1.1E+00	7.1E-01	9.0E-03
Lin. weighted avrg. k=8	8.5E-01	1.9E-02	6.5E+00	3.6E+00	1.1E+00	7.2E-01	9.0E-03
Exp. weighted avrg. k=2	5.8E-01	1.2E-02	4.3E+00	2.4E+00	9.5E-01	7.0E-01	9.3E-03
Exp. weighted avrg. k=4	5.8E-01	1.2E-02	4.4E+00	2.4E+00	9.6E-01	6.8E-01	9.0E-03
Exp. weighted avrg. k=6	6.0E-01	1.2E-02	4.4E+00	2.5E+00	9.6E-01	6.8E-01	8.9E-03
Exp. weighted avrg. k=8	6.0E-01	1.3E-02	4.5E+00	2.5E+00	9.7E-01	6.8E-01	8.9E-03
Mean value	2.4E+03	5.3E+01	2.1E+04	1.0E+04	3.0E+00	1.4E+00	1.6E-02
Median value	2.6E+03	5.6E+01	2.2E+04	1.1E+04	3.1E+00	1.5E+00	1.6E-02

4.3. Problem decomposition and optimum number of clusters

As mentioned before, the aim of the current clustering analysis is to identify typical household daily profiles of energy usage, return temperature to the DH network, and temperature difference between the supply and the return fluid. For the sake of simplicity, the current study is restricted to only working days: all weekend days and official bank holidays are removed. It is assumed that households have a more regular pattern of energy usage during working days compared to rest days. In addition, the data set is divided into four time periods: Spring (March-May), Summer (June-August), Autumn (September-November), Winter (December-February). For each household, the seasonal typical daily profile is thus calculated as the average of all the days over the respective season time period.

In order to perform the K-means clustering analysis, an optimum number of clusters has to be chosen beforehand. Three popular methods, i.e., the Elbow method, the Silhouette method, and the Gap method, have been tested to estimate cluster compactness and clustering quality as a function of the total number of clusters. It is chosen to use the Silhouette method to assess the optimum number of clusters for the analysis of energy usage profiles, return temperature profiles, and temperature difference profiles. One can see in Figure 3 that the optimum number of clusters (highest average silhouette width) is 2 for all the cases. However, it is decided to set the number of clusters to 4 for the analysis of all the measured variables.

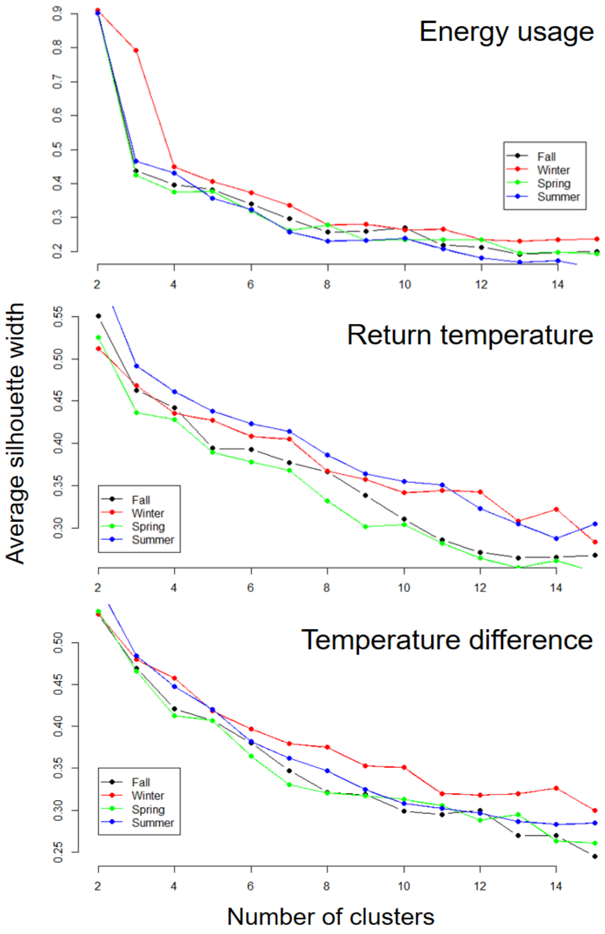


Fig. 3. Average silhouette width (clustering quality) as a function of the total number of clusters.

4.4. K-means clustering results

After all the above mentioned pre-processing steps have been completed, the K-means clustering algorithm can be run. One can see in Figure 4, Figure 5 and Figure 6, the results of this clustering analysis.

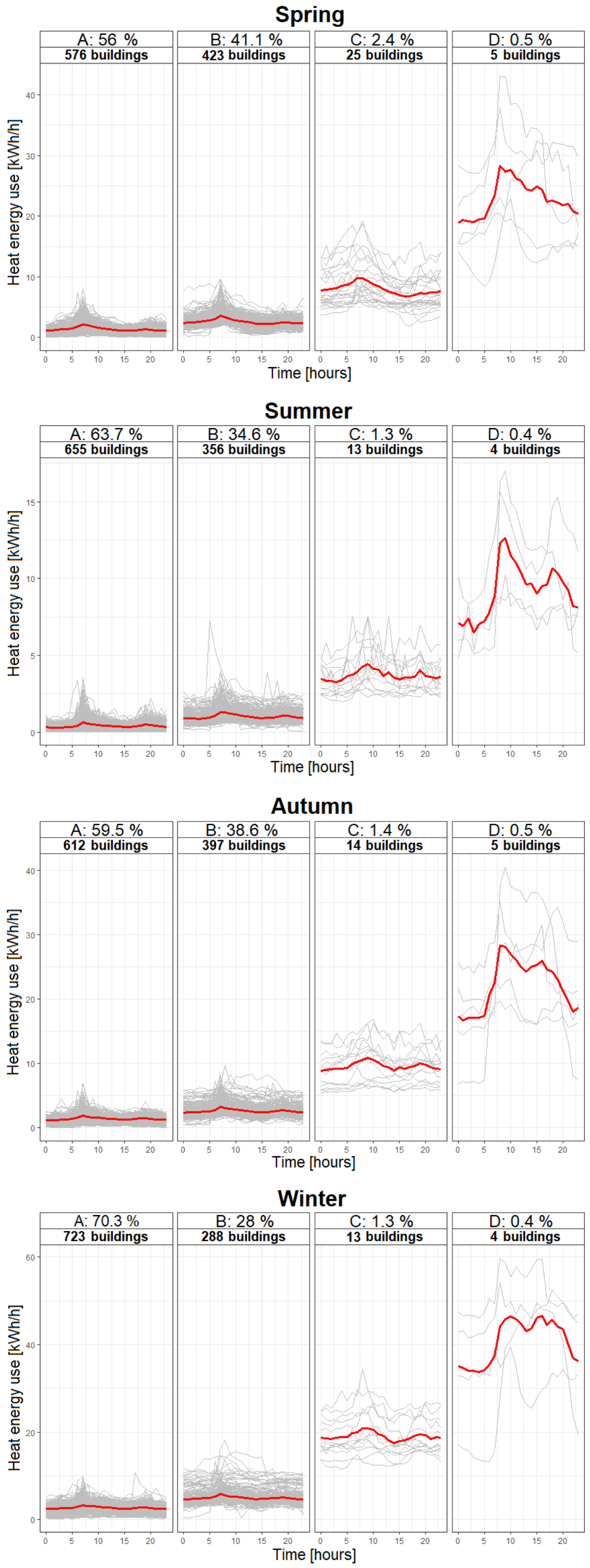


Fig. 4. K-means clustering analysis results for the daily profiles of the heat energy use for working days only during the four seasons.

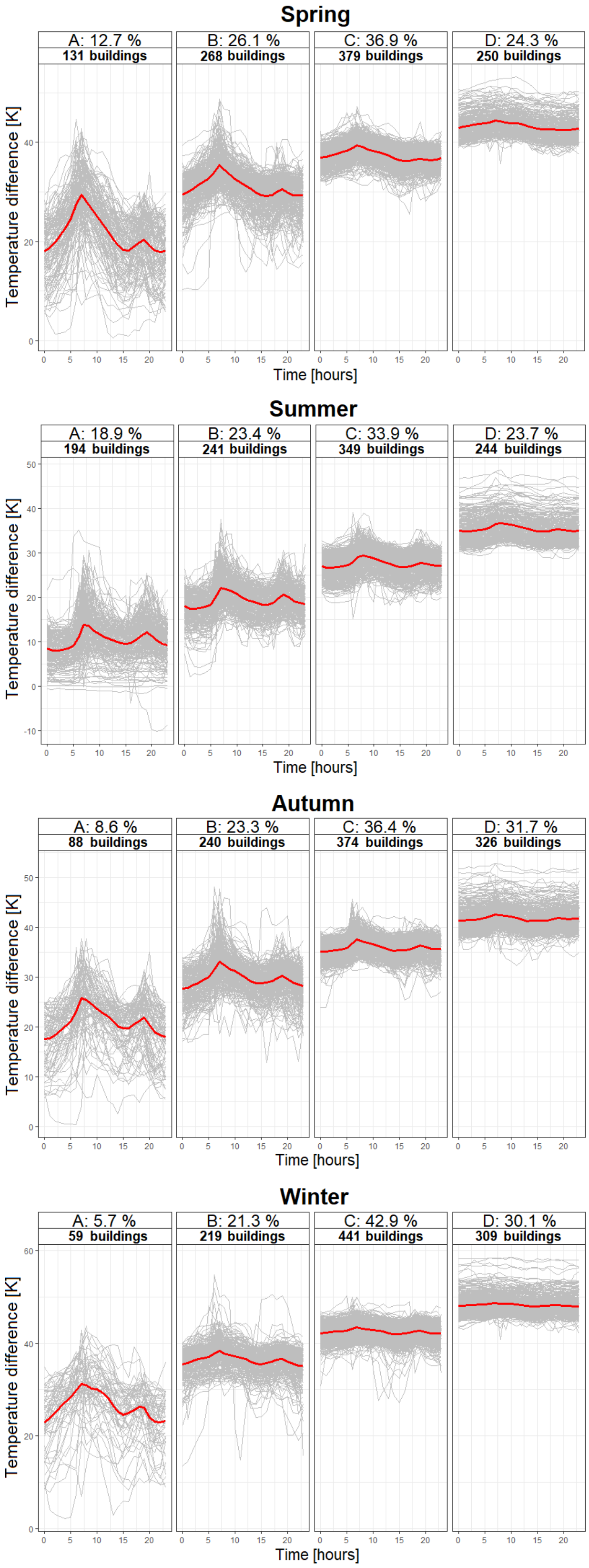


Fig. 5. K-means clustering analysis results for the daily profiles of the temperature difference between the supply and the return fluid for working days only during the four seasons.

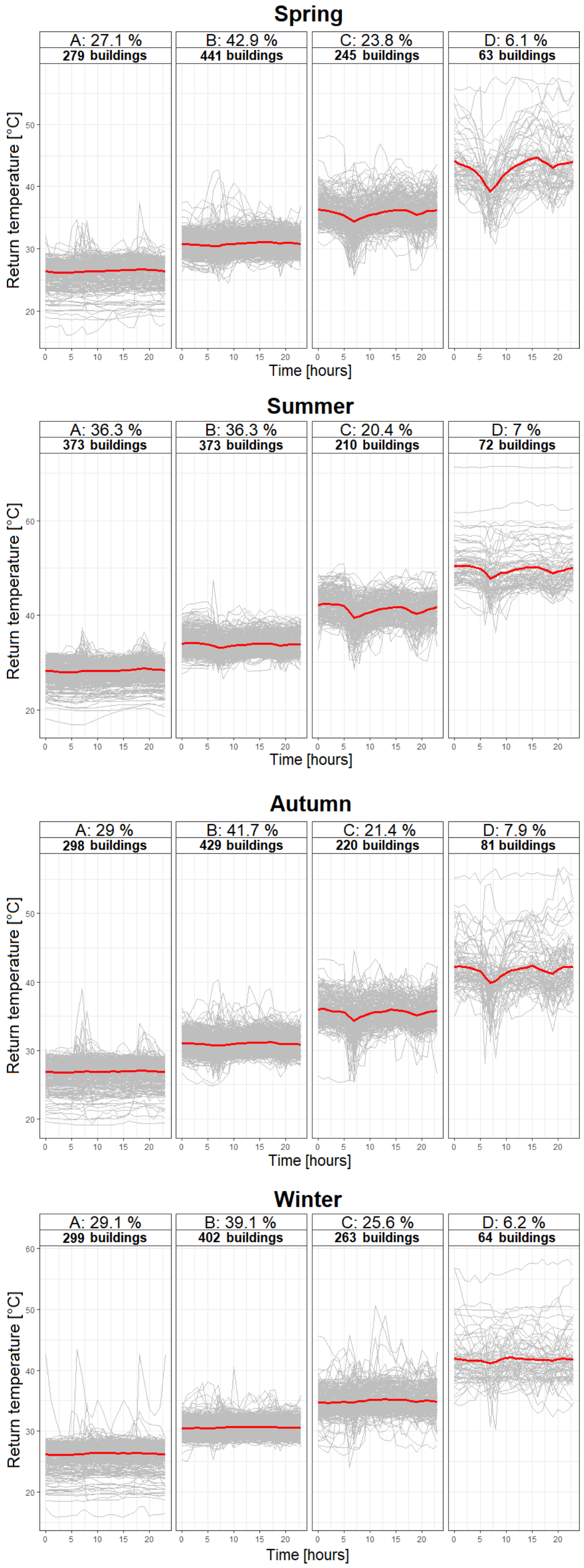


Fig. 6. K-means clustering analysis results for the daily profiles of the return temperature to the DH network for working days only during the four seasons.

For each type of daily profile, namely, heat energy use, temperature difference between the supply and the return fluid, and return temperature to the DH network, the 1028 buildings are classified into 4 different groups (clusters): A, B, C, and D. The subplots present the daily profiles of all the buildings (grey lines) in their respective group. The centroid of each cluster is illustrated by the thick red line. The distribution of each group within the total population of buildings is indicated on the top of each subplot.

One can observe in Figure 4 that the vast majority of the buildings belong to the clusters A and B, which present a rather similar daily profile heat energy use with a significant morning peak (morning shower time on working days) and a smaller evening peak. The same trends can be observed in clusters C and D, but with a much larger overall energy use. One explanatory hypothesis could be that the buildings in clusters A and B are single-family houses (as expected), whereas the buildings in clusters C and D are larger residential buildings with several apartment blocks but a single main smart energy meter. Those results are in agreement with the general description of the study case city and are very similar to what was observed in another Danish city by Giannou et al. [4].

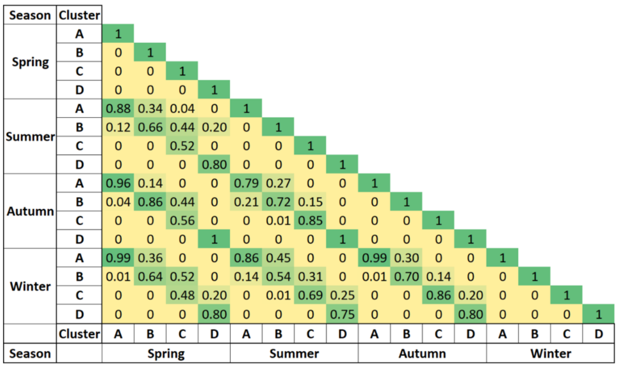
One can see in Figure 5 that concerning the daily profile of temperature difference between the supply and the return fluid, the distribution among the different categories is more balanced. For all seasons, the categories with a relatively low overall temperature difference (clusters A and B) present 2 clear peaks in the morning and in the evening, which is similar to is observed on all heat energy use profiles. However, clusters C and D have a significantly larger temperature difference but with a much more regular profile.

The return temperature profiles presented in Figure 6 appear as a direct consequence of what has been previously described for the temperature difference profiles. Because the supply temperature is expected to be rather stable throughout the entire study case DH subnetwork, the return temperature from the building should be directly correlated with the temperature difference. It is thus logical to observe that clusters A and B have low and stable return temperature profiles, whereas clusters C and D have higher return temperature profiles with more variations and 2 clear temperature drops at morning peak and at evening peak.

4.5. Correlations in between clusters and measured variables

In this section, a simple correlation analysis is performed by calculating the percentage of shared buildings between two clusters from different seasons or different measurement variables. An element of the correlation matrices is marked 1 if 100% of the buildings are shared between the two clusters, and it is marked 0 if there are no common buildings in between those clusters.

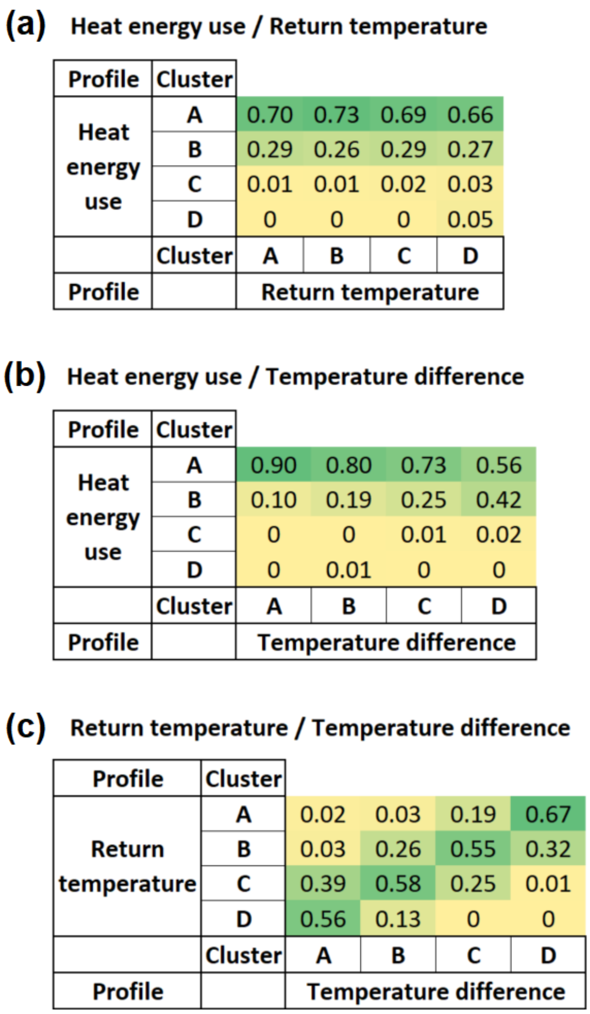
One can see in Figure 7 the matrix of correlations between the different clusters for heat energy use profile during the four seasons. One can notice that most of the terms of the matrix are zero or close to it, except the terms corresponding to the same cluster. This clearly indicates that most of the buildings do not change heat energy use categories in between the different seasons.



Season	Cluster																
Spring	A	1															
	B	0	1														
	C	0	0	1													
	D	0	0	0	1												
Summer	A	0.88	0.34	0.04	0	1											
	B	0.12	0.66	0.44	0.20	0	1										
	C	0	0	0.52	0	0	0	1									
	D	0	0	0	0.80	0	0	0	1								
Autumn	A	0.96	0.14	0	0	0.79	0.27	0	0	1							
	B	0.04	0.86	0.44	0	0.21	0.72	0.15	0	0	1						
	C	0	0	0.56	0	0	0.01	0.85	0	0	0	1					
	D	0	0	0	1	0	0	0	1	0	0	0	1				
Winter	A	0.99	0.36	0	0	0.86	0.45	0	0	0.99	0.30	0	0	1			
	B	0.01	0.64	0.52	0	0.14	0.54	0.31	0	0.01	0.70	0.14	0	0	1		
	C	0	0	0.48	0.20	0	0.01	0.69	0.25	0	0	0.86	0.20	0	0	1	
	D	0	0	0	0.80	0	0	0	0.75	0	0	0	0.80	0	0	0	1
	Cluster	A	B	C	D	A	B	C	D	A	B	C	D	A	B	C	D
Season		Spring				Summer				Autumn				Winter			

Fig. 7. Matrix of correlations between the different clusters for daily profiles of the heat energy use (working days only) during the four seasons.

One can see in Figure 8 the matrices of correlations between the different clusters of daily profiles for heat energy use, return temperature from the building, and temperature difference between the supply and the return fluid. The correlation analysis is performed for the working days of the winter season only.



(a) Heat energy use / Return temperature

Profile	Cluster				
Heat energy use	A	0.70	0.73	0.69	0.66
	B	0.29	0.26	0.29	0.27
	C	0.01	0.01	0.02	0.03
	D	0	0	0	0.05
	Cluster	A	B	C	D
Profile		Return temperature			

(b) Heat energy use / Temperature difference

Profile	Cluster				
Heat energy use	A	0.90	0.80	0.73	0.56
	B	0.10	0.19	0.25	0.42
	C	0	0	0.01	0.02
	D	0	0.01	0	0
	Cluster	A	B	C	D
Profile		Temperature difference			

(c) Return temperature / Temperature difference

Profile	Cluster				
Return temperature	A	0.02	0.03	0.19	0.67
	B	0.03	0.26	0.55	0.32
	C	0.39	0.58	0.25	0.01
	D	0.56	0.13	0	0
	Cluster	A	B	C	D
Profile		Temperature difference			

Fig. 8. Matrices of correlations between the different clusters of daily profiles for different measurement variables for the winter season only and for the working days only.

One can observe in Figure 8 (a) that there is no particular pattern for the correlation matrix of the heat energy use and the return temperature. The distribution of the return temperature clusters within the heat energy ones is very similar to that of the energy profile among all buildings, which would suggest that there is no particular correlation between those two variables.

On the other hand, one can see in Figure 8 (b) that there is a slight trend of buildings with a low heat energy use (towards cluster A) having also lower temperatures and vice versa.

In agreement with what was assumed in the previous section, one can clearly identify in Figure 8 (c) the correlation between the temperature difference and the return temperature from the building.

5 Conclusions

In this paper, the authors have presented the different steps necessary for the pre-processing and cleansing of the raw data from smart energy meters in order to perform a K-means clustering analysis. The K-means clustering method is fairly simple to use, fast to compute and can be applied to time series, which makes it ideal to analyze, categorize and identify typical profiles or patterns in large measurement data sets from vast building clusters.

The study case of the current investigations is a small Danish town connected to the large district heating network of the local regional capital Aalborg. 1665 smart energy meters have been installed in all the dwellings of this locality, and are recording the total heat energy use, fluid flow, supply temperature, and return temperature to the district heating network.

The preliminary results from the study of a one-year data collection of those 1665 smart energy meters (out of which 1028 are kept for analysis) identified 4 distinct typical daily profiles of heat energy use, temperature difference between the supply and the return fluid, and return temperature from the building. Those profiles present clear patterns such as morning and evening peaks of heat energy demand, which lead to corresponding peaks in the profiles of temperature difference, and drops of the return temperature from the building. Those results are in agreement with previous observations of district heating systems in Denmark. In addition, a simple correlation analysis showed that the buildings of the study case tend to stay in the same heat energy use category from one season to the other. Finally, since the supply temperature is expected to be homogenous throughout the subnetwork of this building cluster, there is a strong correlation between the return temperature from the building and the temperature difference between supply and return fluid.

6 Future work

As the current paper only presents the preliminary results of this clustering analysis, many follow-up studies will be conducted on the smart energy meter measurement data set:

- Other pre-processing, missing data imputation and clustering methods will be tested.
- Further analysis of the data and coupling with the local weather information and the Danish national building register will be performed to identify building characteristics such as the envelope thermal performance, or the type of heating system installed.
- A challenging task will be to identify certain traits of the occupants' behavior and practices such as the domestic hot water usage profile, and the indoor temperature set point.
- Anonymized socio-economic studies would also be of great interest to assess the correlation between types of occupant and heating usage profiles.
- The "Shiny" package [14] will be used in the R environment to develop interactive web-based interfaces to present and share the processed smart energy meter data with professionals of the building sector and with the DH utility companies.
- Further data analysis could also help to detect problems in the DH network or identify problematic and critical buildings in the network.
- Combined with high-quality GIS information about this DH network study case, this data set analysis will be used to generate, calibrate and validate an urban-scale numerical model based on the Modelica language, and sub-model component libraries and tools developed by the IBPSA Project 1 [15]. This detailed model will be able to simulate both the thermodynamics of the DH network and of every single building, which makes it suitable for testing urban-scale demand-side management and energy flexibility strategies for building clusters.

This work was carried out as part of the InterHUB project, which is funded by Aalborg University's strategic funding for interdisciplinary research. InterHUB project: Intermittent energy – Integrating Households, Utilities, and Buildings (www.interhub.aau.dk). The authors would also like to acknowledge Aalborg Forsyning for providing the raw data and the technical information for this study.

References

1. H. Lund, B. Möller, B.V. Mathiesen, A. Dyrelund, Energy **35**, 1381-1390 (2010)
2. B.V. Mathiesen, H. Lund, D. Connolly, H. Wenzel, P.A. Østergaard, B. Möller, S. Nielsen, I. Ridjan, P. Karnøe, K. Sperling, F.K. Hvelplund, Appl. Energy **145**, 139-154 (2015)
3. Danish Energy Agency, *Regulation and planning of district heating in Denmark* (2015)
4. P. Giannou, X. Liu, A. Heller, P.S. Nielsen, C. Rode, Energy Convers. Manag. **165**, 840-850 (2018)
5. P. Giannou, C. Reinhart, D. Hsu, A. Heller, C. Rode, Build. Environ. **139**, 125-133 (2018)
6. M.H. Kristensen, R.E. Hedegaard, S. Petersen, Energ. Buildings **175**, 219-234 (2018)
7. R.E. Hedegaard, M.H. Kristensen, T.H. Pedersen, A. Brun, S. Petersen, Appl. Energy **242**, 181-204 (2019)
8. Aalborg Forsyning, https://aalborgforsyning.dk
9. Kamstrup A/S, Smart heat meters & devices, https://www.kamstrup.com/en-en/heat-solutions/heat-meters
10. R environment for statistical computing, https://www.r-project.org/about.html
11. K-means clustering analysis with R, https://uc-r.github.io/kmeans_clustering
12. S. Moritz, T. Bartz-Beielstein, R J. **9**, 207-218 (2017)
13. Time Series Missing Value Imputation, R package, https://cran.r-project.org/web/packages/imputeTS/imputeTS.pdf
14. Shiny, R package, https://shiny.rstudio.com
15. IBPSA Project 1, https://ibpsa.github.io/project1